STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Working Paper No. 23
English only

Work Session on Geographical Information Systems
(Brighton, United Kingdom, 22-25 September 1997)

Item (v) of the provisional agenda

GEOMATICS, POSSIBLE SUPPORT FOR NATIONAL STATISTICAL ANALYSIS AND PUBLICITY

Submitted by the Statistical Office of the Republic of Slovenia [1]

[1] Prepared by Tomas Banovec.

I. ABSTRACT

1. The paper presents the idea of setting up georeferenced registers and national statistics in Slovenia. It shows the possible use of some analytical methods of georeferenced data. It also presents the method for setting up a statistical georeferenced database (statistical register), as well as calling for solutions and assistance in developing analytical algorithms for such purposes as elections and definition of electoral districts, optimisation of service locality with gravity analysis, data transmission with the help of localities, and so on. There are sufficiently detailed data in Slovenia, but the use and development of geoanalytical and geostatistical methods are not well co-ordinated. The paper also gives some suggestions and elements of future development in this field.

II. SOURCES AND USE OF GEOREFERENCED DATA

2. The Statistical Office of the Republic Slovenia has at its disposal practically all administrative and other registers and their contents, which are geocoded with the help of centroids of houses from the Register of House Numbers (EHIŠ). These registers are additionally defined with outlines of territorial units, which determine appurtenance of individual geocoded buildings (EHIŠ) to these units.

3. The Slovenian model of geographic divisions is made according to European Union NUTS (**Nomenclature des unites territorielles pour statistique**). The basic idea is that every building is adequately defined with its co-ordinates (centroid), and that all contents of administrative and other registers are linked to this location. So it should be possible for the purposes of analyses and dissemination to take over entities and their attributes at the NUTS 3 level from the NUTS 9 level (centroids), and indirect aggregation at levels 8, 7, 6, 5 and 4 will not be necessary.

4. **Uniform location definition of buildings** (ELON) is needed if we want to avoid the overlapping hierarchies of "superior" units and monitor the code system. ELON enables analyses and use of georeferenced data collected in various ways.

5. The ELON system can be demonstrated by the following example: the author of this paper is 57 years old. During his lifetime he lived in five different states but always in the same building. Territories of municipalities, their names, size and streets' names changed with these states and regulations. And most probably there will be more changes. In addition, names and definitions of streets and house numbers changed several times. Only the building with its real location and ground plan remained the same.

6. The building is defined in various maps and cadastres based on its geodetical ground plan. This definition includes all its physical, demographic and economic characteristics. The building is also defined in registers, censuses, etc. In the above-mentioned case, the descriptive postal address in various records changed but the **position of the building did not**. By introducing **centroids or "geomatics"** for identification of buildings, we can solve the problem also for "parastatistics", since administrative registers are used by several institutions which need unambiguous **identification and location of the building**.

7. Georeferencing of inhabited buildings only for census purposes is too expensive. Therefore, it is very important to maintain registers on these objects and to use data from administrative registers for statistical purposes.

8. Our georeferences are defined by X and Y co-ordinates with an accuracy of one metre. After digitalisation on the map, this point will remain within the ground plan. This precision permits the addition of the definition of the height above sea level from the digital model of

heights (co-ordinate Z) and other operations. It also permits the linking of image elements from satellites whose accuracy is close to one metre on the ground, even for civil use.

9. We follow the recommendation on geomatics and geocoding adopted by the CES in 1994, i.e. uniform location definition of buildings with the co-ordinate method or ELON. In Slovenia the method was proposed at the end of the 1970s and introduced in the 1980s with the agreement between the National Surveying and Mapping Authority and the Statistical Office. The Surveying and Mapping Authority continues work on centroids of land plots.

III STATISTICAL TASKS ON GEOMATICS

III.1 Formation of territorial units

10. For statistics purposes, on the territory (e.g. the state) with total coverage of data (voters, persons entitled to assistance, etc.), it is necessary to group concrete entities (e.g. buildings with voters) so that each new unit will have the value of monitored attributes defined in advance. An example of this task is that sum values of attributes will be equal in all newly defined units. We should also consider the permitted deviations (in %), other conditions and selected absolute limitations and provisions.

11. Compulsory conditions were respected in this process. Buildings (centroids) and their demographic contents should be grouped into territorial units so that:

a) each of the newly defined areas will have equal value of the required attribute (voters for the two-round majority electoral system, etc.);
b) each building (centroid) will be only in one of the new units;
c) each entity (inhabitant, voter, citizen in the CRP, insured person, etc.) will thus belong to one of the new units;
d) there will be no entity which would not belong in any unit;
e) these units will be delimited with the co-ordinate method, with polygons detailed to the level of concrete buildings (EHIŠ, centroid);
f) there will be no part of the territory which would not belong to one and only one new unit;
g) delimitations will be defined to link centroids of two or more different buildings;
h) the polygon outlining each new unit will be territorially coherent so that there will be no enclaves of other units inside the territory of the concrete unit.

12. The following additional recommendations were made:

a) delimitation of contents so that each unit includes the maximum of complete settlements;
b) the derived inside breakdown of local communities at the level of local, village and urban communities should be taken into consideration;
c) the whole territory of each municipality should be included into one unit.
d) the system of administrative units and their branch offices should be taken into consideration;
e) other regional breakdowns (e.g. electoral districts) should be considered if needed.
f) state borders should be set as a polygon and this outline should be considered to be definite;
g) relief, geomorphologic and landscape characteristics of the country should be taken into consideration;

h) the possibility to link polygons of new units through varying size and accuracy of value attributes for which the territory (polygons) is defined - however, each unit can only vary by 5%.

13. Computers should be used for the design and implementation of this data model. In this way we should obtain a data model with the possibility of dynamic changing of parameters (n - units, voters, lists of children, etc.).

III.2 From the already defined or presumed territorial division to the analysis of operation of existing territorial units

14. The existing territorial units have their own managers (municipalities), which very often do not agree with their statistically measured situations. Therefore, it is essential to reach agreement on reliability and applicability of collected indicators for such units.

15. Revision of old and formation of new territorial units. We have formed several territorial administrative units but we still have to form many new units for new districts, provinces, NUTS-3 units, new municipalities and other territorial breakdowns. All this should be supported with data in an independent and comprehensible way. We have geocoded entities (EHIŠ with centroids) which can form functionally and territorially enclosed areas. This analytical tool is essential in support of the decision to form districts and provinces or NUTS-3 units. Not only state and public services but also marketing can thus be optimised (trade, supply, sale of newspapers and magazines, distribution of various materials, purchase power, services for cars and dwellings, municipal services and so on).

16. **Optimisation of state and public services**. The number and size of units, functions, applications and data necessary for optimisation are derived from users' needs. To this end, a **common multipurpose database containing various data** linked with geocoded buildings is needed, also in the case of national statistics with very small territorial units (census districts and aggregated as low levels, territorially broken down data).

17. As mentioned, we monitor and analyse territorially defined units equipped with adequate attributes for various functions. It is essential that rational outlines and optimum satisfaction of individual and common functions and services be defined directly as regards location (co-ordinates) with isochrones (isolines in general) or with similar scanning patterns such as satellite images and photointerpretation, and that this be modelled. We need to develop criteria for optimisation.

18. **Analysis of functions: territorial aspect**. For territorial analysis we should first define the basic functions and expected output for the state. Furthermore, positions of centres in the model and areas of influence in which these centres should optimally perform their services should be specified.

19. **Rationality of computerized public and state services**. Two tasks should mentioned in this context: a) to set up an a common aggregated database for already existing territorial units and prepare the remodelling of their functions (competence of municipalities, etc.); and b) to collect relevant micro and macro data which enable decision- making in each of the territorial changes (the state and its public competence).

III.3 Interpolation and imputation of attributes from the neighbouring units to the new unit

20. To analyse phenomena at a concrete location (centroid of a building), one sometimes has to interpolate or impute using similar conditions in neighbouring units as a basis. With the help of algorithms, we can measure imputed attributes for our own and neighbouring points.

21. For example, long-time georeferenced monitoring of turnover value of land for sale, whereby some transactions are geocoded (building, land plot, etc.), enables the setting up of the network of such points. This can be used for drawing conclusions on how much other neighbouring real estate, with similar or other characteristics, is worth. Interpolation and imputation of value for individual land and other real estate entities requires an adequate **interpolation and imputation model** as was proposed by OECD experts in December 1995.

22. We use OECD recommendations for evaluation of land and dwellings. Such a method is needed for monitoring of sales value of real estate. Our experience from the past showed that it is relatively easy to transfer territorially homogenous data on real estate if they are adequately prepared for such work. Modern GIS tools can do this very quickly. However, there is a problem with evaluation of imputed attributes. Various systems mixing dynamic economic and market indicators with raw data have a very limited value, although they are frequently used. Exchange of experiences with other more experienced countries would be highly desirable.

III.4 Gravity (momentum) analysis on the example of territory

23. Some phenomena have already been measured and analysed using this method. For example, from the census of employed persons we calculated the shift of employment in Slovenia on the basis of certain centroids of settlements for 1991 census and later statistics of employment. The centre of employment moved to Trbovlje and did not change much over the years (200 m), moving gradually towards the desired centre of less developed municipalities in the direction of Ljubljana. Of course, such national analysis is merely informative, but could be suitable for areas of operation of employment offices or for future NUTS 3 and 4. Then data could be aggregated via smaller units (settlements, municipalities).

24. Gravity analysis on the basis of geocoding can be performed for the whole country, for other tasks and regions and for various purposes. It is interesting when starting points (centres of gravity) for various phenomena are defined with co-ordinates on the territory by different contents and time. Gravity analysis is one of the simplest forms of momentum analysis.

25. Some interesting points with regard to this are, for example, comparison of gravity of a certain phenomenon and its movement on the territory and in time. Where is the gravity of all dwellings and residents in Ljubljana going and how fast? Where, on a certain territory, is the central point of fuel requirements for agricultural mechanisation, i.e. is it close to the existing gas station or not? Where will we place a certain public service if potential users of this service are already known or presumed and, of course, geolocated? Is the existing service (centre) located in the right place? Where is the demographic and where is the employment centre of the state, region, municipality of smaller unit? Where is the centre or momentum of all agricultural land of a certain agricultural holding and is the farm close to this centre? The same question concerns the territorial distribution of insurance and contacts with banks' clients, in payment and financial services in general. Where is the centre of population in a region, is the school located in the centre or at least close to the centre of residence of all school-age children? What about the centre of land owners and the geodesic office, tax office and centre of taxpayers.

26. Gravity analyses are fairly simple and are suitable for various purposes; they can be used for checking data but can also be incomplete and can lead to misinterpretation. This method would be applicable in defining the centres of supply of state services. Of course, the question of momentum analysis is a wider one. It was used already by town planners for

Lovry's model in the 1970s for defining supply functions where they used the model of the momentum of inertia.

III.5 Strategies of geocoding at local, regional and state level

27. Recording data on entities at a certain level is an important prerequisite for decision-making at this level. If a local community has decision-making tasks, it also needs to create an adequate **land information system**. The need for data is not defined only by the territory but also by other requirements, depending on which activities on this territory are important to the manager. This can be seen in Article 21 of the Law on Supplements of the Law on Local Communities. Other examples of this task are special protected areas, such as those frequently destroyed by floods, environmental protection areas, areas of possible influence on the nuclear power plant, etc.. The authorities in these areas are very often responsible for monitoring these entities in detail.

28. When geocoding at the local community level, we should draw attention to the basic technical structure of the geolocation:

- of individual sources - **point influence** (transmitters and their radiation and signals, chimneys and gases) on the location of buildings and everything defined as NUTS 10;
- of predominantly **linear objects** (roads and distance from them, state border and the ban on sale of real estate to foreigners closer than 10 km from the state border or the ban on growing vegetables closer than 200 m from the motorways);
- of **aerial objects** (all NUTS 0 to 9) which comprise and include many centroids of real estate (area of equal tax burden on real estate, etc.);
- of **thematic objects** such as: registered phenomena or phenomena otherwise defined with isolines (height above sea level, scanning patterns, etc.).

29. When geocoding, we need to maintain data comparability at the statistical level. For example, border municipalities should be statistically comparable with neighbouring and similar municipalities across the state border. In this respect, the rights and limitation of the use of both the basic recorded and the aggregated data should be carefully defined.

30. In some areas it is necessary to monitor certain strategic tasks which are outside the direct territorial competence of the manager (local community). For example, global division of work and migration of jobs in certain branches towards the far east are treated as possible geolocated consequences at home. In spite of the manager's lack of competence, such information is recorded and analysed so that the local community is better prepared for the inside changes. This should be helpful in planning and analysing the decision-making.

IV. DATABASES ON GEOREFERENCED DATA

31. It has already been mentioned that we are maintaining georeferenced databases. Furthermore, we should permanently maintain the integrity of the most important state, administrative and statistical registers. National statistics should produce **time cross-sections every three months** and store them.

32. If we wished to store all data from registers (Central Population Register and about 40 linked registers, Business Register and about 22 linked registers, a georeferenced register of real estate) every three months, we would probably end up with too much data. At present we cannot find a suitable solution for such storage; however, we expect in the near future to be able to use the new technology for data storage (Data Warehouse).

33. Special agreements should be made for data support of longitudinal statistical surveys, with which we have a lot of experience and which should be currently maintained, but which would be difficult to ensure for all collected data.

34. If special requests regarding cross-sections of attributes on a certain day (voters on a certain day and by location) occur, then a special daily entity cross-section has to be made. We should find a solution for linking other users. It would be preferable that they obtain the right to use the common **cross-section or even the union** of linked data rather than to do it on their own.