Distr.
GENERAL

ECE/CES/SEM.54/11 (Summary)
11 April 2006

Original: ENGLISH
ENGLISH AND RUSSIAN ONLY

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

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint UNECE/Eurostat/OECD Seminar on the Management of Statistical Information Systems (MSIS)
Sofia, Bulgaria, 21-23 June 2006

Topic (i): Changes in statistical processes

INTEGRATED SURVEY TECHNOLOGY (IST) - DIFFERENT APPROACH TO STATISTICAL DATA PROCESSING

**Supporting Paper prepared by Branko Jirecek and Mira Nikic,
Statistical Office of Serbia and Montenegro**

Summary

1. For a long time data proccessing in statistical systems was (and still is) organized according to the "stovepipe" principle, which results in a set of various, often incompatible, solutions dealing with similar problems. A more general approach was designed and implemented in the Statistical Office of Serbia and Montenegro, which deals with proccessing of all the statistical data in a completely different manner.

I. SHORT HISTORY

2. Automatic data processing in our country has a very long history. The first computer bought in the former Yugoslavia was the computer bought and installed for the needs of the Federal Statistical Office (as it was previously named). For a long time this Office was the leader in progress and development of IT, because innovative techniques in data processing were always implemented here.

3. Recently, there was a huge decline in the activity of the whole Office, as well as in the IT department. The latest “cut down” in the State Union administration left only 4 (four) people in

the IT department. This certainly is not a situation where a lot can be expected, but, nevertheless, we present some results that have been achieved.

II. CURRENT SITUATION

4. The most usual way of processing the data of a statistical survey is the so-called, "stovepipe" principle, where the complete process is performed – from data entry to publishing the final results separately for each survey. The processing of every survey data consists mainly of the following three phases:

(i) Data entry
(ii) Logical control (data "cleaning")
(iii) Tabulating (dissemination, publishing)

5. The development of IT meant that various surveys had various, at that time "popular", implementations. Depending on many factors, such as the importance of a certain survey, (un)willingness for change etc., we had (and still have) the situation where various surveys are IT "handled" in a huge variety of ways – from surveys where data is stored in "plain" sequential files to those where they are in (unfortunately) various databases.

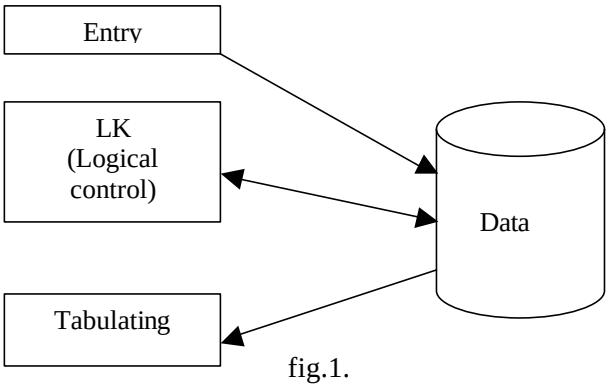

fig.1.

6. For a long time, IT was exclusively done on IBM mainframe platform(s). Naturally, the development of PCs caused some surveys to be "down-sized" to this platform, and, of course, we have "mixed" solutions, just to make the situation even more complicated.

7. We, therefore, have a situation where the processing of statistical data is done on various platforms, implemented with various software tools, with data stored in various ways, separately for every survey, and even worse, for every phase, as shown in fig. 1. This is a very simplified picture, which is done deliberately, because this was the biggest issue that we found all the surveys had "in common" and from where we had to start.

IV. BASIC CONCEPTS OF INTEGRATED SURVEY TECHNOLOGY (IST)

8. It should be mentioned that, although we contacted relevant people from several statistical offices to solve the above-mentioned problems, no one seemed to be interested. Hence, we had to start "from scratch" and "on our own". The only exception we found was EUROSTAT's Euro Trace, but, as it is not able to deal with most of our surveys, we also heard about it too late, since at that time we had already developed the main foundations of our idea.

Nevertheless, it was very helpful to exchange experiences with the people working with Euro Trace.

9. During all these years we have been writing, more or less, only three applications referring to data entry, logical control and tabulating. We felt that it was high time for these three to be finally “written” so that we would not ever (at least in the way we did until now) have to “worry” about them again.

10. Although there is strong opposition to universal applications, we had some rather successful experiences in the past. We would like to mention TABS, a universal tabulating program developed for the IBM mainframe in the Federal Statistical Office and GoDar, a data entry and logical control system for the same platform, developed in the Statistical Office of the Republic of Serbia. Although both of these solutions were outdated a long time ago, we were positively encouraged with the existence of this kind of tradition.

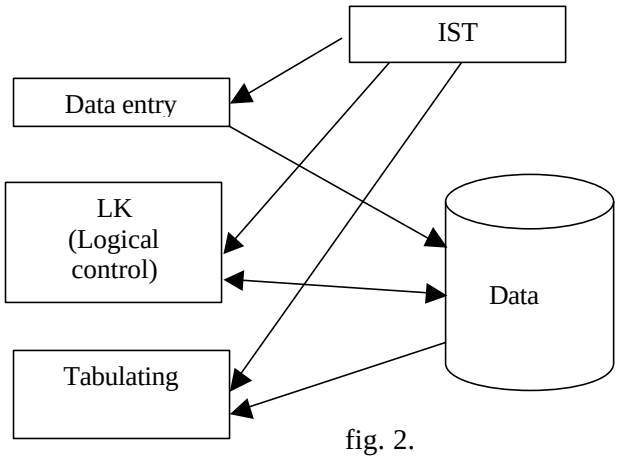

fig. 2.

11. It was obvious that data had to be stored in a relational database. The problem was which database (DBMS) to choose, because we already had two in use (IBM’s DB2 and Microsoft’s SQL-Server), without any knowledge about the possible expanding of this number. For this reason we had to make our solution independent of the physical implementation of the database, which, on the other hand, made the limitation of using only the most basic set of SQL commands, with absolutely no enhancements or improvements of any DBMS.

12. The main idea, as shown in fig. 2, was to try “organizing” the surveys so that they would “take care” of their own data and to process them, instead of the current situation where every survey is a “world for itself”. Technically speaking, IST shown in figure 2, is a database where all the relevant survey data, both operational (such as data description, conditions for logical control, definitions of the output tables) and descriptive (methodologies, field descriptions, etc.) is stored. The TU-11 survey in the field of tourism statistics was the first pilot project and it is currently working “on IST”.

13. However, there is still a lot of work on the whole idea, many things are to be improved, written and/or rewritten, but we are glad that it is brought to an operational level, which means that the whole idea was correct.

V. IMPLEMENTATION

14. IST, as already mentioned, works with any relational database. It would perhaps be better to say that it works with any database, because there is no limitation that data must be in only one database, the data from one survey can be in different (even physically different) databases,

which is not so hard to expect, because this way we are planning to have unique registers, classifications and other “consulting tables” for all surveys.

15. There was a problem with the programming language in which this would be implemented. Finally, we decided to do it in Microsoft Access, in its VBA. The main reason for such a decision was that Microsoft Office is already installed on every PC in the Statistical Office, and users (statisticians) are familiar with it, so they may be able to do some “elementary” operations in Access (certainly much easier than in, for instance, DB/2).

16. The user (statistician) is supposed to choose the survey he wants (and has the right) to deal with, and either input data, logically test and clean it, or produce outputs. The logical control phase is the most interesting one, where we allowed the user either to test and change the data “locally” (on his PC) or “globally” (directly in the database where the data is stored, enabling more people to simultaneously perform this operation).

VI. CONCLUSION

17. Although there is still a lot of work to be done, we are satisfied with the progress we have made up to now. The whole system is functioning in the Statistical Office of Serbia and Montenegro and the work of implementing it in the Office has already started. It has also been offered to MONSTAT (Statistical Office of the Republic of Montenegro) and we hope that the implementation in that Office will start soon as well.

18. We are open to all sorts of discussion and are willing to cooperate with other statistical offices in this matter, hoping to share experiences and gain further knowledge.
