

GISTAT: a geospatial information provider for statistical activities

Enrico GIOVANNINI
President, Istat, Italy

Pina Grazia TICCA
Istat, Italy

Abstract. The paper describes GISTAT the Italian geographic information system for statistics, that is done in terms of a “geospatial information provider” for statistical activities. At present time the basic system is available to provide data to internal users and it is ready for functional and performance tests for Internet users. Collaborative processes with internal and external users/institutions, that are now in a study phase are illustrated. The integration of GISTAT toward the national agency responsible for ruling the “territorial data cataloguing” (National Repertoire for Territorial Data - RNDT) for public institutions (DigitPA) is an issue in which Istat is very strongly focused. The goal is to support the national geospatial interoperability and since the RNDT metadata model is a subset of the INSPIRE model, that integration will contribute also to the European geospatial interoperability.

Moreover, GISTAT has been chosen by Istat as the infrastructure to disseminate census 2011 geospatial data and results (from country level up to enumerations areas level). Moreover in the next future Istat is planning a new system based on integration of registries and censuses to release structural and territorial data. This will be a new very powerful perspective for the development of GISTAT.

1. **The geographic information system**

GISTAT is the Italian geographic information system for statistics; it was firstly created by collecting and harmonizing the census mapping cartography that Istat designs and updates every 10 years to execute the housing and population census. The digitizing process started after the 1991 census, while the census itself was supported by paper maps; for the 2001 census a file-based GIS was designed and setup to support the census.

After the 2001 census the first geographic database was designed and stored inside an Oracle DBMS, it was a great result one of the most important national geographic database for that time; it consisted of a seamless geodatabase implementing the *national territorial continuum*. A first WebGIS was also published to support census data dissemination, providing interactive thematic mapping.

In 2007 Istat started the design and implementation of a complete GIS; that has been the base for the 2011 census. It's named GISTAT and it is now evolving toward a Spatial Data Infrastructure (SDI) now available on the Intranet for internal users. This task is also included in the broader framework of tools available for official statistics, the National Statistical Programme (NSP).

It is based on the principles of a Service Oriented Architecture (SOA) and re-usable component (GeoWebServices) with the purpose of sharing, disseminating and make interoperable geospatial information.

1.1 The geographic database

The spatio-temporal geographic database is one of the key features of the entire system; it stores and manages inside an Oracle DMBS a big amount of geographic data coming from the censuses, from internal analyses/processes and from external sources. It is a national geographic database and it is an "admin borders – centric geodatabase", that because the census mapping cartography is designed within the administrative boundaries; therefore Istat is nationwide, a main actor in the process of administrative boundaries digital cartography updating and dissemination (Even if only for statistical purposes and not purely cartographic). On top of that country level cartography the digital census mapping is designed and updated.

That activity has been done back in time starting from the first census in 1861 and since 1971 geodata have been digitized and then managed by a GIS.

Census mapping cartography is designed to describe urban areas and industries, rather than rural areas, inside each municipality (LAU2). Istat provides the rules to design/update the census cartography, supervises more than 8.000 municipalities, builds and harmonizes the geographic database. The main census executors are the municipalities and that's stated by the law.

Inside each municipality, the main territorial units are the urban localities, classified into:

- a. “**big localities**”: groups of houses, not far more than 70 meters one from each other, connected by roads. The locality should have public services (schools, most important railway stations, pharmacies,...)
- b. “**small localities**”: groups of houses, not far more than 30 meters one from each other, but without public services
- c. “**productive localities**” (industries): a locality in a non-urban area, with at least 10 firms or 200 employees, with an area of at least 5 hectares.

The remaining part of the municipality is delimited, but not classified; it represents approximately the non-urban areas and is called “scattered houses”.

Each locality is then divided into census Enumeration Areas (EA); the EA is the minimum enumeration and dissemination census area. The EAs are delimited and updated according to the guidelines provided by the regulations annexed to a state law on population registers (it’s named “*regolamento anagrafico*”); those regulations are oriented to maintain the comparability with the previous censuses, when possible, and to follow the international recommendations on urban areas.

A codification process was also done to store the morphological attributes, such as: rivers, lakes, desert, mountains, islands,.... Furthermore there is an on-going process to enhance such a codification to evaluate the generation of a Land Cover map.

Moreover inside the geodatabase many derived layers are stored, such as: “local market areas (LMA)”, election districts, morphological urban aggregates ..., and also data coming from external sources such as street networks, and aerial photos at various time periods.

1.2 *The technological platform*

Another key element is the technology; the Institute choose a sound and modern technology based on ESRI GIS software. On top of that a platform has been developed to support the system. The platform implemented is based on a modern SOA/SOE architecture with the purpose of sharing, disseminating and make interoperable geospatial information. It is composed of:

- a. a data server to extend Oracle functionalities with more GIS capabilities
- b. a GIS application server to distribute geospatial data; GeoWebServices have been designed and published through GISTAT using standard WMS (GeoWebServices)
- c. a web server to make GISTAT available and searchable on the internet.

The platform is composed of a development environment and a production environment accessible through a “load balancer” and designed to be scalable and extensible to support future enhancements.

2. GISTAT as a geospatial information

Geospatial information available at Istat is much more of what stored inside the geodatabase; GISTAT is evolving both in archiving and sharing that information. The objective is to provide geospatial information to support statistical activities for internal and external users. For that purpose a prototype Website is available on the Intranet; GISTAT acts also as a starting point to navigate through the geospatial information in different ways: through data download, GeoWebServices and through a FlexViewer. The geospatial information available concerns geographic files, static and dynamic thematic maps, geoprocessing tools, documents, ... The process of collecting, harmonizing and cataloguing that information is a on-going process, as well as the design of a geoportal to enhance the prototype website in making geospatial information available and interoperable.

2.1 Data download

Geospatial data are available for download on Istat website; at present time the data available concern the administrative boundaries at different years, localities and enumeration areas for several censuses, points of interest (railway stations, airports, harbors, hubs).

The thematic maps produced for 2001 census, those included in the official census books have been collected, organized and addressed from the website. Metadata are not still available but the folders structure and filenames work as a catalog; those maps are available through PDF files. The link to interactive thematic maps is also available.

2.2 GeoWebServices

GeoWebServices for census geospatial data have been designed and published through GISTAT using standard WMS; the web services are available for 1991, 2001 and 2011 census including census variables (1991, 2001). GeoWebServices have been optimized for user interactions. They can be searched and accessed from different GIS clients and used to perform geospatial analysis.

2.3 *FlexViewer*

A basic GIS viewer (BT.Viewer) has been implemented to interact with those GeoWebServices. The objective is to provide users geospatial information and easy geographic tools to analyze statistical data in a spatial way.

GISTAT-BT.Viewer is now in a beta release and it is available for internal users; there is a plan of a beta test on the Internet for invited and registered users.

3 **Next steps**

3.1 *Geospatial interoperability*

To promote interoperability a first step is to provide metadata and discovery services; Istat is working on that direction. There is also a on-going collaborative process with DigitPA, the national agency supervising the adoption of the digital code for public institutions; in terms of geospatial information DigitPA defined the rules for the “cataloguing territorial data” for public institutions and implemented a prototype National Repertoire for Territorial Data (Repertorio Nazionale Dati Territoriali - RNDT) geoportal to which public institutions have to contribute. A pilot project has been started to integrate GISTAT metadata and discovery services into the RNDT geoportal through the RNDT geoportal. In that way GISTAT GeoWebServices could be “discovered” and used by the geospatial community to integrate spatial statistics.

The goal is to support the national geospatial interoperability and since the RNDT metadata model is a subset of the INSPIRE model, that integration will contribute also to the European geospatial interoperability.

3.2 *Integrating GISTAT into the corporate datawarehouse*

GISTAT has been chosen by Istat as the infrastructure to disseminate census 2011 geospatial data and results (from country level up to enumeration areas level); therefore the integration into DIDAC, the project to disseminate 2011 census data has been designed. That is a very ambitious project, based on the concept of a full integration of GISTAT into the corporate datawarehouse (named I.Stat). It's a great opportunity of growth that will leverage the GIS and integrate it with other statistical information system.

3.3 *The Geoportal*

The final step to make effectively GISTAT a geospatial information provider for statistical activities is the design and implementation of an Istat geoportal that will make available geographic data or georeferenced data, providing catalogue, discovery and download services.

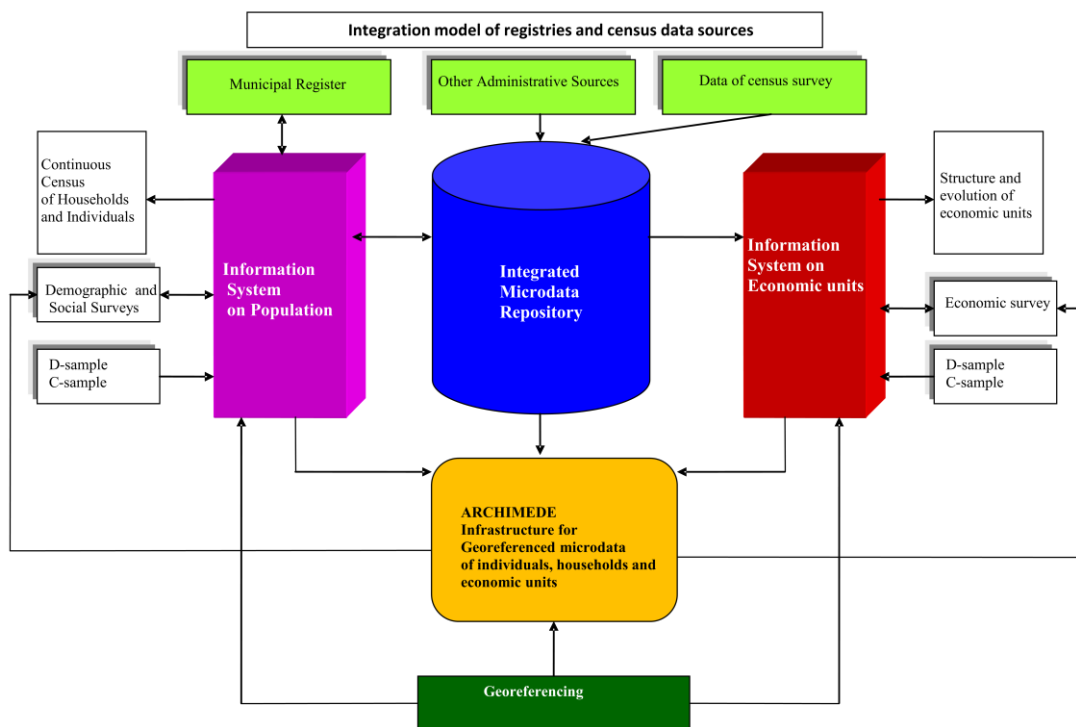
The integration of the RNDT model and INSPIRE model will guarantee the contribution of GISTAT to the European Statistics System.

3.4 GISTAT and the continuous census

Istat has completed the 2011 census and is planning a new future framework. In the Institute's plans, from 2015, the census will be carried out continuously, maximizing the potential of administrative dataset and supplemented by sample surveys for data integration and verification of coverage.

The scheme to be undertaken is as follows:

Fig. 1 - An integrated system of administrative dataset and continuous census data for the production of structural and spatial information



All available files would be properly normalized (in their component of address and street number) and geocoded to the census enumeration area and then disseminated, not only through statistical tables, but also through GISTAT.