

Distr.
GENERAL

CES/SEM.47/29
18 January 2002

ENGLISH ONLY

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

**COMMISSION OF THE
EUROPEAN COMMUNITIES**

CONFERENCE OF EUROPEAN STATISTICIANS

EUROSTAT

**Joint UNECE/Eurostat Seminar on Integrated Statistical
Information Systems and Related Matters (ISIS 2002)**
(17-19 April 2002, Geneva, Switzerland)

Topic I: Application of web technology to integrate statistics

USING THE WEB IN CAPI APPLICATIONS

Contributed paper

Submitted by the Central Bureau of Statistics, Israel¹

I. INTRODUCTION

1. The year 1999 was a special year for the Israeli Central Bureau of Statistics (CBS). We moved into a modern new building. The move brought with it new opportunities for advanced methods and technologies to be applied. One of which was a technological infrastructure upgrading that was applied successfully including all the subjects related.

2. This translocation was followed accordingly with two major strategic principles: one at data security policy level and the second in a data collection strategic plan.

3. **Data security policy** - The ICBS should go on, as a policy, with a strict security system that is based on the physical separation between the two parallel communication networks. Until 1999 the Bureau of Statistics in Israel had the best security system for keeping outsiders from accessing the sensitive and concealed databases. It was simple - there was no direct communication between the organizational network and the Internet. The web was a standalone server and was not connected to the organizational network. As we moved into the new building, the meaning of the data security policy was a complete disconnection between: (i) the **Intranet** network including the organization's main data bases is disconnected from the outside world, and (ii) the **Internet** network connected to outside users and to the Internet.

4. The Intranet network became a well-developed network with a lot of services for the inside users, while, the Internet network was established as a relatively limited network with minimum services. The ICBS encourage the inside users to use the Intranet network as much as possible. The use of the Internet network

¹ Prepared by Shifra Har (shifra@cbs.gov.il).

should be only for what is really needed and could not be done in the Intranet network (and is permitted by the organizational security policy).

5. **Data collection strategic plan** - A significant change in the concept of data collection management in all surveys. Until 1999 data collection were carried out through regional offices and the data was sent to the center as requested by the surveys. The new concept was to create a new and modern **national** center for data collection that centralizes the data collection management in one place, and thereby minimizing the regional offices activities until complete closure in the future.

6. Meanwhile we started to run a modified Gap system adapted to our own specific needs. This security tool enables our organization to connect the two networks virtually although they remain separate.

II. THE NATIONAL DATA COLLECTION CENTER

7. Thus, at that time an interdisciplinary committee based on professional employees, was assigned to examine all the different aspects of the changes in the policy. The new one is based on the concept of a national data collection center. The committee that consisted of four bureau unit experts (surveys unit, statistical methodology unit, subject units and information technology unit) recommended the establishment of such a center as soon as possible. The next stage was to examine carefully all the different aspects resulting from this new policy and find the best principal guideline for establishing the optimal national data collection center.

8. Although the committee's responsibility included all aspects of the national data collection center (economical organizational and technological points of view), in this paper I will concentrate only on the technological aspect.

9. The committee came up with several main guidelines. **First Principle** - the center will be responsible for all data collection types and all the various technologies in the organization, either mail questionnaires, paper questionnaires, telephone interviews, Face to face interviews or using the web for questionnaires and interviews. Indeed, in the first version, the center will be suitable to the technologies we recognize today but should be adapted dynamically to new technologies in the future.

10. **Second Principle** – A central control system should be developed for every survey carried out in the national data collection center. Until 1999 several systems for data collection were developed on different software and every control system was dedicated to its particular survey. According to the new policy, one general system should control all new surveys that will be developed. Such a system enables the survey unit, who is responsible for the field activities, to guide the interviewers and manage their work. It also enables this unit to observe closely the survey's progress in the field. This tool is supposed to improve the unit's work efficiency and enable easy human resources management and therefore allow efficient resource utilization. Although there are great advantages in using this tool, we cannot force the subject units to use it, we believe that the advantages and the low cost will lead to our aim to create one control program to all surveys.

11. **Third Principle** – For efficiency, every interviewer should carry out several surveys. In order to enable the interviewer to manage the surveys he needs a shell tool. This tool will allow him to manage his daily work. He will be able to plan his interview route, and receive all the necessary statistics.

12. **Fourth principle** - In an individual surveys, when we are looking for an individual and the first address is found to be incorrect, the system will provide alternative addresses. In order to be able to do it, the system will give the interviewers several alternatives addresses and will lead him to choose the address where

he is most likely to find the person he is looking for. These addresses are gathered from several files in the data collection center, such as: drivers files, telephone files, and censuses files.

13. **Fifth Principle** – The data collected from all over the country should be available to all employees in the various ICBS units who are working on and are authorized to access and to update the data. The data should be available immediately upon arriving at the center.

14. For each survey there is a responsible team that includes representatives from the relevant subject unit in addition to the constant representatives of the following units: survey unit, statistical methodology unit and information technology unit. Therefore, access to information should be permitted only under the right definition groups of users for each survey.

15. In addition to the general principles mentioned above, there are three more basic principles that relate specifically to the CAPI (Computer Assisted Personnel Interview) systems. **Sixth Principle** – Every CAPI survey, requires information transfer from the center to the interviewer outside the organization, using a BI-directional pass. Therefore, this communication has to be a universal tunnel, that will enable a back and forth information transfer from the interviewers (collected data) to the center and back (survey plan and other information). The new developed tool has to be compatible to all the already existing systems in the organization. Since there are CAPI systems based on various technologies, the system should be independent of the development technology of the survey.

16. **Seventh Principle** – The use of the Internet as a medium for data transfer between the interviewer and the center (and vice versa) in CAPI systems. Here we have to consider seriously, the severe security problem and solve it properly with the satisfactory security tools, so that in spite of everyone's free accessibility to the Internet, the process of data transfer will be highly secured.

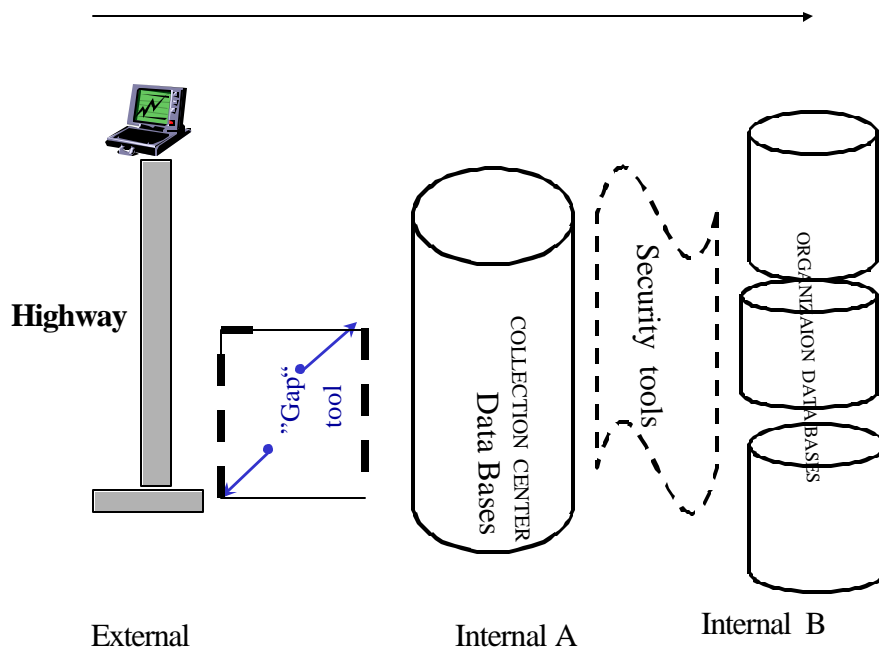
17. **Eighth Principle** – the information transfer between the interviewer and the center should be carried out with a standard modem within the Laptop computer by either connecting to a telephone line at the interviewer's home or using his cellular telephone as a modem (without any need for a telephone line).

III. THE UNIVERSAL PASS BETWEEN THE INTERVIEWERS AND THE CENTER VIA INTERNET IN CAPI SYSTEMS

18. Since the main issue of this paper is the universal pass between the interviewer and the national center via the Internet, we will focus on the sixth and seventh principles. Each interviewer in the organization has his/her own private "area" in the data collection server. There are two kinds of "areas": for sending data and information to the center and for receiving data and information from the center.

19. Every interviewer contacts the center and after a security check he/she is navigated directly to their private "area". After passing the security check the data is moved to another server in the center.

Fig. 3: The information transfer procedure



20. On the left-hand side of the figure, we can see information delivery from the interviewer's laptop computer via the Internet, passing through the Gap system into the data collection center (in the center of the figure). On the right-hand side of the figure we can see the internal network (Intranet) of the organization that allows access to the data collection center via an authorizing mechanism.

21. When the interviewer enters his/her "area" they will view their own home page. In the homepage three main activities can be executed: data transfer (bi-directional), view the general messages that were sent to all the interviewers and the personal messages sent only to him. The personal messages can be viewed only by the recipient.

IV. WHAT TYPES OF INFORMATION ARE INCLUDED IN THE NATIONAL DATA COLLECTING CENTER?

22. In the center, several types of data are gathered. Some of it is **raw data** collected in the field per survey (exactly as in the questionnaires), and another part, the administrative data is used by the organization as a whole and the survey unit specifically.

23. The **interviewer diary** is the collection of his activities in detail. For example: the number of working hours, the number of travel hours, difficulties with finding addresses etc. Such information enables us to calculate the interviewer's salary, and at the same time to follow up on the integrated work of all the interviewers in the field in order to get the global work plan of the survey.

24. **Unsuccessful visit**— can happen in at least two cases. In the Household survey, the interviewer found the address but did not manage to perform the interview, or in an individual survey, he found the address but

could not find the person he was looking for. In both cases the interviewer will note the activity in his notebook (paper), and later, at home enter it into his laptop computer

25. Lately, we have developed a system that enables the interviewer to enter some required data using a mobile phone, in an unsuccessful visiting. At this moment it is valid only for closed questions because of the technological limitations, so it is running as a pilot system so far. We intend to go on with development until we will reach an efficient system.

26. **Supporting files** for control (census, enterprises) – these files are installed once for all the surveys and used for controlling each of the surveys. These files exist already in the organization's database and can be very helpful to the tests we are doing in the center

27. **Administrative files** – In the center we accumulate in the server information about the questionnaire stages. Each stage has its own status number. These files allow us to oversee and control each of the questionnaire work-flow stages and therefore we are able to react quickly to an event. This status describes who is responsible for each stage? Whether the questionnaire had already been transferred via the Internet? When was it? How many times was it redelivered? What was the address? Who are the present responsible interviewers? Did the interviewer already start the interview? etc. These data are used by the field unit and by the technology unit on an everyday basis.

V. CONCLUSION

28. Our vision from 1999 about the national data collection center has been realized almost completely. The national center was established and the majority of our new surveys are performed in the technology we had recommended. Most activities in the organization today are fit for the center's purposes exactly as was defined. Several new Capi systems started with the new guidelines. The development is still going on and it seems that it will never end. New technologies are studied currently and the most suitable will be assimilated in this national center.

References

Har S., Manor N., Using A New Solution for Interactive Communication Between the Two Physically Separated Networks in The Central Bureau of Statistics Israel, UNECE/EUROSTAT Work Session on Electronic Data Reporting (Geneva, Switzerland, 13-15 Feb. 2002).