Merging statistics and geospatial information, 2012 projects - Hungary

Statistics Explained

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

Final report 18 March 2014

Problem

Within the national statistical office (NSO) there were 20 disparate and unrelated databases containing spatial data.

Objectives

The aim of the project was to connect this disparate set of statistical data and spatial data, combining information from the business register and population statistics through an address register and address directory to produce a geocoded data set to be used for creating maps.

Method

The business register in the HCSO was derived from a variety of mainly administrative sources, including the Office of Government Issued Documents, the Registry Court, the tax office, the Treasury, or register questionnaires.

An address register existed within the Hungarian central statistical office (HCSO) containing valid and approved addresses in a hierarchical structure and format: settlement, area, public space, real estate (house number, lot number), parcel (building, staircase, level, door number). However, its initial purpose was to support the conduct of population surveys rather than to have a complete register of the population. Alongside this, the HCSO also had an address directory which had other — non-approved — addresses. This information came in an unstructured and non-harmonised textual format from a variety of (mainly administrative) sources.

The central element of the geostatistics system developed by HCSO is an address directory, which identifies and manages addresses in the approved register, as well as addresses that have yet to be approved. The addresses used in business register and other satellite registers were adapted from textual addresses to a system of address identifiers that were linked to the address register or the address directory. These address identifiers are, in turn, linked to geocodes. By doing so, the HCSO has built-up a completely new system which provides the possibility to seamlessly merge geospatial and statistical information, by assigning geographical coordinates to addresses that are linked to statistical units.

Geocodes were available for the address list used for the census, with geocodes stored at the level of house numbers in the address register. Geocodes for new addresses and changes to addresses were bought from a Hungarian map company. For those addresses that were not approved, but stored in the address directory, a database function was created to identify the closest address from the address register.

This work allowed the HCSO to implement a uniform address management system across various registers, whereby the identity of addresses was established and a geographic code assigned. The functions of the address directory were modified so that it could manage any address coming from any register. This was done by: breaking down addresses into standardised elements; devising systems to correct syntax, synonyms and errors; establishing connections to addresses already verified and included in the address register; devising a system to approve addresses not yet included in the address register.

At the end of the process, registers no longer contained addresses but rather address identifiers, to which the textual addresses were attached from an address book. This allows much easier control of addresses in terms of any maintenance and also means that a wide range of data may be visualised by drawing on the geostatistics from this harmonised system.

Results

A geostatistical system has been developed linking information from business, population and social statistics to the address register or address directory, which in turn is linked through a system of geocodes to a map-making functionality.
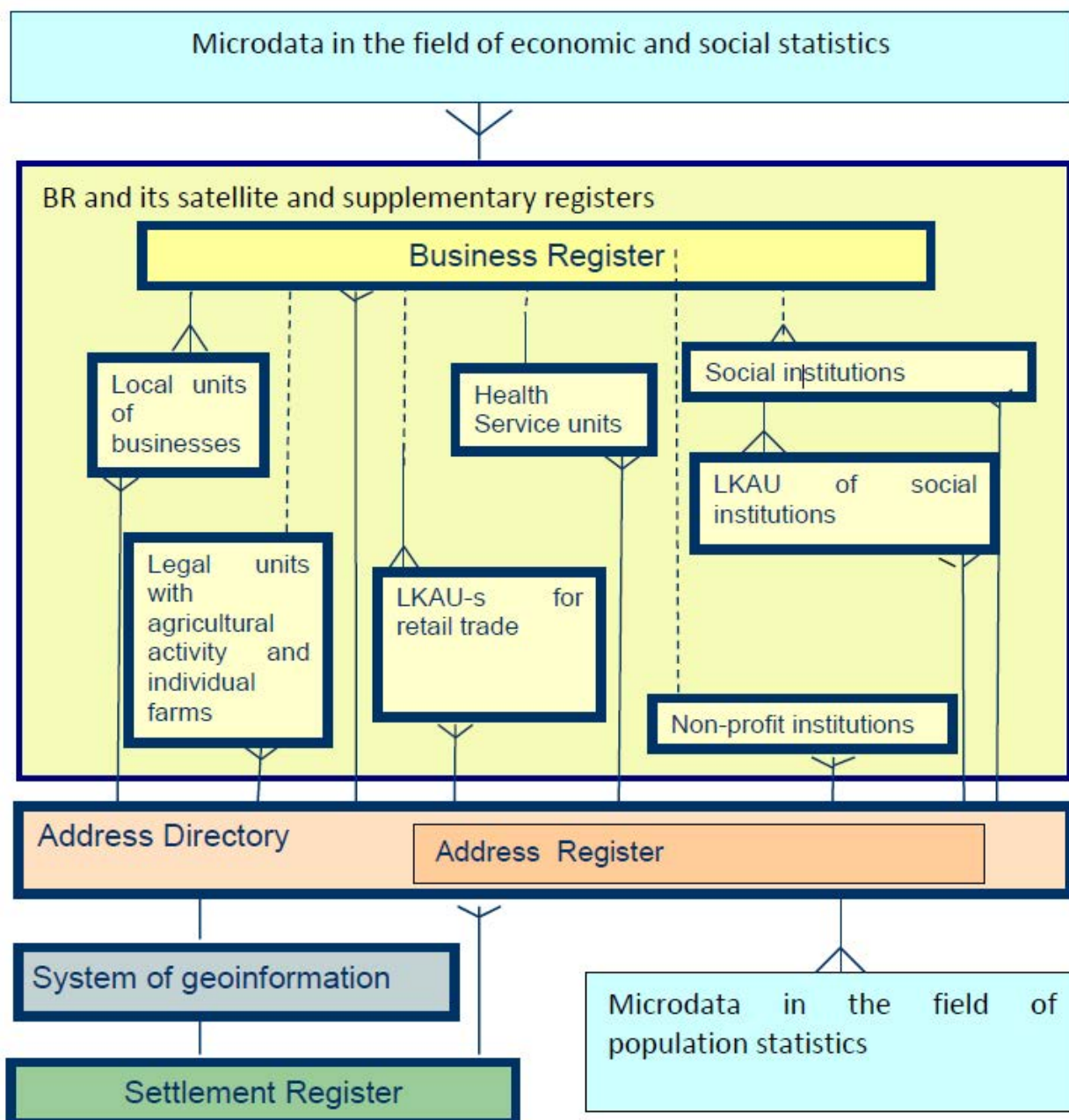
Figure 1: Example of the geostatistical system created within the Hungarian Central Statistical Office

A web interface was developed whereby users may choose from a number of broad statistical themes (such as population, social, economic or environmental statistics), then select one of the available indicators for that theme and one of the (up to four) levels of territorial typologies within the country (regions, counties, districts or settlements). The resulting map can then be personalised by selecting one of four criteria for determining the class boundaries and colour scheme. Users have standard navigation tools (zoom in/out and pan) as well as being able to mouse over any polygon of a statistical division in order to view the code and name of the area and its value for the selected indicator. For all except the most detailed level (settlements), clicking on a specific polygon generates a data tabulation for all areas at the same level of detail.
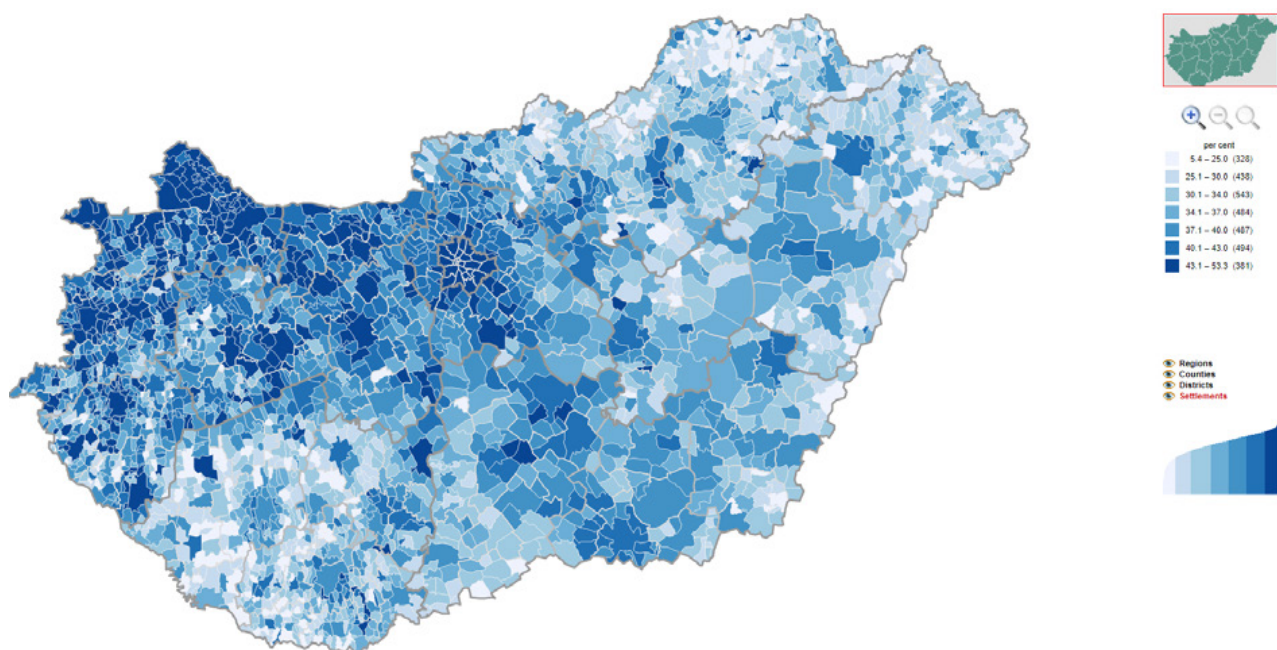
Figure 2: Example of the web interface developed by the HCSO showing results of the 2011 census detailing employed persons as a share of the resident population, by settlement

Furthermore, three sets of metadata were developed, one concerning the geocodes (source, data and accuracy), one concerning the source maps (name, source, date, accuracy, other characteristics) and one concerning thematic maps (source map, data source, date, method of creation, INSPIRE and ISO 19115 themes, terms of usage).

Methodology

- Final report
- Presentation
- Mapping portal

External links

- Hungarian Central Statistical Office