Integrated Data Processing System (EAR)

Hajnalka Debreceni

Hungarian Central Statistical Office (HCSO), Budapest, Hungary; Hajnalka.Debreceni@ksh.hu

The main purpose of the EAR system is to prepare final production outputs based on datasets created in the course of data preparation and data entry processes and to support the data processing activities in general using standard protocols for any statistical domain.
The development of the EAR system started in 2008 with an extensive on-demand survey within the Hungarian Central Statistical Office and is now used for several data processing activities in the HCSO. The result of this survey was a collection of methods used for data processing, which was the basis for the development of the EAR system. The main motivation was to increase the efficiency of data processing using standard methods and solutions and to provide common documentation for the data processing phase of the statistical business processes.
Based on our experiences with the EAR system, one of the main outcomes of the EAR is the increase in quality of data processing tasks due to standardization, full integration with the HCSO metainformation system and the automatically generated documentation of data processing activities. The users of the EAR system are statisticians; they design, redesign and manage their data processing activities. There are more than 200 users in the production environment from all subject-matter domains.
The HCSO considers the EAR system as a key importance system in the business processes with a lot of functionality to support statisticians in the design and management of the data processing tasks.
The functionality of the EAR is twofold. It is

- a framework - it supports the design and the management of data processing activities, using standard methods and data process definition structure
- a set of standard elements – the EAR procedures/functions are standalone standards, defined in a common way to support the definition of data processing activities

Keywords: data processing, integrated systems, standardization

1. Introduction

1.1. Methodological aspects

The Hungarian Generic Statistical Business Process Model (ESTFM), which is the adaptation of GSBPM, consists of several data-production processes (*Annex 1.*) and the Processing (VI) which is one of these process phases.

The main purpose of the EAR system is to support the data processing activities in general using standard protocols for any statistical domain.

Back to the ESTFM, there are other related sub-processes, besides the processing phase:

- II.8. Design data processing
- V.3. Review and validate
- VII.2. Seasonal adjustment
- VII.3. Validate outputs
- VII.4. Apply disclosure control
- IX.3. Archiving

Since the EAR is an IT tool, the Build (III.) process phase has been related, not to the functionality but the deployment.

There are overarching processes characterizing the whole of statistical data production processes (Outlier treatment, Product quality measurement).

The ESTFM is a model having the structure of a matrix, i.e. the sub-processes of the process phases not necessarily follow each other in a straight order while the implementation of the processor they may occur repeatedly during data-production process and there may be loops. Not necessarily all sub-processes take place during a statistical data production, so the various data-production processes may vary significantly. The model strives to describe all statistical processes of the HCSO using combinations of its existing components.

This sub-process means the cleaning of the data collected in the previous phase and their preparation for analyses. In the case of regularly produced statistics this process phase is a

permanent part. Its sub-processes are the integration, control, correction and transformation of microdata, be them data from statistical data collections or from secondary data sources. These sub-processes can be repeated several times, even concurrently with process phase Data analyses (VII). The underlying reason for this is that the need may arise for data to be reinterpreted, which implies the repeated application of the phases of processing. The processing phase covers 5 sub-processes.

1.2. Technical aspects

The HCSO intends to support all of the process phases by integrated systems, like EAR system. Before EAR there were several unique solutions for data processing, the workflow wasn't properly transparent and documented.
The development of the EAR system started in 2008 with an extensive on-demand survey within the HCSO. The result of this survey was a collection of methods used for data processing, which was the basis for the development of the EAR system. This is an external development, a 3-tier (not running in a browser) application, Java based, underlying Oracle database.

The milestones of EAR:

- 2008 – Specification of requirements
- 2009 – Requirements analysis and system engineering
- 2010 – Development and testing
- End of 2010 – Start of implementation
- …
- Today – The system is now used for several data processing activities in the HCSO and is currently being extended to all production processes at all subject-matter domains.

The EAR has been developed and supported by a contracted partner. The maintenance of this system is a common task of the external developers and the internal system managers.

The EAR administrators' tasks, besides keeping touch with the developers are supporting users, educating, making users' guide and tutorial, taking part in deployment, testing and allocation of privileges inside EAR.

The data security is very important, so the security management is a key part of the system. It is twofold, authentication and authorization. The authorization defines who can see what. The system checks the privileges according to the database access rules. The EAR user has to be an active employee. The access level defines the functions, what a user group can execute in an EAR folder. System logs the allocation and revocation of privileges.

We distinguish two environments in the office. First is the production environment with processing of checked and committed data and the other is the test environment. Its purpose is testing and education.

2. Structure of the EAR system

2.1. About the EAR in general

The goal of EAR is to prepare final production outputs based on datasets created in the course of data preparation and data entry processes. The main motivation of this system's development was to increase the efficiency of data processing using standard methods and solutions and to provide common documentation for the data processing phase of the statistical business processes.

The advantages of EAR are an IT system controlled by statisticians which support the production process through more data production periods; for a number of data collections; in a repeatable, modifiable, and documented way.

2.2. Connection to other IT systems

The EAR cooperates with other IT systems; there are input and output interfaces of EAR.

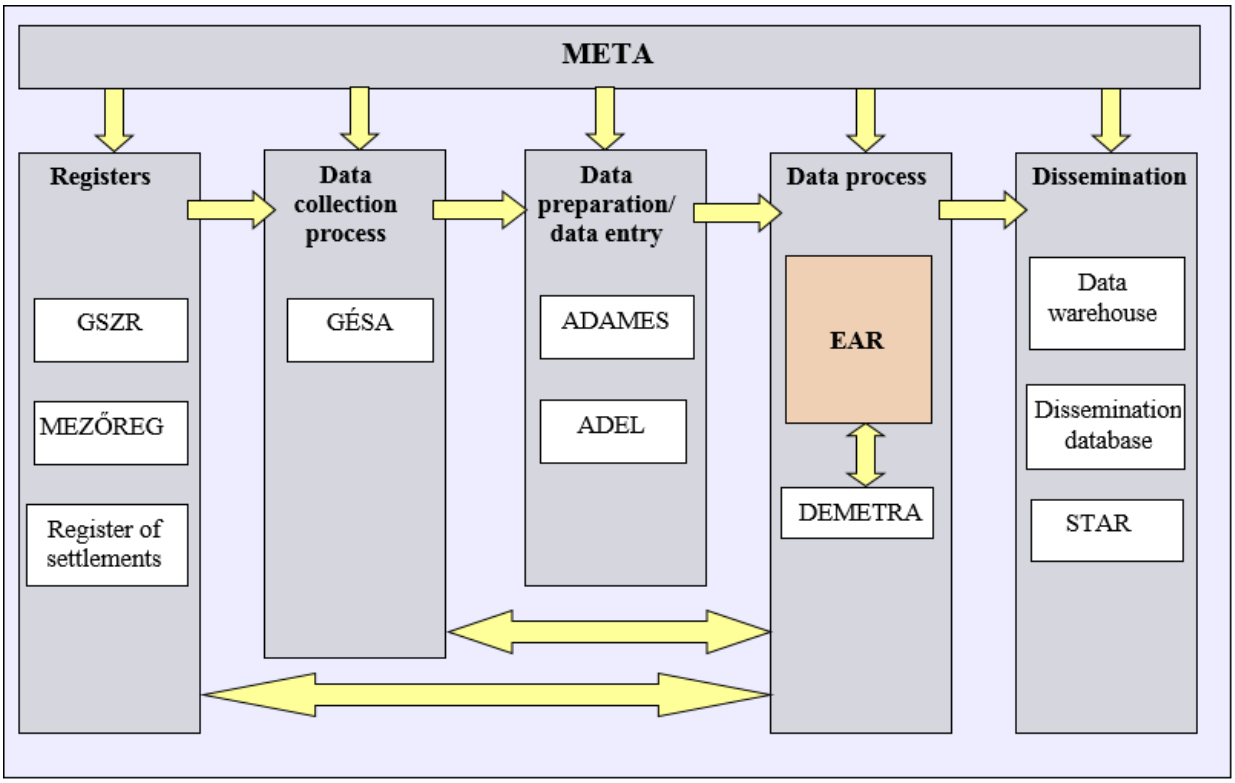

Figure 1 - *Place of the EAR in data-production process*

Abbreviation	Explanation
ADAMES	Integrated Survey Control System for Secondary Data Sources
ADÉL	Integrated Data Entry and Validation System
GÉSA	Integrated Survey Control System for Business and Social Surveys
GSZR	Business Register
META	Metainformation System
MEZŐREG	Farm Register
STAR	Regional Statistical System

Figure 2 – *Abbreviation connection of IT systems*

In addition, there is another joint system, the mail system. The users get an e-mail from the EAR about every own activity, any change in metadata or in cited process plan.

EAR is a metadata-driven application. It means, on the one hand that the EAR has its own metadatabase, which EAR can write and it can store data about processing. This information can be a basis of documentation. On the other hand, EAR reads the information about statistical data from the Metainformation system.

3. Functionality of EAR system

3.1. Framework

It supports the design and the management of data processing activities, using standard methods and data process definition structure.

The display objects - like menu, tags with different functionality and colours - help the users in planning of a processing cycle.

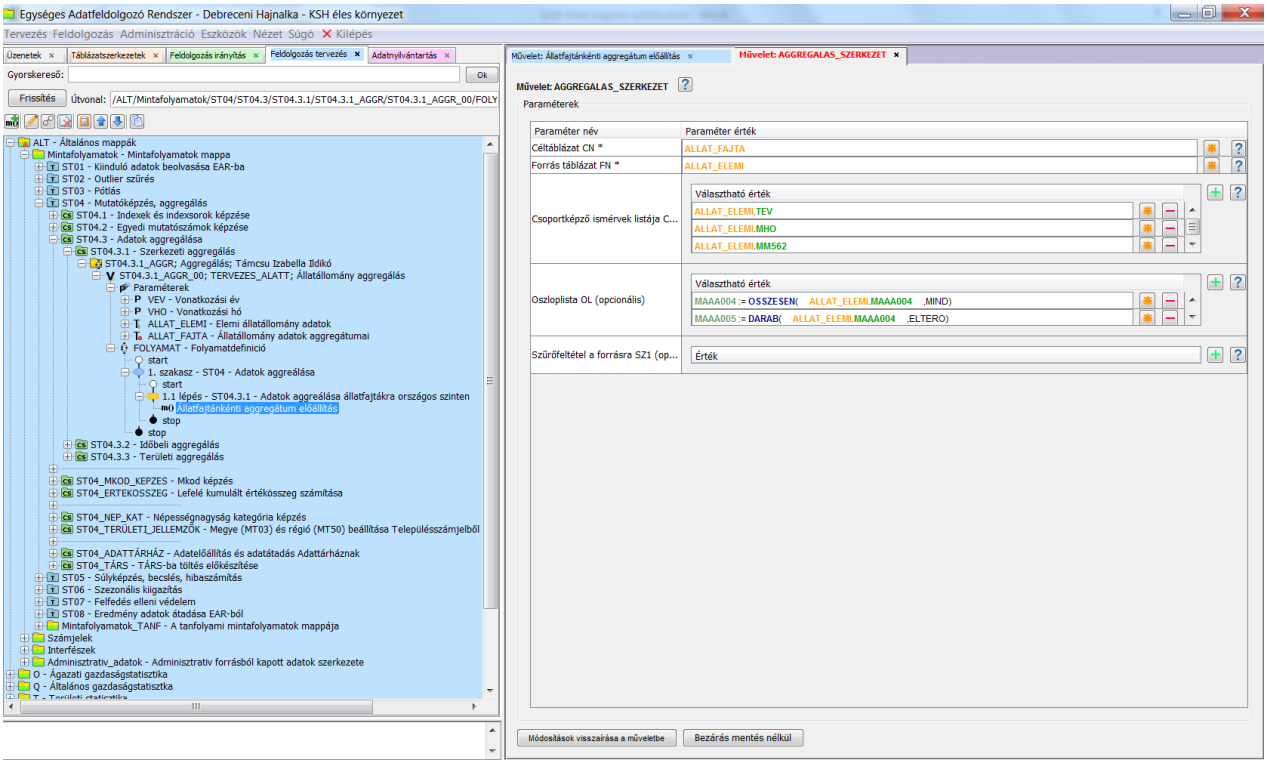

Figure 3 – *Snapshot of the EAR user interface*

Elements of the EAR workflow:

Planning

Table structure - The users prepare several table schemes for inputs and for outputs.

Planning of data processing - There are separate plans for loading data into production environment, making aggregates and finalizing output.

Processing

Processing control - Processing plans can be run, independently from frequency.

Data recorder - The result tables can be queried.

3.2. Set of elements

The EAR procedures/functions are standalone standards, defined in a common way to support the definition of data processing activities

The hierarchical structure of these elements:

Process stages

Process methods

Procedures and functions which execute the above methods

List of standard process stages:

ST01 – Loading data into processing environment

ST02 – Outlier treatment

ST03 – Imputation

ST04 – Deriving new variables, making aggregates

ST05 – Calculating weights, estimation of population parameters, standard errors and variances

ST06 – Seasonal adjustment

ST07 – Statistical disclosure control

ST08 – Finalizing output

The definitions of process stages harmonise with the ESTFM. The workflow, mentioned in chapter 1.1, appears here, the operational model is the same.

Every level (Process stage / methods / procedures) has two kinds of use cases, basic or statistical. Basic means, it can be applied at any process stage, it can be used anywhere in the process plan (e.g. making aggregates). Statistical means, it can be applied only in a specific process stage. Like imputation with environmental value, it belongs to imputation process stage. The lowest level of elements are procedures and functions. These elements like Lego cubes underlie the whole processing.

4. Conclusion and future plans

The HCSO considers the EAR system as a system of key importance in the business processes with a lot of functionality to support statisticians in the design and management of the data processing tasks.

Based on our experiences with the EAR system, one of the main outcomes of the EAR is the increase in quality of data processing tasks due to standardization, transparency, full integration with the HCSO metainformation system and the automatically generated documentation of data processing activities. The users of the EAR system are statisticians; they design, redesign and manage their data processing activities. There are more than 200 users in the production environment from all subject-matter domains. The number of EAR implemented processing is different at each statistical domain. The entire agricultural data processing and the annual calculation of GDP have been migrated to the EAR system. The EAR is in use at several topics - like environment, tourism, transport, social, health care, culture and internal trade – for the present. Our aim is to introduce the EAR system for the rest of data processing solutions within 2 years.

The system is ready but it can be fine-tuned according to user needs following methodological and technical changes, e.g. the short term aim is to apply product quality indicators in the processing.

5. References

Hungarian quality guidelines (2015), Available: http://www.ksh.hu/docs/bemutatkozas/eng/minosegi_iranyelvek_eng.pdf

Szekeres, B.; Mag, K.; Ábry, Cs.; Nagy, A. and Vereczkei, Z. (2014) Improving the Quality Management System: Projects at Hungarian Central Statistical Office, Paper for the European Conference on Quality in Official Statistics (Q2014), Available: http://www.q2014.at/fileadmin/user_upload/Paper_Improving_the_Quality_Management_System_Projects_at_Hungarian_Central_Statistical_Office_Szekeres_Mag_Abry_Nagy_Vereczkei_final.pdf

Kiss, Z. and Debreceni, H. (2015) System documentation, version 0.3, Available: only in Hungarian in the HCSO

Támcsu, I.; Buzder Lantos, F. and Debreceni, H. (2016) EAR User's manual, version 2.29, Available: only in Hungarian in the HCSO

6. Annexes

Annex 1. Hungarian Generic Statistical Business Process Model (ESTFM).