

DRŽAVNI ZAVOD ZA STATISTIKU
CROATIAN BUREAU OF STATISTICS

Geospatial data on enterprises – challenges to geolocate enterprises and their local units for statistical purposes

Zrinka Pavlović

Head of Statistical Business Register, Classifications,
Sampling, Statistical Methods and Analyses Department

INSPIRE Conference 2016 Barcelona



Enterprises VS. people



- **Lifecycle**
 - Birth, existence, death – demography/business demography statistics
 - Enterprises can be transformed through the years
 - “ same enterprise or not?”
- **Location**
 - Residence/ headquarter address
 - Enterprises can operate in many locations – local units
- **Both in focus of statistics**





Enterprises VS. teritory



- Allocation of enterprises within the country
 - Cities/parts of cities
 - Territorial / climate / relief characteristics
 - Other
- Is “doing business” easier in some parts of country?
- Is more enterprises born, survive longer, grow faster or steadier or die faster in some parts of country?
- Are enterprises more successful in some parts of country?
- Which activities are present or missing in some areas?
- Why?
- Should government put more effort and assets in improving conditions for “doing business” in some areas to stop depopulation?



Statistical observation of enterprises



- Statistics can give a lot of information about business population related to location (number of units, demography, density, activities, employment, efficiency, etc.)
- Precondition :
 - Accurate
 - Comprehensive
 - Good coverage
 - Relevant
 - Good quality

Business register

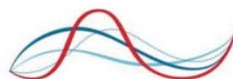


Legal background

- Commission Regulation 177/2008 requires that Business Register must contain geographical location code on local units

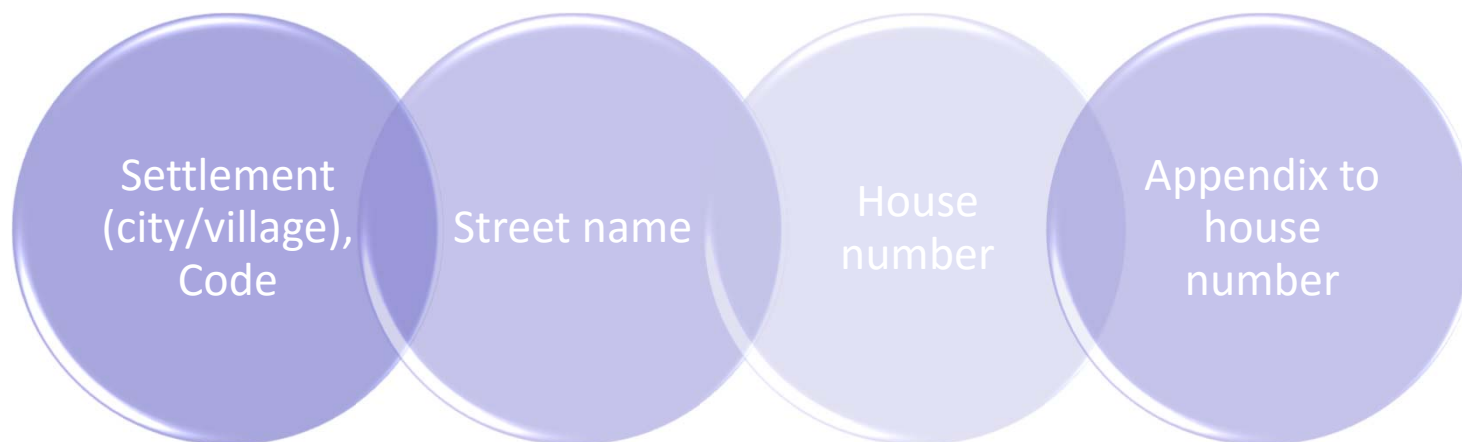
2.11 Geographical location code

Purpose: The geographical location code complements the address and postal codes (2.2) and can be used to derive classifications relating to the geographical location of units at the most detailed level. Other national classifications such as administrative regions, travel-to-work areas, health or education regions etc. can also be derived from it.



Geolocation of enterprises

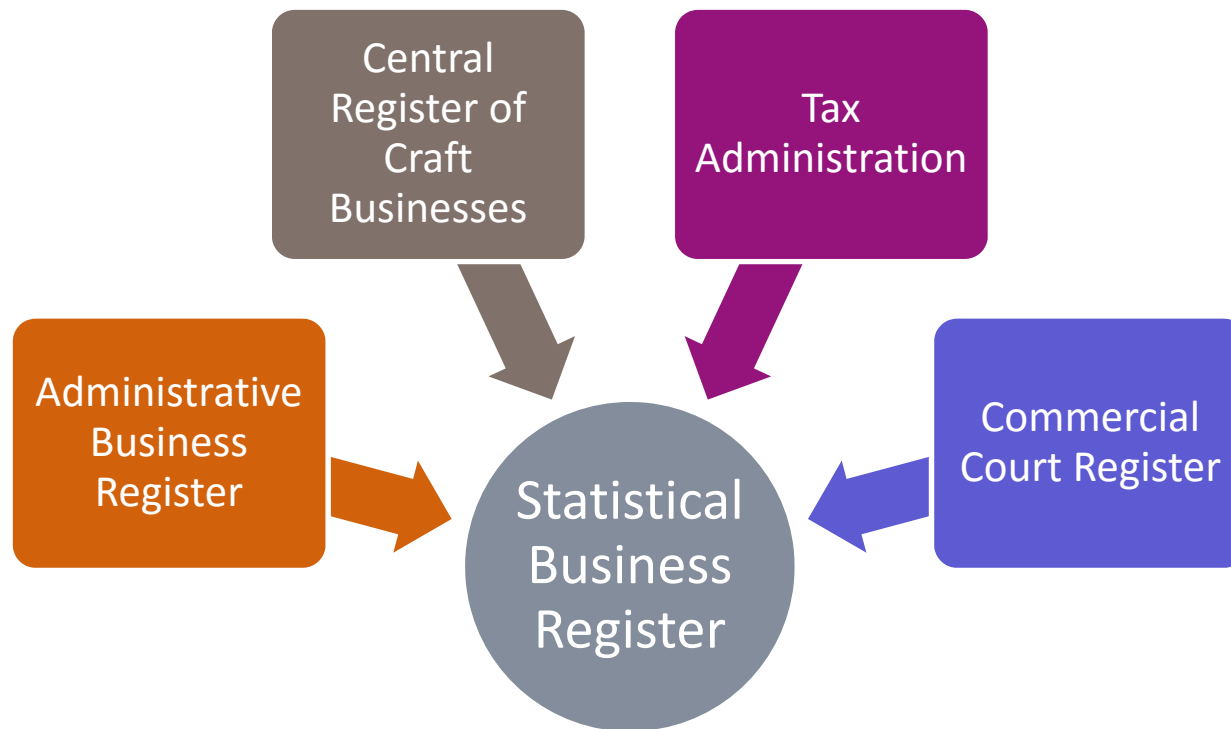
Address data consist of:



Only settlement code is unique and official and used by majority of administrative sources. Other parts of the address are in free text form

Sources of address data

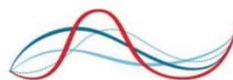
- Statistical Business Register is compiled from several sources which provide address data :





Sources of spatial data

- Central Register of Territorial Units – State Geodetic Administration (SGA) in the Republic of Croatia
- Statistical Spatial Register in Croatian Bureau of Statistics is updated from SGA register
- Not all administrative registers and administrative records are directly connected or updated with SGA register data





Project – merging statistics and geospatial information

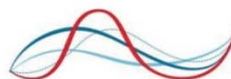
- Goal:
Geolocate enterprises and their local units by assigning them geocode
- Activities:
 - Establish system that will enable every new address entry in SBR to get a unique official street code, street name and geocode.
 - Develop application that will automatically or by clerk intervention match new address entries with official addresses
 - Coding existing addresses in Statistical Business Register
 - Publish geographical presentation of selected SBR data on the CBS website



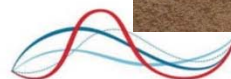
Starting point

- SBR contains addresses for enterprises and their local units that had to be coded:
 - More than 500.000 enterprises (regardless of activity status)
 - More than 600.000 local units (98.000 <> legal unit address)
 - More than 1.100.000 address information
 - 109.215 different forms of street names within certain settlements
 - 68.863 different forms of street names regardless of settlement

- In the Statistical Spatial Register:
 - 51.798 streets in 6759 settlements



Existing situation



Existing situation

- Street names in Statistical Business Register are stored in the version as in administrative sources – not official name of the street as in Register of Territorial Units. E.g. :

Street code	Street name	Sett.code	Sett. name
0721501371	ULICA KNEZA BRANIMIRA	072150	Zagreb

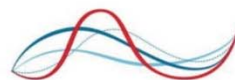
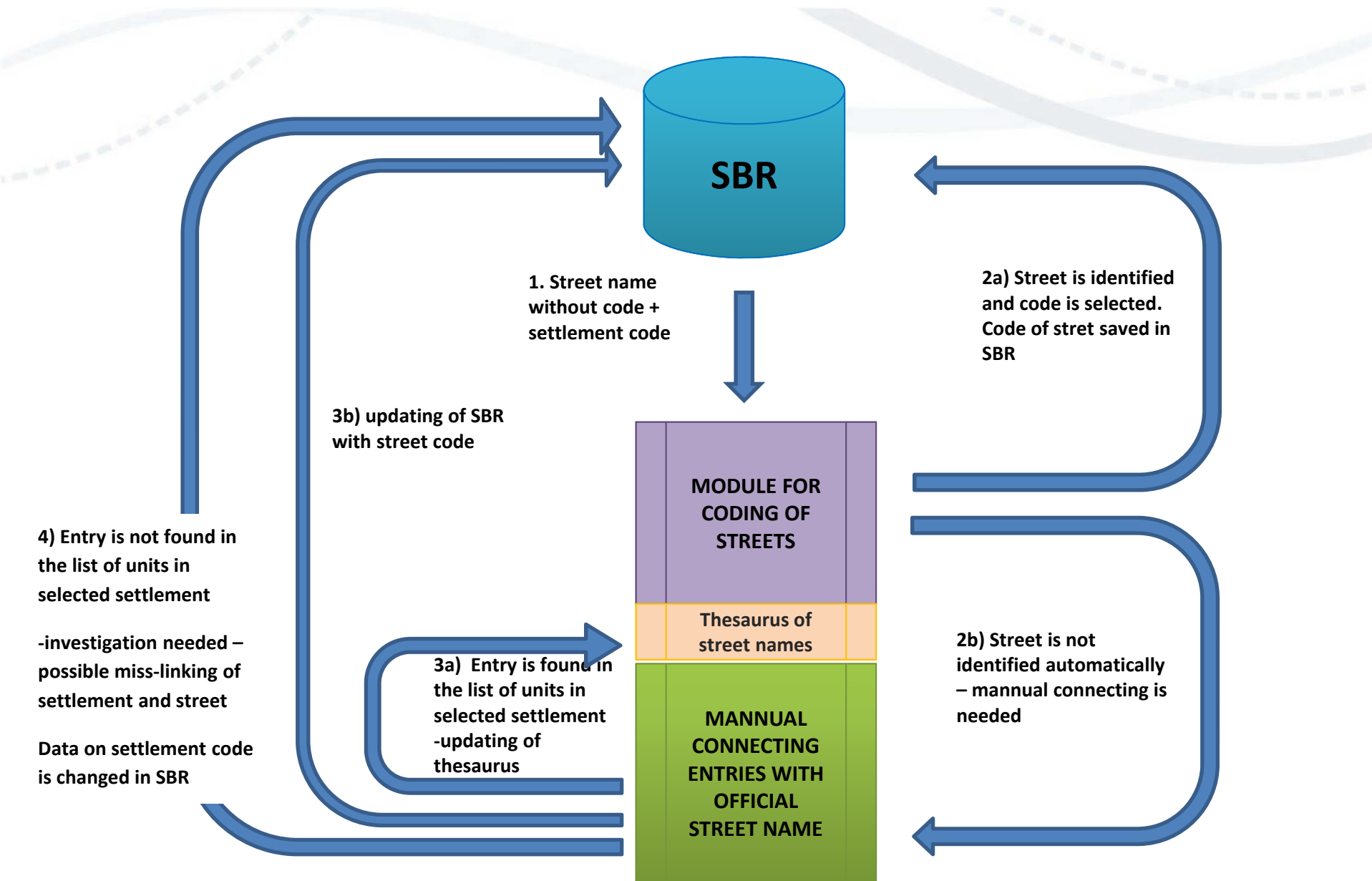
- Many variations of the same street name:

Street:	<input type="text" value="Kneza Branimira"/>	Street:	<input type="text" value="K.BRANIMIRA"/>
House number:	<input type="text" value="20"/> Appendix: <input type="text" value="/1"/>	House number:	<input type="text" value="28"/> Appendix: <input type="text"/>
Street:	<input type="text" value="Branimirova"/>	Street:	<input type="text" value="Nova Branimirova-Retkovec"/>
House number:	<input type="text" value="60"/> Appendix: <input type="text"/>	House number:	<input type="text"/> Appendix: <input type="text"/>

Conditions:

- In order to assign code to these forms of name of the same street in the same town, all this forms should be stored in the Thesaurus.
- Matching of names must be 100% because **one letter** means difference.

ULICA	MARIJA	BABIĆA	Ulica M. Babića
ULICA	NENADA	BABIĆA	Ulica N. Babića
ULICA	MATIJE	BAKIĆA	Ulica M. Bakića
ULICA	VOJINA	BAKIĆA	Ulica V. Bakića



Challenges faced during project:



- Complex spatial situation with many incorrect address entries:

- Streets that does not belong to settlement as registered

- St



- Ho

ter

- Various
appear

- St

- Ho

- Appendices are separated with different characters





Challenges faced during project (cont.)

- House numbers and appendices in different format , non – existing house numbers in the Register of Territorial Units
- Non existing unique and comprehensive data base of historical street names
- The most difficult cases – units that are no more active
- Too many clerical work to investigate all problematic street names = focus on active units

Way forward.....



- Normalising of data from a both data bases needed (removing special characters, double spaces, normalising usual abbreviations and some very common street titulars).
- Automatic matching at the beginning of the project: a little bit above 40 % for unique street names with little possibility to increase automatic procedure since only 100 % match of cleaned data counts (one character might mean difference).
- Clerical matching the rest of non-coded streets
- Thesaurus filled with normalised data and pairs of street names (Stat. Business Register and official street name from Stat. Spatial Register) and consecutively supplemented with new manual entries
- Many dead units with non existing addresses – disregarded
- Searching for files with changed street names in several institutions (in larger cities)





Results:

- IT Application developed for coding streets with thesaurus of streets
 - Stand-alone application consisting of three modules:
 - Street data-base from Statistical Spatial Register
 - Module for matching street names that enables:
 - Transmission of non-coded streets from the source data-base
 - Automatic coding of streets
 - Module for manual matching of street names
 - Thesaurus of street names
- 98% Addresses of active enterprises coded with street codes



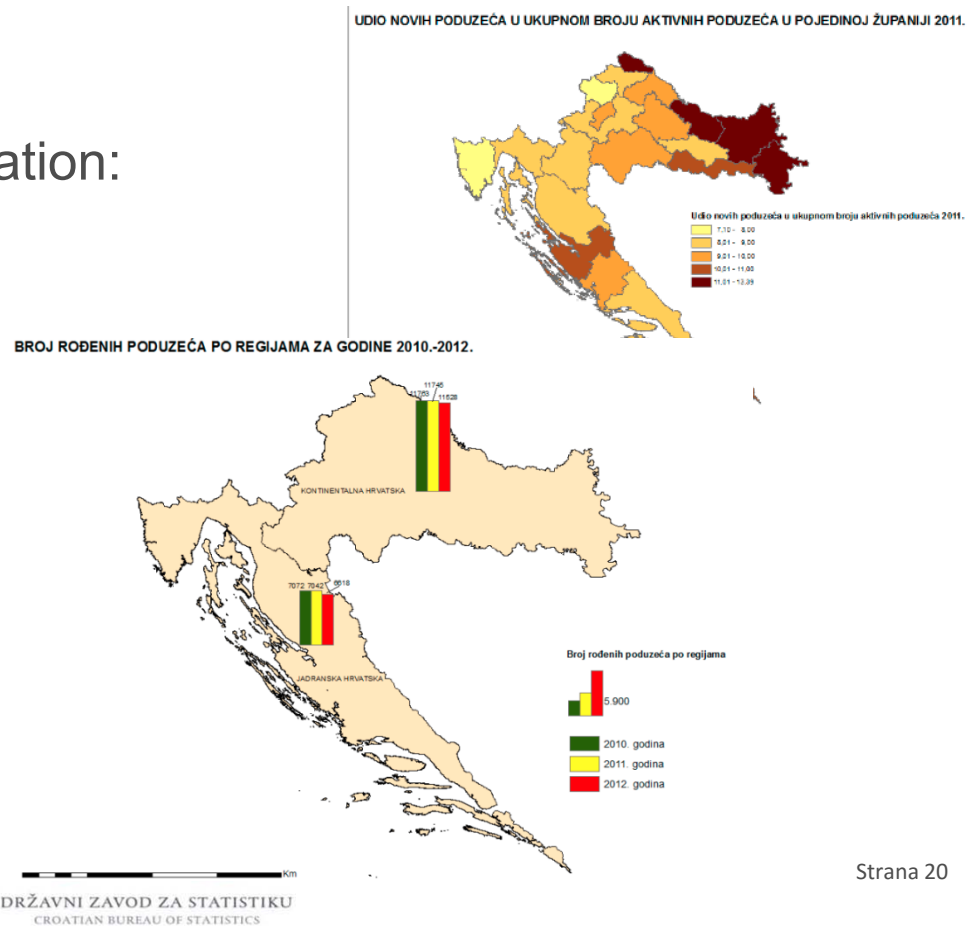


Results (cont.):

- Automatic procedure developed for assigning geocodes from Statistical Spatial Register data base, based on street code and house number – by the moment of entering in the SBR data base
 - Batch updating
 - Manual updating
- Upgraded Statistical Business Register with additional attributes

Geographical presentation of Business Register Data

- Business Demography Statistics convenient for geographical presentation
- Available tools for presentation:
 - ArcMap 10
 - Geostat Portal of CBS





Lessons learned

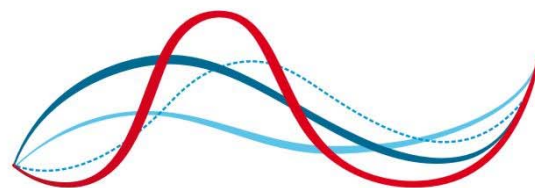
- Addresses of enterprises are in many cases problematic – identified main types of mistakes – new system enables detecting mistakes and gives opportunity to correct them
- Accuracy of spatial data varies between counties
- Spatial registers need to be improved in following years, historical data needed.
- Administrative sources should be contacted in order to put much more focus on accuracy of their address data (link of settlement and street) – connection to official Register of territorial units

Potential reuse



- Module for coding streets with thesaurus and functionality for manual matching and supplementing content of thesaurus was developed as separate application
- Envisaged use for other users in CBS – e.g. Farm Register, Population Register, Population Census
- Possible availability for other institutions – administrative sources of SBR → in-coming data of better quality?





DRŽAVNI ZAVOD ZA STATISTIKU
CROATIAN BUREAU OF STATISTICS

Thank you for your attention!

Zrinka Pavlović

Head of Statistical Business Register, Classifications, Sampling,
Statistical Methods and Analyses Department

Tel: +385 (0)1 4893 514
E-mail: pavlovicz@dzs.hr