STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 26
English only

CONFERENCE OF EUROPEAN STATISTICIANS

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

Topic (iv): Users' experience with online modes

EXPERIENCES WITH AN INTERNET BASED DATA COLLECTION SYSTEM: WEBEDI AT THE HUNGARIAN CENTRAL STATISTICAL OFFICE

Submitted by the Hungarian Central Statistical Office[1]

Contributed paper

I. INTRODUCTION

1. Next month following the announcement of a pilot proJECT for electronic raw data collection (in "e-CORD at the Hungarian Central Statistical Office", invited papers MSIT, Geneva 2001), the Joint Efforts of the Hungarian Central Statistical Office (HCSO) and Telecom in Hungary resulted in a fully operational system called WEBEDI.

2. The idea of the project was formed in March 2000, followed by a contract between HCSO and MATÁV signed in October. The first respondents started to report in March the following year.

3. In Hungary the county offices of the CSO are traditionally responsible for data collection for economic surveys. A centralised "registering and monitoring" system called GESA and a "data input and validation" system called ADEL have been developed and put into operation. Both systems are work with a centralised data base running under Oracle.

4. The projection of the future of data collection shows the co-existence of paper based and electronic questionnaires, at least for a decade. GESA and ADEL are relatively well integrated and both designed according to the most up-to-date principles. The four years old EDIFACT reporting of the largest companies became part of the integration. The web-based solution naturally accommodated itself into this environment and was designed to strengthen uniformity rather than emphasise differences.

5. Recruiting electronic respondents to report via the Internet was limited to three counties only (out of 19) to be able to keep organisational aspects of the project under strict control as one of the basic objectives. It was important to involve a few of the county offices to make the pilot part of the "real life" and to obtain feedback of the day-to-day operational difficulties.

6. The project proved to be a success. The decision to include this Internet based data reporting facility in the future for as many surveys as possible has been taken. In 2002 all respondents of the integrated economical survey in the country were approached and monthly, quarterly and annual reporting via the Internet was offered to each.

[1] Prepared by Tamas Koltai.

7. The pilot showed the feasibility of the chosen technology together with the weaknesses of the implementation. It is part of the success that in the so-called “e-government” project launched by the Hungarian Prime Minister’s Office, the system was accepted as the basis for future developments and gained a considerable amount of financial support within the scheme.

8. Sign posts for the reader:

?? The system: we have to live with for a while
?? Respondents: we have to live with for even longer
?? Good points: what we are going to insist to
?? Missing links: on-line and off-line
?? The future: integration, co-operation, independence

II. BASIC FEATURES OF THE SYSTEM

9. The questionnaires are defined in XML/XSL. All questionnaires are handled by the same XSL program. The specific elements of the questionnaires are written in XML with respect to the forthcoming IQML standards.

10. The electronic questionnaires are “impersonalised” by the data of the GESA system. It is part of the translation (to HTML and Javascript) happening within MATAV’s e-commerce system.

11. A client sees and handles a specific HTML form similar to the original paper questionnaire. During data entry validation, rules prevent the client from reporting syntactically and logically false data. The rules are limited to simple checks like “numeric”, „obligatory”, “control sum”, etc.

12. Incoming data are converted to RDRMES format before the transmission to HCSO. RDRMES is received by EDISC, a message handling system. Output from EDISC is in questionnaire specific format for the Oracle Loader.

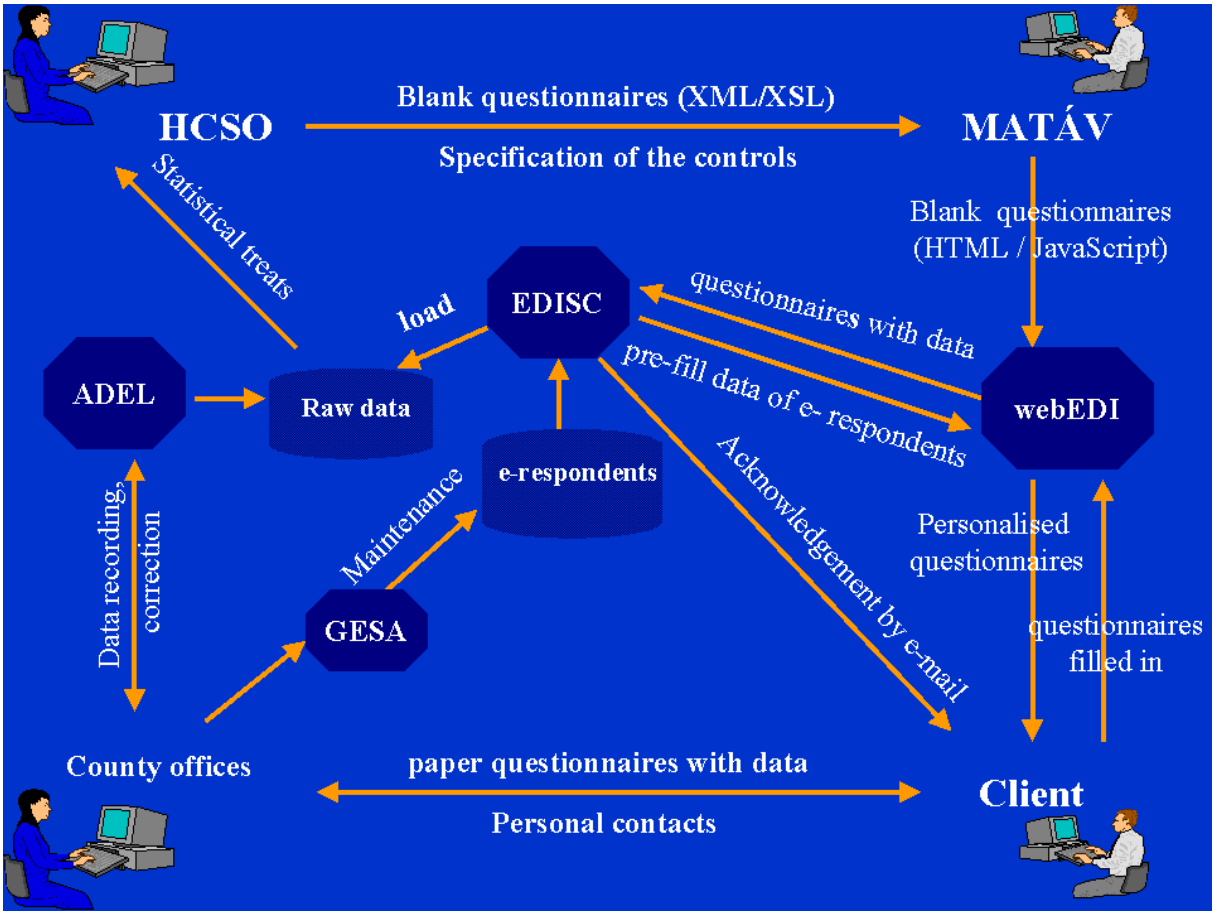
13. Electronic responses loaded into the appropriate relational tables are identical to those used with the paper-based ones. The tables belong to the Oracle data base of the ADEL application.

14. Questionnaires filled-in fully, partly or left empty can be saved both on the client and the server side. Saved (stored) questionnaires can be (up-) loaded and “continued”.

15. The authentication and data integrity can be guaranteed by electronic signature. The certification of the signature is delivered by MATÁV on active card basis.

16. Data security is provided on the Internet by the standard HTTPS protocol, while the communication between the Internet provider (MATÁV) and the HCSO is established via a leased line and secured by using asymmetric keys.

Workflow in the system

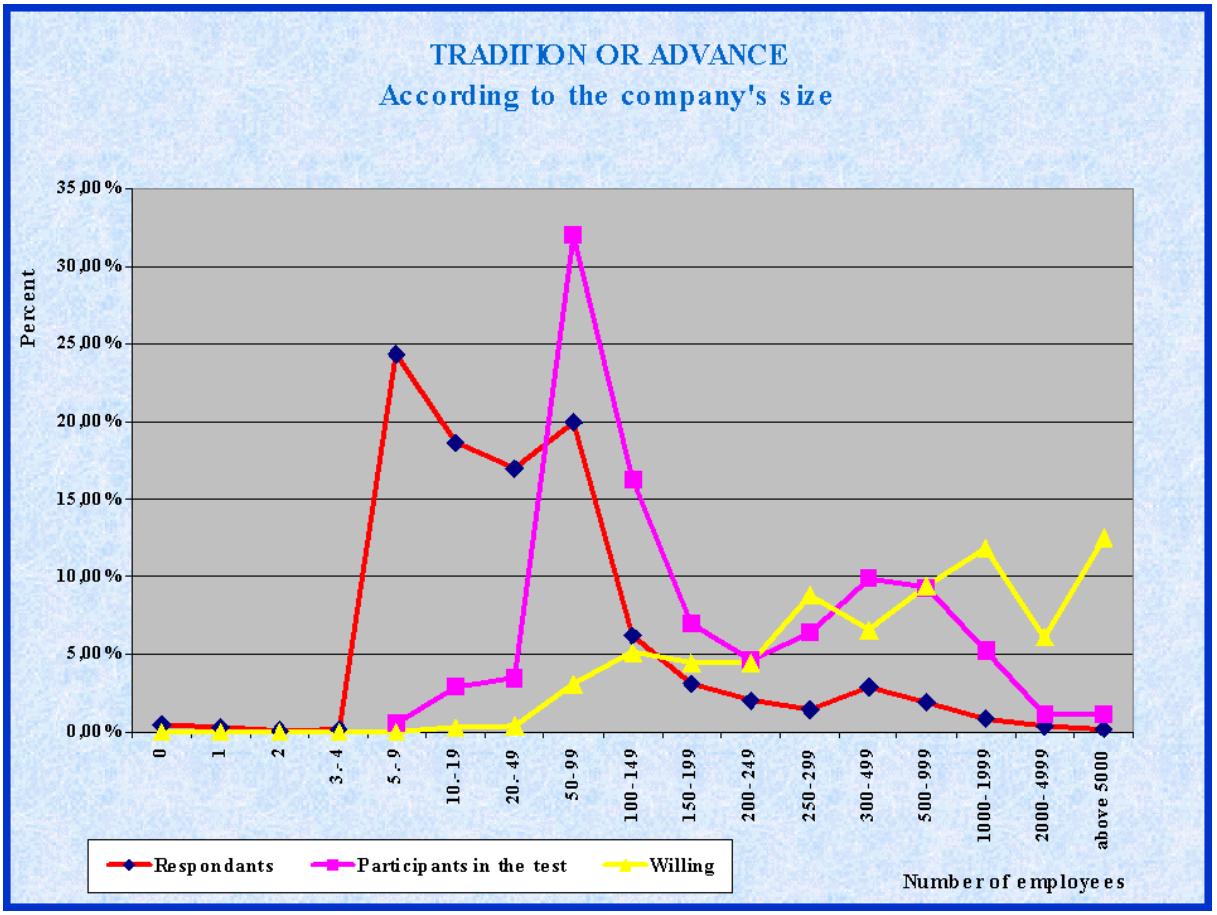


III. RECRUITING E-RESPONDENTS

17. Out of the 8700 respondents in the three countries, 2000 were approached by letter. The 2000 were selected as the best candidates by the by the local offices: ca 10% returned some kind of answer; ca 9% became partner in the pilot project. Setting up standard e-mail communication is easy; setting up standard WEBEDI connection is more difficult and requires a lot of consultation via telephone. Reasons: browser settings/our software. The best partners are medium-size companies.

Respondents of 3 counties in the integrated collection of data

Number of employees	Respondents in the integrated economic survey		Registered as partner in electronic reporting	
1	24	0,27 %		
2	11	0,13 %		
3 - 4	14	0,16 %		
5 - 9	2135	24,31 %	1	0,58 %
10 - 19	1633	18,59 %	5	2,91 %
20 - 49	1493	17,00 %	6	3,49 %
50 - 99	1757	20,00 %	55	31,98 %
100 - 149	549	6,25 %	28	16,28 %
150 - 199	269	3,06 %	12	6,98 %
200 - 249	180	2,05 %	8	4,65 %
250 - 299	124	1,41 %	11	6,40 %
300 - 499	259	2,95 %	17	9,88 %
500 - 999	170	1,94 %	16	9,30 %
1000 - 1999	76	0,87 %	9	5,23 %
2000 - 4999	33	0,38 %	2	1,16 %
above 5000	16	0,18 %	2	1,16 %
Total:	8783	100,00 %	172	100,00 %
			1,95 %	


TRADITION OR ADVANCE
According to the company's size
Percent
35,00 %
30,00 %
25,00 %
20,00 %
15,00 %
10,00 %
5,00 %
0,00 %
0
1
2
3.-4
5.-9
10.-19
20.-49
50-99
100-149
150-199
200-249
250-299
300-499
500-999
1000-1999
2000-4999
above 5000
Respondants
Participants in the test
Willing
Number of employees

IV. E-REPORTING

18. The level of successful e-reporting is significantly higher in the smaller local offices than in the larger ones. A total of 40% of the participants became regular e-respondents. A very large percentage of reports are either on paper or are electronic. The figures are improving with time: timeliness is much better, deadlines can be brought nearer to reporting period. The number of corrections (re-reporting) is less than with paper.

Evaluation of the returned messages (04. 09. 01.)

County	Respondants	Became electronic respondant		Cancelled		Doesn't send		Unreliable sender	
Budapest	108	36	33%	6	5,55%	20	18,5%	40	37,03%
Kecskemét	41	22	54%	2	4,88%	5	12,2%	16	39,02%
Szolnok	23	10	44%	1	4,34%	5	21,74%	9	39,13%
Total:	172	68	40%	9	5,23%	30	17,44%	65	37,79%

County	submissions in May				sumissions in June				submissions in July			
	total	bef. 16	urged	corr.	total	bef.16	urged	corr.	total	bef. 16	urged	corr.
Budapest	53	16	10	3	53	8	3	0	54	16	4	3
Kecskemét	24	3	3	0	26	6	4	2	28	6	3	1
Szolnok	12	4	0	0	15	2	1	0	15	2	2	0
Total:	89	23	13	3	94	16	8	2	97	24	9	4

V. E-RESPONDENTS

19. E-respondents:

?? are handicapped by their own safety measures;
?? suffer from the poor internal communication within their company;
?? tend to mix their statistical and technical problems;
?? forget or lose everything (identity code, password, error-list, etc.);
?? don't perceive the difference between e-mail and webedi;
?? accept the infantilism of a pilot system;
?? regard e-CORD as a common success.

Causes of not sending
04. 09. 2001.

Causes	No.of co.	%
Internet problems	8	26,67%
Doesn't have IT specialist	2	6,67%
Poor internal communication	6	20,00%
Responsible employee left	3	10,00%
Didn't get support enough (from us)	2	6,67%
Lack of time	4	13,33%
They prefer paper	2	6,67%
Couldn't be contacted	3	10,00%
Total	30	100,00%

VI. BALANCE OF THE PILOT

20. The positive side:

?? The XML/XSL technology is easily applicable for implementing electronic questionnaires. It is relatively easy to generate questionnaires, therefore it can be the technology of a questionnaire design and implementation tool for statisticians;
?? The role of a metadatabase and respondent register is growing;
?? E-commerce type web applications can be completed by formalised e-mail communication;
?? Paper-based and electronic reports should be handled in a homogeneous way;
?? Respondents who were able to change from paper to Internet are happy;
?? There is a relatively large number of respondants who were not able to change, although wanted to.

21. The negative side:

?? Electronic registration is needed;
?? Companies are not interested in their own data security;
?? Better integration with respect to validation with the paper based collection;
?? Off-line functions are to be extended;
?? Better assistance in “automatic” filling in, e.g. Excel interface, DB interface, package interface;
?? Help desk must be set up;
?? To be accepted, higher education & support is needed for both the respondents and the statisticians in the county offices responsible for data collection;
?? Marketing & PR is necessary, aimed at respondents, accounting companies and software producers.

22. The financial side:

?? The cost of the pilot was roughly 10 mFt, that is USD 40.000;
?? Expenses were divided equally between the developing partners;
?? The total cost was one to twenty to a planned EDIFACT project in 1999;
?? The system developed during the pilot can be put into operation countrywide with no extra cost;
?? The budget to operate this system in 2002 for the integrated economic survey countrywide is about 18 mFt, i.e. 72.000 USD;
?? Most of the methods and software developed in the pilot are re-usable in the e-government project.

23. The directions of development:

?? The data collection system is to be re-designed. A “mixed media” application should provide all functionalities we need to cover paper-based, e-mail based, Internet-based and perhaps other (e.g. telephone-based) data reporting. The database of respondents, reporting, surveys and raw data reports should be fully integrated, perhaps on XML basis. This database should closely co-operate with the metadatabase;
?? A computer aided “questionnaire/survey design” application is to be developed. Part of this application is about generating pieces of codes which are written (programed) manually at present;
?? It is feasible to develop a so-called “statistics module” or object, which is an interface between in-house information systems and the Central Statistical Centre. Software packages like accounting, payroll, or more integrated ones like SAP, ORACLE Finance should be able to embed such a module. Software producers should be involved and made interested in this kind of cooperation;
?? Accounting companies must be regarded as special data reporting sources and special attention is to be paid to interact with them;
?? The webserver functions should be platform and provider independent. Parts or perhaps the whole on-line data collection system should be the subject of web hosting.

The “dream” system