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 (i): Web technology in statistical information systems

THE USE OF A DATA CAPTURE TOOL IN STATISTICAL SURVEYS ON ENTERPRISES

Contributed Paper

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

Summary

I. INTRODUCTION

1.   Statistical surveys normally involve three respondent classes: families, institutions and enterprises. Each of them has its own different characteristics, which can influence the operational procedures and the choice of the most appropriate data capture technique.

2.   Families are used to interacting with interviewers, so CATI/CAPI techniques are more utilized than others while innovating the survey process. Institutions are usually better equipped with advanced computer technologies to allow a more efficient data capture. Enterprises fill in statistical forms that are mainly sent and returned by mail or by fax. The technical equipment of the enterprise may be very heterogeneous, depending on the size of the company.

3.   The choice of electronic forms and the use of the Internet cannot be imposed on the users. The best way to renew the data collection phase is to adopt a mix-mode technique, which allows users to choose by themselves the response method: either traditional (paper forms) or innovative (electronic forms data).

II. TELEFORM SOFTWARE

4.   Teleform is a Cardiff product that needs a Windows server platform (NT, 2000, 2003 SERVER). It was chosen by the Data Capturing Laboratory-IT Directorate of ISTAT for its versatility in managing multiple data capture channels. The information capture system is capable of processing thousands of paper forms per day with accuracy, speed and efficiency. It generates electronic forms in HTML or PDF, which can be administered by web or by e-mail. It allows the fax-server use for mailing and receiving paper forms,

---

[1] Prepared by Rossana Balestrino (rossana.balestrino@istat.it); Patrizia Larese (patrizia.larese@istat.it)].

which can be processed digitally by the optical character recognition machines included in the system. TELE*form* works seamlessly with production-level scanners, fax servers and Internet to capture, verify, process and index data.

5.   Three modules form the product::
- *The Designer* to define forms;
- *The Reader* to manage the form mailing, receiving and reading;
- *The Verifier* utilized by the operators to verify the quality of the data captured

6.   The Teleform versatility requires high management costs to maintain suitable service levels. The complete application needs 6 servers (Teleform server, WEB server, Input Mail server, Output Pop server, FAX server, DB server) linked in a network.

III.   PROCESSED SURVEYS

7.   Timeliness in the data collection phase is crucial for enterprise surveys and especially for short-term statistics because data have to be available within deadlines fixed by the European institutions. The Teleform application, due to its features and respondent typology, has been considered suitable to optimize the data capture for these surveys.

8.   During the last few months a few projects were realized in the statistical units that are autonomous for the development solutions.

9.   Some descriptive data of the surveys treated in an experimental way are reported below:
- The monthly survey on detail sales uses 2 forms for three pages. The theoretical number of respondents is 2,700. It was undertaken by e-mail and by web.
- The annual survey on telecommunications uses three forms for nine pages. The theoretical number of users is 534. It was undertaken by web, fax and fax-server.
- The monthly survey on employment, working time, and salary, uses two forms for three pages. The theoretical number of users is 1,500. It was carried out by web, by fax and by fax-server.
- The quarterly survey on vacancies and work hours uses one form for four pages. The theoretical number of users is 8,000 for every quarter. It was done by web, e-mail, fax and fax-server.

IV.   PERCEPTION OF THE STATISTIC USERS

10.   The results of the experiment were not perceived in the same way by all statistical users. Some of them were not completely satisfied and confident in the new technique. The IT Directorate, responsible for addressing and coordinating the data processing function and systems, decided to evaluate the software capacity and identify constraints and requirements for a massive start in production.

V.   THE SOFTWARE EVALUATION

11.   The tests confirmed the Teleform versatility about the different channels of mailing and receiving forms. The application develops a standard multi-channel solution for a *thin* form (1-2 pages with a few controls on form fields) at low cost. It does not require any application development because the executables are automatically generated by the form created with the Designer module.

12.   The processing of the paper forms by the optical character recognition machines is an important capacity. The tests on the quality of the automatic recognition of characters gave good results, in comparison to other available optical reading software from the market.

13.   On the other hand, Teleform was not planned for statistical targets. It is not able to process complex forms, such as those for investigating social phenomena. It is not possible to save partially the hand written

forms. To use the electronic form, it is necessary that the users PCs have particular requirements. In a network, Teleform, as a Cardiff requirement, works with a network card high quality at 100 Mbps in the server and in the client.

VI.   MANAGEMENT AND ORGANIZATION COSTS

14.   The complex architecture of Teleform needs minimum requirements to perform well. The network requirements are the most critical because they have to grant the communication between the modules. A complete Teleform application needs six active servers and good network performance.

15.   To renew the data collection process, ISTAT, in addition to deciding which specific data capture tool to use, has to support the costs of the installation and maintenance of an adequate infrastructure. It has to install monitoring and management tools. Moreover, to use Teleform, statistical units should accept changing the organization of their human resources.

VII.   PROSPECTS

16.   The use of Teleform is recommended with *thin* forms and a computer network at 100 Mbps. New procedures will be set up to monitor the different service levels of the processes. The more complex forms require a further testing activity of other software products, both commercial and available in the statistic al official community. In particular the BLAISE software is a standard product present in ISTAT for CATI/CAPI data capture. A test of the Internet BLAISE module has been scheduled.