TOWARDS A NAME FOR EACH TOWN

A challenge related to the grid-based degree of urbanisation

Lucie Lecomte and Hugo Poelman

INTRODUCTION

The European system of urban/rural territorial classifications relies upon basic types of 1 km² grid cell clusters[1]. By definition, these clusters are independent from any existing administrative unit. This is an important asset as it avoids distortions due to historical, administrative and/or political factors. Nevertheless, when analysing indicators at the level of individual grid-based clusters it is very useful to be able to attribute a name to the clusters.

High-density clusters (or urban centres) represent the major cities, each with a population of at least 50,000 inhabitants. There are more than 800 urban centres in Europe. Most of these can be associated with a particular local administrative unit. This means that it has been quite straightforward to allocate a name to each of the grid-based urban centres.

The class of urban clusters (moderate-density clusters) contains suburbs and towns. While suburbs can be associated to urban centres, grid-based towns do not have any a-priori name. Still, providing names to these units can be useful to facilitate referencing them. Hence, the challenge is to provide (unique) names for more than 9000 towns throughout Europe.

A REFINED CLASSIFICATION WITHIN THE DEGREE OF URBANISATION TYPOLOGY

Within the framework of the degree of urbanisation typology, containing three classes, a refined typology has been created, adding an additional level of classification at grid cell level[2]. This second level contains 6 classes[3]:

- Cities: this is the original degree of urbanisation class, i.e. settlements of at least 50,000 inhabitants in a high-density cluster of grid cells (> 1500 inhabitants/km²)
- Towns: these can be *dense towns*, with a density of more than 1500 inh./km² and a population between 5,000 and 50,000, or *semi-dense towns* within a population over 5,000 and a density of at least 300 inh./km² provided they are located at more than 2 km from cities or from dense towns.
- Suburbs: these are cells belonging to urban clusters (i.e. clusters of cells with a density of at least 300 inh./km² and a total cluster population of at least 5,000 inhabitants) that are not part of cities or towns. In other words, they must be contiguous with or within 2 km of a city or a dense town.
- Villages: settlements with a population between 500 and 5,000 inhabitants and a density of at least 300 inh./km².
- Dispersed rural areas: rural grid cells with a density between 50 and 300 inh./km².
- Mostly uninhabited areas: rural grid cells with a density between 0 and 50 inh./km².

[1] Dijkstra and Poelman (2014)
[2] An example of the use of this classification can be found in: Dijkstra, Poelman and Ackermans (2019), p. 16-17
[3] Or 7 classes if dense towns and semi-dense towns are considered as separate classes.

CRITERIA FOR NAMING TOWNS

In trying to allocate names to towns we applied some basic criteria:

- Each town should have a unique name. Homologues can occur only when located in different countries or regions.
- Names already used for cities (urban centres) cannot be used for towns.
- Towns should inherit the name of the local administrative unit (LAU) in which they are located whenever possible and appropriate.

METHODOLOGY

DETERMINE THE RELATIONSHIP BETWEEN LAU UNITS AND TOWNS

First we determine the spatial relationship between local administrative units and towns. We create a boundary dataset for all towns and urban centres. This layer is intersected with the boundaries of the LAU units. For each intersecting area we obtain the identifiers of the LAU and the town (or city), the surface area of the intersection and the total area of the relevant LAU and of the town (or city).

NAMING TOWNS WITH LAU NAMES

Where the link between a town and a LAU is strong enough, the town is named as the LAU.

Three criteria must be met:

1) No city in the LAU.

2) No other town in the LAU – only 1 town in the LAU.

3) > 90% of the town's area is shared with the area of the LAU.

As a result 51% of the towns were named.

GEOCODING TO FIND MISSING NAMES

Eurostat-GISCO provides a geocoding service powered by Nominatim [4] and relying upon volunteered geographic information from OpenStreetMap[5]. The geocoding process is run on the centroids of all towns that are still missing a name. The result is a draft table providing one or more names for the point locations: "city", "town", "village", "suburb", "hamlet", "locality" or "city district". The geocoding process provides up to 4 names per town. Names provided under the heading "suburb" or "town" are the most frequent ones.

Choosing which name to keep requires information on the different names suggested. We need to know how many names are suggested, which names were already attributed to a city/town and which names are suggested for other towns as well. We will discard names that are already attributed to a town/city and names that are suggested for other towns.

The table with the geocoding results is restructured in order to allow for the necessary selections. Names already used for a town or city, and names suggested for several towns at once should be discarded. All remaining names should be taken into account when choosing a name.

If only one name was suggested and it is not yet used nor suggested for any other town, it is allocated to the town. This step provides names to almost 1600 towns.

[4] http://nominatim.org/
[5] https://www.openstreetmap.org/

Amongst the suggested names the same name can be suggested twice or three times (under various headings). In these cases, if the name has not yet been used nor suggested for another town, it is allocated to the town. This provides 272 names.

Next we choose the second or third name when the first (and second where applicable) name is already used or suggested for another town. This step allocates names to more than 900 towns.

Finally, using names not yet used nor suggested for other towns, the first and second names are concatenated to provide more than 1200 names to towns. This still leaves us with more than 500 towns without any name.

The remaining ones have been examined by comparing the location of the towns with a general topographic base map (mostly OpenStreetMap, occasionally ESRI basemaps). A name retrieved from these base maps has been manually allocated to the towns that did not have any name. During this examination process some names that had been allocated from LAU units or by geocoding have been revised if the base map suggested a more plausible name.

Conclusion

By using data from official sources and from voluntary geographic information we have been able to provide names for all grid-based towns. About half of the towns have received the name of the local administrative unit in which they are located. For the other towns the quality of the naming process mostly depends on the reliability and the representativeness of the OpenStreetMap data. In addition, the process using geocoding data relies upon the hierarchy of names suggested by the geocoding tool. Finally, manual allocation of names has been done by "best effort". Still, without field checks, local or regional knowledge or comparison with possible other geocoding results it remains hard to assess the final quality of the names that have been allocated to the towns.

The use of alternative national or regional data sources and/or alternative base maps might also help to improve the quality of the naming process, but currently such analysis goes beyond what we can reasonably envisage.

Despite these limitations all towns can now be referred to by means of a name, which makes explaining related data and indicators much easier. An interactive map viewer allows exploring and evaluating the naming of the towns and allows users to provide feedback and suggestions for improvement.

References

Dijkstra, L. and Poelman, H., 2014, A harmonised definition and rural areas: the new degree of urbanisation. Working paper 01/2014, European Commission, DG Regional and Urban Policy, Brussels, http://ec.europa.eu/regional_policy/en/information/publications/working-papers/2014/a-harmonised-definition-of-cities-and-rural-areas-the-new-degree-of-urbanisation

Dijkstra, L., Poelman, H. and Ackermans, L., 2019, Road transport performance in Europe. Introducing a new accessibility framework. Working paper 01/2019, European Commission, DG Regional and Urban Policy, Brussels, https://ec.europa.eu/regional_policy/en/information/publications/working-papers/2019/road-transport-performance-in-europe