WP.2
26 April 2004

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (ii): Development of IT strategies in statistical offices

MODEM and DEMO-GEODEMO: an integrated official web based system for the collection and dissemination of demographic data

Submitted by the National Institute of Statistics (ISTAT), Italy [1]

1. INTRODUCTION

A new web based statistical informative system to product demographic data on resident population in Italy has been recently implemented by ISTAT. This integrated system is composed by two different modules, completely developed by using Internet technology : the first one, called MODEM, provides the opportunity to collect data coming from four different surveys carried out by ISTAT every year on more than 8.000 Italian municipalities. These data are referred either to the whole population – flows and stock data with the distribution by sex, age and marital status – and to the foreign component - with the distribution by age, sex and citizenship. The second module, called DEMO-GEODEMO, is the electronic data dissemination system, based on a web warehouse, which allows users to build dynamic tables, choosing variables of interest and the territorial level of representation, for the same information collected through MODEM. The latest evolution of the system has been the integration with a cartographic component, called GEODEMO, using GIS technologies.

MODEM system drastically reduces information issue times and strengthens and consolidates the data quality improvement process, while DEMO-GEODEMO satisfies the growing demand for statistical information on demographic matters in real time and provides data at the greatest level of detail.

2. AUTOMATION OF THE DATA COLLECTION STAGE: THE MODEM SYSTEM

2.1 Contents and advantages

ISTAT, regularly carries out four annual surveys on the resident population at the Population Registers' Office in each Municipality:

· Demographic flows and resident population by sex (ISTAT form P.2)

· Demographic flows and resident foreign population by sex and citizenship (ISTAT form P.3)

· Resident population by age, sex and marital status (ISTAT form POSAS)

[1] Prepared by Costanza Giovannelli- ISTAT- Social Statistics Department, e-mail:*cogiovan@ISTAT.it;*Vincenzo Patruno – ISTAT Methodological Statistics and IT Department, e-mail: *patruno@ISTAT.it;* Marina Venturi- ISTAT- Social Statistics Department, e-mail: *venturi@ISTAT.it*

· Resident foreign population by age and sex (ISTAT form STRASA)

For these four surveys an electronic version of the forms has been created on a web platform using a CAWI technique; the Municipalities have therefore been given the possibility of choosing whether to use the traditional paper form, or the new method of filling in and sending data using the ISTAT web data collection.

The advantages provided by using the web for producing statistics are obvious and can be summarised as follows:

· Reduction and progressive elimination of the paper questionnaires and filling in of the latter by hand;

· Possibility of entering consistency and correctness checks directly at source, with operators' ability to removing errors directly;

· Use of online guides for filling in forms;

· Control of the questionnaire's route, with consequent reduction of the interview (the person interviewed can be asked only the parts of the questionnaire that must be answered in reference to other filter questions);

· Elimination of sending times;

· Reduction of costs linked to recording and editing of data;

· Possibility of creating user-friendly applications with pleasant graphic interfaces which do not require any particular computer knowledge by the user;

· Independence from hardware platforms; a web questionnaire can be accessed by any personal computer that has an Internet connection and a browser.

The above advantages translate into a substantial reduction in time needed for collecting, revising, recording and checking data, with the possibility of having available information in real time, with obvious improvements in quality regarding the possibility of carrying out more detailed, efficient controls at the moment in which data is collected.

2.2 The data collection site and the application's architecture

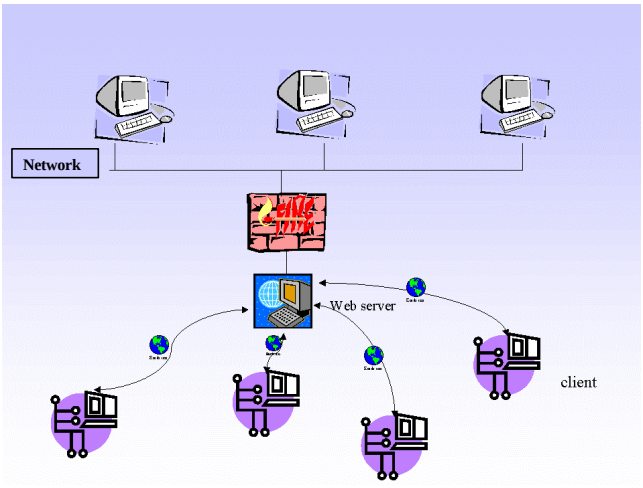
A data collection site based on Internet technology includes two main components:

· A centralised web server

· A browser installed on the clients' stations at the survey office or departments

In our case, in line with the current configuration of the ISTAT computer network, the Web Server is located in an exposed Unix/AIX machine, i.e. accessible from outside the ISTAT network in observance of the security regulations adopted by the Institute to maintain integrity and confidentiality of data.

Figure 1 – Data collection network



The application architecture is, as stated above, client/server and therefore: the data and all the applications are located on the server, but whereas the input/output applications are always carried out on the server, the user interface for acquiring the form and the applications for formally controlling the correctness of the data entered, are downloaded onto the client (i.e. on the PC connected to the site) at the time when connection is made, and is therefore carried out locally.

This allows lower occupation of the network on which the data travels, speeding up the collection operation and the sending of the form.

We can therefore say that, as shown in the following figure, the demographic data collection site (forms P.2, P.3 , POSAS and STRASA) has been created on a three-layer architecture made up as follows:

· The client, which can be a simple PC with an Internet connection and which connects via a Netscape or Explorer browser to the site **http://modem.ISTAT.it.**

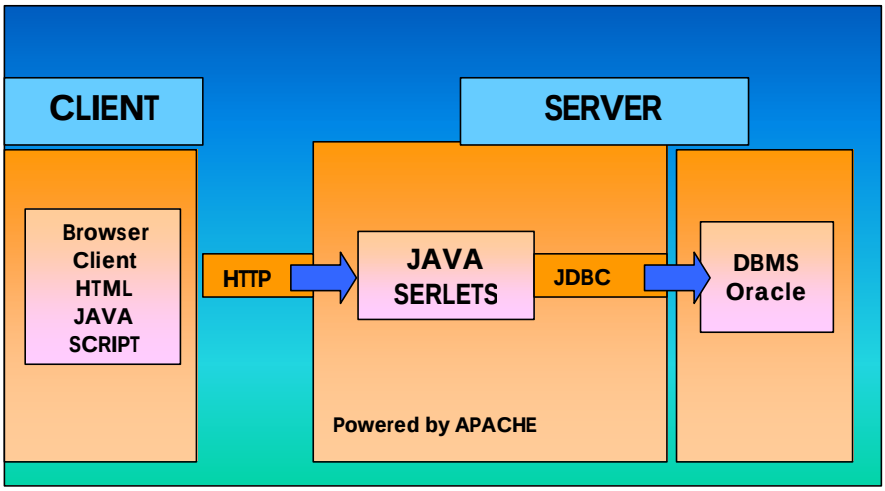
· The web server, which in our case is Apache where the HTML and JAVA servlets are located.

· The DBMS ORACLE server, which we use to store data.

The client requests the service to the Web Server Apache calling for HTML pages and JAVA Servlets. The latter make the connection to ORACLE via JDBC and provide the client with an answer in the form of HTML pages.

To conclude, as said above, all the Input/Output management phase and dialogue between client and server is governed by the JAVA applications, while the data are organised according to a relational structure in DBMS ORACLE. The next page shows an example of the conceptual data model.

Figure 2 –Application's components

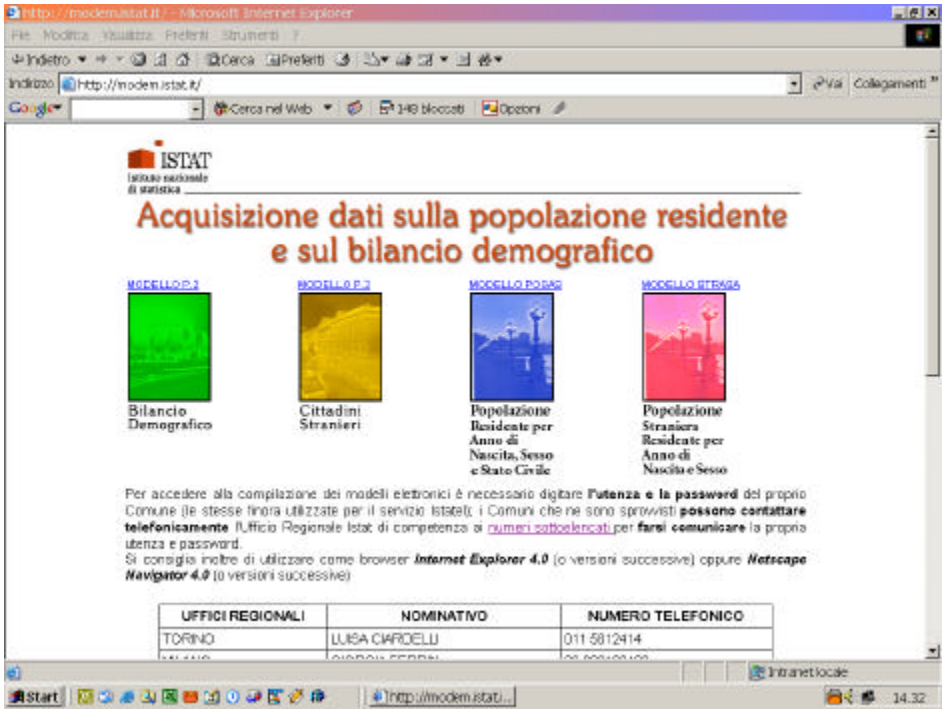


2.3 The electronic questionnaires: structure and basic characteristics

For the four surveys the whole application includes:

· A homepage containing, in addition to the names of the various experts responsible of demographic statistics in the regional ISTAT offices, the links that permit access to the four different electronic forms.

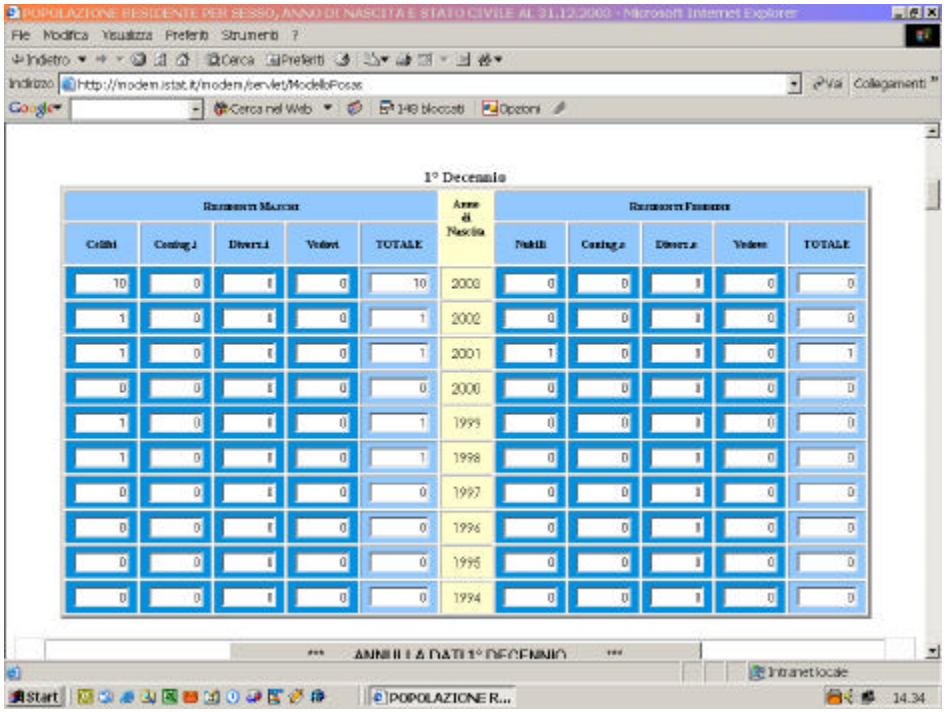
Figure 3 – Homepage



· A second page requests indication of the user ID and the municipality's password; these change from Municipality to Municipality and allow access to only one's own personalized questionnaire

· A third page, the core of the application, which contains the actual questionnaire

Figure 5 – Third page of ISTAT form POSAS



· A final page, which appears on the screen after the user has sent the questionnaire to ISTAT, confirms the correct receipt of the data.

There are two ways to fill in electronic questionnaires; users can insert manually elementary data, or they can upload data from an electronic file, previously prepared, according to a standard format, without compiling the electronic form. This is particularly important because almost all municipalities have their data stored into local databases.

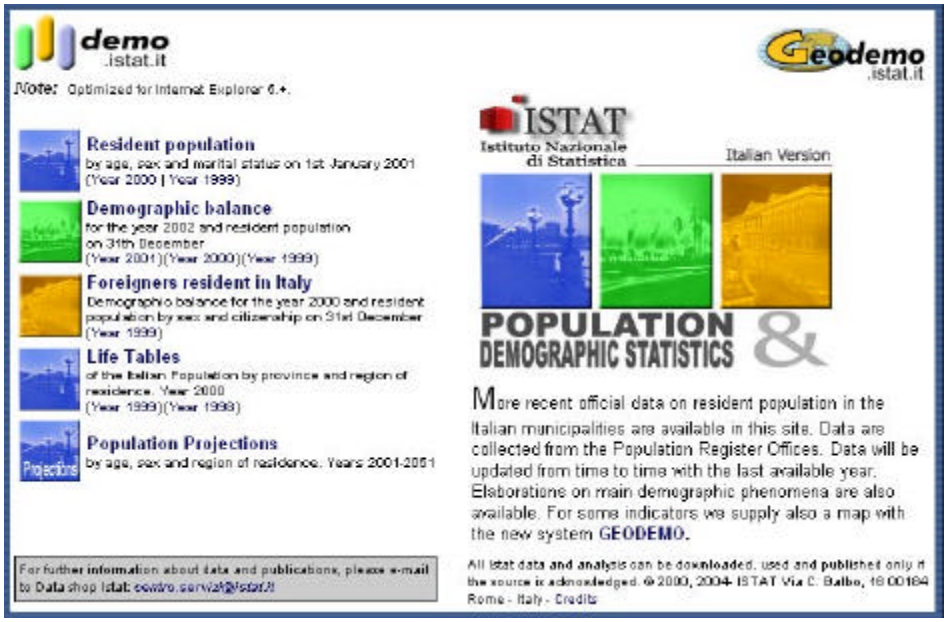
The web based surveys, carried out three times, during the last collection of data, have been used by about 2.500 municipalities, with an increasing quota of respondents: infact, the first year the system was used by about 1.000 municipalities, and the second year about 1.500 municipalities preferred to adopt electronic forms instead of traditional paper forms.

The quite high number of municipalities that have used the application is due essentially to the simplicity of the system. In fact, any particular computer knowledge is requested to the users, neither any application has to be installed on the PC.

3. DEMO and GEODEMO

DEMO is a system for quering data collected through MODEM system described so far. It makes it possible to access data in a very easy way to each single survey, choosing the analisys variables you are interested in and selecting a specific territorial level (the whole of Italy or a geographical area, region, province, or even single municipality). The home page shows clearly the links to the sections regarding the above-mentioned surveys.

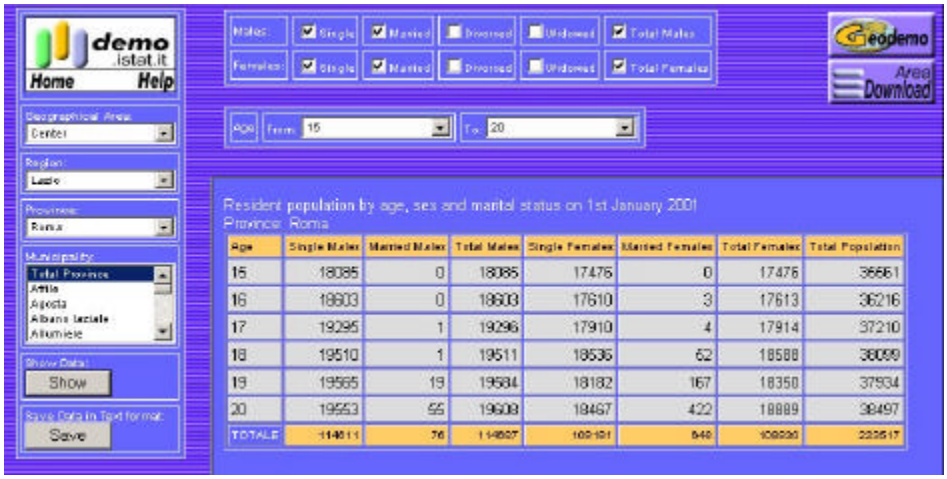
Figure 6 – DEMO-GEODEMO Homepage



3.1 User Interface

The interface was conceived to be of easy and immediate comprehension by the user. Clicking on the desired of application a new page appears which has all the tools necessary for the user to select the parameters and variables of interest. The novelty of the application lies precisely in the fact that all the selection criteria are concentrated on a single page (Fig. 7), unlike the majority of web applications developing in depth and for which, before arriving at the information desired, it is necessary to click various times; since the number of mouse clicks represents the distance separating the user from the information desired, the more clicks are necessary, the farther away the desired information is; consequently, the search time also increases, as well as making the search disorienting and time-consuming for the user. Therefore, the fact of having concentrated the choice possibilities within a single page has made the tool extremely simple to use, while reducing the data access times to a minimum.

Figure 7 – User Interface for the POSAS Survey



The interface has thus been constructed using three frames, two for organising the queries on the database and one for showing, in tabular form, the results obtained. The handling of the territorial hierarchy up to the municipal level is of definite interest.

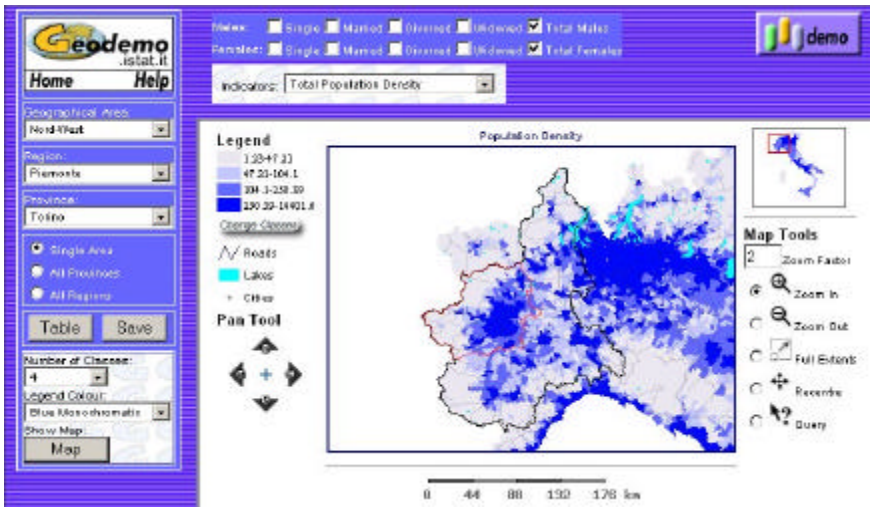
3.2 *The Data Structures*

Data were organised in a relational database. In reality, the tables have been constructed so that they contain not only the elementary data at the maximum level of detail, but also the various aggregations corresponding to the various territorial domains. Any data structure derives from *dimensional analysis* techniques, the same techniques normally used to build up Data Warehouses. We've defined the *dimensional model* and then we've implemented it in a relational database. The database is thus redundant, to the benefit, however, of the performance of the entire system.

3.3 The cartographic component

The latest evolution of the system has been the integration with a cartographic component using GIS technologies. We called **GEODEMO** this system extension**,** to underline the possibility to explore and to report also the territorial dimension of data. Using GEODEMO web interface, users can switch from tables to maps and vice versa using the same page to choose parameters and to display results (Fig. 8).

Figure 8 - Population density for Torino province

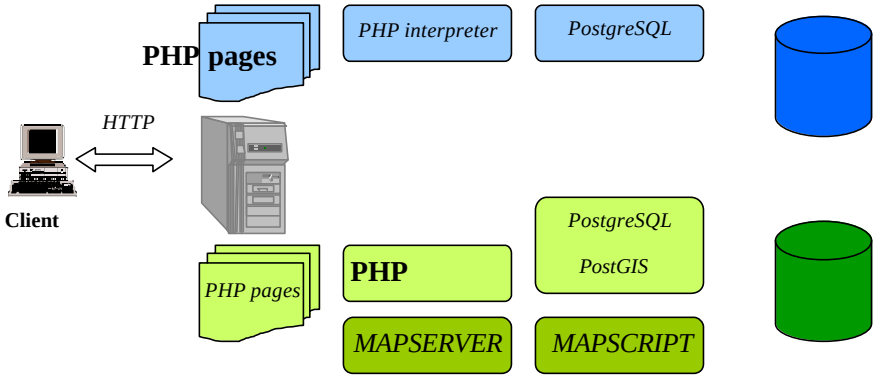


3.4 *DEMO-GEODEMO Architecture*

From the architectural point of view, we can think the system as composed from two different data bases: 1) a relational DB containing all data about population. We designed it according to the Data Warehouse theory to obtain a Data Base specifically structured for querying and reporting. We use this database to generate all tables you can extract from the system and we implemented it using **PostgreSQL** as DBMS. In this case we used **PHP4** as scripting language to build up all applications to capture inputs, to query the DB and to display results.

2) a spatial database containing information in the form of digital co-ordinates (these can be points, lines, or polygons) and in the form of attributes (i.e. demographic information). We use this database to generate on the fly all cartograms and to query all desired administrative boundaries. We implemented it with **PostGIS,** an extension to the **PostgreSQL** object-relational database system which allows ***GIS*** objects to be stored in the database. We needed also two components to complete the software environment: **MapServer** (a development environment for building spatially enabled Internet applications) and **Mapscript** (a set of libraries to "talk" with Mapserver using ***PHP4*** as scripting language). All Software solutions adopted for Demo-Geodemo are **Open Source**, included the Operative system installed on the Server (Linux-Debian).

Figure 9 – DEMO-GEODEMO Architecture



References

[1] Antonia Manzari, Donatella Marsiglia, Sabrina Prati, Marina Venturi "*ISTAT vital statistics: methodological, organizational and technological innovations in data collection and data editing phases*" ISI, Istanbul, 1998

[2] C. Adamson, M. Venerable, "*Data Warehouse design solutions*", Wiley Computer Publishing, 1998.

[3] Mauro Masselli "*Transferring Technology in Data Collection*" NTTS, Crete, 18-22 June 2001

[4] Gad Nathan "*Methodologies for Internet Surveys and other telesurveys*" NTTS, Crete, 18-22 June 2001

[5] V. Patruno, M. Venturi "*The Demographic Web Warehouse on Resident Population in Italian Municipalities*", NTTS, Crete, 18-22 June 2001

[6] Costanza Giovannelli, Marina Venturi "*The MODEM project: Web Data Collection on Resident Population in the Italian Municipalities*" – AMRADS Workshop on Automatic Data Collection - Rome, 15-17 May 2002

[7] V. Patruno, M. Venturi "Geodemo: The Web Warehouse on Demographic Statistics and Analysis with Geographical Information Systems" European Population Conference EPC 2003 - Warsaw, 26-30 August 2003

- - - - -