STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 7
English only

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of Statistics and Geography
(Tallinn, Estonia, 25-28 September 2001)

Topic (i): New opportunities created by cooperation and partnership

THE INE GEOGRAPHICAL INFORMATION SYSTEM

Submitted by the National Statistical Institute of Spain[1]

Contributed paper

I. OBJECTIVES OF THE GIS PROJECT

1. The first steps of the Geographical Information System (GIS) of the National Statistical Institute (INE) are linked to the search for a method to improve the cartography of the census section.

2. The census section is a territorial division of the municipalities connected with the concept of population. It usually varies between 500 and 2,500 inhabitants. As the term implies, the census section is very closely connected with the Electoral Register and it may by no means exceed 2,000 electors (if this occurs, the town councils should split the municipalities). Neither may it contain fewer than 500 inhabitants, unless the municipality is made up of a single section. When a section of a municipality that contains two or more sections consists of fewer than 500 electors, the town council must integrate that section into another census section. Sometimes, when the town council considers that a section is too large, it may opt for its division; or, on the contrary, it may decide to merge two sections whose population has greatly diminished. In either case, the town council must by law inform the INE of such changes for their verification.

3. The surface area of the census section is very heterogeneous, and may contain anything from three or four blocks in big city areas to several square kilometres in municipalities whose population is widely scattered.

4. The census section is a basic territorial unit at the INE, both for the collection and the dissemination of data. The design of the General Population Survey, which is also used for the main INE surveys (Labour Force Survey, Family Budgets, etc.) is a stratified two-stage random sampling based on the census sections as first step units. The randomly selected households within each section are the second stage units. The availability of maps for each census section, clearly showing the topographic elements that mark its boundaries (streets, roads, rivers, buildings, etc.) and the corresponding topography, is fundamental for the unambiguous identification of the dwellings belonging to each section.

5. So far, the INE has used printed maps provided by the town councils, since the latter were obliged to update them before each Population Census or Register and send a copy to the INE provincial statistical offices, together with the updated street maps. This documentation is used by the data

[1] Prepared by Milagros García-Tenorio and Luisa Muñoz.

collectors of the different INE surveys. In fact, these maps greatly differ concerning format and content and are often faulty - some are mere freehand sketches.

6. Consequently, for each census section it was necessary to obtain an accurate, clear, reliable and permanently updated cartography to be used in field work, as well as to maintain historical data for the successive changes. This requisite led to the idea of using a GIS tool that would also offer other advantages such as the execution of spatial analyses, the geocodification of addresses or the achievement of thematic maps with a view to dissemination.

7. The objective of standardized printing of census sections has been put on hold for the time being because of the difficulty in obtaining a uniform base cartography with a sufficient scale for the whole country. The reason is that priority is given to setting up digital boundaries for the more than 30,000 census sections for the georeferencing of the next Census Population results, which will allow the relevant spatial analyses and creation of thematic maps to be included in electronic or paper publications.

II. CARTOGRAPHIC SOURCES

II.1 Cadastral Office (CO)

8. The cartography that is best adapted to the implantation requirements of the INE GIS is the cadastral cartography. As mentioned above, the initial purpose of the Cartographic Area, later called Statistical Geographic Information Unit (SGIU), was to obtain a uniform printed cartography of all the census sections. It was envisaged that it should contain the streets and their numbers, roads, rivers and any element enabling an easy identification of boundaries and, therefore, of dwellings in every section.

9. The only cartography that satisfied these requirements in almost all the country (the Basque Country and Navarre have their own cadastres) was the one elaborated by the CO for tax purposes. The objective of this entity's cartography, with scales of 1:1000 or 1:500 for urban areas and 1:5000 for non-urban ones, is to identify all the urban and rural properties as well as the names of their owners.

10. Another advantage of using this cartography at the onset of the INE GIS project was its being part of a GIS computerization and incorporation process at the CO. This was the result of the fact that, since the mid-eighties, the CO has been implementing its data bases under the Geographical Information System SIGCA, specifically designed to this end. This led to the assumption that the cadastral cartography would shortly be available in digital format.

11. SIGCA contains two huge databases. One is alphanumeric and includes the names and addresses of the owners in each unit. The second is geographical and includes parcels and sub-parcels, as well as a number of layers such as street axes, urban equipment, blocks, hydrographic features, roads, etc. and their corresponding attributes. The INE works with this second database to capture of census sections boundaries.

12. Before starting this work on a regular basis, a pilot test was conducted in a single municipality, that of Mostoles in the province of Madrid. The aim was to digitise the boundaries of the census sections (112 sections) and to print the individual maps for each section. The success of the test proved the feasibility of the project at national level.

13. If the total computerized cadastral cartography had been available, the result would have been satisfactory for the whole country. However, the CO are only beginning to computerize their data bases with the big cities, doing so progressively but slowly, so that even today many small municipalities have not yet been included in the process. This is why the INE has had to look for alternative cartographies.

14. Nevertheless, the appropriate cartographic source for INE activities, both at the beginning of the project and at present, is still the cadastral one. Therefore, an INE-CO agreement was signed in 1997,

according to which the CO committed itself to supply the INE free of charge with a copy on magnetic support of the computerized urban and non-urban cartography that would become available after being generated and loaded into the SIGCA. The INE would also be provided with information on possible subsequent variations.

15. Since then, the Cartography Unit regularly received the available information requested from the CO, thanks to which the Unit was able to complete more than 70% of the municipal sections covered by the said entity.

II.2 The National Geographical Institute (NGI)

16. To tackle the GIS project, the INE had to collaborate from the very outset with the NGI, since this agency keeps track of the changes that arise in the layers of the municipal boundaries - essential information to start the identification of a municipality's census sections.

17. The remaining cartography, with relatively small scales, seemed in principle to be of no interest for a GIS project mainly concerned with urban issues. In fact, the use of 1/25000 or 1/50000 maps proved to be inadequate to identify population entities, roads, streets or blocks but, for lack of a more accurate cartography, this has become the solution for completion of the non-urban part of the boundaries for many municipalities' sections. They can still be used for capturing boundaries of urban sections if nothing better is available.

18. In order to receive this cartography free, an agreement with the NGI is envisaged and, in fact, has already been drafted based on a previous general agreement between the INE and the NGI for the execution of geographical and statistical products with common values added, and for the commercial dissemination of geographical and statistical information.

19. This subsidiary agreement to the general one for collaboration aims at the creation of a thematic digital cartography based on Map 2:25.000 (MTN25). It is meant to include the municipal, provincial and autonomous community boundaries as well as those of the districts and census sections generated by the INE and, as far as possible, the street maps of urban areas. At the moment, it is under consideration.

20. The signature of other subsidiary agreements is also envisaged to yield products for the dissemination of data from different INE Censuses or statistics based on the Basic Numeric Cartography 1:200.000 (BCN200). The CNIG would contribute the BCN200 cartography on raster format, the municipal, provincial and autonomous community boundaries proceeding from the Database of Municipal Boundaries, the location and name of population entities contained in the Database of Population Entities as well as the land use included in the Land Use Database, set up and maintained by the CNIG, according to the Corine Land Cover. The INE would contribute the municipal census results and would be set the task of producing thematic maps to be incorporated into the digital cartography.

21. This thematic cartography, together with the associated statistical data and the applications that enable their consulting and analysis, will be disseminated on CD-ROM. At the moment, the first project is the production of a CD with the main 2000 Agricultural Census results by municipalities. A second project is liable to be carried out with the Population Census results, whose data will be collected next autumn (reference date November 1st). Point 5 will go further into this matter.

II.3 Provincial Statistical Offices (PSO)

22. For the digital capture of census sections boundaries on a topographic basis, it is essential to dispose of paper maps or sketches showing the elements that determine said boundaries: streets, roads, rivers, etc. This graphic documentation, drawn by the municipalities from different sources during the preparation of Censuses and Surveys, is forwarded to the provincial offices concerned which keep it until the next updating and will use it for the data collection of subsequent surveys. It is, therefore, absolutely

necessary to rely on the collaboration of all the PSO's (52 in total) to send all the requested maps to the Cartography Unit, which will then photocopy and return them.

II.4 Other sources

23. Due to the lack of a digital cadastral cartography in many regions of Spain, it has been necessary to look for other cartographic sources that could be used as a topographic basis to identify the sections' boundaries.

24. After an inventory had been drawn up, contacts were sought with all those provincial or autonomous community organisms which produce or may supply cartography with scales appropriate to the work to be carried out. This led to various types of agreements - formal or less formal - for their free supply, specifically with 19 organisms, subject to different degrees of commitment:

- Formal agreements. These are signed by representatives of the organisms concerned and are published in the Official Gazette. This applies to the National Statistical Institutes of Madrid, the Basque Country and Navarre which, in the context of their preliminary work for the 2001 Demographic Censuses, committed themselves to providing the INE with the boundaries of the digital census sections, the georeferencing and the names of the population entities, as well as the continuous topographic cartography of the autonomous community concerned. An agreement has likewise been signed with SITGA, a Xunta de Galicia Society for the Regional Development of Galicia, whereby the latter bound itself on the one hand to supply paper maps for the census data collection, and on the other to digitally update the layer of census sections and population entities.

25. There has also been collaboration and/or dialogue with the Institutes of Andalusia, the Canary Islands and the Diputación de Toledo, which likewise gave rise to projects for the corresponding formal agreements.

- Less formal/collaboration agreements. With a number of agencies, though no formal agreements have been signed, the INE could achieve collaboration regarding either the collection of census data or merely in relation to cartographic information. These agencies are the following:

  - Gobierno de Aragón
  - Diputación de Zaragoza
  - Dirección Regional de Ordenación del Territorio y Urbanismo del Principado de Asturias
  - Instituto Balear de Estadística
  - Servicio de Vías y Obras de la Diputación Regional de Cantabria.
  - Consejería de Medio Ambiente y Ordenación del Territorio de la Junta de Castilla y León.
  - Diputación de Ciudad Real
  - Diputación de Girona
  - Dirección General de Urbanismo, Arquitectura u Ordenación del Territorio de la Junta de Extremadura.
  - Consejería de Política Territorial, Obras Públicas y Vivienda de la Xunta de Galicia
  - Dirección General del Medio Natural de la C.A. de la Rioja
  - Instituto Cartográfico Valenciano

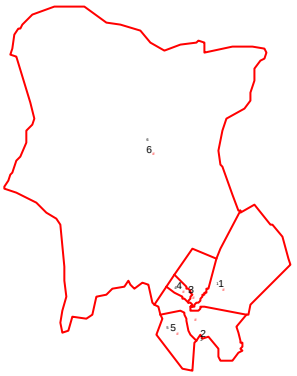
26. In these cases, the agreement almost always lays down the readiness of the agency to send basic territorial cartography to the INE, it being the latter's obligation to supply, when requested, the digital census sections of the municipalities in its autonomous community or province as well as the results of the following Population Census.

III. CHARACTERISTICS OF THE INE GIS

27 The current geographic information available in the GIS of the INE concerns the administrative boundaries: autonomous communities, provinces and approximately 75% of the census sections.

28. As far as municipalities are concerned, there are layers referring to 1991, 1996, 1998 and 1999 with associated data of the 1991 Census and the 1996, 1998 and 1999 Registers respectively. At present, the 2000 layer is being updated in order to incorporate it into that year's Register Data.

29. The Statistical Geographic Information Unit (SGIU) is responsible for the census sections. To date, their boundaries have been captured for 1991 and 1996, the years of the latest Census and Municipal Register respectively, thus these sources' data can be georeferenced. They are now being updated in order to use them in the analysis and dissemination of the 2001 Population Census results, the reference date of which is next 1 November.



Census Sections of Bembibre (León)

30. Capturing is made up of the following stages:

i) Request for information: When this process is started in a fresh series of municipalities, the CO is asked to supply their urban and non-urban cartography. Simultaneously, the corresponding provincial offices are requested to furnish the printed maps of said municipalities' census sections. The CO delivers, in their own ASCII format and for each municipality, five files of the urban and five of the non-urban parts, which correspond to points, lines, surfaces, attributes and texts.

ii) Loading into ArcInfo: By means of an automatic procedure, the files are turned into ArcInfo coverages and a checkup is conducted to verify that all the required layers are available. For the urban part, these are: blocks, parcels, subparcels, street axes, roads, hydrographic features, altimetry and urban equipment. For the non-urban part, they are: polygons, parcels, subparcels, hydrographic features, roads and the cadastral grid. An automatic verification is also made to be sure that each information layer contains all the associated tables and the corresponding attributes.

iii) Capture of boundaries: The capture of census section boundaries is carried out interactively on the screen. The process consists of superposing the layers of cadastral information and comparing them with the sketches on paper proceeding from the provincial offices to identify each section on the screen. When considering densely populated areas, the census section boundaries are usually street axes; in non-urban or scattered sections they are generally roads, rivers or even the very municipal border. These

boundaries are captured from the corresponding layers to generate the new census section layer. When the section boundaries do not coincide with any geographic element in the reference layers or the basic cartography, they are traced manually on a map and digitised later. Currently, much non-cadastral cartography is coming in from other autonomic or provincial public organisms. They use different formats, CAD being the format most frequently received. These files may be treated in two ways: they are either turned into ArcInfo coverages, or are used as background images on which the section boundaries are digitised.

iv) Coding of polygons: The last step is the coding of the polygons in the census section layer with the district and section attributes, as well as the lines, according to the CO specifications.

31. The type of files where information is stored for the time being are under ArcInfo coverage, since this is the software used for capturing census section boundaries, specifically the editing module ArcMap. There is, however, growing awareness of the advantages of storing all information in a database management system, for example ORACLE. For the digitisation of boundaries, other programmes such as Autocad Map or Microstation are sometimes used, but in any case the resulting files are converted into ArcInfo coverages. Arc/View is also used for thematic maps, and printing.

32. As far as the hardware is concerned, the SGIU has a local net made up of a server and 9 workstations, two digitiser tablets, one plotter and two printers.

33. The human team is composed of a statistician who is the head of the unit, two cartography experts (engineers in topography), a computer technician and three computer assistants in charge of capturing the census sections, and one administrative assistant.

IV. THE USE OF DIGITAL CARTOGRAPHY FOR THE DISSEMINATION OF POPULATION AND HOUSING CENSUSES

34. The preceding paragraphs describe the essential role of GIS in the information collection stage of the census. Cartography is also very important in the final stage, i.e. for the dissemination of results and the edition of specific geographic products. All this will be dealt with below. In the first place, mention will be made of the publications, with the help of occasional maps contained in the 1991 Population and Housing Censuses. In this context, the different types of information products and contents are considered and finally an outline will be given of the strategies to be followed for the forthcoming Population and Housing Censuses, reference date 1 November 2001.

IV.1 Geographic products contained in the 1991 Population and Housing Censuses

35. The star product of the latest censuses was a series of 3CD, the FAMILIA CERCA, containing information on three censuses: the 1989 Agricultural Census, the 1990 Buildings Census and the 1991 Population and Housing Censuses.

36. The peculiarity of this product compared to other publications was its very detailed territorial information yielded by an exhaustive processing of the Population Censuses. Besides vast data bases, the CDs in question included a very simple cartography, a set of meta-data as well as an application enabling navigation, analyses, the attainment of graphics and figures and dynamic tabulation. The cartography mainly consisted of the municipal boundaries of Population and Buildings Census, and of agricultural boundaries for the Agricultural Census.

37. The first CD, CERCA municipalities, allows the hierarchic access to a base containing 25 million data. In fact, it is a matrix of 8.163 rows and 2.823 columns, where the former are territorial units (provinces, municipalities) and the latter are variables or categories at different breakdown levels. It facilitates the setting up of dynamic tabulations and the creation of graphics according to the selection that has been made.

38. The second CD, CERCA+100, supplies information on 1,600 variables of the population, buildings and households broken down into over 80,000 small areas which are the districts and census sections, whatever their size, and all the territorial units of the Classification with more than 100 inhabitants (collective and individual entities as well as agglomerations of over 100 inhabitants).

39. Support of this CD, CERCA+100 consists of the boundaries of all the municipalities in Spain, updated to the census reference date (1 March 1991) and derived from the municipal boundaries, digitised at a 1/200,000 scale by the National Geographic Institute (NGI) between 1985 and 1987. Changes that arose through to the census date (mergings, divisions, modified boundaries, etc.) are indicated. Boundaries of all the agricultural regions are also included.

40. Finally, the CD (CERCA+3500) completes the product with detailed figures on population entities and on populations of over 3.500 inhabitants. Cartographic support also comprises municipal boundaries and makes it possible to enter a consultation regarding streets, since written descriptions of the districts and sections where these streets belong are included.

IV. GEOGRAPHIC INFORMATION IN THE 2001 POPULATION AND HOUSING CENSUSES

41. As far as the data collection stage is concerned, many COs will be using INE GIS maps instead of the old fashioned sketches for the census sections routes. In other cases, the maps will be provided by the municipalities themselves or by organisms that have reached some sort of agreement with the INE because their cartography is cadastral or appropriately accurate for the census field work.

42. Much more is expected from the role of cartography in the dissemination of census data. The Census Project envisages a dissemination plan of the results where particular emphasis is laid on products with a geographic support. It refers specifically to 'census data compilations on cartographic support', similar to the above-mentioned 1991 CERCA publication. Here, two products are available: a set of general tables and a pre-established group of more specific and extensive tables.

43. The general tables are highly detailed from the conceptual viewpoint and are elaborated for medium geographic levels (provincial and municipal). They are integrated into a simple geographic interface that enables the creation of maps and data graphics for the main socio-demographic variables. The hardware will probably be CD-ROM, easy to handle and meant to be used by individual users, its price being accordingly affordable. The product will be available by 2004.

44. For the specific tables, the territorial level will be much smaller (sections and sometimes streets or blocks). This is when cartographic support will be greater, including boundaries for all the sections and street maps for some of them. For the attainment of this product, an agreement might be reached with a central or autonomous administration entity.

45. The planned format is DVD-ROM or something similar and the users will be enterprises and public administrations that require a greater geographical detail and may afford higher prices. The product will be available by 2004.