

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**Work Session on Statistical Data Editing**  
(Budapest, Hungary, 14-16 September 2015)

Topic (ii): Managing and supporting changes related to editing and imputation

**Improvement of the quality of statistics by Mezo-validation**

Prepared by Miklósné Paczári and Katalin Kovács, HCSO, Hungary

**I. Introduction**

1. One of the strategic objectives of the Hungarian Central Statistical Office (HCSO) is the continuous improvement of the quality of our products and services. To achieve this objective, beyond the currently available data processing and data checking procedures, mezo-validation is one of the main methods, which is also carried out at the level of microdata.

2. The basic goals of the mezo-validation are the following: improvement of the quality of data collected and published by HCSO; insuring the consistency of data coming from different sources; reconciliation of micro-statistics and statistical domains; making available the collected information in a standardized system.

**II. Objectives of mezo-validation**

**A. Definition of mezo-validation**

3. Micro-validation means the comparison, checking and correction of data at micro level (unit/record level) using individual data or, in particular cases, aggregate data coming from the same or other sources (statistical, administrative), registers.

4. Mezo-validation is carried out at individual level by comparing various administrative, statistical and other data sources concerning the unit of observation, analyzing the consistency and accuracy. Actually, it is a tool that enables us to check the individual data in a wider extent. For this reason, mezo-validation can also be considered as complex micro-validation. The term 'mezo' in mezo-validation does refer neither to a middle step in the checking phases/order nor to the aggregation of the data under checking. The term simply came to common knowledge in our office in this way.

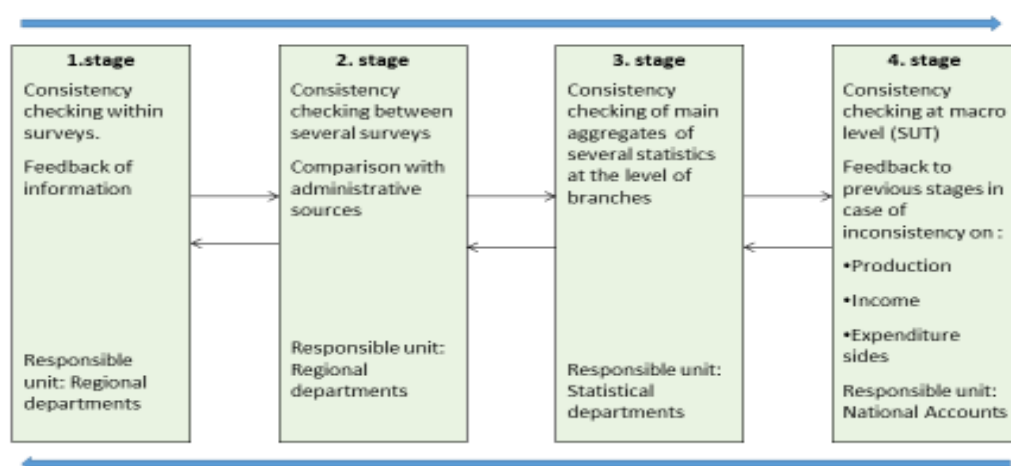
5. Mezo-validation differs from (surpasses) the checking levels/phases applied currently in the office in the sense that it analyses the widest possible range of data and information in a standardized framework ensuring greater consistency between the related statistical domains and the general economic statistics. Mezo-validation differs from the existing checking processes also in timing because it is launched after the primary data processing (either micro- or macro-validation), or even after the first publication. This is partly because mezo-validation can only be carried out once the related data from different sources are available. On the other hand, it is not possible to put it in an additional checking phase due to the shortness of the period between data collection and the first publication. Mezo-validation can be relevant even several times as regards the reference period (e.g. flash and second estimate of GDP).

## B. Life cycle of mezo-validation

6. The elaboration of the concept of mezo-validation and its fitting into the system of quality analysis started in the last quarter of 2010. Then the development and implementation of quality criteria was prepared first for National Accounts between the 1<sup>st</sup> and the 3<sup>rd</sup> quarter of next year. This was followed by the development and implementation of quality criteria for External trade between the 2<sup>nd</sup> half of 2011 and the 1<sup>st</sup> half of 2012 and for performance statistics in the 2<sup>nd</sup> half of 2012. For statistics of services and for investment statistics development and implementation of quality criteria was made in the 1<sup>st</sup> quarter of 2013 and 2014, respectively. The development of quality criteria for statistics of agriculture was the task of the 4<sup>th</sup> quarter of 2014 while the implementation will be launched in the second half of 2015. The development of the quality criteria for institution-based labour statistics is planned for the last quarter of 2015.

## C. The levels of consistency analysis

7. We distinguish 4 levels of the consistency analyses from the micro level to the macro level checks. Feedbacks on inconsistencies are provided to the previous levels.



*Figure 1. The levels of consistency analysis*

## D. Objective of mezo-validation

8. The basic objective of the mezo-validation is the improvement of the quality of data collected and published by the HCSO, making available the collected information in a standardized system. Our medium-term goal is to provide information that can be easily used by the different statistical departments during data revision.

9. The objectives in detail are:

- (a) Improvement of the completeness, accuracy, timeliness and consistency of individual-level data (general economic statistics and other statistical domains)
- (b) Clarification and standardization of information that is available in the Business Register (GSZR)
- (c) Improvement of contact with data providers, increasing the efficiency of contacts
- (d) More effective coordination of data supplies
- (e) Creation of a common information storage surface
- (f) Providing possibilities for analysis

## E. Units involved in mezo-validation

10. The observational unit of mezo-validation is economic organization. The data coming from the activities and the transactions of economic organizations are observed in a complex way.

11. Units involved in mezo-validation have been defined on the basis of the organizational units and indicators in accordance with national accounts, taking into account the sectors' specialities. Considering that about two-thirds of the GDP is produced by the corporate sector, in case of annual data collections enterprises with more than 20 employees, in case of sub-annual surveys enterprises with more than 50 employees have been involved in the validation till now.

#### **F. Scope of data involved in mezo-validation**

12. All statistical and other data have been used – either expressed in value or in volume – that refer to the main economic indicators (production, costs, capital, labour, etc.) or that are in logical connection with them. The scope of data that need to be validated is continuously broadening and can be extended.

13. The necessary data sets on economic organizations include the main business register data (GSZR), the main data for data collection management (Integrated Survey Management System for Business Surveys (GÉSA)), Structural Business Statistics (SBS), foreign trade statistics, institution-based labour statistics, corporate tax, VAT, and other data. The scope of data involved does not necessarily cover all the data spaces of the relating questionnaires. At the same time, in many cases, data and information coming from non-statistical sources are also taken into consideration.

### **III. Improvements**

#### **A. Initial support for mezo-validation tasks**

14. At the beginning of work the members of the working group took part in some basic education in data mapping, got acquainted with basic concepts and with the Hyperion database query software and other applications, and attended an intensive training in accountancy. There have been on-site consultations at the regional departments in the topics of performance and foreign trade statistics and they have studied the practice of the Irish Statistical Office as well.

#### **B. Our improvements and the introduction of the applications supporting consistency analysis**

15. Under a separate contract, from the beginning of 2012, we have managed to receive the electronic balance sheets, profit and loss accounts and the notes to the financial statements that had been collected by the Ministry of Justice but had not been sorted in database. The incomplete and informatically insufficient information have been arranged in database by the Statistical Office after a development work that had taken a quarter. The database can be sorted, makes it possible to process a query and has been made available on the intranet to the staff.

16. A structured web interface has been created which can be accessed from the intranet and facilitates the coordination of mezo-validation and the contact between the staff.

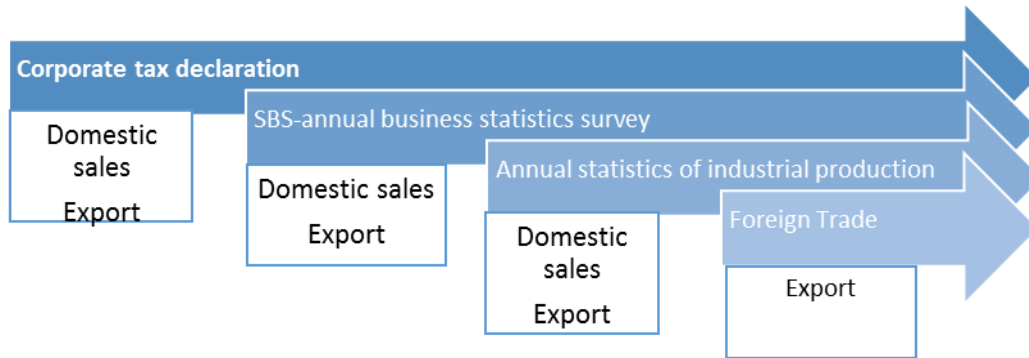
17. On the web interface one can find the minutes of the latest enterprise consultations and checking, description of IT skills necessary for mezo-validation, training materials, reports of the regional departments, professional documentation, and other assistance. There is a link to the electronic annual reports of the top 1000 enterprises (the balance sheets, the profit and loss accounts and the notes to the financial statements), hence, the reports become easily and quickly available instead of the former slow, cumbersome, moreover, during peak hours not operating search engine. Another link to the homepage of the electronic annual reports operated by the Ministry of Justice is on the webpage as well.

18. The IT expert of the Business Statistics Department has developed a form type query application that helps to carry out the mezo-validation task. Within this application, the linking of about 20 databases of different size and content became possible. In addition to this, the IT expert created a tutorial DVD to Hyperion dynamic query software.

19. In this application queries can be processed according to the frequency of the data collection: monthly data, quarterly data, annual data. Within the last one, annual data can be compared to sub-annual data as well as to base data.

20. Through one of the menu option of the application, one can see the annual turnover (split in domestic sales and export sales) coming from the corporate tax, the annual SBS-performance statistics, the annual statistics of industrial production and the foreign trade data collections on one screen.

21. On this screen, the differences are clearly visible, even the sales through foreign companies with domestic VAT registration.



*Figure 2. Comparison of export sales*

The screenshot shows a software interface for data entry and consistency analysis. The main window title is 'Mezovalidálás 2013 Utolsó javítás 2014.09.02'. The interface includes a menu bar with options like 'Tevékenység', 'Szerkesztés', 'Lekérdezés', 'Blok', 'Rekord', 'Mező', 'Súgó', and 'Ablak'. Below the menu is a toolbar with various icons. The main content area is divided into several sections:

- Regisztrar adatok:** Fields for company registration data, including 'Ágazat', 'Létszám', 'GFO', 'Ágak', 'Megye', and 'Kiemelt szervezet'. A 'Kilépés' button is visible.
- Adatgyűjtési feladatok a minőség jegyében, 2013-as adatok (Mezovalidálás):** A section for data collection tasks, with a sub-section for 'Adatok 1000 Ft-ban'.
- TAO (Társasági Adatok):** A table with columns for 'Belföld', 'Export', and 'Összes'.
- Éves Integrált:** A table with columns for 'Belföld', 'Export', and 'Összes'.
- Ipari termékjelentés:** A table with columns for 'Belföld', 'Export', and 'Összes'.
- Külker termék:** A table with columns for 'Belföld', 'Export', and 'Összes'.
- Szolg. külker:** A table with columns for 'Belföld', 'Export', and 'Összes'.
- Külker exp. összesen:** A summary section with fields for 'Külker exp. összesen', 'Éves integrált PHAA082', 'Éves integrált PHAA083', and 'Mindösszesen'.

*Figure 3. Consistency analysis of sales based on different data sources*

22. The code of the companies involved in mezo-validation has been built in the Business Register (GSZR) and in the Integrated Survey Management System for Business and Social Surveys (GÉSA) therefore the query from different databases is getting easier.

23. The following codes have been introduced to distinguish the TOP 1000 companies:

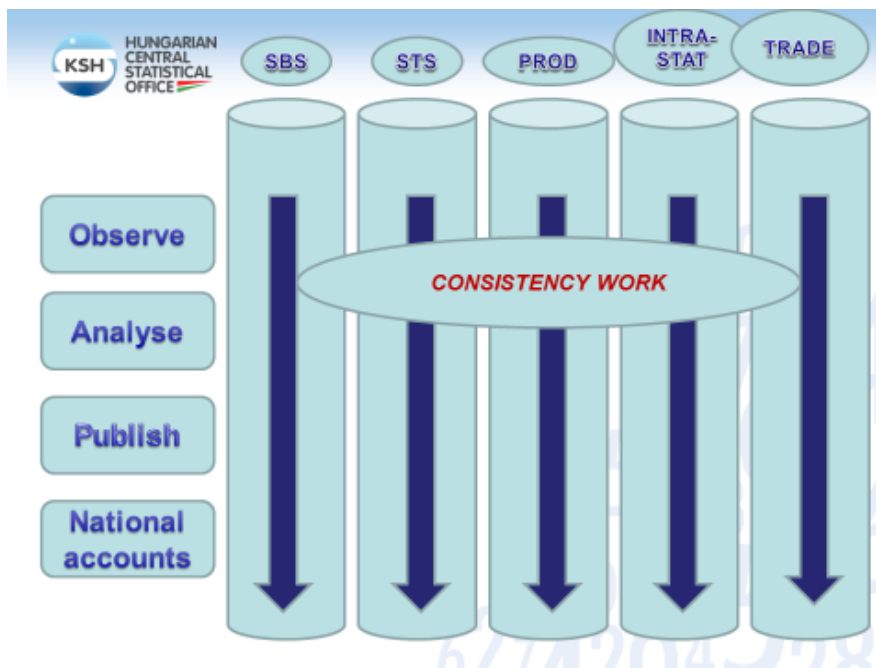
0	Not part of the TOP 1000
1	TOP 100
2	Other companies part of TOP 1000 and involved in mezo-validation
3	Other companies part of TOP 1000

24. Staff involved in mezo-validation has created customized queries in Hyperion.

#### IV. Special tasks within the framework of mezo-validation

##### A. Comparison of turnover from the corporate tax declaration, SBS, the annual statistics of industrial production, foreign trade

25. Task: Within the framework of the comparison of turnover from the corporate tax declaration, SBS, the annual statistics of industrial production, foreign trade, additional check of the annual data took place every year after all data sources are available.



*Figure 4. Consistency analysis of turnover*

26. The objective of the task is to achieve the consistency of data, to sort the acquired information in a structured format (at national level) and to make it available to the National Accounts Department.

27. This means, in details, filling in the Excel tables with the annual data based on the following criteria:

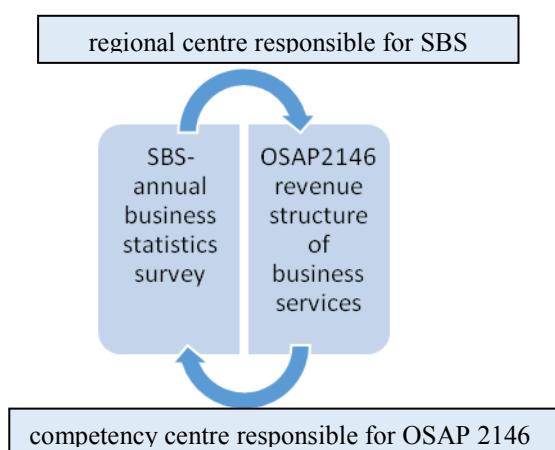
- domestic sales, export and total turnover according to the different data sources (corporate tax declaration, SBS, annual statistics of industrial production, foreign trade),
- reasons for the differences in sales (domestic sales, export),
- information about contract work,
- information about the sales through foreign companies with domestic VAT registration

## B Mezo-validation of statistical domains

### 1) Comparison to the structure of SBS turnover

28. The validation takes place at the data preparation phase of the Sales structure of business services (OSAP 2146) and of the Report about the activity of tour operators and travel agencies (OSAP 1035). Net sales of the reported activities and export sales in case of the Sales structure of business services are compared with the equivalent data of the annual performance statistics. In case of OSAP 2146, the net sales per activities and the export sales are also compared.

29. The comparison of OSAP 2146 and the annual performance statistics is a special two-way control because the data of the top enterprises are checked both by the competency centre and the regional departments responsible for annual performance statistics.



*Figure 5. Analysis of sales structure*

### 2) Investigation of SBS turnover and the data on shops and other establishments (OSAP 1852)

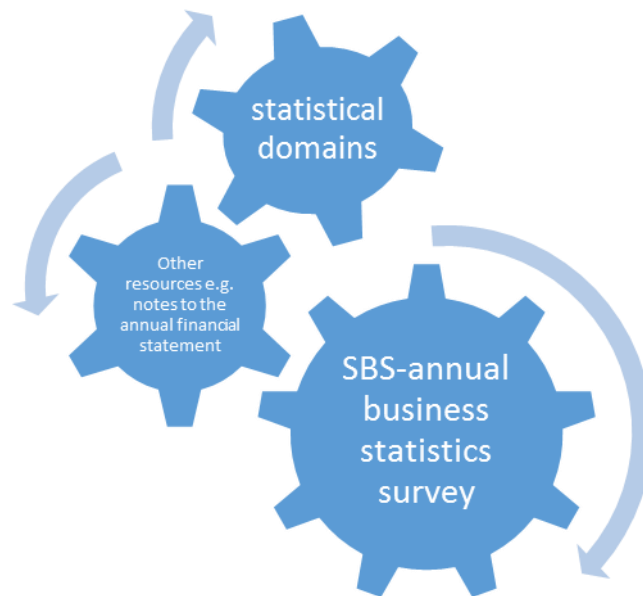
30. The turnover of shops is compared with the net sales.

### 3) Comparison of environment statistics and the corresponding business register data at settlement level

Environment statistics data collections	Object of validation
Annual report on municipal liquid waste (OSAP 1723)	Information about sewage treatment plant and sewerage rate
Water supply, waste water collection and waste water treatment services to settlements (OSAP 1062)	Information about sewerage rate
Report on the waste water collection investments in waste water agglomerations (K001)	Equivalence to investment on OSAP 1062
Basic information on economic units providing infrastructural services in the area of competence of local governments of settlements (OSAP 1659)	Comparison of service providers in case of OSAP 1061, 1723 and 1062

## C Investigation of sales structure

31. The task here is the comparison of sales per activities from the annual performance statistics with the relating data of statistical domains at individual level. The aim is to achieve consistency of the sales per activities in the different data collections.



*Figure 6. Reconciling statistical data collections with administrative sources*

32. Necessary information to carry out the task have been previously identified as
- which questionnaires are used to collect net sales, what is the database identifier; information about the discrepancy of definitions,
  - which of the data from other statistical domains can be used for comparison,
  - characteristics of the individual statistics (gross data: netting, million HUF: change to thousand HUF)
33. There are statistical domains that can be easily used for direct comparison, and there are statistics that can only be compared in terms of size or indirectly. Every other available resource is used, just like the enterprises' notes to the annual financial statement.

## D Investigation of codes for shareholder structure

34. We check the correctness of codes and set the right codes in the register in order to achieve right shareholder structure codes as well as to establish new codes that have been created by the integration of shareholder structure codes and ESA sector codes.

35. Checking shareholder structure codes of all the top enterprises was launched once, in 2013. The codes are checked annually if there is a new top enterprise with unknown shareholder structure code. In these cases the code is set based on the corporate tax declaration or the notes to the financial statements. If during the annual maintenance of the shareholder structure code for all the top data suppliers the code derived from the corporate tax declaration is different from the code coming from the register, those codes need to be validated separately.

## E Investigation of other receipts data

36. Previously, validation of other receipts data of enterprises was not in focus. Investigation has been made only in the case when the other receipts exceeded 500 million HUF, or the other receipts fluctuated extremely compared to the past few years, or in the current year other receipts were not

reported but in the previous years the enterprise had other receipts. The structure of other receipts is also available gathered from the notes of the financial statements.

## **F Investigation of export product structure**

37. Product-level data of export sales from the annual statistics of products (OSAP 1039) are compared with the export product data from the external trade in goods statistics. The products reported in both data collections must be harmonized, taking into account the fundamental differences between the two data collections and nomenclatures.

38. The validation covers those data providers which have reported data both in the statistics for external trade in goods, and in the annual statistics of products. From these data providers those have been brought into focus where the aggregated data at NACE level exceeded 1 billion HUF, and reported very different products. Data from the two different data sources had to be made comparable by the help of the CN-CPA correspondence table from Eurostat's homepage.

## **G Monitoring big investments**

39. The task here is to monitor big investments, and quest investors (on the basis of press information and internet) and obliged them to provide data. This is especially important in case of the data providers of the government sector.

## **H Participation in consultations and audits at big enterprises arranged by the Foreign Trade Statistics Department and the Central Bank of Hungary (MNB)**

## **V. Summary**

40. Although mezo-validation has been defined as posterior validation, the aim is to detect errors at the earliest possible stage. Therefore, based on the mezo-validation experience of the previous year, we pay attention to the coherence already during the primary data processing that will be checked later by mezo-validation. In this way, we can avoid or reduce the additional checking and correction of data.

41. Future tasks and improvement of mezo-validation are the extension of the investigation to a greater scope like economic organizations, company groups, institutions, etc. The ultimate aim is to involve statistical data collections in the most comprehensive way possible.