

# Challenges and Responses ~ Innovation of the Hungarian Statistical System

Laczka Éva

*Hungarian Central Statistical Office, Hungary*

## 1. Background

The first Hungarian independent statistical institute, the National Statistical Office was founded in 1848, but it functioned only some months with the failure of the Hungarian Revolution and War of Independence of 1848. The direct predecessor of the Hungarian Central Statistical Office (HCSO) was established in 1861, at the beginning it performed its tasks as the statistical department of the Ministry of Agriculture, Industry and Trade and became an independent statistical office in 1871. According to the deed of foundation the tasks of the Office were the following:

*„1.§ The objective and tasks of the statistical office are to collect comprehensive and reliable data on the state of the areas under the scope of statistics, analyse, order and publish data so that they could be used for legislative, administrative, scientific and practical purposes.*

*2.§. This entails that the statistical office aims not only at providing faithful and objective picture of the facts and results of state organisation, state theory and social activity of the nation promoting by this the knowledge and information of the nation, but at contributing with the help of the information collected and the regular statistical registering of the country's state to the satisfaction of the needs both of practical life and science.”*

The above mentioned objectives continue to apply even nowadays; the HCSO produces and publishes credible, good quality data on the society, economy of the state, the environment and its changes for a broad scope of users. Besides preserving the values accumulated by its predecessors, the HCSO

strives to comply with the changing requirements of the epoch. The main challenges are the continuous improvement of data quality, the decrease of respondent burden, the modernisation of the forms, methodology of statistics and tools of dissemination. The newly emerging tasks, the continuously developing environment require from the HCSO a flexible structure proper to adaptation.

Hungary joined the European Union in 2004 and became at the same time part of the European Statistical System. This means that the Hungarian Statistical System must comply with EU regulations, statistical standards, the basic principles of statistical data collections and publication, and with codes of conducts and operation which in our case are embodied in the European Statistics Code of Practice as well as Regulation (EC) 223/2009 on European Statistics.

## **2. Objectives**

The purpose of this study is to give a picture of the structure, functioning, latest developments of the Hungarian Statistical System, focusing on the improvement of data quality, the use of electronic solutions in statistical data collection, data processing, better documentation of the processes, including their impact on the decrease of the burden of statisticians and respondents.

The developments presented have brought about significant changes that will influence in the long term as well the direction of further improvements in the HCSO.

Figure 1 presents the process of statistical data collections, processing and production of datasets in the HCSO and it helps the interpretation of the place and role of the developments carried out.

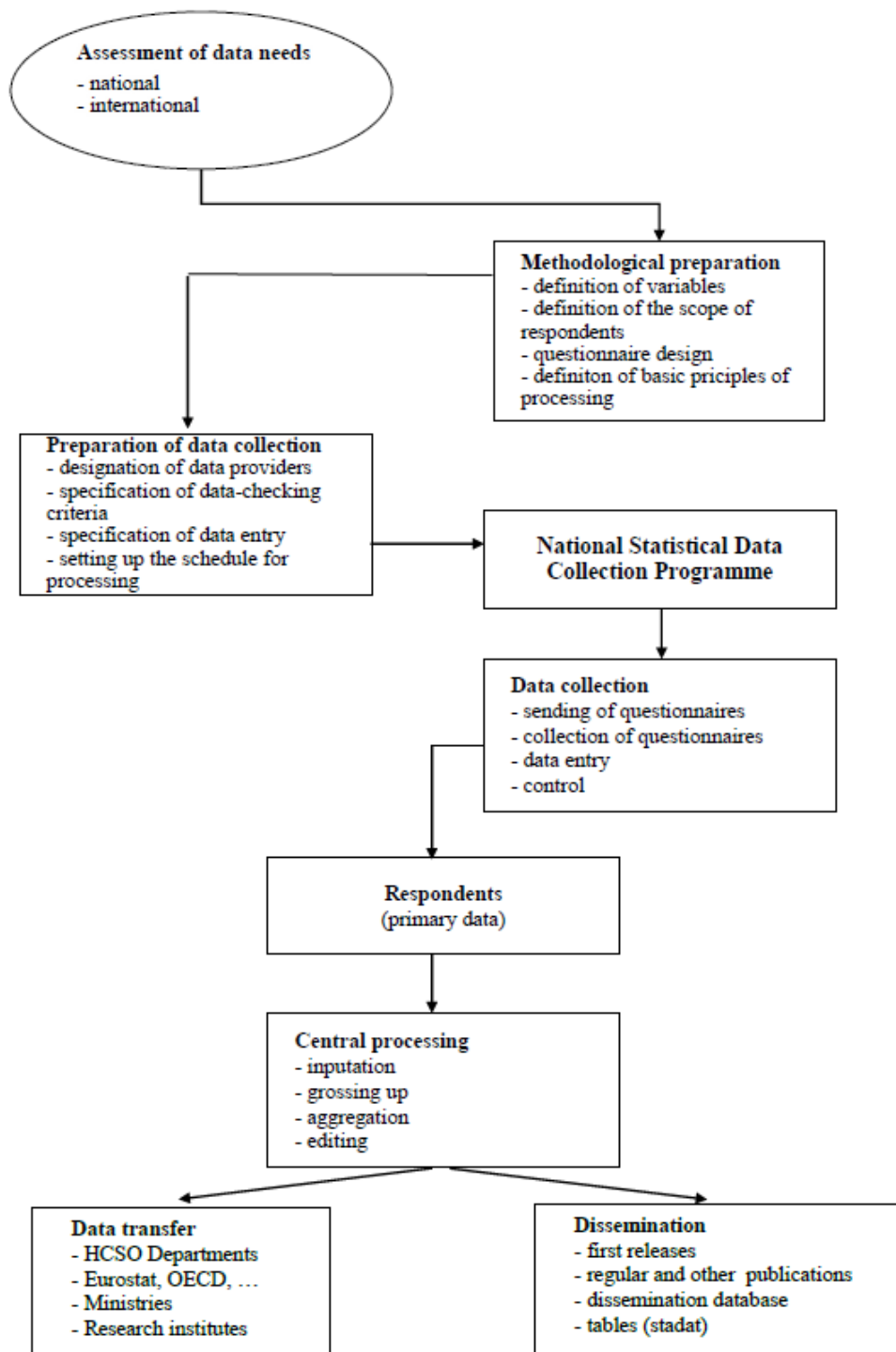


Figure 1

### 3. Data Quality, Levels of Validation

Data validation plays an important role in raising the reliability of statistical data. In the Hungarian practice – as a result of methodological development – we distinguish four successive levels of validation mainly in economic statistics.

- a. **First level:** *validation of primary (micro) data within a given survey/questionnaire*. In the HCSO this level can be considered as a traditional level of validation that can be found in the case of all statistics.
- b. **Second level:** *validation of micro data* within a given specific field of statistics *using several questionnaires and available administrative data sources* (including spatial and time consistency checking as well). As this validation is situated between micro and macro-economic level controls, it is called *“meso-validation”*.
- c. **Third level:** during the *statistical validation process at branch level*, the consistency of statistical data and the available administrative data is examined for the main aggregates of several statistics at branch level.
- d. **Fourth level:** *validation at macro-economic level* using supply and use tables (SUT) data, this makes the system of validation of economic statistics data complete. In practice; it is validation of national accounts data.

While in the Hungarian Statistical System the first and third level data validations have been used for years (certain elements had been functioning already earlier), the second level appeared as a new element a few years ago. The concept, methodology, literature and IT background of meso-validation were elaborated in 2011. A detailed practical methodological guide was prepared for meso-validation, containing among others a “data map” with the list of the data collections, administrative data sources and variables to be covered by the validation process. In the first phase of the development work we performed the meso-validation of thousand enterprises producing the highest gross value added contribution, over 50 % of GDP. In the following two years meso-validation was gradually extended to smaller enterprises and to

almost all field of statistics as well.

The fourth level (validation at macro-economic level) was elaborated in relation to national accounts, relying on international (mainly EU) practices as well. The four-level data validation allowed to put in place a complex quality checking of economic, business statistics data and made possible to avoid the difficulty of treating consistency problems in the phase of national accounts compilation.

The development of the complex four-level consistency checking system was later a great input to the introduction of electronic data collection system (integration of validation in electronic questionnaires), and electronic data processing (the unified data processing system).

#### **4. Electronic Data Collection**

Electronic data collection is one of the most important tools to decrease respondents' burden, which also improves the efficiency of statistical work. The HCSO – even after the turn of the Millennium – sent the majority of the questionnaires to the respondents in a printed form, after filling the forms in the respondents sent them back to the office by post or by e-mail. Electronic data collection in the classical sense of the word was used – mainly on a voluntary basis, as a pilot exercise – only in the case of some data collections in business, economic statistics (this was the KSHXML data collection system), despite the fact that the use of KSHXML was available for 60 per cent of the questionnaires in business, economic statistics performed by the HCSO.

The new electronic data collection system replacing KSHXML, called “KSH-ELEKTRA” had been developed to put in place a more modern, user-friendly, quick, more sophisticated system. The introduction of the KSH-ELEKTRA started at the beginning of 2011, and the possibility to use the KSH-ELEKTRA system was extended to all data collections of businesses from the 1<sup>st</sup> of January 2013. Since the end of 2015, 95 per cent of the compulsory data collections of the HCSO are carried out electronically, through the KSH-ELEKTRA system which also integrates validation (consistency checking) in the most important questionnaires.

#### **5. From HOMBÁR to the Unified Data Processing System (EAR)**

After the turn of the Millennium the HCSO used its resources – including resources aimed at preparing Hungary to EU-accession – to the development of IT tools simplifying data processing and increasing efficiency.

The new data processing systems developed provide a comprehensive documentation of all the

phases of data processing and make compulsory the use of standard data processing processes. The different data processing steps *can be performed by statisticians without special IT knowledge, and data processing time becomes shorter*. The systems are adapted to the standard systems of data preparation, data collection organisation and dissemination of the HCSO. The traditional data preparation and dissemination systems were already supported by metadata-driven IT systems based on unified principles, but in most cases data processing was performed with individually developed programmes. The new data processing systems support the statistical data production process as a unified system generally used. The unified framework data processing system supports and documents the different phases of data processing and administrative tasks.

### 5.1. HOMBÁR

After the turn of the Millennium Hungarian statisticians dealing with agricultural statistics applied successfully for an EU-tender aimed at developing agricultural statistics. The financial resources of the project made possible to review the data processing tasks of agricultural statistics as a whole. The renewal of data processing was justified by the fact that the traditional data processing system was relatively slow, it was not adequate for the efficient implementation of validations, it was not sufficiently well-documented, and last but not least it required important human resources. This was the reason why agricultural statisticians and IT experts decided to develop a special, new data processing tool, system that was called HOMBÁR (it means: Granary).

The concept of HOMBÁR is quite similar to the LEGO game; agricultural statisticians and IT experts created LEGO cubes (statistical operations) of different forms that were specified by agricultural statisticians and programmed by the IT experts. Using a comparison, the system functions in such a way that statisticians chose or assemble LEGO cubes according to their needs whether they compile “a horse” or a “tractor”. This means that if the statistical process changes, statisticians reorder the LEGO cubes (the system can react to the changes in a flexible manner). In this way data processing is led and managed directly by statisticians, the task of the IT team is “just” to secure the IT operation of the system and produce the new LEGO cubes (statistical operations). The development lasted 3 years and further 2 years were needed before the processing of agricultural (survey and admin) data with the new system became a routine for statisticians. The use of the HOMBÁR system made possible to reduce by half the time of data processing and by 30-40 per cent the human resource needs (in the case of IT people) as well as the cost of data processing. The HOMBÁR provided a proper documentation, relation with the databases, and integrated more efficient consistency controls.



## 5.2. Unified Data Processing System (EAR)

On the basis of the experiences of HOMBÁR, in 2008 the HCSO decided to extend the new data processing system to the whole statistical production of the office. As the whole statistical system is much broader than agricultural statistics and is in a certain sense more complex, further development was needed. The new, extended system was called the Unified Data Processing System (in Hungarian EAR). The EAR compared to the HOMBÁR is a more general data processing system, which does not process the data of only one field of statistics, or few statistical domains, but can be used for all micro data collections and data sets of the HCSO. Nowadays in the EAR system more than 100 basic statistical operations and more than 20 integrated statistical operations are available for statisticians.

### 5.2.1. META Database governed system

One of the major characteristics of the data processing systems is that they change frequently; including the scope of the data collected, the concepts linked to the processing, the methodology, needs of dissemination, etc. that can also change through the years.

Consequently, the EAR concept was to be elaborated in a way to follow the changes without

generating each time new IT programming work. The EAR system created had to allow statisticians to describe changes concerning all the fields of our statistics “selecting and ordering the LEGO cubes”, without the need of new traditional IT programming.

On the other hand; one of the most widespread solutions to follow changes in a dynamic way is the use of meta-database which led us to create a META database governed system. The basic principle of the META database governed operation is that the information needed for the functioning of the EAR system (the components necessary for statistical data processing, the different steps of data processing, the tasks in each step) are stored in the meta-database. The running of the EAR system relies on the use of the meta-database that contains the tasks to be implemented and the way of implementation (Annex).

### ***5.2.2. An integrated system***

The EAR system is linked at data level to the meta-database system of the HCSO, the data entry system and the Data Warehouse. It is linked through interfaces to tools for data analysis. The purpose is to fulfil the greatest amount of statistical processing functions, but in those cases when processing can be made more efficiently with existing software, data access or interface must be guaranteed.

### ***5.2.3. A system governed by statisticians***

Statisticians define and execute data processing using the meta-database and the integrated primary data processing steps. In the case of a change in the dataset and/or of the processing process, the modifications are performed directly by the statistician. The contribution of IT team is “only” needed for the creation of the interfaces and primary operations.

### ***5.2.4. A transparent, well-documented and standardized system***

The EAR system stores the processing phases defined by the statistician in the meta-database system from where the documented information can be extracted. It is a unified, standardized processing system in the sense that different data collections, datasets and data from different periods can be processed together.

### ***5.2.5. The tasks of the statisticians***

In the development phase of the EAR the task of the statisticians is to specify the requirements towards the system. It was necessary to specify for each statistical domain the basic operations that were needed during data processing. Already at the beginning the aim was to create a comprehensive set of



basic operations. An important task was to adjust the functioning META-database system of the HCSO to the needs of the EAR system which required the development of the META-database.

In the preparatory phase agricultural statisticians considered that the HOMBÁR system was not a suitable tool to process account type of data, like the Economic Accounts for Agriculture. Thanks to the experiences gathered over the almost 5 years EAR practice, the Hungarian National Accounts team developed a new way (additional governance system) to make the EAR system suitable to process National Accounts data as well connected to an EU-tender aimed at developing national accounts' data. The main objective of the project was to integrate Process Table and Tabular Approach into the National accounts, which resulted in the change of the whole computation mechanism and the move to a new system.

#### **5.2.6. IT tasks**

During the development phase of the HOMBAR/EAR systems, the task of the IT team was to secure the connection with the systems functioning in the office, create the operating IT environment, but the development of the META-database required also the contribution of IT experts.

## **6. Conclusion**

After the turn of the Millennium Hungarian agricultural statisticians decided to modernize the Hungarian Agricultural Statistical System in particular in the case of data processing system. The renewal of data processing was justified by the fact that the traditional data processing system was relatively slow, it was not adequate for the efficient implementation of data validation, it was not sufficiently well-documented, and last but not least it required significant cost and human resources.

The HOMBAR (the first new data processing system) was developed by agricultural statisticians, it was a data processing system governed directly by statisticians (the task of the IT team became "just" to secure the technical IT operation).

On the basis of the good experiences of agricultural statisticians, in 2008 the HCSO decided to extend the new data processing system to the whole statistical production of the office. The use of new systems (HOMBAR, EAR) makes possible to reduce the time of data processing and decrease the human resource needs significantly (especially in the case of IT teams) as well as the cost of data processing. It provides a proper documentation, relation with the databases, and integrated more efficient data validation, consistency controls.

The new system is an integrated system, linked at data level to the meta-databases of the HCSO, from

the data entry systems to the Data Warehouse, dissemination systems of the HCSO.

As the result of the development work nowadays the new data processing system is suitable not only for primary (survey) data processing but also for data processing of macro-data, such as integrated datasets like national accounts.

## References

New Techniques and Technologies for Statistics (NTTS) Conference 2015, Brussels.

Beszámoló a KSH 2010-2011. évi tevékenységéről, (Központi Statisztikai Hivatal, 2012, p, 61) (Report on the activities of the HCSO in 2010-2011 (Hungarian Central Statistical Office, 2012, p.61).

Regulation (EC) No 223/2009 of the European Parliament and of the Council of 11 March 2009 on European statistics (OJ L 87, 31.3.2009, p. 164).

European Statistics Code of Practice, Adopted by the European Statistical System Committee, 28th September 2011.

Homepage of the Hungarian Statistical Association: [www.mstnet.hu](http://www.mstnet.hu).

**Published: Volume 2016, Issue 9 / September 25, 2016**

## EAR in the statistical data production process

