I. Introduction

1. The Integrated Data Entry and Validation System (named ADEL) in HCSO supports the validation and correction of survey data. The system is integrated with other IT systems, it uses standard interfaces for communication. This meta-driven system receives questionnaire data from the Electronic Data Collection System (ELEKTRA), receives datasets from the Data Collection System for Administrative Data (ADAMES), it reads and updates survey control information (stored in the GÉSA system). The ADEL system provides the validated data to the Integrated Data Processing System (EAR) and sends information to the Tables System (dissemination system) about the questionnaires modified in the final correction phase.

2. The system has a standardized year and survey independent framework which consists of the general functions (unified applications for error handling and explaining, for managing contact information, for reporting, etc.). However, the survey-specific validating, correcting applications and reports may be flexibly involved. The system supports the phases of data validation, including the preparatory and the final correction steps as well. The main drawback of the system is that it supports manual correction.

3. The users’ actions are controlled by their access rights (read, update, admin), their department and survey level rights (these are metadata information) and the status of the survey (open, closed).

4. According to severity aspect, the validation rules are classified (into four categories) and this classification defines the method of error handling. According to timing aspect, each validation is classified whether it must be processed in online validation or in the batch (background) validation phase only, and statisticians can make different validation groups as well in batch mode.

5. The system provides product- and process quality management information: logging, collecting information to indicators and enforces compulsory steps in the validation process.

6. The ADEL system enforces the development rules in the database level and in the file system (naming convention, standardized object and folder structure). All information and data are stored in a central (Oracle) database which is protected by strict access rights and all of the functions, corrections are regulated and controlled by the ADEL system.

7. It is an inhouse-developed system and the HCSO develops it continuously. In year 2014 about 130 surveys used the ADEL system for data validation while 145 surveys were included and managed by IT systems and 130 surveys’ data was received by the Electronic Data Collection System (ELEKTRA). It
is not expected to significantly increase the number of included surveys in the next years because some surveys have only one or two data providers and it is not effective to develop the IT background and applications for them.

II. Location of IT systems in the statistical business process in the HCSO

8. In line with the international and national intentions to develop and implement integrated IT systems into the statistical business processes, the HCSO has been developing and using such tools for decades; all supported by the common metainformation system.

9. Mentioning the most prominent ones related to establishment surveys, such tools are the Integrated Survey Management System for Establishment Surveys (GÉSA), the Electronic Data Collection System (ELEKTRA), the Integrated Data Entry and Validation System (ADEL) and the Integrated Data Processing System (EAR). The latest system is the recent addition on this regard and is currently under implementation. These tools cover and support the core of the phases of the Generic Statistical Business Process Model (GSBPM as shown on Figure 1). The IT tools are organised in alignment with the phases of the ESTFM, which is the Hungarian adaptation of the GSBPM.

10. All systems are meta-driven and integrated in the sense that they can use each others output as an input and their interoperability is ensured by common metadata. ADEL is in strong connection with ELEKTRA: after some checks the questionnaires sent through the electronic system can be easily uploaded into the Data Entry System.

---

**Figure 1**
III. The users of the ADEL system and their activities

11. In the **design preparation** phase the statisticians at the **subject-matter** departments are responsible for designing the questionnaires, describing the items with metadata identifications (measures, nomenclatures). They write the metadata information in the common Metainformation system, define the check rules of questionnaires and write them in the database.

12. In the **development preparation** phase the **IT developer** is responsible for designing the database structure of the questionnaires, creating the database tables and other objects and he grants the necessary access rights on objects. The IT developer compiles the validation program codes using a code-generating application and saves the validation program codes in a code library (Pl-sql library). After that IT expert develops the online validation application, designs the screens which must follow the form of the questionnaire. Finally, he makes the required reports if the general reports are not adequate. These developing steps are supported by template applications (online application for correction), built-in applications (application for modifying data providers’ contact information) and standard procedures (error handling procedures) as well. Developers can build their applications with using ready-made modules, procedures and reports.

13. In the **implementation phase** the statisticians in the **local offices** validate and correct the questionnaire data with the ADEL system. In case of data validation error they get in touch with the data provider and consult how to correct the error. Statisticians can request a new questionnaire or if the error is acceptable they accept it with explanation.

14. In the **final correction phase** if the statisticians at the **subject-matter** departments face an error during the data analyses he can make a decision whether to give back the rights of the correction to the local offices or to correct the data by himself. At that time the survey’s database will be opened for final correction and the responsible department (local office or subject-matter department) makes the data correction. The final correction may be available for all the data providers or partial (when users can modify questionnaires marked for correction in advance). After finishing the correction the database will be closed, protected and nobody can modify data later.

IV. Most important functions of the ADEL system

15. After the successful login the users of the system can choose the function from the menu structure of the system to start. The menu structure shows only those functions which are allowed to users while others are hidden. The functions are controlled by the users’ access rights (read, update, admin), their department and survey level rights and the status of the survey (open, closed). The most important functions are included:

A. Data entry, online validation, correction

16. New questionnaire’s data may be entered then saved to the database. The stored data may be retrieved from the database at any time and may be corrected with an online data validation, correction application. If the user has only read only (select) access right or the survey’s dataset is closed (and protected) in the database, the system hides the data manipulation functions. At that time all applications (online validation application as well) start only in read only mode.

17. The user can navigate between questionnaire tables, rows and items he can correct data with the survey’s online data correction application. When the user passess an input item the validations for it are processed. If there is an error the error message text appears in a popup window and the user can handle the error according to its severity category.
18. The display items next to input items can show the previous period data, the index or other information which helps the correction.

19. After appearance of the error message the user can make a decision whether to correct the data, optionally to give an explanation or to save the data without correcting and correct it later.

**B. Batch (background) validation**

20. After finishing the online data correction users can start batch (background) validation sessions. This is necessary because in online mode error messages are only displayed when the user navigates to the data. Moreover, there are validations which are planned to run only in the background because the process would be too slow or the validation would be too complex and the statisticians would not be able to identify the source of the error. The metadata information of the validation defines whether it is active in online mode or only in the batch (background) session.

21. Before starting the background validation session the user can filter the questionnaires to be checked and defines which validations to start (defines the validation group level).

22. Filter questionnaires of data providers to be checked: before starting the batch validation session the user completes a parameter page which gives the questionnaires to be checked. These filter parameters may be properties of data provider, e.g., region code, activity code or the date when a questionnaire was last modified.

23. Validation group level is metadata. Using the group level information the users can start validation sessions which include increasingly more and more validations, in nested structure. For instance, the 1st group of validations include automatic validations run for all questionnaires which were received from the electronic data collection system. If there are validation errors in this phase the system sends the error report to the data provider. The 2nd and higher groups of validations can be started later by the users of the system. The 2nd group level includes the validation inside the questionnaires, the 3rd group level includes the validations with the previous period data, the 4th group level includes the validations with other surveys’ data. Therefore, the higher is the validation group level the more validations are included. The questionnaire is considered validated only if the highest validation group level is completed. (The higher level of validation includes all the levels below it, e.g., if the 3rd group level validation is run then the 1st and 2nd group validations start as well.)

**C. Preparing error report**

24. Every failing validation action inserts an error row into the survey’s error table. When the error is corrected, the error row is deleted. So the error report shows the error rows from the error table. Users can make reports about questionnaire data and/or only the error rows at any time. Before reporting parameters can be given to filter the questionnaires and the error rows.

25. The users can set whether to retrieve only the error rows from the error table or to retrieve the questionnaires data and the error rows as well.

(a) All errors are stored in the error table but users can filter error rows to display on the report, e.g., only the severe errors but not the warning errors.

(b) The error report shows the error message and also shows the identity code, the name of the data provider, the number of the table and the row where the error occurred. Next to the error message other information can be displayed which helps the correction of data, e.g., index of previous period data, or the previous period data itself.

(c) The error reports can be exported to text or Excel format and can be printed if needed.
D. Accept errors with explanation

26. If the error severity level is 'caution' (is acceptable) and the error occurs, the statistician can consult with the data provider and makes a decision to correct the data or accepts the error with explanation. At that time the error stays in the error table, the status of the error row changes to 'accepted' and the system logs the information who accepted the error, when and with what comment. Even though the different surveys have different data structures and different error table structure, the HCSO developed a data structure independent - general - application for accepting errors.

E. Explanation report about accepted errors

27. When the statistician has consulted with the data provider and has accepted the error, the system records who and when accepted the error and with what explanation. Users can ask for report about accepted error rows and their explanations. Users can give the same filter parameters as for the error report case. The explanation report can be exported to text or Excel format and can be printed if needed.

F. Special applications and reports

28. There are year and survey-independent functions in the system but if the statistician needs a special function, the IT developer can develop a modified or a new application and/or report. If an application is modified (and has the same name as the general application) and is stored in the surveys’ folder, the system recognizes the modified application and starts the modified one instead of the general application. When a new application is developed and stored in the surveys’ folder and the needed metainformation is written then the system expands the menu structure with the new function.

G. Collecting management information and displaying them

29. The system logs the following information:
   - who, when and what functions were started;
   - who and when entered (or loaded), corrected, deleted, controlled a questionnaire;
   - who and when accepted an error and with what comment;
   - how many times was a questionnaire opened for correction;
   - how much time a questionnaire took to correct;
   - how many errors and accepted errors are in a questionnaire;
   - how many errors and accepted errors are in the survey’s database.

30. The users can prepare reports with the information content listed above, with filtered parameters, if needed. All reports can be exported to text or Excel format and can be printed if needed.

H. Regulating and controlling the workflow

31. Metadata information is stored by the system including the users’ identity code with their department code and surveys with the responsible departments and users. Furthermore, users can define which data providers belong to them. So when they enter the system, they can choose from the following options:
   - (a) they want to manage only their own data providers;
   - (b) they want to access all of the data providers;
   - (c) they can retrieve and correct other users’ data providers (substitute someone with the same access rights and department code). In this case the system logs activities with the logged user’s identity code but user can manage the substituted colleague’s data providers.
32. For each survey there is one (sometimes two) master user (survey administrator) who has extra rights and can run workflow control functions. These master users can authorize or revoke authorization for receiving data from the electronic data collection system. With their special rights they can open and close the surveys’ database and they can accept severe errors with explanation.

33. When data receiving from the electronic data collection system is disabled then the questionnaires are waiting for loading into the database (into the ADEL system database tables) till the master user allowes data receiving and data loading again.

34. The IT developer, after completing the IT background activities (has designed the database structure, has created database tables, has granted rights for users on objects), initializes the survey, e.g., opens the database using a built-in function of the system. It lets the responsible departments and users know that they can start the data validation and correction. When the statisticians have finished the data correction, the master user closes the database running the ‘close’ function and the database becomes protected, so after that nobody can modify any data. The system enforces compulsory steps for the survey, for instance,
   (a) all validations must be processed for all questionnaires;
   (b) all errors must be corrected or accepted with explanation by statisticians or the master user;
   (c) all previous periods must be closed.

At that time the survey’s database satisfies all defined validation rules and it is protected against any modifications till the database is closed.

V. The characteristics of the ADEL system

A. Integrated with other IT systems

35. The ADEL system uses standard interfaces and there are strict rules and conditions of the connection with other IT systems. Some examples of connections and criteria.

   (a) The electronic questionnaire data can be loaded into the ADEL system only if the survey’s database is opened, the receiving of the questionnaire data is permitted and the questionnaire is the first version. If the data provider resubmits the questionnaire which had been modified by the statistician then the new questionnaire is waiting, the old one is not reloaded. In this case the statistician receives an email that a new questionnaire has arrived and he can view and print it if needed and he can reload the new questionnaire manually into the ADEL system.

   (b) If the survey’s database in the ADEL system is closed (and protected) or the receival of electronic questionnaire data is suspended then the questionnaires are waiting until the conditions change.

   (c) The data transfer from the Data Collection System for Administrative Data (ADAMES) is enabled if the survey database is opened and the user has access rights for the survey. It is currently under development.

   (d) After finishing the data validation the surveys’ datasets are loaded into the database of the Integrated Data Processing System (EAR). This is allowed only if the survey is closed in the ADEL system (if the dataset is validated and protected against modifications).

B. The system is meta-driven

36. The framework application, the standard functions, procedures and the built-in applications are controlled by metadata information. See a few examples below.
(a) The functions of the framework application can be expanded

- The framework application is a menu structure with standard functions but it can be expanded with survey-specific applications and reports. In this case the IT expert develops the new applications (reports), saves them in the survey’s folder and describes the metainformation (application/report name, type). When the user has logged in, the system reads the meta description and recognizes the new functions and modifies, expands the standard menu structure.

(b) The standard error handling is metadata-driven

- The severity level (warning, acceptable, severe, fatal), a feature of the validations is described in metadata tables and used to control the standard error handling processes. For example, it defines how the error message appears on the screen in a popup window or in the message row on the bottom of the screen; how many buttons will be displayed in the popup window. The error message text is metadata as well.

- If someone – who has proper right – changes the error message text or the severity level of a validation, the error handling processes work differently immediately. It means that the IT developer does not have to modify the applications if the severity level or the error message changes, he modifies only the metadata in the database.

(c) Managing the access rights at department and user level

- It is described as metadata, which department (or departments) is responsible for validation and correction of a survey. When a user has logged in, the system controls whether his department has right for the survey and if does not have the proper rights, the user gets an error message and the application exits.

- It is described as metadata that the final correction is allowed (or not) and for what department.

- Users, who are responsible for validating and correcting surveys, are listed in a metatable. If a user, who does not have right for the survey has logged in for correcting data, gets an error message and the application exits.

- For every survey it can be described as metadata which data providers belong to a user. When a user marks he wants to work with his own data providers all the applications and reports filter the questionnaires to his ones.

(d) Nomenclature code validation

- The items of the questionnaire are identified on physical (table, row, column) and on logical (measure, nomenclature, nomenclature code) way and it is saved as metadata information. The standard nomenclature code validation procedure uses this metadata for controlling the code validation. When the entered code is valid the standard code validation procedure returns the denomination of the code, if the code is invalid the user gets an error message.

(e) Batch validation sessions controlled by validation groups

- The batch validation group number is a metainformation of the validations. The system can validate data inside the questionnaire and between questionnaires (previous period or other survey’s data). If the statistician would like to validate data inside the questionnaire first and validate data between questionnaires later, he can do it with using validation groups feature. User can raise the validation group number so more and more validation is processed and with the highest group number the questionnaire is fully validated.
C. The ADEL system is built modularly and expandable flexibility

37. The system consists of the general year and survey-independent framework application with standard functions, standardized applications, reports in the application folder and the survey-specific applications and reports in the surveys’ folders. (Every survey has its own folder.)

38. If needed quite a lot (but not all) general applications and reports can be overridden with survey-specific applications, reports. For instance, the general error report can be overridden if the statistician has a special request. At that time the IT expert develops the modified report, saves it in the survey’s folder and when the users enter the system, the framework application recognizes the modified application in the survey folder and the new one is started instead of the general one.

39. Furthermore, the standard framework application can be expanded if needed. At that time the IT expert develops the new application (or the new report), describes the necessary metainformation and saves the new application in the survey’s folder. When the users login the framework application reads the metainformation, expands the menu structure and the users can start the new application.

D. The system functions are controlled by the access rights and the status of the survey

40. According to access rights the users are classified into four groups, these are roles.

(a) IT developers have administrative ‘admin’ access rights. They create the necessary database objects, grant the users’ rights on the objects and they can write and modify metainformation.

(b) The application framework shows the ‘read only’ functions to users who have only ’select’ rights, e.g., data retrieving, preparing questionnaire data, error or explanation reports.

(c) The users with ’update’ access rights can modify data, can give explanation to error rows (if the severity level of the error is acceptable) and they can prepare any reports.

(d) The master users are the survey administrators with extra rights. They can run workflow management functions (open and close the survey database), and can accept severe errors with explanation.

41. The ’update’ and the ’master user’ roles are managed dynamically by the framework application, it means that the right is activated (switched on) when the user with this role enters and the right is deactivated (switched off) if the user exits from the system. These data modification access rights are activated only by the ADEL system, so the users can not modify data with other tools, e.g., Sqlplus, SqlDeveloper.

42. The users’ activities are controlled by their access rights (roles), their department and survey level rights (metainformation) and the status of the survey (open, closed). For example, when the survey’s database is closed nobody can modify data, the framework application hides the data and error row manipulation functions.

E. Standard error handling method, error categories

43. When in the validation process an error occurs the error row is inserted into the survey’s error table and when the error is corrected the error row is deleted. If the error is accepted by the statistician the error row stays in the error table and the status of the error changes to ’accepted’. These error management operations are carried out by general error handling routines stored in the database as stored procedures.
44. The validation program codes are compiled with an application, they are implemented in Pl-sql programming language and saved into Pl-sql libraries. The online data correction application and the batch (background) validation sessions start the validation program codes from the Pl-sql library. The advantage of this method is that the IT developer implements the code only once, he does not have to implement different program codes in the online and in the batch application.

45. According to severity aspect, the validation rules are classified into four categories and this classification defines the method of error handling. This severity level metainformation controls the standard error handling procedures and it defines how the error message is displayed on the screen, where the user can navigate to after the error message appears and whether the user can write comment (accepts the error) or not. The error message text is metainformation as well.

46. The severity levels are the following:

   (a) Warning error

   • The lowest level, it is only information. If the error occurs, an error row is stored in the error table and it is accepted automatically with the explanation 'Info error'. In the last (message) row of the online validation application screen the error message is displayed and the navigation goes automatically to the next item. Users can ask for error report with or without warning errors.

   (b) Acceptable (caution type) error

   • If the error occurs the error row is saved in the survey’s error table. In the online validation application the error message is displayed in a popup window with three buttons. If the statistician can not correct the data immediately (postpones the correction), he can navigate to the next item with the button 'Next' without correcting the data. If the statistician can modify the error, he chooses the button 'Modify' and with the button 'Accept error' the statistician can accept the error with comment. If the statistician corrects the error the error row is deleted. If the error is not corrected but accepted the error row remains in the error table and it is logged who and when accepted the error and with what comment.

   (c) Severe error

   • If this type of error occurs the error row is saved in the survey’s error table. In the online validation application the error message is displayed in a popup window with two buttons. If the statistician can not correct the data immediately (postpones the correction), he can navigate to the next item with the button 'Next' without correcting the data. If he can modify the data he can choose the button 'Modify'. If the statistician corrects the error, then the error row is deleted.

   • The severe error is acceptable only by the master user of the survey with a general severe error accepting application (displayed only master users). In this case it works as an accepted error row, it is logged who accepted it, when and the status of the error changes to ‘accepted’.

   (d) Fatal error

   • When this type of error occurs, the error message appears in a popup window only with one 'Modify' button. The application stops and the statistician must correct the error immediately. He can not navigate to the next item without modifying the data. If the statistician can not correct the error the application exits, the data and the error row is not saved. (Code validations are included in this error category.)

F. Product and process quality standards are ensured
47. The system enforces users to process the compulsory steps and the dataset to be compliant with all validation criteria. For example, the database can not be closed until there is a questionnaire where the batch validation group level is not the highest and where the errors are neither corrected nor accepted. If the user missed the highest batch validation session or did not close the previous period database, he gets an error message and can not finish the data correction phase. If every criteria are met, that is, all questionnaires are validated on the highest batch validation group level and all errors are corrected or accepted and the previous period database is closed then the survey’s database will be closed and became protected against modifying.

48. The system logs and provides information about who and when started a function, shows the responsible department and users. Users can retrieve information about when a questionnaire was entered or loaded, when was modified, controlled or deleted. The system provides information how many errors are in a questionnaire and how many errors are accepted. The system collects information about corrections, how many times a questionnaire were opened for correction and how much time was spent for correcting a questionnaire.

G. Standard user interface

49. Standard function and user interface layer is provided for users. The standard but expandable menu structure, the general applications and reports, the standard error handling processes, applications and the compulsory developing rules provide that users meet year and survey-independent methods and user interface. The font size and colours, the navigation rules (navigation between tables, rows, items), the placing of input items and display items, the appearing of error messages are standardized.

H. Compulsory developing rules

50. The system enforces to keep the developing rules in the database and in the file system as well. IT developers must follow the naming conventions and there are compulsory database objects (tables, columns) and compulsory folder structure for surveys. The IT developers and the statisticians must describe the necessary metainformation. If the IT developer does not follow the developing rules (uses a non-standard naming convention) or the metainformation is insufficient then the framework application and the standard procedures can not find the vital information and the user gets an error message.

I. Supporting the development

51. Application and report templates and program code-generating application support the IT developers. When the IT developer designs the database objects and data structure of the survey there is a template repository plan in the Designer and he has to complete it only with the special items, columns. An application helps the IT developer to implement validation program codes (in Pl-sql programming language). With this application the IT developer can choose the items to be checked, the operation to be performed and in the last step the application generates the validation program code in Pl-sql language. If a new application or report is needed, the IT developer can use templates and usually can build the application using standard building blocks.

J. Provides online documentation from the framework

52. Users can open the application’s and the surveys’ documentation from the framework. The following documentation is available: Application User’s Guide, Testing Reports, Survey-specific User’s Guide, Descriptions about Changes (if a validation rule or the program code has changed) Description about check rules and their metainformation (retrieving from the database).

K. Technology
53. All information and data are stored in a central (Oracle) database. The retrieving and correcting data is regulated and controlled by the ADEL system. The users are assigned to different roles and they get access rights on objects through their roles. The retrieving (select) and administrating (admin) rights are direct. It means that users can query data not only with the ADEL framework application but with Sqlplus, SqlDeveloper and Hyperion tools as well. The data-modifying role (update) is an indirect role, it means that the ADEL framework application activates the access right when the user has logged in and deactivates it when the user exits. So users can modify data only with the ADEL framework application.

54. Oracle Designer (10.1.2) is used for designing the database objects, the applications are developed with Oracle Developer Form (11.1.2). The reports are made with Sqlplus (10.1.0). The client is Win7, the application server is Red Hat Enterprise Linux Server (5.8), the database server is HP-UX 11.31 (Itanium), the database is RDBMS Oracle Database 11g (11.2.0.4).

VI. Summary

55. The ADEL system is one key element of the Statistical Data Processing systems of the HCSO. Its primary function is to make the questionnaire data available – provided in different ways like paper questionnaires, electronic data sets, etc. – in a standard format for further processing after meta-driven validations. The ADEL supports multi-layered validations and generates control lists if needed – beside the basic functionality – and enables questionnaire data corrections not violating the integrity of data.

56. The strong standardisation of the ADEL system is important not only in use of the final applications but in the development phase of them also for the sake of effectivity and the minimisation of the human (users, developers) errors. The ADEL system – being closely integrated with complex statistical data processing environment – has the key role in assurance of the reliability of the raw data.

57. The ADEL system has been in use for data entry and validation for about 15 years in a constantly changing environment. The requirements of the new subsystems – often with the need of technology change – are fulfilled with the continuous development not violating the original design.