

**USING A MAP OF SHORT TERM STATISTICS  
 IN THE PROCESS OF ORGANIZING BUSINESS SURVEYS**

Short-term statistics (STS) involves collecting, processing and providing timely and multi-disciplinary information in the area of economy. In Poland STS is mainly conducted through the DG-1 survey, which is a large survey (covering over 30,000 enterprises) conducted on a monthly basis over a short period of time, with a complex system of data processing (linkage of data from other surveys, generalization and conversion of data into fixed prices). The need to reduce the respondent burden and the development of new information and statistical technologies call for a concerted effort aimed at optimizing the organization of statistical surveys. One example of such an effort is the development of a map of short-term statistics conceived as a tool enabling the creation of an integrated and ordered knowledge about the survey.

**1. A map of a system of statistical surveys**

A map of a system of statistical surveys should be understood as a generalized, integrated and order description of the organization of statistical surveys containing the following elements:

- *Basic information about surveys* containing: survey scope, target population, survey frequency, survey administration date, data collection methods, survey implementation.
- *A description of input data (X)* – sources and forms of data; input variables, quality assessment criteria (number of refusals, missing data, error estimates).
- *A description of survey results<sup>1</sup>(Y)* i.e. the format (raw data, adjusted seasonally or on a working day basis, trend), form of presentation (indicator, absolute value), date and place of releasing data or making them available.
- *A description of the transformation of input data into results*, which is essential for further, more in-depth analyses. It can be depicted as shown in Fig. 1.

Transformations can vary and an estimate of an output variable  $y_i$  is obtained by transforming one or more input variables  $X$

$$y_i = f(X) \quad (1)$$

One way of describing interrelations between variables, which is useful for purposes of analysing the whole system, is to present them in the form of a matrix of transformations  $M$ .

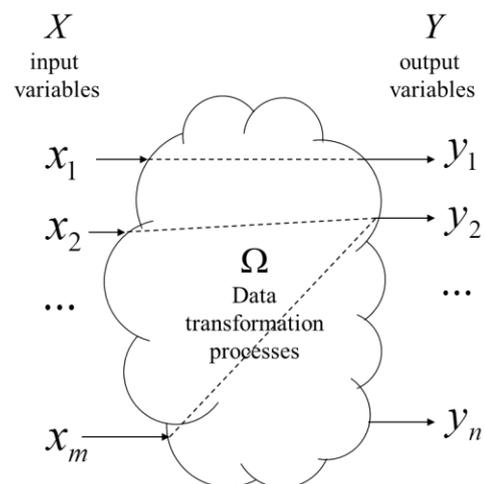


Fig. 1. The process of data transformation

<sup>1</sup> An interesting (and very extensive in its own right) discussion of the definition of statistical surveys and their results can be found in e.g.: [Sundgren 2003] and [Stefanowicz 2007].

$$M = \begin{bmatrix} m_{11} & \dots & m_{1j} & \dots & m_{1n} \\ \dots & & m_{ij} & & \dots \\ \dots & & \dots & & \dots \\ m_{m1} & \dots & m_{mj} & \dots & m_{mn} \end{bmatrix} \quad (2)$$

where  $m_{ij} = \begin{cases} 1, & \text{when input variable } x_i \text{ is used to estimate the value of variable } y_j \\ 0, & \text{otherwise} \end{cases}$

## 2. A map of short-term statistics

According to the above definition, A map of short-term statistics (MSTS) is an information system, which describes and orders key elements related to survey organisation and the distribution of results produced by short-term statistics.

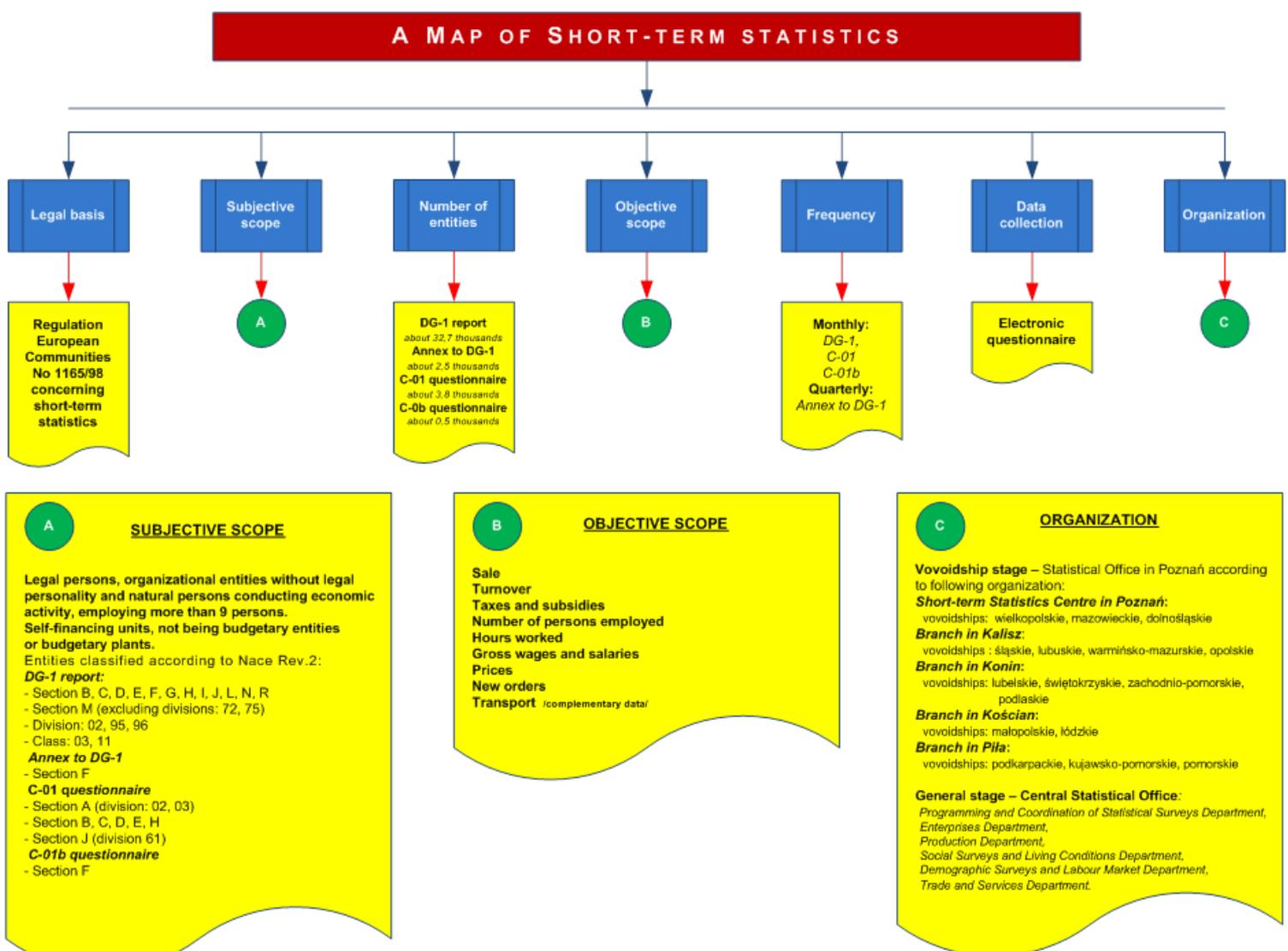


Fig. 2. Characteristics of the map of short-term statistics

Considering the high information value of short-term indicators, they are produced not only for Eurostat, but are also available for national institutions, such as the government and the central bank, and for financial markets, where they are used to analyse the country's current economic situation and make decisions about economic and financial policies.

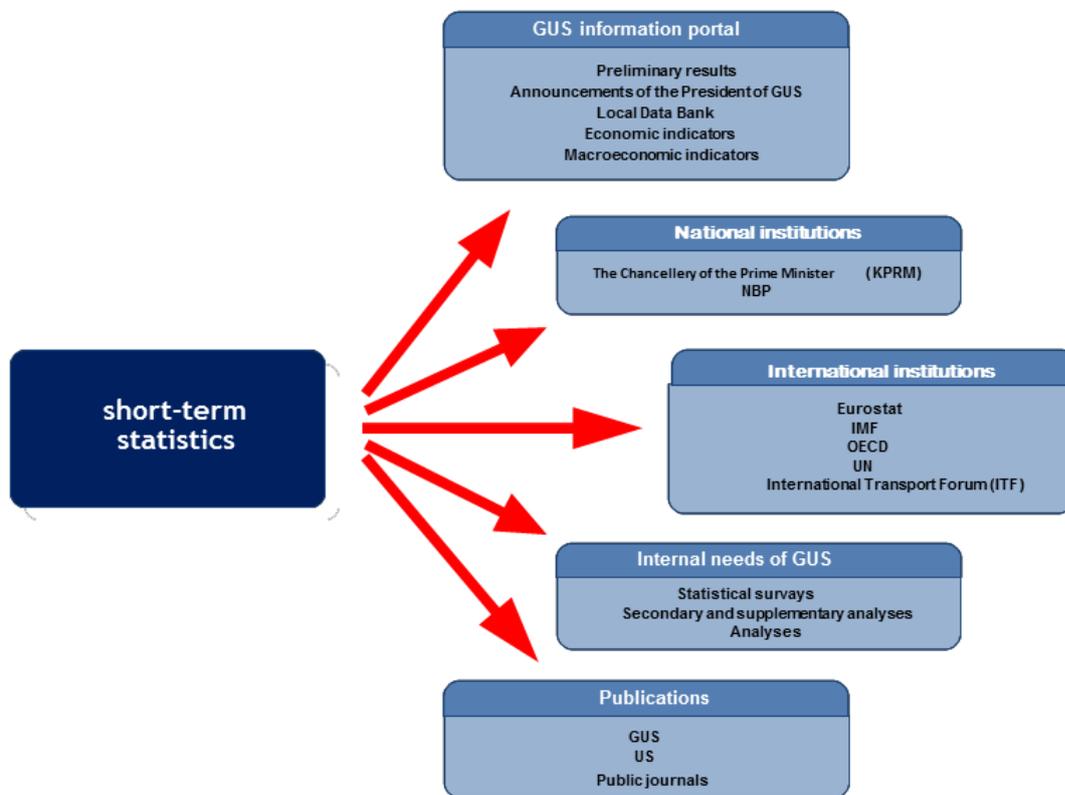


Fig. 3. Destinations for short-term data

The map of short-term statistics contains information about level of detail required by different users and release dates. It is also a kind of guide of GUS publications. The knowledge about users of statistical information is a necessary prerequisite for an in-depth analysis and assessment of their information needs, which provide the rationale for possible changes in survey organisation to meet users' requirements.

Q u e s t i o n n a i r e	Input variable	Output variable												
		New orders	Domestic new orders	Non-domestic new orders	New orders to the euro zone	Turnover	Domestic turnover	Non-domestic turnover	Domestic turnover to the euro zone	Sold production at constant prices - preliminary results	Sold production at constant prices - final results	Sold production at current prices	Labour productivity at current prices	Labour productivity at constant prices
DG-1	Net revenues from the sale of products					✓	✓		✓	✓	✓	✓	✓	✓
	Net revenues from the sale of products to the euro zone								✓					
	Net revenues from the sale of products for the non-domestic market						✓	✓						
	Net revenues from the sale of goods and materials					✓	✓							
	Net revenues from the sale of goods and materials to the euro zone								✓					
	Net revenues from the sale of goods and materials for the non-domestic market						✓	✓						
	Value of manufactured products not classified as sales									✓	✓	✓	✓	✓
	Received new orders	✓	✓											
	Received non-domestic new orders		✓	✓										
	Received new orders to euro zone				✓									
	Excise tax on goods and materials					✓	✓							
	Excise tax on products of own production					✓	✓		✓	✓	✓	✓	✓	✓
	Subsidies								✓	✓	✓	✓	✓	✓
Average paid employment									✓	✓	✓	✓	✓	
Price indices of sold production (previous month=100)								✓						
C-01	Prices of the sale of representative products and services (excluding VAT and excise taxes) in the previous month								✓	✓			✓	
	Prices of the sale of representative products and services (excluding VAT and excise taxes) in current month								✓	✓			✓	
	The value of the sale of representative products and services (excluding VAT and excise taxes)								✓	✓			✓	

Fig. 4. Part of the Matrix of variables - manufacturing

### 3. Using the map of short-term statistics in the process of survey analysis

A map of a system of statistical surveys is useful not only because it brings order to a certain body of knowledge, but also because by introducing additional parameters it can serve various analyses connected with the process of producing and distributing statistical information. For this reason it is necessary to determine for each output variable  $y_i$  certain characteristics, as shown in Fig. 5.

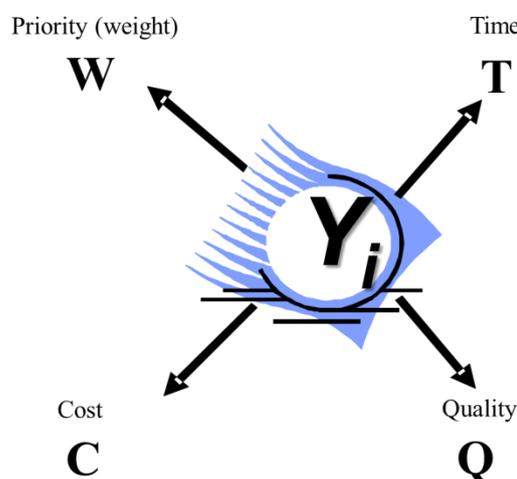


Fig. 5. Assigning characteristics to an output variable

For short-term statistics, the following assumptions have been adopted as to the criteria:

*Priority (weight)* – 3 categories of output information have been distinguished. The largest weight has been assigned to data, which are to be estimated in order to comply with international obligations or national regulations. The lowest priority has been given to supplementary information.

*Release date* – output values have been adopted on the basis of data release practices that have been in use so far, which are described in MSTs.

*The cost of obtaining a variable* – has been calculated by decomposing the total cost of a survey (as described in the Survey Programme of Public Statistics) by means of a matrix of transformation  $M$ .

As a result of calculations, supplementary information has been obtained about the process of implementing surveys in short-term statistics. A few insights are worth highlighting:

- In the DG-1 and C-01 surveys, 62 input variables are used to produce annually over 450,000 reports covering 9 topic areas, 72 output variables and over 100,000 released figures.
- *Average monthly wage* is the largest output variable group, which consists of 19,500 figures.
- Preliminary results (for *Average employment* and *Average wage*) are available as early as the 16<sup>th</sup> business day following the end of the month. 49% of variables used are monthly data, which account for 96.5% of all values estimated in the two surveys. Additionally, 2/3 of monthly output variables have the highest priority.
- The average cost per one report is PLN 42, but the average amount necessary to estimate one value is as much as PLN 185, with considerable discrepancies between the cost of different output variables (*Base amount* being the most expensive one, with the cost of estimating one value amounting to as much as PLN 400,000, and *Average employment* being the least expensive, with the cost of estimating one value at PLN 3). The maximum cost of estimating a single output variable, understood as a group, has been calculated for the *Indicator of producers' prices on the domestic market* to be PLN 1.6 million.
- To estimate *Sales of construction-assembly production by types of building*, as many as 14 input variables are required. The same number of input variables are used only once to calculate output variables, while 6 input variables are used to estimate quarterly or annual data.

#### 4. Summary

In the modern world of information society public statistics not only has to provide reliable and up-to-date information about the socio-economic situation of the country, but also be able to quickly respond to the changing needs of users of statistical data. Changes are also required to make use of modern telecommunication technologies and new sources of data. For this reason, new surveys need to be adapted to respond to changing conditions. A map of statistical surveys is a tool, which can be useful for studying these processes because:

- it provides **essential, integrated information** about implemented surveys, which helps to identify end users and release dates, and describes relationships between different variables. In other words, the map provides a quick answer to questions about what, where, when and how.
- it has **a motivational aspect** – statistical surveys are usually highly complex processes, require the involvement of large groups of employees and a precise division of tasks. This often leads to the isolation of activities and lack of awareness of the overall purpose of individual tasks. The map of statistical surveys offers a wider view of the idea of analysis and data release, raises the awareness of the significance of the results and the overall purpose of the survey. In the end, it provides an answer to the question of who and what the survey is conducted for, which is necessary to justify the enormous expense of time and money required to obtain and process all the information.
- it help to conduct **in-depth analyses**. Given certain assumptions, the map helps to precisely determine the cost of obtaining different variables, quickly identify input variables necessary to obtain high priority information, possible delays in data release dates.
- it can help to **optimize surveys**. A matrix of transformations with precisely described and quantified criteria is a necessary prerequisite to determine the optimal way of data processing in statistical surveys, which can be obtained by solving a problem in mathematical modelling.
- it can be useful to **integrate statistical forms**.

#### Bibliography

- [1] Sundgren B., 2003, *Developing and Implementing Statistical Metadata Systems. A Network of Excellence for Harmonising and Synthesising the Development of Statistical Metadata*, EPROS Project Number IST-1999-29093,
- [2] Stefanowicz B., 2007, *Informacyjne systemy zarządzania*. Przewodnik. Szkoła Główna Handlowa, Warszawa