UNITED NATIONS **E**

Economic and Social Council

Distr.
GENERAL

CES/2001/21/Add.4
4 May 2001

ENGLISH ONLY

STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE
CONFERENCE OF EUROPEAN STATISTICIANS

Forty-ninth plenary session
(Geneva, 11-13 June 2001)

CENSUSES, GEOCODED DATA AND SMALL AREA STATISTICS

Supporting paper submitted by Istat, Italy [1]

I. INTRODUCTION

1. The demand for small area statistics is strongly increasing. Allocation of structural funds, national and local policies add pressure on statistical offices to provide local data. Most of data available at local level are census data, which usually requires long time to be released and early become old. Survey data are often more updated, but are usually not available at local level. Burden on respondents, that cannot be further increased, and budget constrains, do not allow to estabilish new surveys. It must be also considered the trade-off between local data release and confidentiality protection needs.

2. A closer integration of GIS in statistical applications, a better integration of existing data and a more intensive use of all census outputs, is a good strategy to increase quantity and quality of local data. New and more advanced tecnhologies in administrative records linkage, digital mapping, address matching, and a larger storage capacity in faster computing resources environment, offer big opportunities in this direction. Map integration of several geo-data sources like remote sensing images, ortophotomaps, regional technical maps, National Mapping Agency's cartography, Cadastral maps, street-maps of private company for geo-marketing, geographical

[1] Paper prepared by Mr Fabio Crescenzi, senior researcher at Istat

data-bases on streets, railways, facilitate area mapping and allows to increase knowdledge on small areas.

3. Fields in which GIS applications can give good results are many and all of them of crucial interest for statistics. Urban studies, agriculture, environment, with special reference to: inhabited areas design, urban areas growth analysis, development planning, disadvantages 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.

4. Main phases of statistical surveys can be fruitfully supported by GIS. In data collection: design of enumeration areas and enumerator paths; in data check: preparation of plan for data checking; in data presentation and analysis: design of output areas designed using confidentiality data protection strategies; visualisation of functional zones; geo-coding of statistical units on maps; data presentation on maps; presentation of indicators at micro territorial and local level. GIS play a special and crucial role in census mapping.

5. Some gaps are a major obstacle for an immediate wider use of geo-data in statistics: a) lack of awareness of existing geo-data; b) lack of co-ordination between National Mapping Agencies and others 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. A challenge for the next year 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 at reasonable scales
- land cover maps built on common standards (at national and international level).
- updated databases of streets, railways, rivers, lakes, etc. .

6. To fully understand the potential benefit of census collection process it should be considered not only the output of census data itself, but also all products of the considerable fieldwork, like collection of data on streets, address numbers, and others attributes that can be used in address matching and in geocoding of administrative records. All census outputs can be used in developing strategies of small area data dissemination. The digital census mapping database can be used in procedure for area mapping and data access via intranet-internet, travel-to-work flow can be used in functional zones mapping.

7. Main GIS applications at Istat are: (i) census mapping; (ii) production of digital census atlas on CD-ROMs; (iii) geocoding of addresses on maps; (iv) visualisation and check of functional zones (i.e.: local labour areas); (v) developing prototypes for data access via intranet and internet; (vi) Arezzo pilot project for a land cover map 1:25.000; (vii) pilot study on morphological urban agglomerations.

8. A challenge for next years is to move from a first phase, more oriented to use of GIS for data presentation and exploration, to a new phase in which statistical modelling and methods are more integrated in GIS tools. Examples of future applications of GIS are: models for non response evaluation and correction or models of data estimation when source and target zones do not match, new methods and applications of administrative records geocoding.

II. CENSUS MAPPING

9. In Italy, as in other countries, census data collection process provides a unique opportunity to build a complete small area mapping. Census enumeration areas are called “Sezioni di censimento” (Sdc) and Sdc features are defined to allow enumerators to easily identify borders of their assigned region, to correctely count statistical units, to avoid any possible double counting. Natural boundaries for Sdc are streets, railway lines, hydrological features such as rivers and lakes.

10. In 1991 and for the first time, an advanced methodology of census mapping was used. “CENSUS”, a complete digital database in ArcInfo format and scale 1:25.000, was developed on integration of remote sensing images, IGMI maps (Italian National Mapping Agency), technical maps at regional level and municipalities information. “CENSUS” contains administrative boundaries, (20 Regions; 95 Provinces; 8100 Communes, corresponding respectively to level NUTS2, NUTS3, NUTS5 of the European nomenclature) populated localities boundaries (“centri abitati” e “nuclei abitati”) and Sdc (about 330.000) boundaries, coordinates of Sdc centroids. Out of 330.000 1991 census section, about 70% are in “centri abitati”, about 10% in “nuclei abitati” and the remaining 20% in “case sparse”. “centri abitati” and “nuclei abitati” are about 10% of Italian area. “Case sparse” are about 90% of Italian area. Analysing population, 91% is in “centri abitati”, 3% in “nuclei abitati”, 6% in “case sparse”.

11. CENSUS 2000 is the mapping project for 2000-2001 censuses round. It starts from the 1991 mapping, and combine several available cartographic data which provide a large amount of information on populated areas, streets, rivers, railways like digital ortophotomaps, technical maps, vector data, from other government agencies, local authorities or private companies. Main objectives of CENSUS 2000 are: i) to redesign and reduce in size census sections in extra-urban areas to get the integration of all censuses mapping, including for the first time agriculture census. “Case sparse” in 1991 were too big in extension to be used in agriculture and environmental analysis; ii) for the first time, to design productive locality; iii) to improve integration with other territorial databases of public interest; iv) to use data extraction from digital maps of private companies and local authorities to fill enumerators paths. Four types of populated localites are mapped in the first stage: 1) centri abitati (populated areas), 2) nuclei abitati (small populated areas), 3) località produttive (productive localities), 4) case sparse (remaining areas). In the second stage,using techniques differentiated for each type of locality, Sdc are mapped.

III. LOCAL LABOUR MARKET AREAS AND MORPHOLOGICAL URBAN AGGLOMERATIONS

12. One gap in local analysis is that data are usually referred to administrative areas which are virtually without geographical or statistical significance. Using census data on travel-to-work flows it is possible to build territorial units having a clear statistical meaning. Italian labour market areas were built using an iterative procedure designed by New Castle university on the intra-communal travel to work flows matrix. This procedure it is not based on GIS, but GIS was used to analyse and to check local labour market areas.

13. Pilot studies on Morphological Urban Agglomerations in the Milan and the Florence area, were developed. The purpose is to extend the study to the main Italian metropolitan areas.

IV. ADDRESS MATCHING AND SMALL AREA STATISTICS FROM ADMINISTRATIVE SOURCES

14. Most of administrative databases contain addresses which allow, when coordinates of address numbers are known, a complete matching of data to points of a digital map. In any case an address is a respondent identifier and due to confidentiality restrictions, data for use to statistical purposes have to be geographically aggregated.

15. Istat uses a procedure of address matching to link addresses to Sdc codes, so that data of each Sdc can be aggregated. For each Sdc is available, as attribute of Sdc polygon, the list of streets and for each street the ranges of address numbers in the Sdc, called "Itinerario di sezione" (Sdc path). After the addresses orthographic correction and normalization, the procedure links to each address the Sdc code. An important application is the geocoding of ASIA register (the Italian businesses register). Applications will be extended in the next years also to geocoding of records from population and social registers, including schools and hospitals registers.

Sciacca (Sicilia) Orthophotomap and census mapping

- - - - -

REFERENCES

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

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

OPENSHAW, S, ed. (1995) Census User Handbook. Cambridge, United Kingdom: Geoinformation International

UNITED NATIONS (2000), Handbook on geographic information systems and digital mapping, *Studies in Methods*, Series F No.79