

**UNITED NATIONS STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE**

CONFERENCE OF EUROPEAN STATISTICIANS

Work Session on Statistical Data Editing

(Ottawa, Canada, 16-18 May 2005)

Topic (iii): Electronic data reporting – editing nearer source and multimode collections

MODERNIZATION OF THE DATA COLLECTION SYSTEMS AT THE CSO OF POLAND

Supporting Paper

Submitted by the Central Statistical Computing Centre, Poland¹

I. INTRODUCTION

1. Data collection for statistical surveys by the public statistics in Poland is mainly done by paper forms and questionnaires. Regional statistical offices and their branches are responsible for coding, registration and correction of the forms after their collection. It requires a lot of manual work of our statisticians at the regional level.

2. The increasing cost of labour has been observed over the last few years, and at the same time, new technological solutions have evolved in the area of information and telecommunication technologies, which are leading to introduction of changes in both the organization of data collection processes for statistical surveys as well as applied technologies.

Systems for optical data recognition - OCR

3. The system of optical reading of documents (OCR) allows increasing efficiency of the data registration process. This technology has been used in the CSO of Poland for many years. The OCR system supplied by Italian firm Elsag-Bailey was used for statistical data processing since 1998. During 2001 and 2002 the new OCR system was designed and implemented (mostly by internal efforts of the CSO) to process data collected during National Census 2002 (that was the National Housing and Population Census 2002 and National Agricultural Census 2002). During the second half of 2002, 18 data processing centers processed altogether 300 million pages of questionnaires. From 2500 up to 3000 statisticians daily have been involved in the process. Now the second generation of the OCR systems is implemented in the CSO based on software and hardware used during processing of national census questionnaires. Currently almost 30 statistical surveys are processed with the use of this new OCR technology. No further modernization of those OCR systems is predicted.

System of statistical interviewers - CAPI

4. Until recently there were two phases in the processing of statistical interviews: first, the interviewers filled in the questionnaires, and second the data from the paper questionnaires were recorded on PCs in the regional statistical offices. Due to extensive costs and difficulties arising during the creation of a proper technology and organization for the processing of statistical interviews, the process has been divided into many stages. During the first phase (in 2004) almost 400 interviewers were

¹ Prepared by Krzysztof Kurkowski (k.kurkowski@stat.gov.pl).

equipped with the tablet PC and almost 500 more are planned to be equipped with similar hardware during 2005. The control and management of all interviewers as well as applications for information collection are done internally in the CSO. The Microsoft platform is used both for the operating system and the developers tools. There are plans to use GSM – GPRS technology for data transmission.

Reporting Portal

5. The concept of the reporting portal (for respondents), which will be the element of electronic reporting, is presented in more detail in later sections. The project began in 2004 and should be finalized in 2006. Pilot projects for electronic forms created with the use of Internet technology, have been conducted in recent years. During 2005 there are plans to implement 30 different electronic forms on Internet.

6. The reporting portal will offer a new functionality for data collection from statistical units. GUS (Central Statistical Office of Poland) is obliged to offer such an updated functionality from mid-2006.

7. The objectives of the reporting portal are:

- Decreasing time and work on surveys by improving the quality of data,
- Decreasing survey costs related to printing and sending forms,
- Decreasing survey costs related to repeated contacts with a statistical unit to clarify errors (improve data quality),
- Realization of legislative obligations.

Current State

8. For statistical surveys, data is gathered via paper forms. Forms are completed directly by the statistical unit (respondent) that is the subject of the survey or less frequently through specialized pollsters, which deliver the form to the respective regional statistical office or branch of a regional statistical office) after conducting the survey.

9. The forms are delivered to the appropriate statistical unit (respondent) through the post or directly by the representative of the reporting entity (statistical pollster for surveys of a poll nature). In addition, contacts related to the subject of the survey (providing clarifications, error correction, pressure) are performed via telephone or mail.

10. Data from the forms are entered into dedicated processing systems with the use of OCR technology (36 surveys of the highest number of forms) or through manual input (around 220 surveys). The Survey Coordinator is responsible for the Statistical Survey.

Survey scheme

11. Action carried out for a particular statistical survey can be divided into the following stages:
- (i) Establishing the survey card as well as its update on the regional statistical office level;
 - (ii) Distribution of forms to statistical units;
 - (iii) Collecting survey data;
 - (iv) Consolidating data and preparing Survey results.

Desired state

12. The obligation of public area entities to provide public services through electronic means, as well as the widespread access to the Internet, gradual departure from traditional means of acquiring information is assumed, based on the use of paper forms, in favour of electronic technology for acquiring data. In future, the main channel of informational flow between statistics and statistical units is planned to be the Internet.

13. The traditional method (paper forms) will be applied in an auxiliary manner in order to enable the fulfillment of the obligation to transfer data to/from the statistical units that do not possess the technical possibilities (or required knowledge) to take advantage of the Internet channel. The total withdrawal from the use of paper forms is not planned, but rather only withdrawing from their registration through OCR, when its use will no longer be economically viable. Below the possibilities of supporting each of the survey stages with the use of the reporting portal are indicated:

1. Establishing the survey card as well as its update on the regional statistical office level

Selection from BJS (statistical business register).

14. The generated survey card is sent to the reporting portal, since it uses survey card data to send notifications to statistical units about the Survey as well as tracks reminders and contacts between statistical units from the survey card and the regional office employee. After receiving the survey card and after the reporting portal ascertains that the survey has begun (on the basis of the survey calendar), the reporting portal sends notifications on the survey to respective regional office employees participating in the survey process (statisticians). This allows automating the start of the survey process in the given regional statistical office.

Updating the survey card

15. Modifications to the reporting portal survey card pertain to data, which in the current survey process are modified in the regional statistical office, i.e.:

- Removing the statistical unit from the reporting portal survey card;
- Modifying and adding a new statistical unit to the reporting portal survey card;
- Change to the assignment of statistical unit as a unit obliged or not to participate in the statistical survey – this modification pertains to a few surveys;
- Change of statistical unit assignment to the branch – operation rarely executed, in most cases the initial assignment of statistical unit to the branch is enforced.

16. The current reporting portal survey card (for the given regional statistical office) is transferred to the respective logical and computing audit system. This is related to the requirement that these systems must always possess a current survey card.

Reporting portal repositories

17. The reporting portal repositories result from the functionality that the reporting portal should offer to enable the management of the survey card and notify statistical unit of the necessity to carry out the statistical obligation. These include:

- reporting portal survey card – initially loaded from BJS, then modified by authorized regional statistical office employees,
- reporting portal Statistical Unit Database (reporting portal BJS) – repository includes basic data on all statistical units (e.g. REGON, address data).

2. Distribution of Notifications and Paper Forms

18. In order to support effectively this Survey stage, the reporting portal informs via e-mail of the statistical unit's necessity to execute its statistical obligation as well as informs regional statistical office employees (Statisticians) responsible for the survey, of the commencement of the survey and necessity to prepare paper forms for sending.

19. Statistical units are notified only via e-mail, since it is assumed that it will not visit reporting portal web pages on its own initiative. The Statistician should be informed via e-mail and also be

supported by the list of tasks to carry out, available to the Statistician after he/she logs into the Statistician application (AS).

Form distribution

20. If the survey does not have electronic forms, each statistical unit obliged to participate in the survey (in accordance with the reporting portal survey card) should receive a notification of the survey as well as all forms by paper mail. If the survey has electronic forms, then the following possibilities arise:

- If in its profile statistical unit has the “paper mail” status assigned for this type of survey, then it receives the notification and forms through paper mail,
- If in its profile statistical unit has the “electronic mail” status assigned for this type of survey, notification is then sent in electronic form to the e-mail address of statistical unit and statistical unit assigned to the given survey. The notification is also posted on the portal’s WWW pages.

The fact that statistical unit was notified and was sent forms by mail should be registered by the reporting portal database. This allows tracking the actions executed by statisticians as well as generates reports related to the form sending process.

3. Collecting survey data

21. The process of collecting survey data in reporting portal takes into account the following facts:

- Statistical unit may supply the reports via traditional means (post, fax, personally delivery in the regional statistical office), reports that are not subject to OCR are entered in the same manner that statistical unit performs the entry on portal pages (reporting portal registers which regional statistical office Division employee entered the report and on behalf of which statistical unit). Reports that are subject to the OCR System, are registered in the OCR system, from which they are transferred to logical and computing audit systems
- Statistical unit may supply the reports via electronic means – statistical unit may complete the form on the web page or reports sent by statistical unit systems may be sent directly to reporting portal.

Validation in logical and computing audit systems

22. Data entered from forms to the reporting portal reports repository is sent (automatically once in a while or upon request) to logical and computing audit system for further data auditing. The reporting portal reports repository on the other hand is updated with the audit results of logical and computing system over small time intervals – reporting portal “knows” whether the report was accepted or rejected due to an error in order to generate a reminder of a correction.

Clarifying errors

23. In case errors are discovered in the report by the logical and computing audit system, its clarification should take place, which results in a correction of the report. In the case of the electronic channel, the notification via e-mail is used (such as notifications of the obligation to participate in the survey). In the case of the standard channel (telephone, fax, personal visit, post) reporting portal supports the Statistician contacting statistical unit by registering every contact with statistical unit in relation to errors in reports.

Correcting errors

24. Correcting an error in the report is conducted in the same manner as data collection, i.e.:

- Correction of report on web pages by statistical unit and its repeated transfer,
- Sending corrections by statistical unit IT systems,

Supplying manually completed corrections via post mail or fax or personal delivery to the regional statistical office or its branch. The report is then corrected by the statistician in reporting portal.

Sending reminders

25. If the statistical unit does not fulfill its obligation to submit the report or submit the correction of erroneous data on time, the reminding process is introduced. It continues until the statistical unit supplies the required data and until the data collection deadline passes (this deadline is established in the survey calendar). Statistical unit reminding takes place through the following means: standard – telephone, fax and post – performed by statistician or electronically – compulsory through e-mail, optionally through information on web pages. Electronic notifications are carried out automatically by reporting portal or entered by the statistician. The fact that statistical unit is reminded is registered in reporting portal.

Paper and electronic forms

25. Due to the fact that there are two methods of collecting data from statistical units following the implementation of the reporting portal – through paper and electronic forms – ensuring uniformity by reporting portal is addressed in this point, regarding data flow from statistics to statistical units at each stage of the survey.

26. Paper forms applied in statistical research can be divided into two groups:

- Closed forms - are characterized by a static structure with a closed and pre-determined number of available fields.
- Opened forms – that include “drop-down” elements, i.e. unequivocal layout of fields, enabling the description of a given event according to the adopted scheme. The number of events described by particular statistical unit varies, and so such a form has an open structure – adjusting every time to the amount of entered information.

27. Taking into account the present types of paper forms used by GUS, the following critical requirements are raised that reporting portal should fulfill in order to effectively execute the transfer of data via electronic means:

- Requirements related to the access channels

- Due to the widespread access to the WWW network, the primary channel for electronic forms should be the WWW channel.
- Many statistical unit possess IT systems that could easily be expanded to include a functionality for sending Reports directly to reporting portal or generating a set with an appropriate format (then sent from WWW pages) - support of the tool of the XML format.
- In order to maintain conformity with the Act on the electronic signature, the forms' mechanism has provided the possibility of verifying the electronic signature if statistical unit sent the data signed with the help of a qualified certificate.

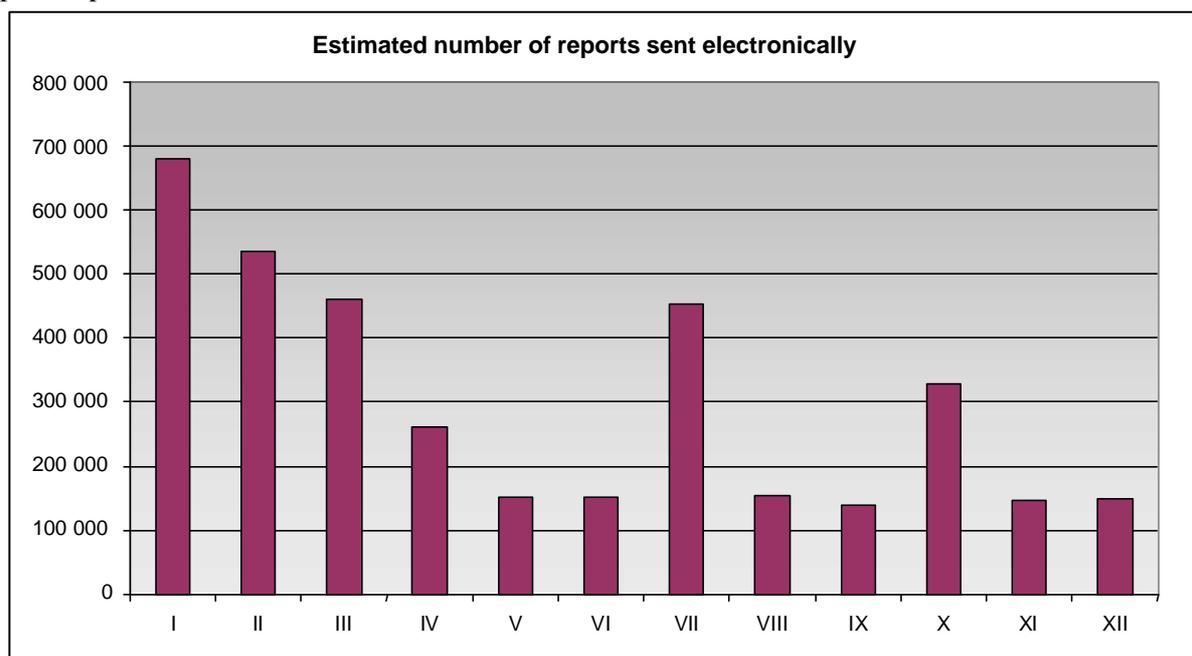
- Requirements related to entering and validating data

- Statistical unit may distribute over time the entry of data into the form. This signifies that the form mechanism enables entering data into the form in installments, during subsequent logins to reporting portal.
- In order to minimize the number of errors in data as well as contacts with statistical unit, the mechanism of electronic forms includes the maximum level of audit at the data entry stage.
- In case errors occur during validation, statistical unit is notified in a clear manner of the errors in data.
- The possibility exists of sending data despite the occurrence of certain errors. GUS permits a situation in which data is erroneous but still takes part in further analysis.

Number of forms

28. The number of statistical surveys executed through post mail conducted on an annual scale ranges from 200 to 240 (GUS also performs surveys through a poll method, which will not be maintained

by reporting portal). The chart below presents the number of electronic reports in the 4th year of reporting portal operation:



Timetable

29. For statistical needs, the data collecting process is characterized by the significant irregularity of data receipt from obligated statistical units, related to the nature of conducted surveys. In practice, the data is transferred within a few days prior to the expiration of the indicated deadline as well as directly following its expiration (often as a result of a reminder on the part of the statistics unit). The chart below presents the estimated distribution of report transfer intensity via electronic means in the first quarter of the reporting year. These months are characterized by the highest intensity of data transfer for statistical needs.

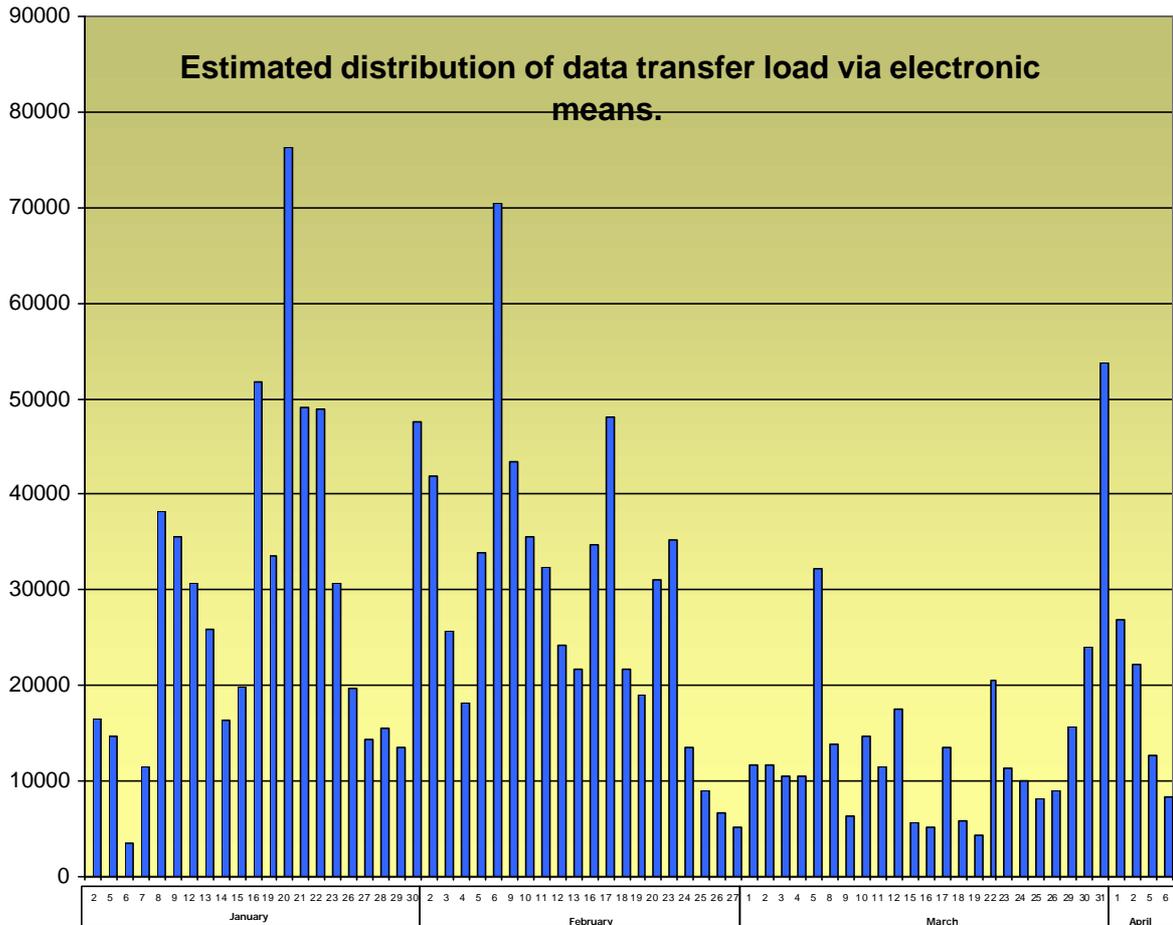
Scalability and reliability

30. Requirements related to reporting portal performance, its reliability as well as fulfilling the timetable are the most important requirements placed upon the reporting portal. The increase in performance of solutions of this class will be achieved with the use of new machines, increasing machine resources – number of processors, RAM, system component configuration (increasing the number of operating processors or threads, connections to the database etc.) as well as optimizing database queries.

Interfaces with other systems

31. The survey conduct process shows that reporting portal should interface with the following IT systems in GUS:

- BJS – in order to download the survey card, statistical unit data as well as update this data,
- Logical and computing audit systems – in order to send this data for further auditing and download audit results,
- Printout systems – in order to supply statistical unit and statistical unit address information required to send paper forms.



Interface with BJS

32. The interface with BJS pertains to:

- Sending the survey card to reporting portal,
- Sending statistical unit data to reporting portal,
- Downloading current statistical unit data in case the survey card is modified.

Interface with logical and computing audit systems

33. Logical and computing audit systems are prepared independently for particular statistical surveys – there are then as many of them as there are surveys, i.e. over 200. The interface with logical and computing audit systems pertains to:

- Sending to systems the current reporting portal survey card,
- Sending Reports to systems,
- Downloading Report statuses and error description from systems.

34. The file interface may be replaced at a later time with replicas at the database level, but only in case the SQL Server database is used for PortalBis (the same that will be used by the new logical and computing audit system).

Interface with printout systems

35. The interface with printout system pertains to sending statistical unit address data to systems.

Implementation of a new logical and formal audit system

36. The use of one national logical audit database is planned, however, it is assumed that for large groups of reporting units, logical audit databases operate at the level of statistical offices. Database servers with Microsoft SQL Server, access to the logical audit application as well as data provided by SQL Server take place through terminal technology by Citrix. The implementation process of SQL Server databases will be distributed in time, as such over the initial period of portal functioning the data collected electronically is transferred to the FoxPro standard database as well as SQL Server database (depending on which type of database maintains the given statistical survey).
