UNITED NATIONS
ECONOMIC COMMISSION FOR EUROPE

CONFERENCE OF EUROPEAN STATISTICIANS

Seminar on New Frontiers for Statistical Data Collection
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (ii): New methods and technologies

TOWARDS A MODERN ELECTRONIC DATA COLLECTION SYSTEM

Contributed Paper

Prepared by György Benoist, Hungarian Central Statistical Office, Hungary

I.   Introduction

1.   Over the past years statistics have been facing several new challenges. Many times the interests and conditions behind these challenges are contrasting, thus it is not easy to find a good solution that satisfies every concerned participants.

2.   All data users – businesses, natural persons, and in particular decision makers (national governments and the European Union) – require more detailed and timely data of the same (if not even higher) quality as globalisation is spreading and the world's economy growing. Economic crises during the past 5-6 years pointed to the weaknesses of the nowadays statistical system and initiated a common way of thinking towards a new up-to-date one that could react rapidly, more adequately and more comprehensively, and being able to give a proper answer to new changes and unexpected events.

3.   On behalf of data providers – especially businesses – there is a strong demand for reducing costs and administrative burden. They are interested in decreasing the obligatory statistical data needs and satisfying them relying on their already existing registrations related to their activities which are subject of other administrative, tax or book-keeping controls. Although burden imposed by statistical data collections amounted only a few percent of the overall administrative burden, the comfort-feeling connected to reduction of statistical burden is rather huge. From this point of view harmonisation between statistical and administrative data needs and timeliness as mush as possible is essential towards reducing the burdens.

4.   Besides the demands mentioned in the previous paragraphs, generally statistical services are facing with continuously decreasing financial and human resources.

5.   The driving forces from these directions indicate, that reducing burden on respondents as well as on statistical services is essential, and cannot be delayed. Traditional data collection methods are not always suitable to satisfy all the new demands and give an acceptable answer for these challenges. There are several instruments how respondents' burden in statistics can be reduced parallel resources saving and effective use, for example application of new data collection techniques, look for new data sources, make steps towards integration of data sets and benefit from the synergies. All these together can be productive. All the efforts towards a reformed

statistical system are in line with endeavours of European Union and national government policies.

6. One of the possible tools is electronic data-collection methods. They are not absolutely new tools since e-mailing excel files and XML-type of system have been used for more than a decade in Hungary for statistical purposes. Unfortunately only this latter method can be really considered electronic data collection method. Nowadays in Hungary strong majority of the statistical data collections is still mailed on printed forms to the respondents, who can fulfil their reporting obligations using the following methods: filling-in and returning the paper questionnaires by mail or fax, returning the Excel-type questionnaire made available on the HCSO website by e-mail or using the web-based data-collection application (KSHXML).

7. Although XML-system was under continuous development[1], the Hungarian Central Statistical Office (HCSO) has decided in its strategy for the period 2009-2012 to develop a rather new and modern electronic data-collection system, which would simplify further and shorten both the data collection and the data-preparation phases. The plans of the new system aimed the introduction of an application that could replace the recently used one eliminating its limitations and simplifies the statistical work for both the respondents and the statisticians. Additionally to initial plans economic crises in Europe intensified the efforts to modify the development towards a cost-effective and more user-friendly system, which applies the latest technology, and not only replaces the old one, but significantly extends its use among the data suppliers and statistical domains. All these ambitions are in line with the endeavours towards the e- administration.

8. The aim of this document is to present the newly developed tool, and to collect during the discussion ideas for possible further fine-tuning.

II.   The new electronic data collection system: ELEKTRA

A.   The project

9. Based on the strategy of the HCSO for the period 2009-2012 initial steps have been made towards the new electronic data collection system. The two basic aims of the project were development of a modern, user-friendly and rapid application that exceeds the limitations of the used XML-system[2] and able to deal with electronic questionnaires at least amounted to those together returned via e-mail and XML at that time. Besides it should be qualified to design questionnaires and to contain all already used functions.

10. Planning of the new system has been started in September 2009. The first milestone was finalising the specification that listed and clarified all the operational requirements. Based on this document and taking into account the features – including the experiences gained – of used XML-system the actual work could start at the beginning of 2011. The new system was named ELEKTRA. Second milestone was installing the modules of the system in the new hardware environment (September 2011). Testing the system has started immediately by participation of both IT experts and professional statistical staff.

11. Meanwhile due to further reduction of the HCSO budget and the decision of the Hungarian government on reducing administrative burden on enterprises significantly, statistical processes carried out in HCSO had to be overviewed. In order to improve electronic data collection activity and to enforce burden-reducing demands raised by data providers and the government, the ELEKTRA-project was comprehensively examined. This complex review clearly showed that the new system should overrun the original informatics requirements and features planned. Additional professional and statistical considerations should be added and combined with the

---

[1] KSHXML operates since 2004.

[2] Limited questionnaire size, rather complicated programming of the validation rules, filling-in is slow, data import is restricted to rich client application, desktop is not user-friendly, communication via governmental portal is not possible.

initial ones, as it is never sufficient to introduce a technological solution on its own. Decision was made that utmost statistical questionnaires should be included in the new system and ELEKTRA should operate as unique and obligatory tool for statistical data sending for businesses from January 2013.

12.   These emerged statistical requirements surpass significantly the former plans; hence adaptations of the system under development and additional efforts had to be made. One hand it became necessary to strengthen the initially suitable hardware environment and infrastructure as their capacity – although it was over-dimensioned 5-times – seemed not eligible receiving the foreseen amount of forms and handling the respondents according to new expectations[3]. On the other hand softwares had to be checked whether they will operate smoothly if the number of the data collections and questionnaires increase significantly.

13.   Presentation of ELEKTRA has started in October 2011. In the first round about 50-60 businesses already sending data electronically have been invited to return their opinion. One month later the system in operation has been presented on an open day event and from that time several opportunities were used to present it to the future partners and statisticians. HCSO lays emphasis on the communication towards the data providers in order to increase willingness and capacity of respondents to provide their data via ELEKTRA.

B.   Characteristics and parts of ELEKTRA

(a)   Main features and structure

14.   ELEKTRA is a web-based data collection tool that do not need install on the working station[4]. It is available on the HCSO web servers, and can be accessed and completed via web browser with a minimum requirement of at least Adobe Flash Player v10.6 and Internet Explorer 6, and minimum 1 GB RAM is needed on the PC. When these minimum requirements were established not only the average IT-environment of the businesses already reports, but also information on the possible worst situation were taken into account.

15.   To reach the objectives of the project a modern data collection system was imagined, which

   (a)   is attractive for both the respondents and statisticians,

   (b)   is suitable to handle utmost statistical data collections (expansion compared to the KSHXML),

   (c)   ensures data-import on the respondents' side (tables can be exported and re-imported; previously only installed version of KSHXML supported this function),

   (d)   makes possible the automatic reception of data and their upload into the data-preparation system in HCSO,

   (e)   fulfils all expectations in data quality,

   (f)   operates reliable, secure and cost-effective, and

   (g)   ensures also the access via the governmental portal to our partners, the respondents.

16.   According to the plans at the beginning hardware elements are functioning alike those of the XML-system and it was not necessary to increase their measurement. Compilation of configuration was based on estimations, whereas in-situ charging could not be carried out or the charging-test could have led to a misleading result due to the fact that each statistical

---

[3] Compared to the formerly planned about 80 different forms and annually about 100 000 questionnaires, the new expectations are around 140 different questionnaires and estimated 1 million returned questionnaires.

[4] KSHXML, the former electronic data collection system in Hungary had both off-line and on-line versions.

questionnaire represents a different charge according to its complexity and size, transmission and procession methods. With the extension of the requirements significantly bigger infrastructure and environment were needed: for example extra memory and auxiliary processors were added, bandwidth for connection was widened, and further development of the TIR was carried out.

17. Basically the work flow did not change in comparison with XML; the necessary phases are the same than before (*Figure 1.*). The new features and/or elements of ELEKTRA are:

    (a) renewed on-line functions

    (b) modernised respondents' desktop with attractive design, ergonomic operation, more functions (task and calendar)

    (c) questionnaire-designer and -completing module

    (d) simple connection possibility to other data collection tool[5] without using again ID and password

    (e) possibility to enter via governmental portal.

Figure 1. The simplified overall architecture of ELEKTRA
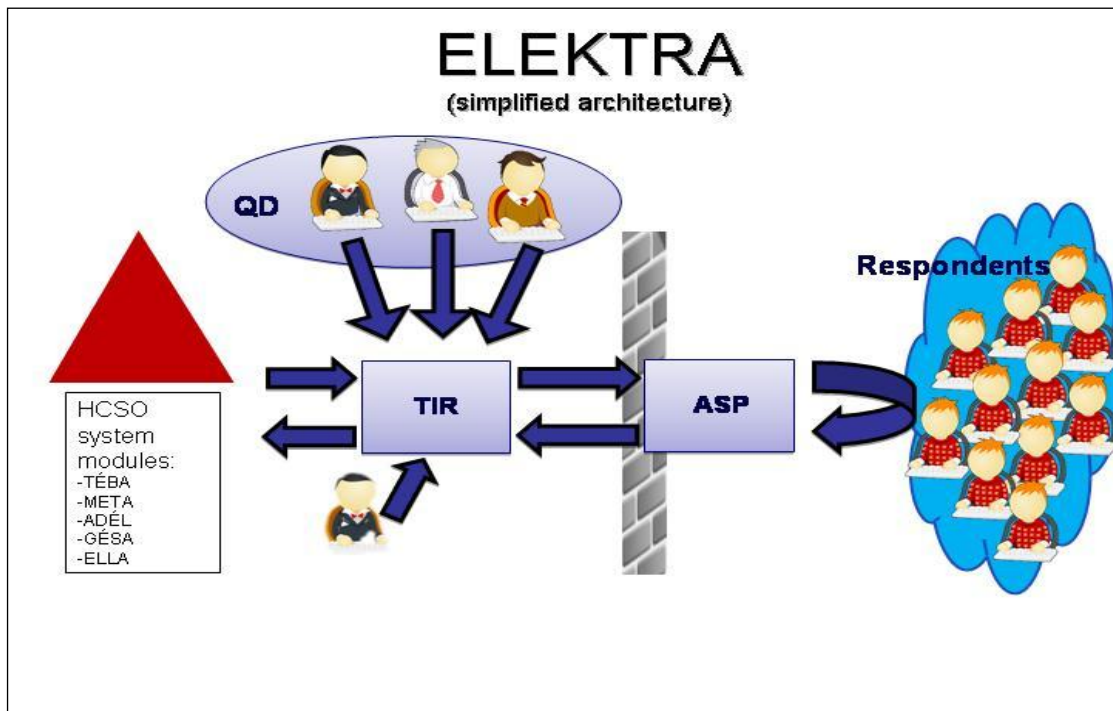
(b) Modules of the ELEKTRA-system

18. During the development the following modules have been adapted significantly or new ones were developed.

19. **TIR** [„activity management unit"]
This is the central module gathering all management and monitoring functions, accessible directly only to HCSO staff. Still having open interfaces for future modules.
Tasks connected to this module are:
- management and operation of the system

---

[5] Application that collects data on foreign trade; operates similarly to KSHXML.

- registration of the data respondents (both businesses and natural persons) in client database
- dealing with work-flow (personalization of the questionnaires with name/address and/or data, sending them out and receiving them etc.)
- data exchange with ELEKTRA-modules and other systems of HCSO (database connections with META[6], GÉSA[7], ADÉL[8] TÉBA[9] etc.)
- monitoring of the operation and reporting
- CRM [helpdesk supporting] function: registration of respondent and their problems, helpdesk, technical preparation and execution of new recruiting and actions; offers a better quality correspondence with the partners

20. **Questionnaire-designer** (QD) [*Figure 2.*]
For the sake of reducing costs (free of charge) and already existing, open-coded software is used based on NetBeans Rich-Client Platform (RCP) technology. It allows creating bigger and more complex questionnaires than before, sometimes with several repeated tables (for example reporting data of local units). Main functionalities are: editing in graphical WYSIWYG (with Visual Library API) and having a complex window-handling system. It is connected directly to TIR with Spring HTTPInvoker protocol, and indirectly to the other modules, like META, ELLA[10].

Typical functions:
- pixel accuracy design of chapters, panels and components (from csv file too), as well as repeating chapters and panels (length of questionnaire can be different in case of different respondents)
- "mouse-ready", drag & drop, undo/redo, cut/copy/paste functions
- more than one questionnaire can be opened and used at the same time, it is possible copying and inserting to and fro between them
- visual and hierarchical synchrony
- controls and calculations (Java script)
- consistency checks on the questionnaire
- testing within the QD module

Figure 2. The view of the questionnaire-designer (example)

---

[6] META – name of the database for metadata (indexes, nomenclatures, description of data collections and forms, etc.)

[7] GÉSA – name of the data collection organiser and monitoring system in HCSO

[8] ADÉL – name of the platform in HCSO that handles returned data (paper and electronic); place of data entry, checks and corrections

[9] TÉBA – interface system between KSHXML/KSHEMAIL and ADÉL/GÉSA that registers returned forms, stores them and unloads data into the ADÉL

[10] ELLA – database for uniform description of validation rules in HCSO
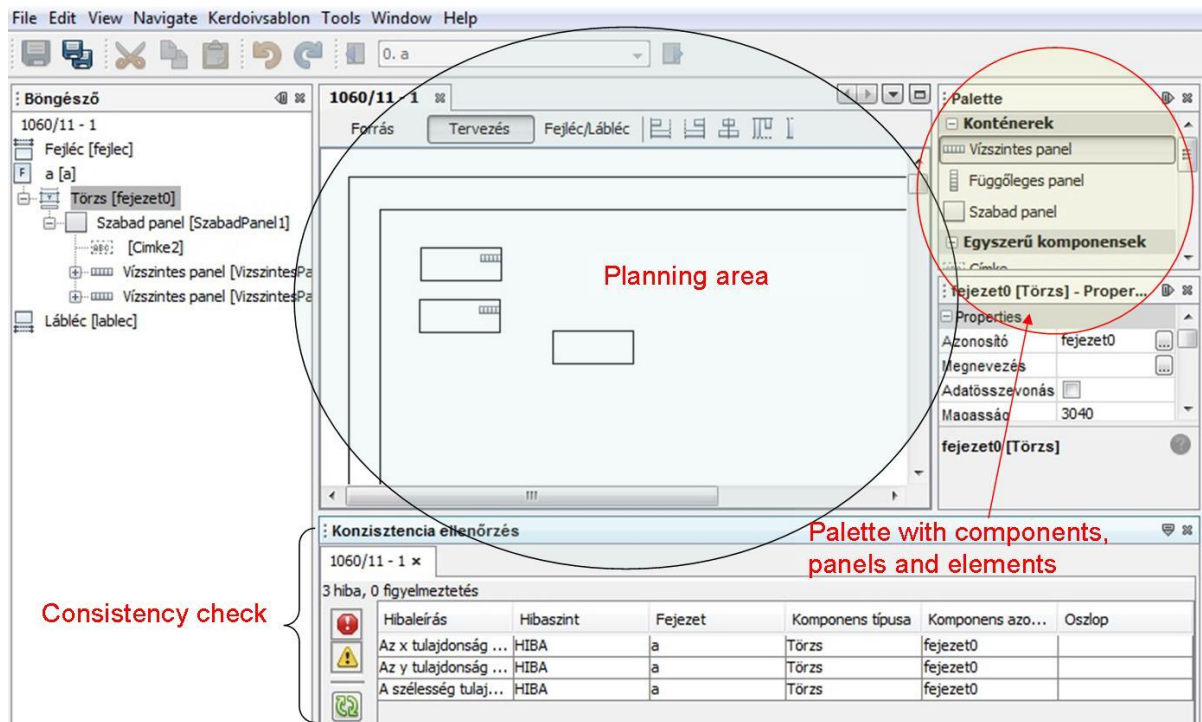
21. **Respondents' client and server module** [application server] (ASP) ´
    This module contains the respondents' functions; functionally it is a "mirrored TIR". It has an up-to-date new design (similar to e-mail program), and user-friendly desktop [*Figure 3*.]. Reaction time is short and respondents can fulfil their data providing obligations using the same tool in case of different questionnaires. Its main tasks are dealing with the questionnaires, data and competences of the data providers and users, allowing the selection among the actual forms and dealing with encrypting.

    Functions:
    - assisting registration (respondents have to register on-line when using ELEKTRA the first time);
    - assisting access via government's portal;
    - serving the questionnaire-completing module;
    - temporary data-storage between QCM and TIR;
    - allowing export/import activity (no need for manual data entering) and printing;
    - runs encrypting program and stores the encrypted data; and
    - transmission of data to TIR.

Figure 3. View of the respondent's desktop (example)
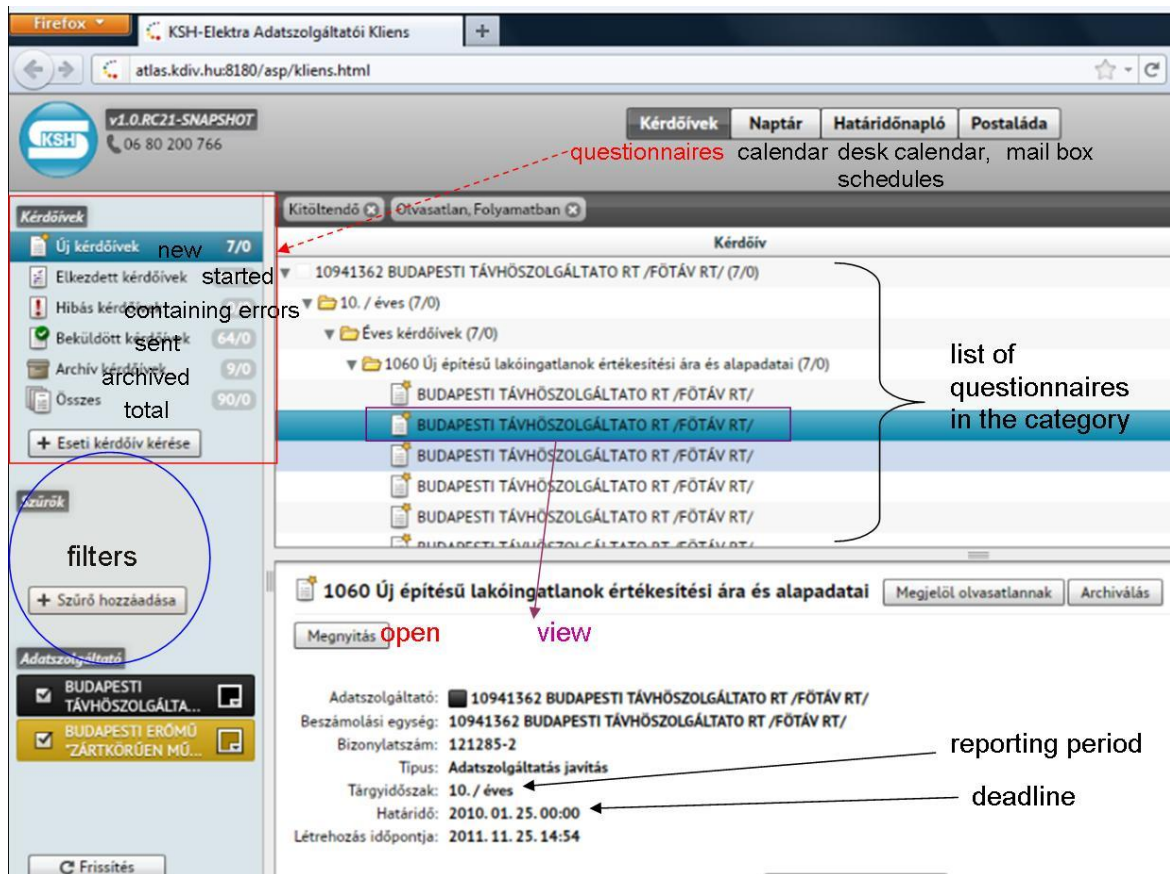
22. **Questionnaire-completing module** (QCM)
This module should be unambiguous, modern and elegant with ergonomic and pleasant functionally in order data providers would use it with satisfaction when filling-in the questionnaires. FLEX frame-system, a new, open-coded, but reliable technology was used during the development of this module. Adobe Flash plug-in, generally available in almost all browsers, is needed for running FLEX. There is no need installing the application on the PC, calculations and controls are based on JavaScript, exporting and/or importing data is possible in xls, csv, json, xml and pdf formats.

C. Ongoing works

23. In order ELEKTRA would operate properly and could serve a wide range of data collections several additional works are going on or still foreseen. In order to finish all these tasks in time a close cooperation between statistical and IT experts is needed.

24. Introduction of ELEKTRA gives a unique opportunity to overview the whole statistical data collection system, and where and when appropriate eliminate unnecessary cells or re-consider tables keeping in mind reducing respondents' burden and cost-effectiveness. Simpler is the work when the questionnaire is already available in electronic format also, and more effort is needed when only on paper version exists. Reporting deadline could also be subject for reconsideration in order to better to other administrative (e.g. taxation) ones.

25. The revision of the recent data collections are not limited only to re-design the questionnaires, but manuals and other handouts have to be rethink, META information should be clarified and refined, in certain cases definition of indices and indicators must be harmonised, as well as the supporting systems have to be maintained and adapted, too. For example structure of the questionnaires, indexes and nomenclatures as well as their (sub)variations used have to be described in the META system, validation rules to each questionnaire/table/cell are going to be documented in a standardized manner in ELLA database, and error-messages could be made

more interpretable also for data suppliers etc. Another task is creating the links with the data processing system.

26.   ELEKTRA offers the possibility of elaborating bigger and more complex questionnaires, thus more sophisticated on-line controls and data validation are possible towards better quality data and better consistency. The limit of these controls is only the reaction time of the system: if it would be too long, respondents would feel it more burdensome.

27.   A data collection system could work properly when data providers could check – and if needed revise – their data from the previous reporting period(s). In order to ensure this possibility and meet the requirements, the questionnaires still subject of revision are also re-designed in ELEKTRA. Returned data will be migrated in these electronic questionnaires and made available for the respondents, who can revise them on-line. Other questionnaires, which might still needed for better coherency of data, but not subject of revision, will be transformed in pdf format files and made available in ELEKTRA.

28.   Besides testing, still ongoing task is the continuous correcting of the errors, verifying these corrections, adapting the system functions to most ergonomic requirements – based on feed-back from data suppliers and statistical staff.

29.   Another important aspect is making the electronic data collection legally binding for all respondent, however special attention should be paid on involvement of smaller units, who might be not able fulfilling its obligations (due to having an older PC or software version). Experiences heard on international forums[11] warn us to make this transition carefully.

III.   Timing of introduction and expectations

30.   Introduction phase of ELEKTRA starts in October 2012 when the first batch of data collections[12] will be made available electronically. In the first round data suppliers already reporting via XML can use it for data transfer – at the beginning on a voluntary base. This "testing-period" would provide indispensable feedback refining the system before the January 2013 "general" start.

31.   By Introducing ELEKTRA our positive expectations are the following:

   (a)   size or complexity limitations of electronic questionnaires are solved;

   (b)   data collection method is more user-friendly, filling-in and checking is simpler and faster (built-in automatic validation, controls and feasibility checks);

   (c)   less follow-up contacts (phone calls or e-mails) are needed, correction and imputation are reduced;

   (d)   security of data transmission grows significantly;

   (e)   more efficiency and better quality of data, whereas capacity from manual data entry will be moved towards quality controls and validation;

   (f)   period will be shorter between reporting period and dissemination; and

   (g)   CRM module helps monitoring more effectively the respondents and provides a useful helpdesk opportunity.

---

[11] Last time: „Facilitation of Data Transfer from Enterprises to NSIs" Luxembourg, 22-23 March 2012.

[12] The first 38 questionnaires are the data collections still running in 2012, generally monthly and quarterly ones.