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

EUROPEAN COMMISSION EUROSTAT

Work Session on Electronic Raw Data Reporting
(Geneva, Switzerland, 6-8 November 2006)

CENSUS DATA COLLECTING BY INTERNET. SPANISH EXPERIENCE IN 2001

Adolfo Galvez Moraleda, INE, Spain

Introduction

Spain has conducted census operations regularly from the second half of the 18th century. During the last century, censuses of dwellings and population have been undertaken once a decade.

In 2001, the National Statistical Institute of Spain (INE) relied on a strong technological component throughout the entire census process. Main issues to be pointed out in this field are:

    a) Questionnaires were "pre-filled" with personal data from the Population Register. They were carefully designed to allow digitalisation into a high performance system, which enabled most of the information to be processed within three months after collection.
    b) The option to complete questionnaires on the Internet was made available for the whole population.
    c) The dissemination system is based in datawarehouse technology. This includes OLAP navigation techniques that allow users to design tables according to their needs and analyse the information presented requiring more details and at the same time guaranteeing statistical confidentiality.

Census was based on Spanish Population Register (Padrón). Padrón is established in each municipality to record individual residence information. This information is limited to only a few characteristics: name, sex, date and place of birth, citizenship and educational attainment. In 2001 Census the questionnaires were "pre-printed" with the Padrón data to be confirmed or updated by respondents.

More than 30.000 interviewers worked on data collecting, firstly distributing paper forms and after a short time picking up them from the dwellings.

Census data collecting by Internet

1. Main characteristics

Main drivers on this project were:

    - Internet data collecting method would be available for the whole population.
    - Security proceedings would be essential to accomplish the task.

- We needed encourage internet use.
- Making the system as "user-friendly" as possible.

In the course of developing some challenges had to be faced up to and solutions were adopted for all of them:

a) Authenticating the users:

Two kind of users may be identified:

- Those who need to update some information of the data that came pre-printed in the questionnaires.
  These people must have an electronic certificate (X.509 class 2 from FNMT – a Spanish certification authority) to process their forms.

- Those who only complete the questionnaire, thereby confirming the Padrón data.
  For these people, two personalised keys were included in the envelope that were distributed with census forms.
  These keys were different for each dwelling and they were stored in a RDBMS together with personal data that were not printed in the questionnaires (name of the parents in the same way that appear in the identification card of each citizen). This personal information had to be provided to begin census completion by internet. Alternatively, they can choose to identify themselves by an electronic certificate.

b) Flexibility on census completion

It was possible for the users to save partial information and continue the fulfilling task later. When finished, users were given a ticket to confirm that the questionnaire was completed successfully.

Users could choose the language they want to complete the questionnaire. All co-official languages were supported and also English, French, German and Arabic languages.

c) Quality of the information

Validation rules were implemented in order to ensure the consistency on the information provided by respondents.

d) Accesibility

The system was designed to be usable for people with disabilities.

e) Coordination with our field network of interviewers

Supervisors received weekly files via e-mail with the information about questionnaires fulfilled through the Internet (passive mode). Besides this, they can contact with the system via web interface to make queries about this subject.

2. System's software and hardware

The technology and architecture of the system ought to be flexible, scalable, reliable, and safe. High performance and high availability were also compulsory characteristics.

The data collecting system was implemented following the three-tier paradigm (data, business logic and user interface) and it was developed using Java components.

From the client side the application was tested using different browsers and no additional piece of software needed to be downloaded by the clients. Only web browser with HML 4.01 and Javascript 1.4 support were required.

Hardware components covering different aspects included:

- Security : Firewalls, intruder detection and antivirus appliances were set up in different locations on the network topology.
- Performance: Load balancing elements, clustered servers with redundant components.
- Confidentiality: Server with SSL3 certificate.
- Integrity: Back-up server

3. Results and analysis

Despite the system was sized to support high load of concurrent users and was designed to have a good time to answer, only 13.818 households fulfil the questionnaire through this channel. This was approximately $1‰$ of total households. From them, 29.9% has used electronic certificate to authenticate to the system.
Besides this, 16238 households had tried to modify personal data but they gave up doing use of electronic signature.

No basic and essential cause have been found to explain such low response rate for this experience. Instead of this, it can be thought that multiple reasons can be influenced on it.

a) **Lack of publicity**
The action has been deemed insufficiently advertised. Anyway, throughout the collection period specific campaigns were launched and no increase in collection were detected on later dates.

b) **Delay in starting the action**
The possibility of using Internet started a week after the action of data collecting by paper.

c) **Coordination problems with census interviewers.**

Deficient information for some interviewers derived from the confusion in the payment of the questionnaires collected by Internet caused that they don't encourage people to fill the questionnaires electronically. As soon as the problem became known, serious efforts were done in order to correct it. So, after a few days everybody had good information about the situation and instructions to encourage people to use internet channel.

d) **Difficulties with the electronic signature.**

This reason cannot account for the low response rate in the households for which the use of the electronic certificate was not compulsory. But it is true that the process to obtain this certificate was not very easy for most of households in 2001 in Spain.

e) Lack of incentive to fill in the questionnaire by Internet instead of paper.

The paper forms were easier to complete than by internet and the incentive of filling electronically (the geographic distribution of surnames delivered to the user) was not enough attractive for the users.
Besides this, people must to wait to receive paper forms with the envelope with identification keys to answer by the internet.

f) Scarce presence of Internet in the households

Perhaps the presence of the Internet in households was lower than expected.

g) Some technical issues

Technical difficulties regarding with passwords, questionnaires with several households,… could be influenced on some users' resignment. Planning with a large testing period would be advisable.

4. Learnings

Despite of the low response rate, Internet has revealed as a useful way of collecting census data.
Ten years after, probably some of the reasons pointed upon will be surpassed easily but probably a mixture of procedures will be needed to approach data collecting in the next future.
To prioritise Internet (and the telephone) as the first possibility to fill in the forms by households, it would be advisable to offer them chronologically before the traditional methods (like Singapore in 2000) with enough publicity of this matter.
Only the households not using those channels will be contacted as it has been done in last census.

Madrid, October 2006