

**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

**ELECTRONIC DATA COLLECTION SYSTEM DEVELOPED AND IMPLEMENTED
IN CENTRAL STATISTICAL BUREAU OF LATVIA**

Supporting Paper

Submitted by the Central Statistical Bureau, Latvia¹

I. INTRODUCTION

1. Traditional methods of data collection at CSB are based on paper questionnaires, which can be filled in by interviewers or respondents. Respondents receive a paper questionnaire via post, fill it in and return to CSB. Interviewer collect data doing personal interviews using paper questionnaires, CAPI or CATI technologies.
2. Operation with paper questionnaires consists of procedures listed below:
 - Questionnaire design;
MS Word is in use for the questionnaires design in CSB.
 - Questionnaires printing;
CSB has its own offset printing facility, but nevertheless it is resources consuming procedure.
 - Some of the printed questionnaires have to go through the pre-printing process in which data of previous period are printed in;
 - Sending the printed (in some cases also pre-printed) questionnaires to the respondents;
Each questionnaire has to be put in the envelope. An exact address label has to be glued to the envelope and all package of prepared envelopes has to be delivered to the post office.
 - In case of not receiving filled in questionnaires at the deadline shown in the questionnaire, reminders have to be sent to respondents;
To organize that procedure all incoming questionnaires have to be registered and reminder letters prepared and sent out.
 - Collected data from paper questionnaires has to be retyped into CSB Data Management System, data checking procedures have to be done;
The first data checking is done by the operator immediately after data has been typed into system by launching data checking application, if the system found incorrect data operator has to contact exact respondent using phone.
The second and final data checking has to be done on full data set and it takes reasonable amount of time and server resources.
 - More than 120 persons from total of 540 CSB employees ensure timely execution of the processes listed above.

¹ Prepared by Karlis Zeila, Vice president of CSB of Latvia (karlis.zeila@csb.gov.lv).

3. Implementation of the electronic data collection is the only way to decrease expenditures and raise collected data quality. This task has been set up during the development and implementation of the CSB Integrated Metadata Driven Data Management System¹ (further in the text Data Management System - DMS). Real design has been started at the end of 2002 and completed at August 2003. Implementation took longer and the first e-surveys really have been launched at March 2004.

II. ELECTRONIC DATA COLLECTION, CSB APPROACH AND KEY REQUIREMENTS

4. There are several approaches known how to collect data electronically, see table No 1.

Data carrier	Communication channel	Data security options	Process control & management options
Floppy disks or CD	Ordinary mail service in the Post office	Use of encryption software on both CSB and respondent sides	Not available
e-mail message with attachments	e-mail system via internet connection	Use of encryption software on both CSB and respondent sides	Low
www	Internet connection, with some specific software installed on the PC work station	www provided security	High

Table No. 1.

5. Taking into account that at disposal of CSB of Latvia at the time of Electronic Data Collection (EDC) system development was implemented DMS, the EDC system development approach and key requirements to that has been defined as follows:

6. General requirements for CSB EDC system:

- Use www solution for CSB EDC system design (further instead of term “electronic questionnaire” use the term “WEB form”);
- EDC system should be integrated in the DMS;
- Layout of web forms has to be as much similar to paper questionnaires as it is possible, to ensure simple transition to web based data submission for the respondents;
- Functionality of the EDC system particularly on the respondents side has to be very advanced to raise interest to use it instead of classic paper forms;
- All processes have to be metadata driven from common DMS metadata base and automated as much as possible;
- The same design tool MS Word for both paper and web forms has to be used.

7. EDC system based on web data submission has to provide functionality listed below (key requirements):

- On the main screen of the respondent’s side ensure possibility to see the list of web forms as well as paper questionnaires with the deadlines and status of each form which respondent has to fill in;
- Ensure Pre-loaded data option in the web form for respondent’s descriptive data (e.g. respondent’s name and address);

- Ensure launching of the validation procedure on the respondents side using only validation algorithms attached to each form by CSB statistician;
- For periodical surveys respondent has to have a possibility to see and correct data of previous periods;
- Special alarm message for CSB statistician has to be generated in case if change of data by respondent in previous periods has been done;
- System has to ensure registration of respondents and to define access rights for them;
- Ensure possibility to use classifications codes (NACE, PRODCOM and etc.) on the respondents side with the search option;
- Ensure that respondent is always informed that CSB has accepted respondents submitted data;
- Response control system has to be able automatically generate and send reminders to respondent in both cases when deadline is very close and deadline is over;
- The respondent must be able to print questionnaires, for internal use;
- Help facilities. Should be available on both respondent and CSB sides;
- Security: it must be guaranteed that WEB forms data could be accessible only for the responsible bodies on respondents side (user name and password should be introduced for “ login” in the system) as well as statisticians on the CSB side. Respondent should be able to change password;
- To block login in case of use of wrong password or user name several times in turn;
- Accordingly with CSB IT strategy, which standardize on Microsoft products, EDC system interface should be compatible with Microsoft technologies currently presented on the market.

III. HOW THE EDC SUBSYSTEM WORKS

8. The core element for the EDC system is electronic questionnaire or in our case we agreed on WEB based data entry and validation form or simply web form. For different surveys are different web forms. These forms are available to respondents and are as a replacement for paper questionnaires. Responding procedure is organized as respondent connection to CSB EDC system which is partly located on CSB WEB server.

9. Implementation of the key requirements listed in previous section contributes to increase of data quality and gives to respondent a set of advantages to make his life much easier when providing statistical data to CSB.

10. Integration of the EDC within DMS in a quite simple way we can see on the figure No.1.

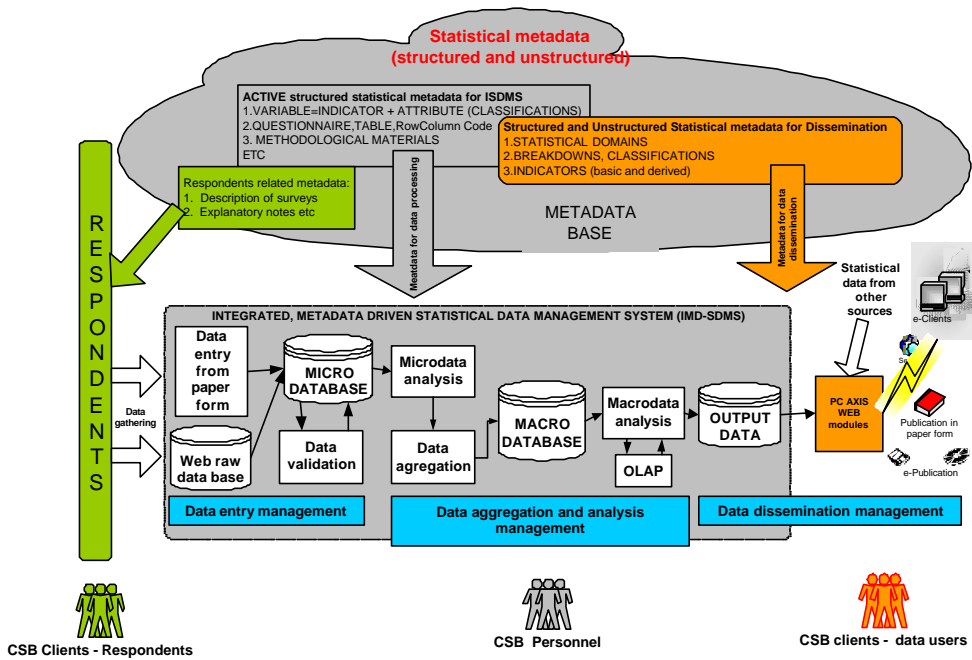


Figure No.1.
High level diagram of DMS in integration with EDC subsystem

11. On this very high level diagram EDC appears like extra database named “Web Raw Database” connected with Meta and Micro databases.

12. How data from respondent’s PC become available in CSB Data Management System we can see on the Figure No.2

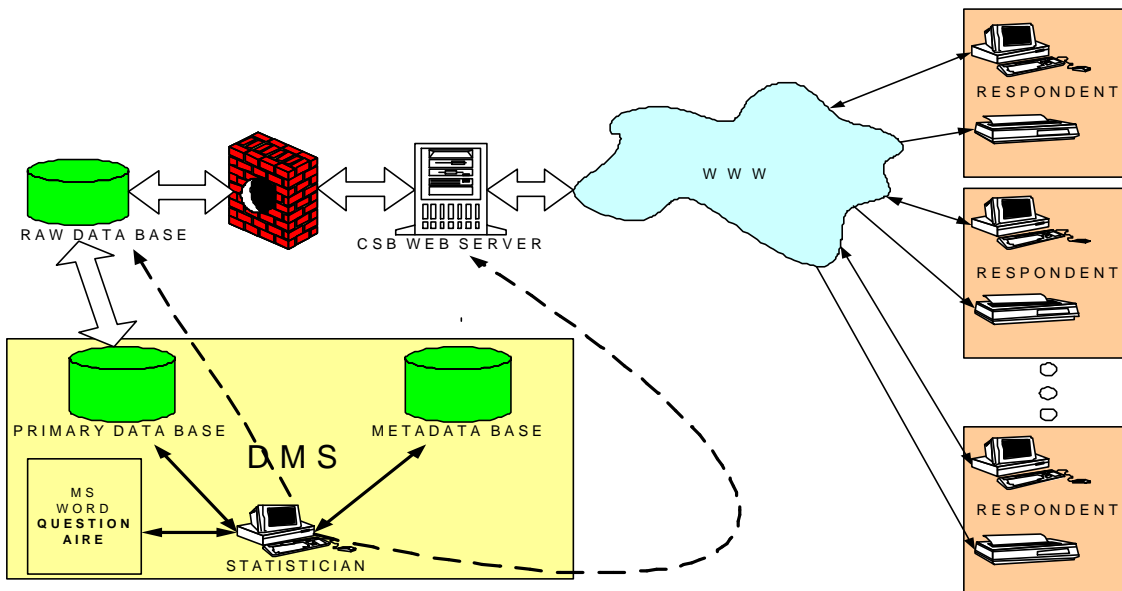


Figure No. 2.
Dataflow from Respondent to CSB DMS

13. As it was mentioned before EDC system is not a tool for web form design, is tool for attaching of the exact survey questionnaire layout, which has been designed using MS Word for printing on the paper, to DMS Metadata base by linking the exact fields of the questionnaire to exact indicator descriptions in the Metadata base. Description of this procedure is shown on the Figure No.3.

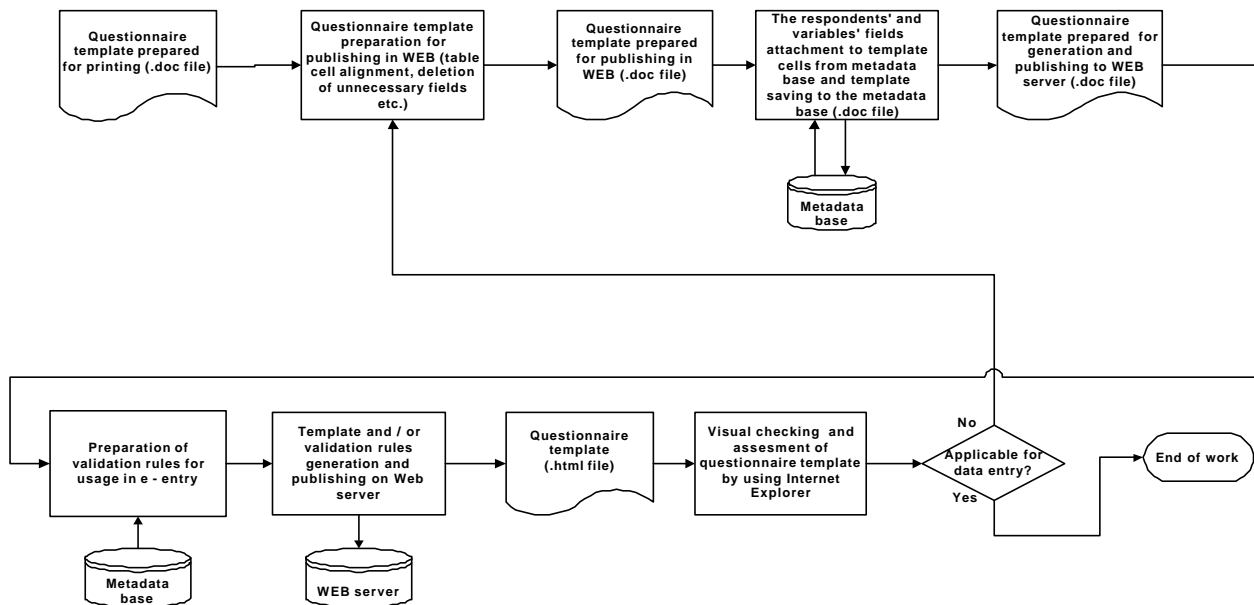


Figure No.3.

Questionnaire linking with Metadata base and transforming from MS World to HTML format

14. EDC system consists of three software modules:

- HTML forms generation from Word documents and their publication to a WEB server;
- Module, which provides with the data exchange between respondent and CSB (including the data entry and validation);
- Respondents, questionnaires and data administration module.

15. HTML forms generation module provides with the Web forms design:

- By using the same metadata base both for the DMS manual data entry and for the electronic data entry forms;
 - By using the same Word templates both for the questionnaires print out and for electronic data entering;
 - Macro commands are installed in the MS Word, which allow to add fields from common metadata base;
 - Statistician can add respondent and data fields to Word template from metadata base;
 - By using the common validation rules description, any rules can be added to electronic questionnaire that are defined in metadata base;
 - Word document is saved in HTML format and published on WEB server (Described above process is illustrated on Figure No 3)
 - Any survey, data of which, we would like to collect electronically has to be described in EDC and DMS system common metadata base.

16. Module, which provides with the data exchange between respondent and CSB ensures data entry and validation (previously described validations in metadata base) from respondent workstation, as well as data “sending” to Primary data base in CSB:

- Each respondent has a unique user name and password. It is possible for respondent to change the password;
- Respondent can see a list of questionnaires for him to be sent in a particular period – both in electronic and paper form;
- Respondent has access to information from CSB Business Register about its address, name, phone etc. If any of these attributes has been changed, the respondent has mark the changes in special field;
- Respondent has access the required classifications for the questionnaire, with search possibilities by the code and by the name;
- Respondent has access to data from previous periods for review, analysis and correction;
- Data entry can be made in several steps, saving the data each time;
- Before the “sending” of data to CSB data validation should be performed and mistakes proved.

17. Data administration module:

- In CSB statisticians are provided with applications which ensure:
- Adding respondents from CSB Business register to the system;
- Adding the published questionnaires on the WEB server to particular respondents in particular periods;
- Viewing the respondent’s entered data to any survey at any time;
- Preparing different reminders to respondents and sending them automatically or manually.

IV. CONCLUSIONS

18. Analysing information obtained from March 2004, when EDC system has been launched in the CSB of Latvia, we can point out several advantages and achievements as well as some restrictions. Advantages and achievements would be as follows:

- 37 different business statistics surveys are implemented within the first year of EDC system exploitation,
- 38% of response rate achieved for the first three surveys implemented,
- 10% of average response rate achieved for all surveys implemented,
- Comments from respondents, which have started to use the system, are quite positive, due to advantages described in paragraphs above,
- No extra staff was necessary in both statistical and IT sections for system implementation and maintenance,
- User interface is very user friendly and does not require special training neither on respondents’ side or on CSB side.
- Some restrictions we would like to mention are as follows:
- Being developed as a part of Data Management System it can operate only with surveys described in the DMS common metadata base. In DMS there are 67 business statistics surveys at the time being.
- Use of MS Internet Explorer is a restriction as well because a lot of enterprises are moving to open source software usage (Mozilla for instance)
- Implementations of large surveys are problematic in existing EDC system version due to inconveniences when scrolling the web form. This will be improved in the next version.
- Sometimes respondents are suffering from unstable work of the communication channels or Internet services providers,

- Hot telephone help desk had to be established in CSB and system administrator takes over the technical assistance functions especially in cases when respondent is not enough advanced to operate in internet or in quite often cases in IT environment,
- It is not possible to fill in the same web form from several workstations or by several persons on respondent' s side simultaneously.
