STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 5
English only

CONFERENCE OF EUROPEAN STATISTICIANS

Joint UNECE/EUROSTAT Work Session on Electronic Data Reporting
(Geneva, Switzerland, 13-15 February 2002)

Topic (i): Management, organisation and policy issues

E-CORD USING INTERNET, SCANNING AND MOBILE TECHNOLOGIES IN THE CZSO

Submitted by the Czech Statistical Office[1]

Contributed paper

I. DATA COLLECTION WITHOUT ASSISTANCE AT THE RESPONDENT'S

1. A possible breakdown of the various statistical surveys conducted in national statistical agencies is by data collection method. One group includes surveys in which respondents prepare and transmit data without being helped by the statistician, and the other group deals with surveys taken of respondents by the interviewers. It might be of interest to note that the numbers of the Czech Statistical Office (CZSO) staff increasingly engaged in data capture and checks are split evenly between the two groups (approximately 380 each). Obviously, the numbers do not include such events as censuses, particularly the housing and population census, the periodicity of which is over one year.

I.1 Processing of statistical questionnaires

2. The CZSO ensures the collection and processing of approximately 120 questionnaires with a periodicity of one month to one year. They are split into the following groups: industry, construction and investments, trade and services, agriculture, forestry and the environment, banking and non-production sector, social statistics, demography and wages. The questionnaires are processed in seven nation-wide data processing departments located in capitals of the regions. At these places, questionnaires from exhaustive and/or sample surveys are gathered from respondents throughout the whole country, either on paper or in electronic form. The departments carry out the primary processing with checks and corrections using LANs and the data are in turn transmitted via WAN over long distance to Prague to be centrally processed in a database system.

3. Given the large number of questionnaires processed in parallel, the frequent modifications to the questionnaires and especially to their processing projects, and the extensive volume of data processed, it was essential to consistently standardize the procedures involved in the process, to formalize individual steps accordingly and to develop specialized tools.

4. Using information technology, each project of the processing of statistical questionnaires can be divided into the following stages:

- ?? technical preparation of statistical surveys
- ?? electronic collection of statistical questionnaires
- ?? primary data capture and checks
- ?? storage of data and grossing-up

[1] Prepared by Jan Matejcek.

?? processing of tabulated outputs.

5. The technical preparation of statistical surveys encompasses a sequence of interlinked and intertwined activities carried out in close cooperation between the individual CZSO departments. Within each framework, links are established between subject-matter requirements and technical aspects of the implementation of each project. The preparation is split into the following steps:

?? development of technical project;
?? technical design of the electronic questionnaire;
?? software preparation of individual components;
?? debugging and testing of software.

6. The key software product supporting technical preparation is the ProjektMan (see part 1.2). Electronic collection of statistical questionnaires is dealt with in part I.3.

7. Due to the complex structure of a statistical questionnaire today, and given the requirements of methods employed to assess the quality of data files produced, the data capture programme is fairly sophisticated. Such a programme is operated in a network mode with the necessary cooperation of other programme components. High demands are also made on standardization of these tasks regarding work with basic populations and samples, standard installation and control of the tasks and procedures for efficient provision, exports and imports of data files. Currently, the CZSO mainly uses the standard programme platform DataMan for primary data capture and checks, which is incorporated into the CZSO information environment through connection to the database system Oracle on the one hand and to the system ProjektMan on the other hand.

8. Having been checked and corrected, the primary data are loaded into the basic database operated under the database system Oracle in the production servers of the central processing in Prague. This database is designed to gross up individual data files and to serve as a source to generate statistical outputs. Grossing-up methods and their programme implementation are complemented by a programme protection of individual data. Users from the CZSO can also access the database through a number of standard products to make individual selections from the database. A public database is under preparation. It is designed for immediate work of users outside the CZSO and is in the initial stages of conception and pilot project.

9. The processing of tabulated outputs for verification is part of data capture and is generated in the product DataMan. A generator of tables is employed to develop software for the processing of statistical tables from the Oracle database. This is a product made for programmers, which considerably reduces labour input and contributes to unification of the programme tools used. The generator of tables contains a module for conversion of tables into html format, which makes it possible to use web technology means for storing and viewing. The module also generates a search tree automatically using defined characteristics of individual tables.

I.2 ProjektMan

10. ProjektMan is the basic SW product, which supports the technical preparation of surveys in the CZSO. In its framework, comprehensive preparations of the questionnaire design and the processing of technical projects are carried out.

11. As a highly specialized graphic editor, ProjektMan supports the generation of statistical questionnaire sections, including required check links, explanatory notes, and respective technical attributes of individual items in the questionnaire. The defined sections are in turn used to make up a statistical questionnaire for a given year and periodicity of survey. Currently, ProjektMan also generates texts for the CZSO decree on the programme of statistical surveys, which is brought out in the Collection

of Laws. The output of this step of ProjektMan application includes matrices of questionnaires and of the decree for the print shop.

12. All information on the statistical questionnaire, including its graphic form, is accepted as the basis of a technical project. The technical project is then complemented with other required information such as the definitions of outputs, classification of check runs, description of checks, time schedule of programme preparation, time schedule of processing, grossing-up method, archiving of data, etc. Standard projects are developed on a fixed defined outline. They mostly take the form of a free text with syntactically defined sections, which permit the generation of parts of software directly (DataMan) or of relevant parameters (Oracle). This is the method used to support the generation of capture programmes, populations/samples of units to be surveyed, check runs, descriptions of the files, and the like.

13. The projects, questionnaire sections and questionnaires themselves are stored in the database of technical projects. As a rule, the database embraces the entire set of all projects in force for a given year of surveys, which implies that no addenda to the projects are made. The database is accessible to all users connected to the CZSO network. The ProjektMan and all necessary files can be installed on independent PCs. A certain amount of synoptic information, such as processing time schedules for any task, is available from the database, always with the option of print.

I.3 Electronic collection of statistical questionnaires

14. To streamline the process of data collection, the CZSO laid down a project of electronic provision of questionnaires. The primary task of this project is to raise the quality of collected data while decreasing labour input and respondents' burden. It is an advanced technological system, which uses the available PC equipment of respondents and a fast spreading means of electronic communication.

15. The electronic questionnaire is a specialized software which, when installed in the respondent's PC, generates a complete graphic representation of a particular statistical questionnaire accompanied by explanatory notes and code lists, ready to be loaded with statistical data. It is a matter of course that the installation includes detailed instructions for work with the product. The respondent can amend the data entered, until the questionnaire is completed correctly. The data being entered are continuously checked by a system of logic links, and no longer in the CZSO during processing as was the previous practice. This eliminates the time-consuming and costly activity linked with post-correcting the data reported on already collected statistical questionnaires during processing at the CZSO. Check sums, which constitute a standard part of the paper questionnaire and which are calculated and filled in by the respondent, are undertaken automatically by the software.

16. The completed questionnaire, checked on the PC screen, can be provided to the CZSO in several forms chosen by the respondents depending on the equipment they have. The simplest method is to print the questionnaire on paper and mail the paper document. Another way is to copy the questionnaire in a standardized format onto a diskette and mail the diskette. Still another form is to transmit the file via e-mail.

17. Software for 2001 was expanded to offer the possibility of providing statistical questionnaires via Internet, also with the aim of minimising the respondent's burden resulting from activities connected with the installation and administration of SW. However, the price for this simplification is the necessity of minimising checks on the respondent's PC, because the interactive method of running these checks would considerably extend the time of connection to Internet. This variant designated as EPV-WEB is thus appropriate for simple questionnaires only.

18. An important aspect is to secure individual data – i.e. to ensure that they are protected. This is why special cryptographic procedures for data transmitted in electronic form are part of the software delivered.

19. There are two ways for respondents to obtain the software they need free of charge: either on request (SW is sent on diskette or e-mailed) or by downloading from the Internet (the CZSO website includes software for individual statistical questionnaires).

20. In the first stage (1995-96), the CZSO ran an extensive survey of respondents to find whether respondents would be interested in this method of data collection and to obtain information on available computer equipment. Surprisingly, the interest in the application of electronic collection was high. In the second stage (1996-97), fundamental functions of electronic data collection were defined. The decision to retain the form of statistical questionnaire as the basic form of communication with the respondent turned out to be conceptually important in the process of verification. Software for a statistical questionnaire for quarterly business statistics was developed during this stage (1997), used to test technical aspects and organization of the whole system in real operations with respondents.

21. The third stage (since 1998) saw the incorporation of the EPV into the standard technological environment used in the CZSO to design statistical surveys. The EPV was supplemented step-by-step with other functions requested by both respondents and the CZSO staff in charge of processing. For processing in 2001, the CZSO offered 43 statistical tasks in the classic EPV and 27 tasks in the EPV-WEB version on its Internet page under 'Electronic data collection'. Two tasks have been processed using both methods.

22. From the viewpoint of building a unified technological environment for the programme preparation of statistical surveys, we were aware that electronic data collection could in no case be an isolated independent branch but that it had to be a regular component of the environment. It was imperative to maintain full automatically ensured compatibility between the software used by the CZSO for processing and the EPV software used by the respondent.

23. For the reasons above, the programme module that generates files for installing the EPV of a given questionnaire for the respondent was fully integrated into the ProjektMan programme system. This implies that the generation of programmes for the EPV and programmes for the processing at the CZSO are identical as far as requested functions are concerned, and that the specification of the functions is given by the description of statistical survey in the technical project.

24. The module for the generation of software for the EPV fully employs the specification of a statistical questionnaire in ProjektMan and transforms the questionnaire into a form appropriate for filling in the questionnaire on PCs, including the extent of the field to be filled in, prompts for individual items and definitions of the structure of the output file. The whole block of check links is taken over automatically. The programmer's only task is to ensure that relevant code-lists and other necessary files are attached. Attached to the programme generated in this way are auxiliary functions such as installation procedures, printing of the questionnaire, conditions of use, prompts, etc. The average size of the programme compressed is approximately 1.5 MB.

25. The current percentage of questionnaires provided electronically (1 to 25%, depending on the type) does not correspond to the level at which respondents in the Czech Republic are equipped with computing machinery. Virtually all businesses subject to reporting duty have PCs and over 90% of them are connected to the Internet. There are several causes behind this discrepancy. One of them is the low stability of statistical surveys in the period of the harmonization of Czech statistics with the EU. Once the harmonization is completed (in 2002), the respondents will be able to efficiently connect the production of requested statistical data to their information systems. Commencement of the activity of certification authorities for electronic signature in recent months creates another prerequisite for expanding the EPV to its expected level.

II. DATA COLLECTION WITH ASSISTANCE AT THE RESPONDENT'S

26. In statistical surveys, which call for the statistician's assistance at the respondent's place, the CZSO staff either complete the questionnaire with or take the questionnaire from the respondent. Included in these field surveys are consumer price statistics, household budgets, labour force sample surveys, agricultural statistics, and some censuses. All of these surveys are very demanding in terms of capacity and, consequently, costs which is why it is especially desirable to make them effective using the means offered by computing and communication technology.

II.1 Field surveys with the reporting duty

27. By declaring the programme of statistical surveys for a given year, the duty to provide relevant statistical data to the CZSO staff is imposed on respondents – mostly businesses. Collection of the data by the specialized staff on the spot is usually more appropriate, because another burden of the respondent is avoided and the data obtained are more unbiased.

28. The basic surveys with the reporting duty include:
- ?? measurement of consumer prices of goods and services;
- ?? harvest estimates;
- ?? censuses of population, various businesses and equipment.

29. The measurement of consumer prices is split into (i) monthly surveys of a rather wide range of goods and services and (ii) weekly surveys of a limited number of food products. The hitherto used technology of input data collection relied on data recorded on paper medium – initial record card. The interviewers collect prices of the goods being surveyed, recording them in cards first and then checking the price data and loading them into the PC at their district office. The processed file of input data is e-mailed to the central workplace for price statistics in the town of Hradec Králové within the time schedule, where input checks are run and the data processed further. The CZSO conducts the monthly survey of consumer prices in forty-two administrative districts and the capital city of Prague, and about sixty interviewers are involved in the collection. A new solution is under preparation at present – see II.3.

30. Harvest estimates are prepared monthly, from spring to autumn, by harvest inspectors, who are external experts. The inspectors pass harvest data over to district agricultural statisticians for checking, loading into the PC where district summaries are made, and then sending the data to the town of Ceské Budejovice for central processing. Questionnaires on crop-covered areas are processed in a similar way. While other questionnaires of agricultural statistics, whether they concern livestock or crop production or the agrocensus conducted once every five years, are completed by the respondent directly, and are collected with the assistance of about sixty district statisticians of the CZSO. In 2002, the optical character recognition technology will be gradually introduced into the entire agricultural statistics domain. The technology was already used in part and proved successful in the latest agrocensus.

31. The housing and population census 2001 is also based on the scanning of questionnaires, which will be finished in January 2002, as laid down by the time schedule. As far as small-extent censuses are concerned, the CZSO conducted a census of hotels and restaurants in 2001, and a census of construction entrepreneurs without employees and of water supply and sewerage systems is planned for 2002. All of the censuses are built on the scanning technology.

II.2 Field surveys without the reporting duty

32. Unlike the previous group of statistical surveys, these field surveys are not supported by the reporting duty imposed on respondents to provide data. The surveys are thus based on voluntary cooperation of a sample of respondents – households. The basic surveys of this group include:
- ?? labour force sample survey

?? survey on incomes and expenditures of households (household budgets).

Apart from these, other surveys of the ad hoc type are conducted, e.g. poverty survey, continuing vocational training survey, leisure survey, etc.

33. The labour force sample survey is taken quarterly. It includes a total of 32 thousand households and 120 field interviewers. Dual technology had been used for data input by the end of 2001. About a half of interviewers used paper questionnaires in face-to-face interviews, recording respondents' answers in the questionnaire to load them later in the office into the PC to obtain the electronic form. The other half of the interviewers used notebooks for the interview with a programmed interviewer-respondent dialogue and loaded the answers straight into the notebook. In both cases the provided and checked data were e-mailed to Prague for central processing. In early 2002, the remaining interviewers were also equipped with notebooks and the paper questionnaires were no longer used. The programme for the controlled dialogue is created in the programme product Blaise developed by CBS Netherlands. Currently, notebooks Fujitsu Siemens Life Book B-2175 are mostly used.

34. The survey on incomes and expenditures of households has so far been based on paper input documents – monthly diaries of incomes and expenditures of households. A total of 3 650 households of various groups are included in the survey. Completed diaries are collected by 140 CZSO workers, who check the recorded data, make corrections and convert the data on the PC into electronic form. The produced files are e-mailed to Prague for central processing. The input data document – the reporting household diary – has 24 pages and is rather time-consuming and demanding to complete and load in the PC. It has not yet been decided whether the CZSO will take the road of adjusting the diary to be suited for optical character recognition or of providing the respondent households with handhelds or light notebooks or whether the Office will accept any other way of work simplification.

II.3 Mobile technology in the statistics on consumer prices

35. The collection of data on consumer prices through mobile technology is, at the time of writing, still a pilot project. The proposed solution makes it possible to collect data by means of Compaq handhelds of the iPAQ Pocket PC H3630 type. The implemented operating system Microsoft Pocket PC 2000 cooperates with the fastest processor now available for the area of mobile devices – Intel Strong ARM, 206 MHz, 32 MB RAM and 16 MB ROM. The key criterion for the selection from several variants was the ergonomics of the handheld. The advantage of the handheld is its very bright colour display, which can be read under all light conditions ranging from complete darkness, usual lighting conditions in the shop to direct sunshine. The handheld is provided with an infrared port for wireless transmission of data, which is used for connection with the Nokia 6210 mobile phone.

36. GSM communication is effected straight from the field by mobile phone under the fast data transmission protocol HSCSD. Transmission to the central workplace is done in two ways:

?? with the help of commuted line and modem,

?? via Internet network.

At the central workplace, the data are loaded into the Oracle database to be further processed there by the method used before mobile technology became available.

37. Before the mobile technology became available, the transmission of data and the following production of initial verification outputs were concentrated within very tight deadlines. This left a large margin for error, as only a very short time was available for checking 90 000 prices and related items, and for verification through phoning respondents. There was no possibility for the interviewers to readily change survey requirements (changes in representatives of goods and responding shops, etc.) and to reflect the changes immediately in the data capture programme.

38. Since the availability of mobile technology, the interviewer collecting prices in the field uses built-in registers (goods, shops), nomenclatures and check links in an interactive way. The interviewer

can ask the iPAQ to offer him the survey according to the procedure implemented during a previous period, which significantly facilitates and upgrades work when a substitute has to be invited to stand for the regular interviewer. The interviewer's duty is to transmit the collected data on the day they were provided. To do that, the interviewer makes infrared iPAQ – Nokia connection and clicks the data transmission icon of the iPAC. At the same time, this process of logging-in for data transmission updates all changes relating to the interviewer concerned. The updating can include the register of price representatives or respondents, methodological instructions, as well as sending a voice message to the interviewer.

39. The new technology made it possible to immediately identify the price using the following three indications simultaneously:

- ?? who the interviewer is;
- ?? where the data were provided; and
- ?? when the price was provided (time is attached to each price the moment the price is provided).