STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Working Paper No. 7
English only

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

Item 5 of the provisional agenda

# APPLICATIONS OF GEO-DATA TO STATISTICS AT ISTAT

Submitted by National Statistical Institute, Italy [1]

CONTRIBUTED PAPER

[1] Prepared by Fabio Crescenzi.

## I. INTRODUCTION

1. In the near future we expect a strong increase in demand and importance of statistical information associated with geographical data, both at national and international levels. Existing gaps are a major obstacle for an immediate broader use of geo-data in statistics. This paper attempts to summarise some of these gaps: a) lack of awareness of existing geo-data; b) lack of co-ordination with National Mapping Agencies and other official mapping bodies; c) lack of efficient and user-friendly data-interchange and communication procedures; d) redundancy in data acquisition and data storage; e) insufficient update of data; f) lack of guidelines for meta-information; g) price highly variable; h) problem of copyrights.

2. One of the major challenges is to standardise geo-data for interchange and to allow free access to geographical data for aims of public interest and some kinds of products, including:

- maps of administrative borders on reasonable scales;
- land cover maps (at national and international levels) built on a common classification and scale;
- updated databases of streets, railways, rivers, lakes, etc..

3. Main phases of statistical surveys can be fruitfully supported by GIS. In sample design: design of homogeneous areas to stratify statistical units; in data collection: design of input areas and interviewers paths; in data check: preparation of plan for data checking; in data presentation and analysis: design of output areas including strategies to protect data confidentiality; visualisation of functional zones; geo-coding of statistical units on maps; data presentation on maps; and presentation of indicators at micro territorial and local level. GIS plays a special and crucial role in the design of a census geographical database.

4. A challenge for the coming years is to move from a first phase, more characterised by use of GIS oriented towards data presentation and exploration, to a new phase in which statistical inference and statistical modelling can be more integrated in GIS tools. Examples of future potential uses of GIS are: models for non-response evaluation and correction or models for data estimation when information available for geographical areas does not completely overlap the ones in which we are interested.

5. The fields in which GIS application can make an important contribution are many. For example, urban studies, agriculture, environment, with special reference to: inhabited areas design, urban areas growth analysis, development planning, disadvantaged areas, homogeneous land cover areas design, industrial areas, heritage sites areas, civil protection, protected areas, national parks, green space, transport facilities and other infrastructures location. All these fields are of crucial interest for statistics.

6. Several geo-data sources can be considered. To avoid overlapping with official mapping bodies, the integration of existing maps should be a priority. Maps to be considered are: remote sensing images, ortophotomaps, regional technical maps, NMA's cartography, cadaster maps, street-maps of private company for geo-marketing, geographical data-bases on streets, railways. Existing problems are the integration of different products and the updating of maps.

## II. APPLICATIONS OF GEO-DATA TO STATISTICS

7. The main uses of GIS at the Italian National Statistical Institute (Istat) are the following: (i) design of census input areas for 2000-2001 censuses round; (ii) production of

CD-ROMs to present geographical and statistical data and to build thematic maps; (iii) geocoding of addresses on maps; (iv) visualisation and checking of functional zones (i.e.: local labour areas); (v) developing prototypes for geo-data access via intranet and internet; (vi) Arezzo pilot project for a land cover map 1:25.000; (vii) experimental design of morphological urban agglomerations.

## II.1 Geographical Istat database: CENSUS and SISTER projects

8. The lack of an official map used for statistical purposes pushed Istat to launch the CENSUS projects in 1991 with the aim to design its own geographical bases. Here are the main phases of CENSUS building:

- Identification of 3 kinds of areas: centri abitati (inhabited agglomerations with a meeting place for the population such as a church, a shop, etc.), nuclei abitati (inhabited agglomerations without a meeting place for the population), case sparse (remaining areas). This result was obtained using advanced methodologies (integration of remote sensing images, technical maps at regional level and municipalities information).

- Sub-division of these areas into basic territorial units (census sections), following physical borders such as streets, railways, rivers, etc.. In the biggest towns census sections are building blocks.

- Acquisition of all administrative borders, inhabited areas borders, census sections borders.

- Census sections were used as input areas for population and enterprise censuses, but not for agriculture census.

- Building of the National Street database (list of streets and related street numbers enclosed in each census section).

- Procedure of borders validation at municipality level.

- Construction of a six classes land cover map.

9. CENSUS is a digital ARC/INFO 1:25.000 database. It contains borders of about 323.502 census sections, inhabited areas (“centri abitati” e “nuclei abitati”) as detected in 1991 and all administrative borders (NUTS2-Regions (20); NUTS3-Provinces (95); NUTS5-Communes (8100), 1991 figures), coordinates of a reference point for each census section. Out of 323.502 1991 census sections, about 70% are in “centri abitati” (about 230.000 in absolute figures), about 10% in “nuclei abitati” and the remaining 20% in “case sparse”. Analysing surfaces, “centri abitati” and “nuclei abitati” form about 10% of Italian surface. “Case sparse” form about 90% of total Italian surface. Analysing population, 91% is in “centri abitati”, 3% in “nuclei abitati”, 6% in “case sparse”. “Case sparse” are too big in extension to be used in agriculture and environmental analysis.

10. SIS.TE.R. is a system for street recognition and attribution of census section codes. It is a software developed by Seat on the National Street Database, the alfa-numeric database of all Italian streets as derived from 1991 census of population.

11. The main challenge for the 2000-2001 census round is to redesign and reduce in size census sections in extra-urban areas to obtain the integration of all census territorial bases

(including agriculture). Other aims are to design industrial agglomerations, to improve integration with other territorial databases of public interest, and to make a broader use of digital maps of private companies in designing interviewers' paths.

## II.2 CD-ROMs to present geographical and statistical data and to build thematic maps

12. GEOSTAT is a CD-ROM made by Istat, Esri Italia and Seat containing ARC/VIEW data publisher, regions (NUTS2), provinces (NUTS3) and communes (NUTS5) borders, more than 150 variables (mainly from 1990-91 censuses) and co-ordinates of the main streets, railway, rivers and lakes. The cost of GEOSTAT (including VAT) is approximately 500 ECU.

## II.3 Geocoding of addresses on maps

13. Demand for local and updated statistics is already high. New requirements are coming from European structural funds policies and add new pressures to statistical offices in providing data responding to both requirements. Quite often data comes from sample surveys and results in updated data which is not significant at a local level, while total survey gives data significant at the chosen local level which are not updated.

14. Burden on respondents cannot be further increased. Many bodies ask respondents for the same data, causing disaffection to statistics. When local statistics are obtained, the confidentiality issue is a further obstacle to be considered and data cannot be released without some protection strategy.

15. For all these reasons, Istat is moving towards a more effective use of administrative archives for statistical purposes. An important example is ASIA (the Italian register of enterprises), built by the linkage of several administrative registers. Linkage of different archives is used for data validation and only when relevant differences arise is a form is sent to respondents.

16. To perform statistical analysis at local level and to represent phenomena on maps, geo-coding of statistical units is required. Administrative archives usually contain addresses, so a tool is needed to represent addresses on maps or to attribute an address to a micro-zone. To do this, Istat uses SIS.TE.R. (System for street recognition). Once a statistical unit is associated with a census section, it can be represented on maps using coordinates of census section reference point.

## II.4 Visualisation and check of functional zones

17. One of the major difficulties in using census data for local analysis is that they usually refer to territorial units of administrative nature, which are virtually without geographical or statistical significance. The solution is to identify territorial units having a clear statistical meaning. Istat did this in the local labour market areas research. Italian labour market areas were built using an iterative procedure designed by New Castle university on the intra-communal work flows matrix. This procedure is not based on GIS, but GIS was used to analyse and to check local labour market areas.

## II.5 Prototypes for geo-data access via intranet and internet: GEOSERVE project.

18. GEOSERVE is a project developed under the 4th Framework Research Programme of the European Union in the sector "Telematics for Administrations". The GEOSERVE aim is to develop a network of Geo-data Access Services for European Administrations and Data Providers. The Services will support the identification, selection, and timely access to geo-data for data users on a European scale. For the provider of geographic data, the system will enable the international marketing of geo-data to a large user group. The objectives of the project GEOSERVE are to:

- harmonise the requirements of European data providers and data users;
- develop a concept for a European geo-data access services network;
- guarantee that the user needs and user participation drive the project;
- build demonstrators for the network service nodes and for the following clients: data provider client, geo-data user client and info-desk client.
- assure openness and flexibility of the system by interfacing with various Geographic Information Systems, geo-data formats, geo-viewers and application areas.

19. The project is driven by user demands. A consortium of leading IT industrial partners, service organisations, geo-data provider administrations and data user administrations will develop the concepts and services and will validate and test them with pilot applications. Pilot sites will demonstrate regional and international exchange of geo-data. The GEOSERVE system will provide Geo-Data service nodes, Geo-clients and added value functions on Internet and Intranet. The demonstrators in Finland, Greece, Germany, and Italy will link various international and regional data providers and users. They will support scaleable geo-clients for data providers and data users in administrations ranging from citizen information kiosks to professional user clients.
Openness of the system will be reached by including and serving a variety of geo-representations and transfer of format standards in use in European member states and regions. The project will pay great attention to monitoring and actively promoting the development of standards in the GIS and Geo-transfer domain.

20. Important European users are involved and participate in the project: surveying institutes of Finland, Lower Saxony (Germany), providers of thematic data such as the Ministry of environment of North Rhine Westfalia (Germany), ISTAT, Decision Systems Limited of Dublin (Ireland) and Institut Géographique National, France. All these users contribute as providers of geoservices and geodata. Major Validation Sites will be built in four Countries: Germany, Finland, Greece, Italy.

21. For numerous tasks in public administrations, geographic data has become an important part of decision processes. In many application fields data needed for specialised tasks is a combination of geographical, demographical and other non graphical data. The user groups (data producer, data enhancer and data user) and the need for user-friendly access to geographic information are rapidly growing with the widespread introduction of Geographic Information Systems (GIS) and geographic viewer applications integrated into standard office communication and software environment and the increase of Internet applications. Data acquisition of geographic basis data is mainly performed by surveying institutes: raster data, vector data, satellite images, aerial photography. Additionally, there are many other data providers on the market who are doing business with enhanced data: e.g. they produce thematic data packages or convert data to various GISs. In general, there is a big pool of geodata already collected and available for the users.

22. Due to this variety, data is usually spatially distributed and within the responsibility of different data providers. Therefore it is difficult for the user to access data. The issue is that a common European "market place" between producers and users is missing. A network-based system will be developed to provide user-friendly Geo-Data-Services for Identification (metadata), Selection, and Access of Geo-Data and of Geo-Services. For the data provider the system will be used for marketing the data to a large user group. Brokers will be able to efficiently sell Geo-Data and services with the GEOSERVE system.

23. Through GEOSERVE, data providers obtain a means to distribute their information and geodata effectively to the customer. Data users in administrations and businesses will use GEOSERVE to identify, select, order and obtain the geographic information they need for their task.

### II.6 Pilot Project for a land cover map

24. In the next few weeks a land cover map, realised in the context of a pilot study on Arezzo, will be finalised. A 1:25.000 land cover map will be released based on a classification similar to Corine Land Cover level III, but extended to increase knowledge on urban areas. The purpose is to test results to build an Italian land cover map.

### II.7 Design of morphological urban agglomerations

25. Experiments on Morphological Urban Agglomerations were done for the Milan area and the Florence area, giving good results. The purpose is to extend building of urban agglomeration to all the main large Italian cities.

## REFERENCES

ISTAT (1997). I Sistemi Locali del Lavoro 1991. *Argomenti*, Volume 10

ISTAT, ESRI ITALIA, SEAT (1997). Geostat: dati cartografici e statistici a livello nazionale, regionale, provinciale e comunale. Roma, 1997, (CD-ROM)

ISTAT (1997). La progettazione dei censimenti 1991, *Volume 1*