# OpenStreetMap Administrative Polygon Shape Files with Hierarchy

2019-11-11

Michael Reichert <michael.reichert@geofabrik.de>

## Table of Contents



## Version History and Copyright

| Version | Created At | Changes |
|---|---|---|
| 1.0 | 2017-12-11 | Initial document |
| 1.1 | 2019-11-11 | Describe clipping to coastline and simplification, partial rewrite of the document. Add int_name field, document geomtype field better. |

Copyright in this document is held by Geofabrik GmbH.

## 1 Preface

The OpenStreetMap (OSM) project (www.openstreetmap.org) has collected an enormous amount of free spatial data and the database is growing every day. Many people want to use this data for their own GIS projects but have been hindered by the use of a non-standard data format in the OSM project.

Geofabrik has pioneered a mapping from OSM features to a traditional GIS structure with their “OSM Data in Layered GIS Formats” specification. That specification covers most low-level aspects of OSM.

This document describes an three add-ons to the “adminareas” layer of the “OSM Data in Layered GIS Formats” by Geofabrik:

- An additional mapping of detailed OSM structures concerning the hierarchy of administrative units (from countries over counties to municpalities) is added to the layer. The hierarchy is represented by additional fields for each feature, referencing upper-level entities by their OSM ID. See section 4 for details.

- Administrative boundaries in OSM include territorial waters. However, many use cases require geometries that represent the actual country outlines, i.e. boundaries should be clipped to the coast. This option does not add any additional fields.

- Boundary polygons having many vertices are a challenge for some use cases and need a lot of disk space. We can simplify these boundary lines.

The options are fully backward compatible. You are free to choose only one of the three.

# 2 Introduction

## 2.1 Versions

This is no static document. New versions of this document are likely to appear from time to time. If the definition of layers or features is changed significantly, the layers will get new version identifiers.

Layer names will use version numbers with leading 'v' and without the embedded '.': So version 0.1 of this document will use the suffix “v01”, version 2.7 will use “v27” (minor versions above 9 are not allowed).

## 2.2 Map datum

All coordinates are unprojected WGS84 (EPSG:4326).

## 2.3 Character encoding

All strings are encoded in UTF-8.

## 2.4 Attribute names

All attribute names are lower case and are less than 11 characters long so that they are not truncated in shapefiles.

## 2.5 Layers

Administrative polygon shape files with hierarchy by Geofabrik consist of only one single layer called adminareas.

## 2.6 Common attributes

Most tables/shape files by Geofabrik will have the following columns/attributes:

| Attribute | PostGIS Type | Descriptiosteps/gis-roads-v06.pln |
|---|---|---|
| osm_id | VARCHAR (10) | OSM Id taken from the Id of this feature (node_id, way_id, or relation_id) in the OSM database. In case several features in the OSM database are joined into one feature, this is one of the Ids. This ID is not necessarily unique because one OSM object can result in several geometry objects. Also note that when doing shape file exports, this will be exported as a VARCHAR type since shape files don't support long integers. |
| lastchange | TIMESTAMP WITHOUT TIME ZONE | Last change of this feature. Comes from the OSM last_changed attribute. Reflects changes in the attributes of a feature; changes in the geometry will not necessarily change this. |
| code | SMALLINT (2 Bytes) | 4 digit code (between 1000 and 9999) defining the feature class. The first one or two digits define the layer, the last two or three digits the class inside a layer. |
| fclass | VARCHAR(40) | Class name of this feature. This does not add any information that is not already in the “code” field but it is better readable. |
| name | VARCHAR(100) | Name of this feature, like a street or place name. If the name in OSM contains obviously wrong data such as “fixme” or “none”, it will be empty. |
| int_name | VARCHAR(100) | English name of this feature (OSM tag name:en=*, fallback: int_name=*). If there is no English/international name, this field is empty. It is not guaranteed that this field contains ASCII characters only. |
| geomtype | VARCHAR(1) | This field is available for backward compatibility with the standard shape file. It indicates the OSM object type (N: node, W: way, R: relation). This layer contains features derived from relations only. Therefore, this field is always set to “R”. |

The code and the combination of layer name and fclass always contains the same information.

## 2.7 Spillover Shape Files

When a certain layer becomes too large for one shape file (shape files are limited to 2 GB in size), it will automatically spill over into additional shape files. A shape file named “gis_osm_adminareas_v10_1.shp” will have spillover shape files names “gis_osm_adminareas_v10_2.shp”, “gis_osm_adminareas_v10_3.shp” and so on.

# 3 Codes and Feature Classes

This layer is similar the the “boundaries” layer but it contains polygons built from boundary lines. This layer does have a name attribute.

Additional attributes:

| Attribute | PostGIS Type | Description | OSM Tags |
|---|---|---|---|
| postalcode | VARCHAR(10) | Postal code for this administrative area. Postal codes are only available for a few administrative areas, they are not necessarily unique. | postal_code=* |

The following feature classes exist in this layer:

| code | layer | fclass | Description | OSM Tags |
|---|---|---|---|---|
| 1200 | adminareas | | | boundary=administrative |
| 1201 | adminareas | admin_level1 | | + admin_level=1 |
| 1202 | adminareas | national | National border | + admin_level=2 |
| 1203 | adminareas | admin_level3 | | + admin_level=3 |
| 1204 | adminareas | admin_level4 | Usually a border of the first level below national; NUTS-2 (Germany: Land, France: région, UK: England/ Scotland/ Wales…). | + admin_level=4 |
| 1205 | adminareas | admin_level5 | | + admin_level=5 |
| 1206 | adminareas | admin_level6 | Usually a border of the second level below national; NUTS-3 (Germany: Kreis, France: département, UK: county…). | + admin_level=6 |
| 1207 | adminareas | admin_level7 | | + admin_level=7 |
| 1208 | adminareas | admin_level8 | Usually a city or borough boundary. | + admin_level=8 |
| 1209 | adminareas | admin_level9 | | + admin_level=9 |
| 1210 | adminareas | admin_level10 | | + admin_level=10 |
| 1211 | adminareas | admin_level11 | | + admin_level=11 |

*Note that due to editing errors introduced by OpenStreetMap contributors and also due to lack of data in some areas, there is no guarantee that these areas are complete; there may always be missing bits. Contact Geofabrik if you are interested in a redacted data set.*

The exact meaning of the admin_levels 1 to 11 varies between countries and is documented in detail on the OpenStreetMap Wiki:
https://wiki.openstreetmap.org/wiki/Tag:boundary=administrative

# 4 How hierarchy is modeled

If requested, administrative boundary shape files contain information about the hierarchy.

## 4.1 Structure

Our administrative shape file contain hierarchy information about the relationship between two administrative units. Each feature has attributes pointing to features further up the hierarchy.

For example, if a municipality and the county, state and country it belongs to are available in OpenStreetMap, the municipality will have attributes with the IDs of the county, state and country (see examples at section 4.2).

Many countries don't use all levels in OpenStreetMap. For example, many features at level 4 will have a parent3 field of NULL.

In particular situations, several levels can be empty. For example, two of sixteen German states lack all levels between 4 (state/Bundesland) and 9 (subparts of municipalities/Stadtbezirk). In such a case, the fields parent5, parent6, parent7 and, parent8 will be NULL for a feature at level 9.

The shape file contains a column parent_osm_id which contains the ID of the next higher feature to circumvent this problem.

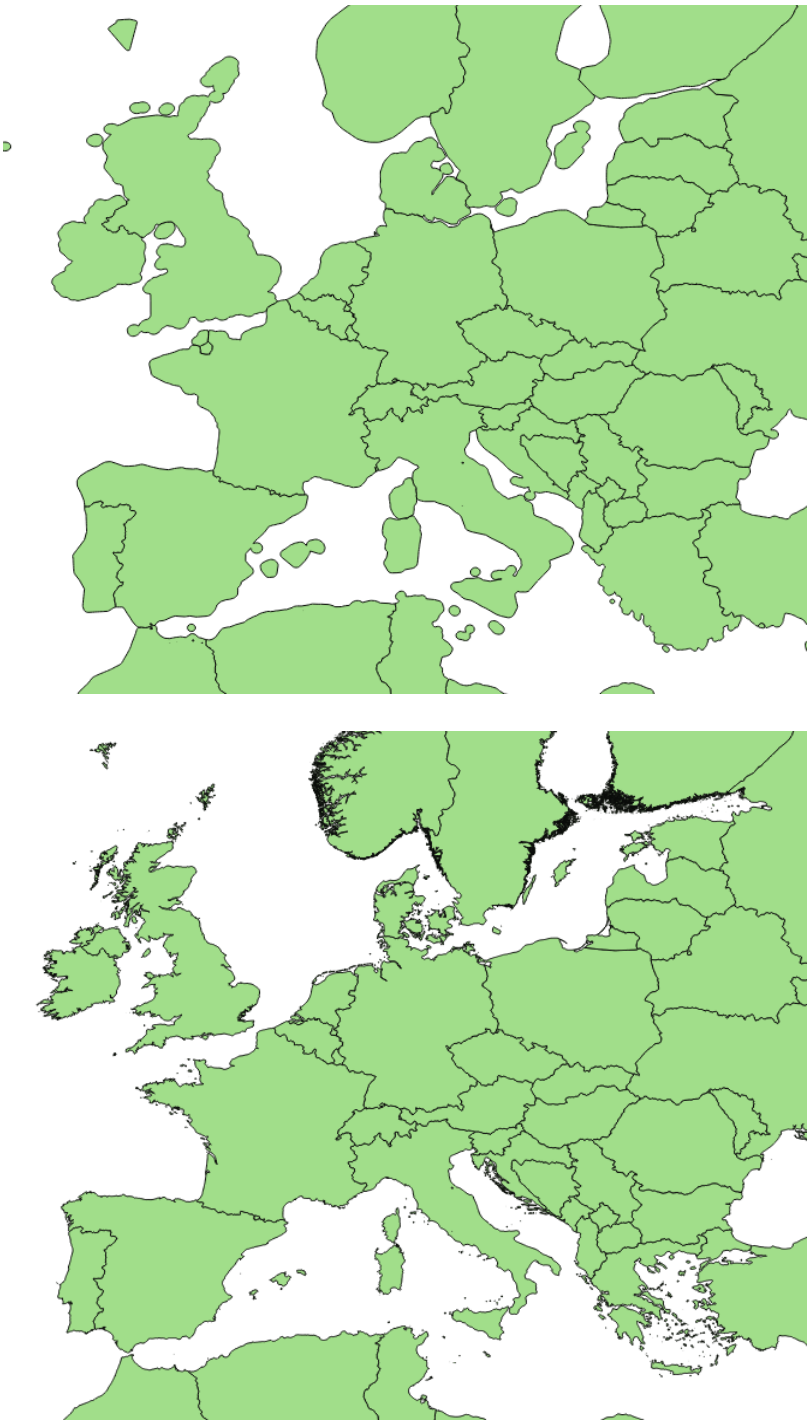
| Attribute | PostGIS Type | Description | Note |
|---|---|---|---|
| parent_osm_id | INTEGER | OSM ID of the adminstrative unit of the next higher administrative level this unit belongs to | refers to osm_id |
| parent_code | SMALLINT | Code of the feature parent_osm_id refers to | equal to code of the referred feature |
| parent2 | INTEGER | OSM ID of the adminstrative unit with code 1202 which this adminstrative unit belongs to | refers to osm_id |
| parent3 | INTEGER | OSM ID of the adminstrative unit with code 1203 which this adminstrative unit belongs to | refers to osm_id |
| parent4 | INTEGER | OSM ID of the adminstrative unit with code 1204 which this adminstrative unit belongs to | refers to osm_id |
| parent5 | INTEGER | OSM ID of the adminstrative unit with code 1205 which this adminstrative unit belongs to | refers to osm_id |
| parent6 | INTEGER | OSM ID of the adminstrative unit with code 1206 which this adminstrative unit belongs to | refers to osm_id |
| parent7 | INTEGER | OSM ID of the adminstrative unit with code 1207 which this adminstrative unit belongs to | refers to osm_id |
| parent8 | INTEGER | OSM ID of the adminstrative unit with code 1208 which this adminstrative unit belongs to | refers to osm_id |
| parent9 | INTEGER | OSM ID of the adminstrative unit with code 1209 which this adminstrative unit belongs to | refers to osm_id |
| parent10 | INTEGER | OSM ID of the adminstrative unit with code 1210 which this adminstrative unit belongs to | refers to osm_id |
| parent11 | INTEGER | OSM ID of the adminstrative unit with code 1211 which this adminstrative unit belongs to | refers to osm_id |

## 4.2 Example

| osm_id | code | name | parent_osm_id | parent_code | parent2 | parent3 | parent4 | parent5 | parent6 | parent7 | parent8 | parent9 | parent10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 122604 | 1208 | Chicago | 122576 | 1206 | 148838 | NULL | 122586 | NULL | 122576 | NULL | NULL | NULL | NULL |
| 122576 | 1206 | Cook County | 122586 | 1204 | 148838 | NULL | 122586 | NULL | NULL | NULL | NULL | NULL | NULL |
| 122586 | 1204 | Illinois | 148838 | 1202 | 148838 | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL |
| 148838 | 1202 | United States of America | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL |
| 164778 | 1210 | Zehlendorf | 55734 | NULL | 51477 | NULL | 62422 | NULL | NULL | NULL | NULL | 55734 | NULL |
| 55734 | 1209 | Steglitz-Zehlendorf | 62422 | NULL | 51477 | NULL | 62422 | NULL | NULL | NULL | NULL | NULL | NULL |
| 62422 | 1204 | Berlin | 51477 | NULL | 51477 | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL |
| 51477 | 1202 | Germany | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL | NULL |

# 5 Excluding territorial waters

Polygons of countries (code 1202) and often the next level below (1203 or 1204) contain the territorial waters on sea (usually 12 nautical miles). Geofabrik can clip all polygons to the coast if requested. See the following comparison of a shape file with and without clipped geometries.



The top image shows unclipped boundary polygons directly from OSM, the bottom image shows them after clipping.