

# Merging statistics and geospatial information, 2014 projects - Poland

Statistics Explained

This article forms part of Eurostat's statistical report on *Merging statistics and geospatial information: 2019 edition*

Final report 30 December 2016

## Problem

Three specific problem areas were identified, each of which was linked to combining statistical data and spatial data within the context of real world examples, highlighting areas where policymakers and decision makers could benefit from exploiting new forms of data. The first concerned an evaluation of changes in the number of school children and the demand for school places to assist local government planning. The second was a study analysing the distribution of economic activity in urban centres in relation to infrastructure developments such as multi-modal transport accessibility. The final study was designed to fill a gap by providing a tool to visualise statistical data using topographic data.

## Objectives

**Action 1** : this part of the project was designed to develop a methodology for studying and visualising the extent to which places in primary schools in 2014 reflect the need for such places (whether there is an excess of available school places or an insufficiency) and to project the situation for 2020 taking into account the demographic situation. In addition, teaching conditions were examined in terms of differences in the average number of pupils per teacher. Local governments face the challenge of adapting their network of schools and local infrastructures to changes in numbers of children while respecting binding standards.

**Action 2** : aimed to develop a methodology for analysing the accessibility of selected economic activity centres to multi-modal transport and the distribution of economic entities (hereafter called businesses) in relation to certain infrastructure elements.

**Action 3** : aimed to develop a methodology for assessing the suitability of a database for topographic objects for visualising statistical data and the production of spatial statistics.

## Method

**Action 1** : this action focused on data for the voivodship of Mazowieckie (the voivodship containing Warsaw). Poland has in recent years seen a reduction in its total number of births due to lower fertility and postponement of motherhood, directly resulting in a progressively smaller number of pupils in schools; in Mazowieckie a low point in the number of pupils was reached in 2010, after which the number started to increase again, at least in part influenced by the lowering of the age for the start of compulsory education. Demographic projections indicate that

the number of pupils will continue to rise until a peak is reached around 2020 after which it will decline again.

The work was done in two stages, a pilot stage for a narrower geographical area during which the methodology was tested, followed by the implementation of the methods for the whole voivodship. The pilot study involved defining assumptions and the data required, obtaining, correcting and merging the data, calculating indicators and analysing the results. The following stage followed a similar path to that for pilot study, with updated data and verification rather than development of the assumptions and methods.

The following data sources were used: data from the Ministry of National Education for pupils/students, teachers, schools, school rooms and primary-school identification data; town hall and municipality (gmina) data for district schools; social study frame data for population data. Whereas data for Warsaw was available electronically in a uniform structure, the data for rural areas was more varied in structure and format and required standardisation; data from various sources were merged and gaps filled. The information on the boundaries of school districts (school catchment areas) were sometimes out of date and so not all school districts could be linked to a complete set of addresses. A range of other issues were identified where information in various data sets did not match, such as outdated information where more than one municipality had a school district covering a particular locality, or outdated information concerning changes in the status of schools (moving from public to private management or simply closing).

The calculation of data for 2014 and the estimates for 2020 took account of a number of factors. To estimate the requirements in 2020 it was necessary to take into account changes planned for the starting age of education in the school year 2016/2017. Furthermore, based on 2014 data, a relationship was calculated between the number of children attending a given district school and the number of school-age children actually residing within this school's district. Another adjustment needed was for schools that had opened in 2014 or recent years: these schools had an atypical age structure of pupils as they mainly had children in the first school year(s), whereas by 2020 they could be expected to have children across all school years as each cohort moved on annually. In a similar way, schools whose district (catchment area) had changed were also treated separately.

As well as calculating the number of pupils and the availability of places, several indicators were compiled:

- number of classes (called branches in Polish) per room used for conducting lessons;
- number of pupils per teacher;
- percentage of children transported to school because their homes were too far from district schools.

Where these indicators showed atypical results, the underlying data were examined in detail to see if there were reasons for the results or possible errors.

The final step for this action was to produce a visualisation application for the indicators. The outlines of school districts were established for 2014 and 2020 by aggregating census districts to school districts; adjustments were made for known changes to school districts planned to have been implemented by 2020.

**Action 2** : the focus of this activity was on the cities of Białystok and Toruń which were target areas for a study of transportation. An analysis was conducted into the distribution of economic activity with reference to infrastructural and the multi-modal accessibility of transport (road, rail and tram-based). For analytical purposes, grid systems were used: assigning businesses to grid cells enabled information on these businesses and their characteristics (legal form, form of ownership, main type of activity) to be aggregated by location and also to relate them to the availability of infrastructure such as networks for transport, sewerage, water supply, district heating and gas distribution.

A preliminary step was to identify the functional areas for the two cities. The following indicators were developed:

- the share of hired labour commuting into Białystok and Toruń within the overall number of persons commuting to these cities from each gmina (municipality);
- the share of persons leaving particular gminas to work in Białystok and Toruń within the total number of people leaving those gminas for work;
- the share of de-registrations from Białystok and Toruń in the overall number of de-registrations from these cities by gmina of the present registration.

Based on this, two functional areas were identified, being composed of territorial units that had minimum values for all three indicators. This was done in such a way that the functional areas were spatially continuous, only containing

territorial units that bordered each other.

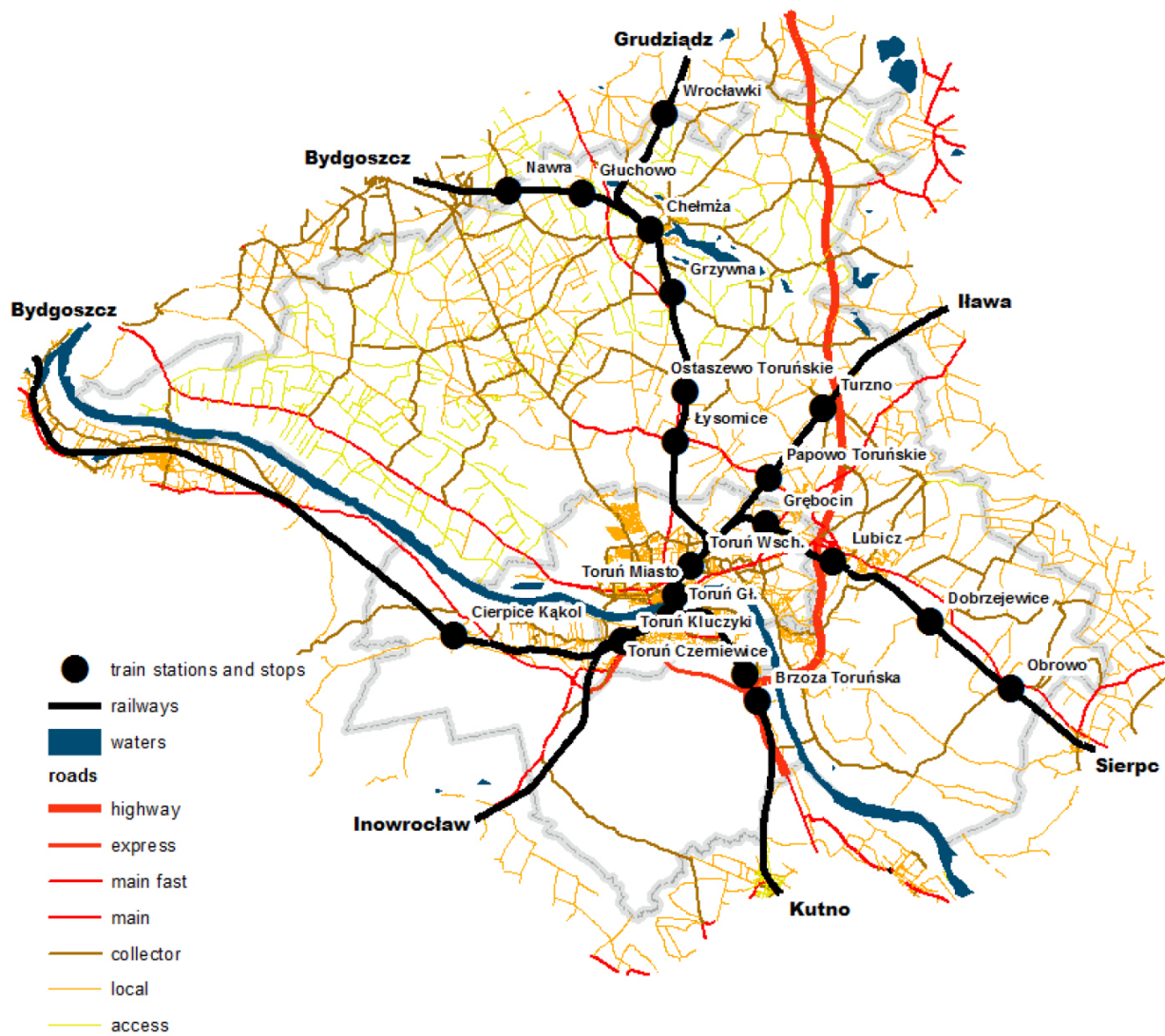
The source of vector data on technical infrastructure was the database of topographic objects (BDOT) from the mapping agency (the head office of geodesy and cartography). An orthophotomap and a vector layer of roads and streets (DRUL) used in work related to the maintenance of spatial address databases (as part of the TERYT register), were applied to evaluate the degree of land cover with different objects. Spatial data about businesses were sourced from the statistical unit database (BJS) for the two cities concerned. Public transport connections (buses and trains) to the cities of Toruń and Białystok from their respective functional areas were acquired from Blue Ocean Business Consulting.

Three different analyses were then performed.

The data from BDOT was converted for use in the software adopted for this project and evaluated. For several of the networks (such as district heating, water supply or sewerage) the networks shown in the data were found to be largely incomplete and further investigation showed that these only concerned over-ground sections of the networks. Other sources were investigated but all required payment for establishing datasets. From 2018, information on complete networks are expected to be available in a national database and will be available free of charge to the statistical office. The completeness of data concerning rail and tram networks was evaluated and it was established that the data were complete; in practice, the data for the tram network were not used. Road data from BDOT were also evaluated and used. The data from the statistical unit database were analysed and it was decided to exclude a small number of businesses, such as those under bankruptcy or liquidation proceedings, as well as those related to private partnerships. The distribution of businesses according to legal form, ownership form, size and economic activity was analysed. For the presentation of data five economic activities were selected: manufacturing; construction; distributive trades; transportation and storage; and professional, scientific and technical activities. To compare the information from different sources for roads and businesses, the linear data for roads were converted into density data (kilometres per km<sup>2</sup>) and the point data for businesses (and their characteristics such as levels of employment) were also converted into density data (ratios per km<sup>2</sup>).

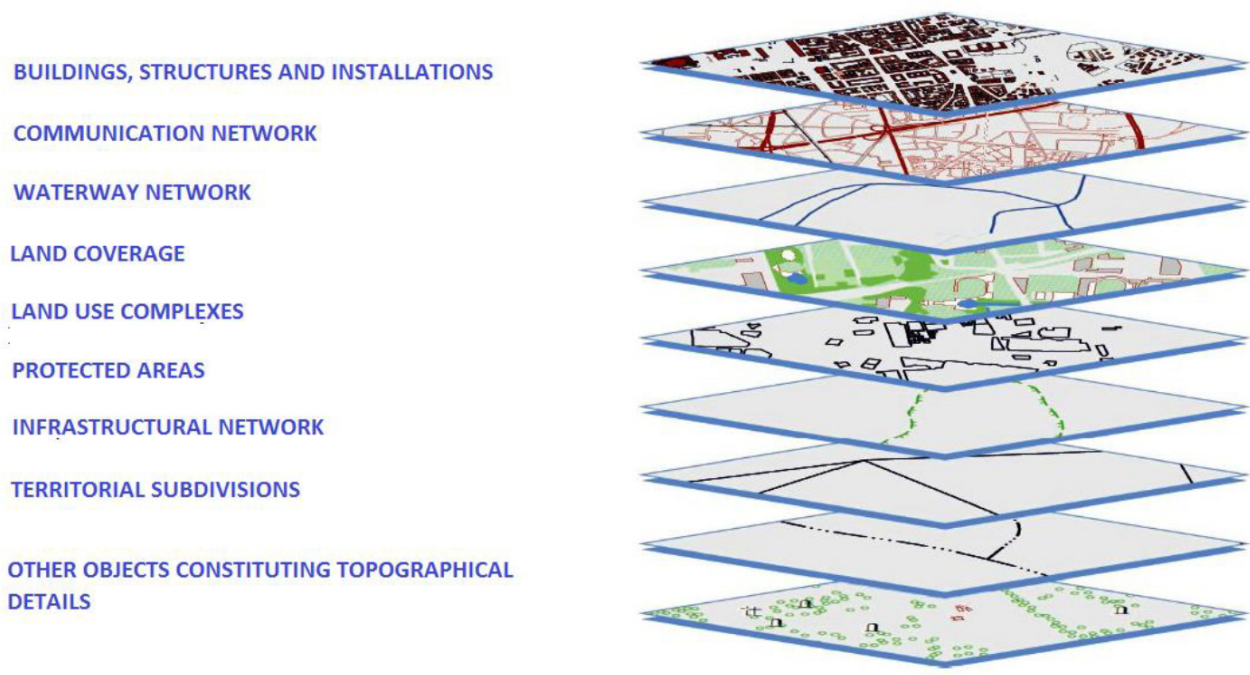
For the second analysis, which was based on transportation accessibility for bus and train connections, the focus was on access time, looking at time to and from the two cities as well as the number of connections between them. An analysis was made of bus and train timetables between (to and from) the two cities and the municipalities within their functional areas, looking at three two-hour intervals in the morning (between 4 and 10 a.m.) and in the afternoon/evening (between 2 and 8 p.m.). This analysis resulted in approximately 5 thousand connections (for the two cities combined), of which the vast majority were bus connections. For each locality within the functional areas, an index was then compiled of the average commuting time for each two-hour time interval, based on the total time for all connections and the total number of connections. These averages were then adjusted to allow for the closeness of the time intervals to the peak commuting time and also to allow for time intervals in which there were no connections. Finally, the index values were summed across all time intervals to produce an overall index for a locality. This index was compiled separately for buses and trains and as a joint index.

The third analysis focused just on the city of Toruń and analysed accessibility by road and by rail. The information from BDOT on rail tracks was filtered to exclude tram rails as well as rail tracks without passenger traffic, for example ones that were no longer used or were only used for freight. The remaining track information was supplemented by manually adding information on stations and stops. The railway tracks were then split up into segments between stations/stops and the travel time for each segment was added. The information from BDOT on roads was refined to include only usable road sections and again travel times were added to road segments based on the maximum speed limits and a weight for encountering traffic difficulties. The road and rail networks were then visualised for various road and rail classes.



**Figure 1: Visualisation of the transport network**

**Action 3** : this final action focused on data from the Database of Topographic Objects (BDOT), initially based on 2012 data and then updated with 2015 data. BDOT data covered nine categories of object classes: water networks (natural and manmade); communication (transport) networks; energy infrastructure networks; land cover; buildings, structures and devices; land use complexes; protected areas; territorial divisions; other objects.



**Figure 2: Visualisation of feature class categories in BDOT10k**

The second source studied in this final action was the land and building register (EGiB), an official register containing information on real estate. It contained information on land, buildings and premises and indicated (among other characteristics), the names and address of owners of real estate, the cadastral value of real estate, and information on lease arrangements. In vector form, the register did not completely cover all urban areas in the three voivodships or all of the rural areas in the 10 voivodships, while the descriptive part of the register was also incomplete.

The final source was orthophotomaps.

The first step was to perform an analysis of data quality in the BDOT, looking at data for single-family houses which formed part of the buildings, structures and devices class in the dataset. The focus was on the Mazowieckie Voivodship, in other words the Polish capital city region, the largest NUTS level 2 region in Poland in terms of area and population.

The vector map objects from the BDOT could only be compared on the basis of an orthophotomap, given the incomplete EGiB coverage. Evaluation was performed by way of selecting a sample comprising 10 % of grid squares. The sample selection process involved an earlier identification of the spatial diversification of residential buildings in the voivodship, followed by drawing fields divided into a few types of areas: city centres (6 % of the sample), suburban areas (47 % of the sample) and rural areas (47 % of the sample). Data were verified in terms of:

- completeness of objects entry;
- compliance of object boundaries with the orthophotomap;
- correctness of entered attribute values;
- correctness of the identification of objects.

The evaluation resulted in a report containing test field ranges and a list of errors.

Based on the initial assessment of the quality of the BDOT data, the second step was to compile a spatial analysis of data selected from the source. Single family houses were selected: for each of these the location of the centroid was identified and the area added as an attribute.

## Results

**Action 1** : on average in 2014, one in eight schools had less school places than children (who should be in education) residing in a given district. The largest proportion of such schools was recorded around Warsaw. Conversely, an excess of school places (compared with the number of pupils) was recorded for most schools, with some having several times more places than school-age children: these were mostly in the north-eastern and southern part of the Voivodship.

An apparently insufficient number of places for children residing in a school district does not mean that these school places were not available: in fact, a real insufficiency was recorded for only half the schools where the number of children residing in a given district exceeded the number of places. A number of factors underlie this, one of which is the fact that parents are not obliged to send their children to schools in districts where they reside.



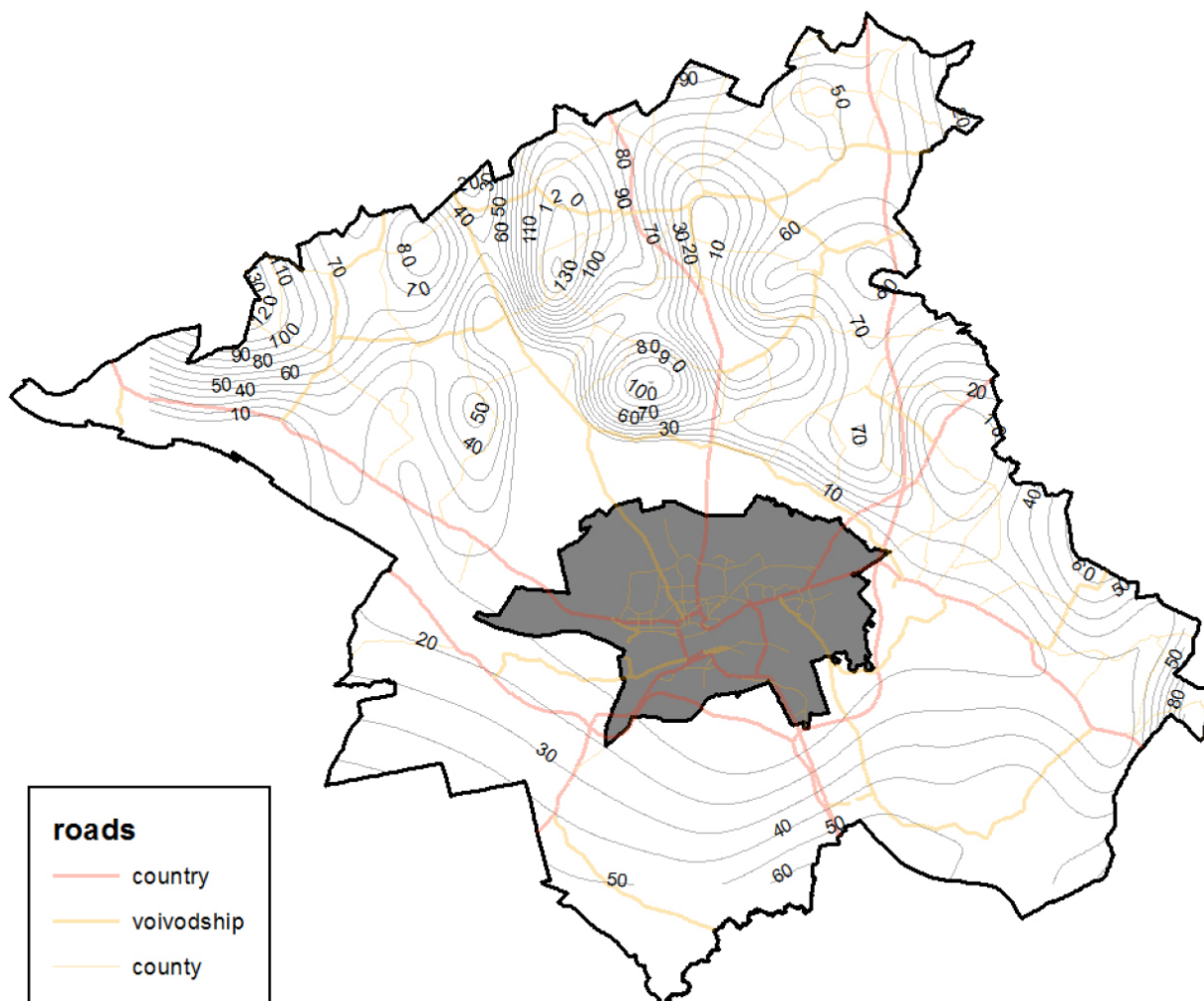
**Figure 3: School districts by type of gmina and actual availability of places in 2014**

A large proportion of schools have an excess number of places. In many rural gminas (municipalities) the number of pupils per teacher was very low. In extreme cases there were an average of 2-3 pupils per teacher, with the average for all schools in the study being 11 pupils per teacher. Closing schools in rural communities often results in pupils having to travel further to school: in some gminas 90 % of children had to commute to schools and in one case the proportion was 100 %.

Concerning projections for 2020, the number of schools with insufficient places for children residing in school districts is expected to increase, with the situation likely to be most severe in the central part of the Voivodship, as well as in the north-eastern gmina of Łyse, in which the closure of four primary schools is expected to affect teaching conditions adversely. Conversely, the percentage of schools with an excess number of places is expected to increase mainly in southern parts of the Voivodship. Pupil transport issues are likely to increase as more and more schools face closure.

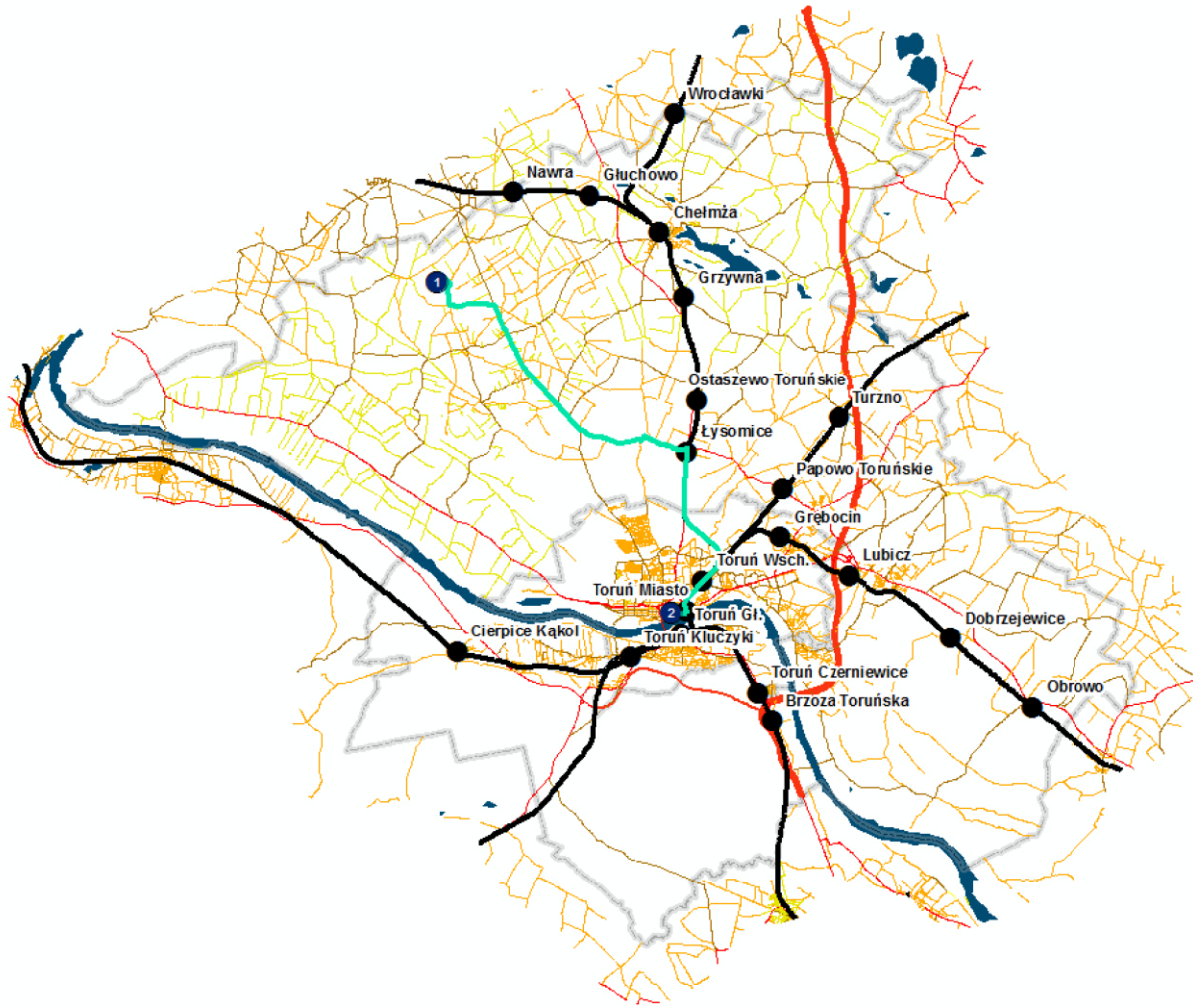
**Action 2** : the original intention when looking at business concentration was to compare this with the concentration of a range of networks, but in the end the only reliable and comprehensive comparison available was with the road network. Areas of the two cities were identified which had above average business density and areas with above average density for road networks and these were mapped, with various analyses for different types of businesses (by legal form, ownership form, size and economic activity). Overall there was little correlation between the location of businesses and the distribution of roads in either city although the mapping of the business data did show local variations depending on the different criteria used, particularly in Białystok. A further analysis was performed which identified where the higher density of businesses and the highest density of employment was in each of the two cities: while these concentrations (businesses and employment) were relatively close to each other in both cities — particularly in Toruń — they did not coincide exactly.

For accessibility by bus and train, maps were plotted showing either the road or rail network accompanied by isolines representing accessibility index values. These isolines were not simply concentric circles around the city centres, but were curved to reflect on one hand the location of major roads and rail lines and on the other hand the frequency of rail services at different times of the day. Accessibility indices were plotted for the morning (going into cities) and for the afternoon/evening (leaving cities), which also made it possible to contrast these two situations and thereby identify locations with relatively balanced or unbalanced accessibility in these two directions.



**Figure 4: The accessibility from Toruń to the localities within the functional area between 2.00 p.m. and 8.00 p.m. according to the bus transport accessibility index (roads)**

For the final analysis, where road and rail networks were combined with information related to the travel time for each segment of these networks, it was possible to identify the most efficient (in terms of time) routes between locations, for example between areas outside of the city into the city centre. When combining information for road and rail transport these identified the best of several nearby stations to drive to in order to take the train into the city in the shortest combined travelling time.

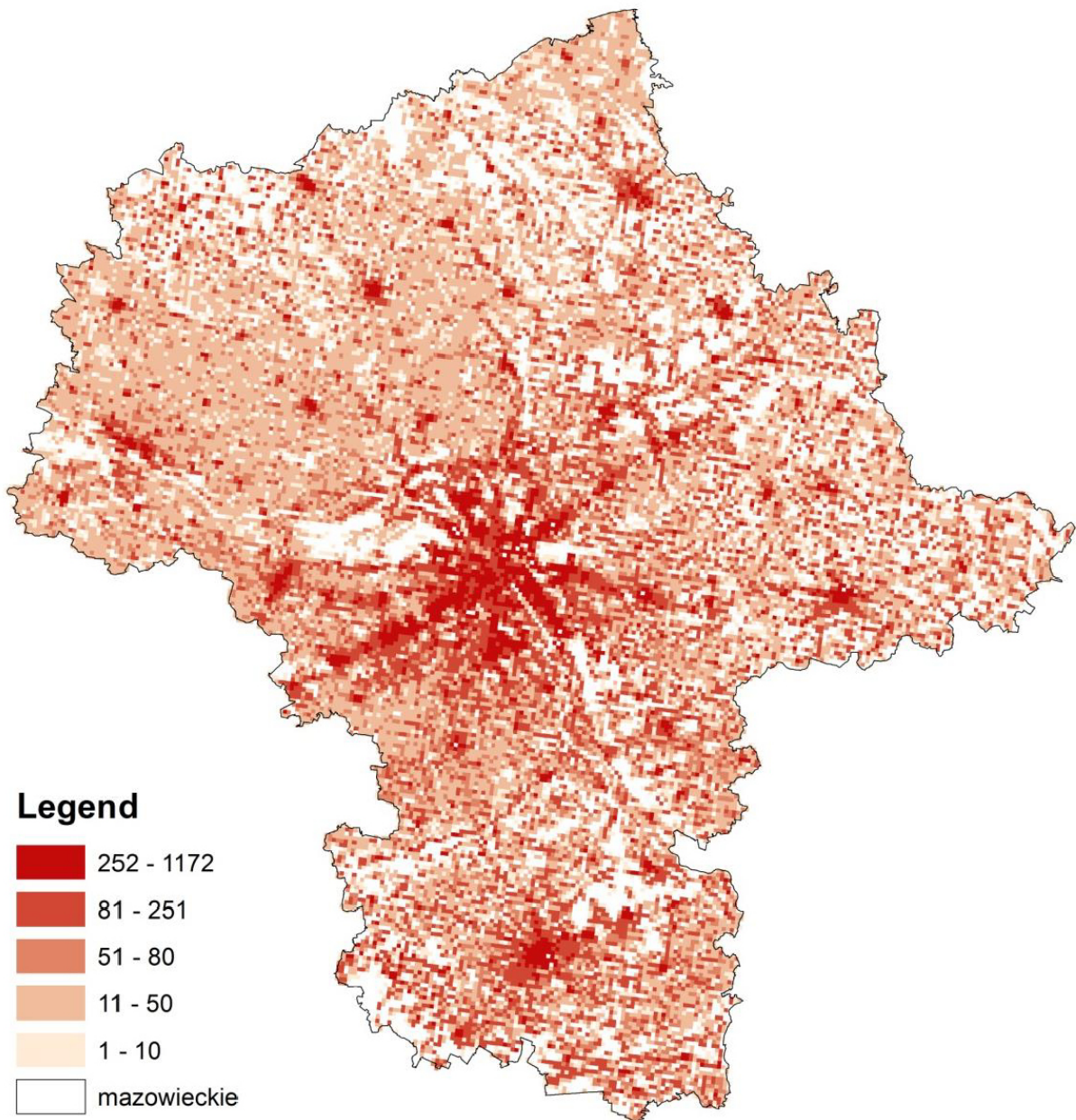


**Figure 5: The most efficient travel method to Toruń from the north-western part of the functional area (based on the network model)**

**Action 3** : no counterparts were found in the BDOT dataset for 68 objects (0.18 % of all identified objects) in city centres, 538 objects (0.45 %) in suburban areas and 202 objects (0.54%) in rural areas. Equally various attributes were examined, such as the number of storeys and the area and the proportion of errors was also regarded as low. It was concluded that the dataset fulfils the criterion of usefulness for statistical purposes.

Having prepared a dataset for single family houses the information on the number of houses and their area were summarised within grid cells and the results presented as density indicators within a choropleth map.





**Figure 6: Number of single-family houses/km2**

## Methodology

- [Final report](#)

## External links

- [Statistics Poland](#)
- [Mapping portal](#)

View this article online at [https://ec.europa.eu/eurostat/statistics-explained/index.php/Merging\\_statistics\\_and\\_geospatial\\_information,\\_2014\\_projects\\_-\\_Poland](https://ec.europa.eu/eurostat/statistics-explained/index.php/Merging_statistics_and_geospatial_information,_2014_projects_-_Poland)