STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Working Paper No. 26
English only

Work Session on Geographical Information Systems
(Ottawa, Canada, 5-7 October 1998)

Item 9 of the provisional agenda

REGISTER DATA, POINT-BASED STATISTICS AND GIS IN THE SWEDISH AND THE ESTONIAN STATISTICAL SYSTEMS

Submitted by the Statistical Office of Estonia and Statistics Sweden [1]

INVITED PAPER

[1] Prepared by Inge Nael and Lennart Nordström.

I. POINT-BASED AND POLYGON-BASED DATA

1. The administrative system structures of governments reflect political, social and cultural value systems and their heritage. Obviously, these systems differ from country to country and influence the design of the statistical and spatial systems in the nation. Consequently, they also have a bearing on how Geographic Information Systems (GIS) can be used in statistics.

2. Broadly speaking, there are two ways in which GIS can be used in statistics, each one dependant on the national administrative set-up in question. The first development line is a *polygon-based approach*, where statistics for small- and large-scale areas are linked to polygons, over-laid onto a topographic background map in a discrete system. The second development line is a *point-based approach*, where statistical micro data are linked to x/y co-ordinates in a continuous system.

3. Most of the world's nations follow the polygon-based approach. The enumeration area - or equivalent - of the population census becomes the smallest territorial unit for which statistics are made available and the statistical and geographical systems are geared to this level of detail. To a great extent, statistical micro-data are collected in this field.

4. Due to their common history and similar administrative systems, Sweden and Finland are two examples of countries that follow the point-based development line. Seven-digit co-ordinates locate the position for statistical micro-data and the smallest territorial unit becomes not a polygon, but a point identified with one-metre precision. Statistical data from administrative registers are mainly used.

5. While Sweden's and Finland's systems are an antipode to polygon-based systems, Estonia does not yet have the tools for a statistical system based on administrative records. To work with a spatial system using point-based data is, however, basically the same goal in Estonia as in Sweden and Finland and co-ordinates for the location of census households will be collected during the census enumeration phase.

6. As will be seen in this paper, there are similarities and differences in the way Sweden and Estonia are incorporating GIS into their statistical systems. In spite of the different points of departure, the end goal, a point-based statistical system, is basically the same for the two countries.

II. ADMINISTRATIVE REGISTERS AND POINT-BASED STATISTICS IN SWEDEN

II.1 The Origin of the Swedish Population Registration System

7. Already in the 16th century, the compilation of Swedish population data was of crucial importance for the martial activities the contemporary kings indulged themselves and the nation in. The first parish register, registering births and deaths, was launched in 1571 and an unbroken sequence of more extensive population registration started from 1686 onwards.

II.2 Personal ID Numbers (Sweden)

8. Taking a giant leap in time, personal ID numbers were introduced in Sweden in1947, well before computers were a usable reality, and two decades later the first generation of an EDP-based population register was created by the Swedish Tax Board (1968). The personal ID number is one of the key information units that Swedish GIS builds upon.

II.3 ID-numbers and Cadastral Information - to Give Demography a Location (Sweden)

9. The significance of Swedish population registration for the scope of this paper is not only the long time series of demographic data accumulated in the system. The registration of demographic data for each individual is linked to a specific cadastral unit, a land parcel, which is the formal place of residence. The registration changes as individuals move during their lifetime and it is the responsibility of each person - or hospitals/midwives/parents in the case of births and hospitals/doctors/the deceased's family in the case of deaths - to supply updating information to the local tax offices.

10. The entire nation is subdivided into such parcels. Land parcels vary in size, from plot sizes of less than a thousand square metres around detached suburban houses to many thousands of square metres in national parks.

11. Cadastral information, linked to demographic information on individuals and their personal ID numbers, gives demography a location. The possibility to work with spatially referenced ID numbers is one of the key elements in our GIS that I will come back to later on in this paper.

II.4 Sweden's Statistical System - a Historical Background

12. Statistics Sweden is part of an administrative system with roots in the 18th century. Statistics itself was given an institutional framework in 1749 through *Tabellverket*, re-named *Statistiska centralbyrån* (Statistics Sweden) in 1858, two years before the first population census in 1860.

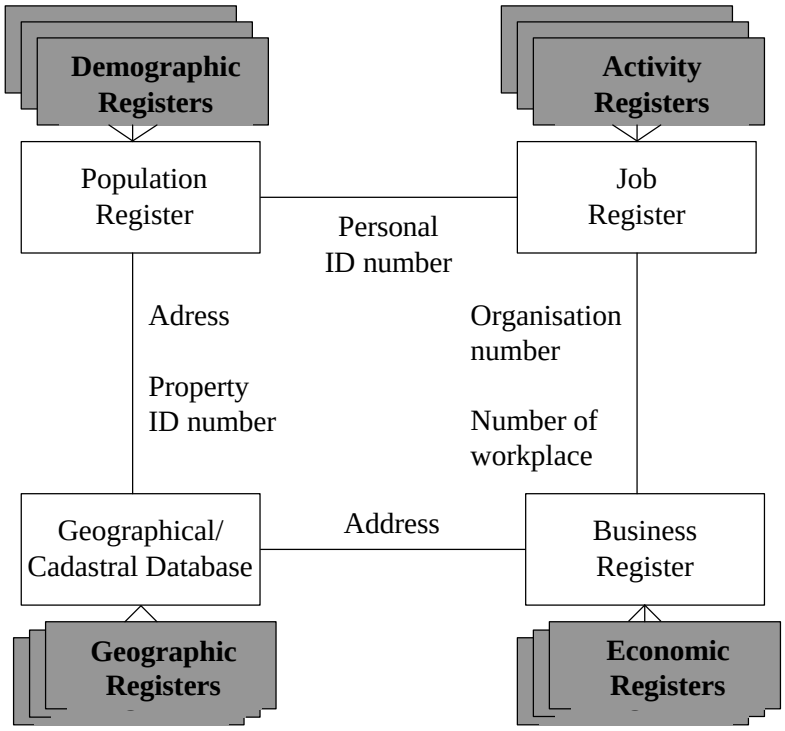
13. A century later, in the 1960´s, a centralisation trend swept over Sweden. Economy-of-scale and quality reasons were two of the main factors behind centralisation and statistics formerly produced by line ministry statistics was generally placed within Statistics Sweden's domain of responsibility. It is probably correct to say that this trend and the timing of the centralisation movement has been beneficiary for Swedish statistics, not least for the harmonisation of the national statistical system and, eventually, for the GIS solutions of today. A reverse decentralisation trend in the ´80s has not eliminated this.

II.5 Towards a Register-based Statistical System in Sweden?

14. Data collection from administrative sources is a prominent feature in the Nordic countries' statistical systems. Denmark and Finland have maybe come even further than Sweden and Norway in building a register-based statistical system, but the preconditions are basically similar.

15. Statistics Sweden's statistical system consists of four basic registers which can be inter-linked. Many other registers can be linked to one basic register. These registers cover the demographic, labour market-related, economic an geographical dimensions in statistics, as illustrated below. The geographical information in the GIS module can thus be linked to a large number of registers in the system.

Graph 1: Basic Registers and Identifiers in Statistics Sweden's Register System[2]



16. Administrative registers, however, need some fine-tuning in terms of links/identifiers, consistency, documentation and methodology in order to make up an ideal register system; this is the focus of attention for a special project at Statistics Sweden. The Geographical Database is presented in more detail in *section II.8.3* of this paper.

II.6 A Register-based Population Census in Sweden?

17. The National Tax Board's population registration system [3] described above is the major source for Statistics Sweden's compilation of current population statistics. The register is continuously updated with information from the Tax Board's local units. Theoretically - and where limited to demographic variables - a Swedish register-based population census would consequently be possible to take at anytime. In reality, the role of later field-based population censuses has been to complement register data with information on, for example, profession, household composition and housing.

---

[2] Based on "The Role of Administrative Registers in an Efficient Statistical System - Methodological Problems and Quality Issues", Wallgren/Wallgren, Statistics Sweden 1996.

[3] Registret över totalbefolkningen (RTB) - the Register of Total Population.

II.7 The Provision of Digital Land Information (Sweden)

18. In most countries, surveying organisations have a longer history than their statistical counterparts. An embryonic National Land Survey of Sweden was founded as early as 1628 and the production of analogue paper maps on various scales has been extensive ever since.

19. The processing of digital land information started in the 1960s and today's databases encompass a full range of digital topographic maps, cadastral maps, orthophotos, elevation information, etc. in thematic layers, geared to be used in a GIS.

20. Being the organisation responsible for registration of property, the National Land Survey is also hosting a digital cadastral register of all individual land parcels/cadastral units in Sweden, including the x/y point co-ordinates for each unit and for buildings. A geo-referenced register for apartments - crucial i. a. for a register-based population census in 2002 [4] - has been suggested and draft plans are presently being evaluated.

II.8 The Statistical/Spatial Interface: What Can Be Geo-referenced? (Sweden)

II.8.1 Demographic and Related Variables in Registers

21. The essence of the presentation of the administrative set-up in Sweden given above is that *administrative, statistical and geographical data can successfully be merged, using the ID of the cadastral unit as the common key or identifier*. As all persons are registered on one specific cadastral unit - referenced x/y co-ordinates - we have a formal location or at least a workable proxy for the location of the night-time population.

II.8.2 Local Units in the Business Register

22. Statistics Sweden hosts the Swedish Business Register. It contains extensive information for each local unit, for enterprises and concerns, while a derived public version of this register contains less variables and less detailed information for the variables for each local unit.

23. Each local unit is registered on a particular street address. Using the links created in the geographical database between land parcel and street address, local units and their day-time population are geo-referenceable in the same manner as the night-time population.

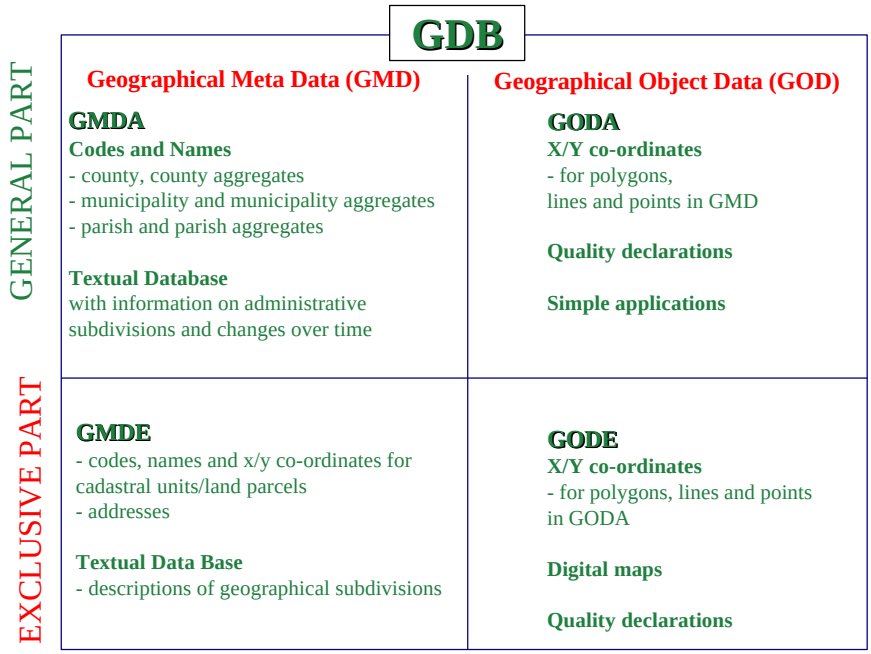
II.8.3 Statistics Sweden's Geographical Database

24. Statistics Sweden has developed a Geographical Database for all types of geographical information to complement the register system's statistical databases. The database stores codes, points, lines and polygons for all administrative areas down to parish level and various customised data sets as seen in the system overview below. It also contains co-ordinates for cadastral centroids and/or for buildings.

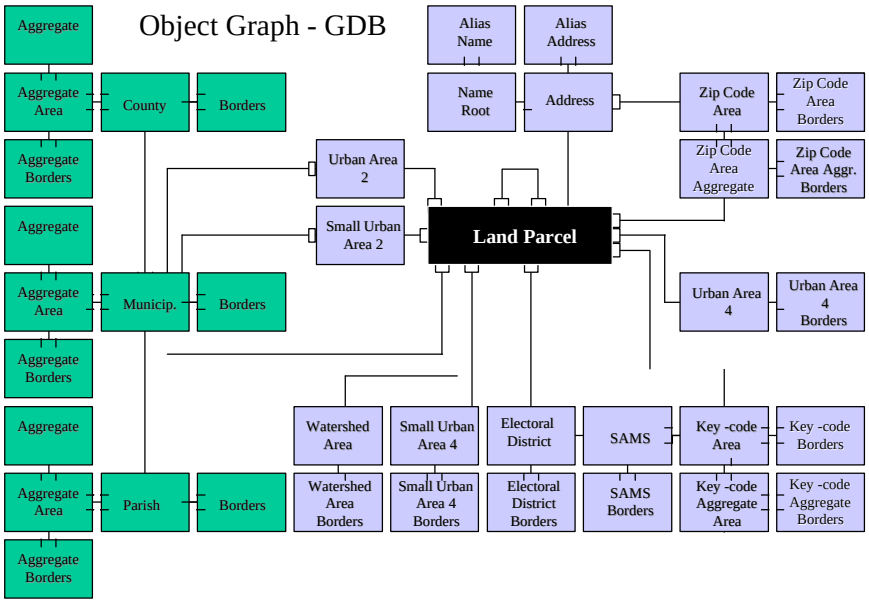
---

[4] The approach for the 2002 Population Census is still being discussed.

Graph 2: Statistics Sweden's Geographical Database



Graph 3: Statistics Sweden's Geographical Database - Object Graph



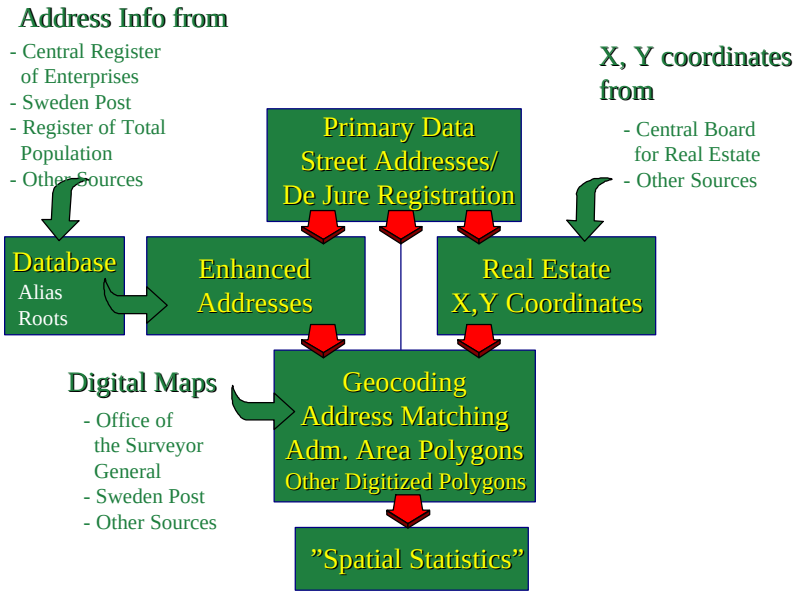
25. As a remedy for the address inconsistencies in registers, we have created a national street database, with built-in modules for correction of spelling and abbreviations in register data. There is also one module for transforming so called *alias names* (e.g. "the railway station", "the stadium", "the airport" of one particular city) to geo-referenceable street information.

26. Still, in practice, acceptable geo-coding quality might involve prohibitive amounts of labour. This is especially true for the Business Register's local units.

II.8.4 Geo-coding of Micro Data from Surveys

27. Micro data in Statistics Sweden's register system can normally be geo-referenced. This is also true for statistical micro data in surveys which have a reference to location, such as an address or a cadastral ID.

Graph 4: The Geo-coding Process for Survey Data



II.9 Quality and Precision (Sweden)

28. When looking at the precision of the formal location, the restrictions of certain assumptions must obviously be born in mind: the location of a given centroid might, for example, differ substantially from the actual location of a given building. Furthermore, one cannot assume that information given by a respondent in a survey should automatically be tied to the formal registration point. Street addresses, or again building co-ordinates, might be better for representing the location of a statistical result.

29. In spite of the importance of the cadastral unit for geo-referencing statistical information based on administrative data, the identification (name and number) of "their" unit is rarely known or remembered by individuals. Information submitted by individuals in surveys or censuses tends to be expressed as street names and numbers in urban settings or villages/farm names/post boxes in a rural context. Some, in practice rather weak, links for matching cadastral unit and street address are available in the administrative registers *per se* as the population register contains both information types, but the lack of formal and universal address standards in EDP-based registers, with misspellings and abbreviations in abundance, poses serious consistency problems.

III. ADMINISTRATIVE REGISTERS AND POINT-BASED STATISTICS IN ESTONIA

III.1 Personal ID Numbers (Estonia)

30. In Estonia, personal ID numbers were introduced in 1989, along with the re-issuing of Estonian passports after regaining independence. The structure and use of personal ID numbers in Estonia are similar to those of Sweden and Finland. PID numbers are kept in a central database. All residents of Estonia receive personal ID numbers upon issue of passports (with the exception of new-born children who, since 1992, receive PID numbers on certification of birth). The population database also includes PID numbers for non-residents, if they have been issued a visa for more than 1 year (in certain cases, such as the owning of real estate in Estonia, their native PIDs are used).

III.2 ID-numbers and Cadastral Information - to Give Demography a Location (Estonia)

31. In Estonia, the Soviet system of registration of the place of residence was set in force between 1940 and 1989. The system guaranteed reasonably precise records of population location, and the data was considered highly reliable. This system was one of many which were abandoned after re-gaining independence, but unlike other system changes, no replacement was created for the abandoned system. Therefore, there exists no legislative enforcement for registering the place of residence and registration is done more-or-less on a voluntary basis. It is estimated that about 30% of the population which has moved from previously registered place of residence is no longer registered at all.

32. The draft of a new law for registration of place of residence has been ready for quite a long time, but has been slowed down due to political discussions. There is a possibility that the new law will identify an official postal address, rather than an official place of residence.

33. In Estonia, the cadastral database is being developed. Land Reform, initiated in 1993 and estimated to be completed in 2001, has many tasks to fulfil simultaneously:

registering the whole country in the Cadastre and the Title Book,
restitution of pre-occupation land ownership
optimising land use
fuelling economy via the real estate market.

34. Due to a broad scope and an accuracy- and precision- demanding legislation (is very similar to the German system in terms of the Cadastre and the Title Book), the act of Land Reform is not very rapid. As of now, approximately 25% of land has been registered in the Cadastre (and the Title Book).

35. Therefore, to date, the Statistical Office of Estonia cannot use cadastral information for relating statistics with geographic space. Instead, building centroids will be used.

III.3 Estonia's Population and Housing Census in the Year 2000

36. The Estonian Population and Housing Census 2000 will be carried out using the traditional method of enumerators. This is not to say that this practice will be followed

indefinitely, but the traditional path has to be observed as long as administrative registers in Estonia do not reach the desired quality level. In this respect, Swedish and Estonian censuses are very different.

37. Both systems are, however, point-based - or more precisely, Estonia is approaching a point-based system. This means that the post-Census analysis can use similar methods and solve similar tasks.

III.4 History (Estonia)

38. The Soviet system destroyed many maps in 1940 in Estonia; the second wave of increasing map secrecy was in the 1960s. The history of land survey goes as far back as that of Sweden, for the simple reason that Estonia was under Swedish rule governed by Swedish king in the 17th century.

39. Filling in the gap of availability of spatial data (created between 1940-1989) requires intensive activity. This is accomplished many authorities joining forces; the Estonian National Land Board takes care of 1:10 000 – 1:20 000 basic mapping, the Estonian Border Guard has contributed in mapping areas close to borders on the same scale, the Estonian Maritime Board creates navigation database and maps, and the Statistical Office of Estonia is involved in creating 1:5 000 to 1:2 000 databases for urban areas. More resourceful municipalities carry out their own large-scale utility mapping.

40. Such decentralisation of mapping activities is not very common in other European countries. If administered well, it permits rapid mapping, taking various interests of different customers and governmental agencies into account. At the same time, though, there is always a risk of poor coordination, duplication and waste of resources.

41. The Statistical Office of Estonia has paid more attention to attribute data than mapping of line-work. At a later stage, if high-positional accuracy data from mapping organisations and the Statistical Office’s attribute data is joined, a valuable GIS database for the whole of Estonia can be created.

III.5 The Geographical Database of the Statistical Office of Estonia

42. The mapping program for the Census 2000 was initiated in 1995 by digitising building centroids of the 1989 Census in rural areas. The geographic base was the 1:50 000 Base Map. The Base Map was created by digitising SPOT orthophotos and was done by the Estonian National Land Board with support from the Swedish Government.

43. Later, the map was used for reforming Estonia’s administrative division – the number of settlements grew from 3,500 to 4,500. The reform was governed by the Ministry of Internal Affairs and was carried out by local government. The Statistical Office’s task was the contribution of paper maps and the creation of a digital database of old and new settlement borders.

44. Currently, 1:50 000 paper maps have been distributed to local governments for updating both maps of rural areas and compilation of population lists. As a result, the GIS database of Estonian rural population should be ready by mid-1999.

45. In 1995, the mapping of urban areas was initiated. The task is to create 1:5 000 maps, but in many cases the data will be as precise as 1:2 000. If available, digitised files are imported, or else existing city plans on paper and orthophotos are being digitised. Digital urban census maps include street network, building shapes, centroids and links to other registers IDs. All mapping activities have been subcontracted. The creation of the database should be completed by July 1999. In Q3 and Q4 1999, the database will be used for Census planning and after Q1 2000, GIS analysis and consumer market CD production can start.

46. Later in 1998, after the installation of high-end GIS software in the Statistical Office (procured under PHARE support), the seamless database for the whole country will be created, including:

building centroids (both in urban and rural areas);
vector line-work (roads, streets, rivers, coastline, etc.);
building shapes in urban areas;
linkage of building centroids to building registry and population registry (in urban areas, where available);
linkage to population lists compiled by local governments (in rural areas).

47. The Census GIS database will be used for enumerator area planning and for printing enumerator maps. In addition, WWW service for maps will be created (in Intranet at the first stage).

IV. EXAMPLES OF POINT-BASED APPLICATIONS (SWEDEN AND ESTONIA)

48. Although the location and attributes of individual register objects can sometimes be mapped (e.g. information on local units and land parcels), confidentiality considerations in statistics mean that the results from point-based systems regularly have to be made less detailed at the publication stage. The point-based system’s advantages over a polygon-based system might appear marginal. To a large extent, the benefits, however, lie in the flexibility of a point-based system at the pre-publication stages. The possibilities for in-house analytical work, spatial sampling and causal ”what if” planning, i.e. testing the spatial and statistical consequences of an event or a decision, are a definite advantage. Some examples will be given in connection with the presentation of this paper and during one of the demo sessions at the Ottawa workshop.