

MAIN RESULTS OF THE SURVEY ON BUSINESS CIRCLES' NEEDS OF STATISTICAL METADATA

I. INTRODUCTION

Survey on determination of the business circles', as the largest part of the users of statistical information, needs of statistical metadata was carried out for the first time in the country.

The main objective of the survey was to investigate the needs and to define the main types of statistical metadata, which are needed for business circles in the country, and which can facilitate and help the work with statistical data and their correctly understanding and interpretation. At the same time the survey has for an object to improve the business circles' information concerning: statistical information system; statistical information sources (data and metadata); possibilities of statistical data usage and dissemination; the NSI' information and meta-information systems development directions in order to improve and modernize the ways of dissemination of statistical data and users access to the statistical data.

The survey responses are connected mainly with the problems concerning data presentation and data dissemination. Those problems are connected indirectly with the matter of statistical metadata. That approach is applied in view of the fact that the "metadata" concept is rarely disseminated and it is not enough known to the respondents. In order to prevent complicate terminology explanations, more practical interpretation of the concept of "metadata" was adopted. According to the survey directions the concept "metadata" contains auxiliary information for all users, that:

- orientates in statistical information system and informs about conducted statistical surveys;
- Presents the content and the quality of the statistical data;
- shows the ways of statistical data production and how they can correctly understand and explain;
- informs about the ways of data dissemination and access to the data.

The current survey was carried out on the basis of the sample of 4100 units. The sample is a sub-sample from the sample by which the employed persons, working time, wages and salaries and other labour costs are investigated. The sample was stratified by the forms of ownership, by the Nomenclature of the economic activities A17 and by the size of the surveyed units (according to the number of employed persons). The survey was carried out simultaneously with the regular quarterly Enterprise survey on the number of employed persons, wages and other labour costs for second quarter of 2004 (statistical form E – labour, quarterly).

The needed data are collected by the survey on business circles' needs of statistical information and their load as respondents' questionnaire. Response rate is 98.1%. During the survey the questionnaires, filled in by firms from private and public sectors of the national economy, were received as follow:

- 3592 by private sector enterprises and
- 430 by public sector enterprises.

The following information was received by the survey: about the degree of statistical information usage in the business circles and the respondents' main statistical information sources; on the level of notification about statistical metadata availability and possibilities to facilitate

users' activities by application of new methods and means for statistical information providing and dissemination; on the basis groups of statistical data, for which instructions and explanations by "metadata" are needed; on the suitable forms and means for statistical data and metadata providing and other.

II. MAIN RESULTS

1. Statistical information usage and main sources of statistical data

1.1 The results of the survey present that the statistical information is used in the activities of large part of business circles (67.2%) in the country as dominant frequency of statistical information usage is "Rarely" – approximately for the half of interviewed units (46.9%). (Fig.1).

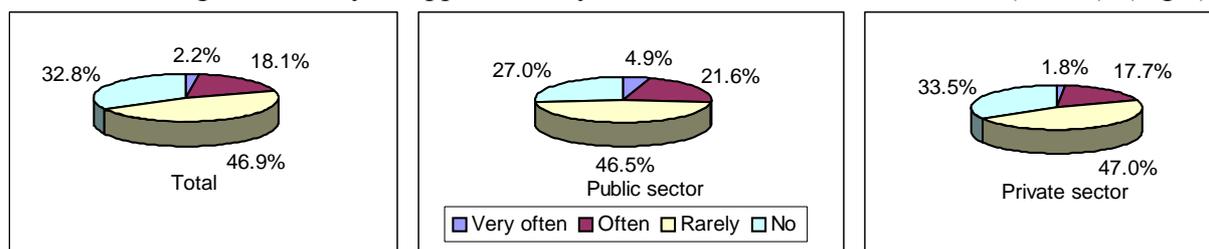


Fig.1. Structure of interviewed enterprises' responses to the question "Is the statistical information in use in your organization (enterprise)?"(Total and by forms of ownership)

Do not utilize statistical information in 32.8% of the enterprises as the rate of the firms is higher in the private sector (33.5%) in comparison with the public sector (27.0%).

Only 4.9% of the enterprises in the public sector answer that they use statistical information very often. 1.8 % of the private sector's respondents indicate the answer "Very often". The firms' estimates in both sectors – public and private – are approximately equal for the answer "Rarely", indicated in 46.5% of the interviewed units in the public sector and in 47.0% of the private sector's firms.

1.2. Answering the question " Which of the following reasons have influence on statistical information usage in negative terms in your enterprise (firm)?" about the half of the interviewed enterprises (46.5%) is indicated that " the lack of acquired habit to work with statistical data" is an important reason. More then 1/3 of the respondents are indicated answer "not need to use statistical data in our activity". A small part of the respondents (17.4%) consider that they "can not find needed statistical information suitable to their activity". At the same time 18.7% of the respondents are indicated "distrust in reliability of statistical information" as a reason negatively influenced on statistical information usage in their organization (firm).

In the answer "Other" as reasons negatively worked on statistical information usage, the respondents are indicated the lack of statistical knowledge; poor qualification; lack of time; lack of financial resources to buy needed equipment.

1.3. More then half of the respondents in business circles indicate that they use "Often" or "Rarely" Internet (60.7%) as main source of statistical information. Considerably less part of the respondents (37.4%) are indicated that they use in their practice the NSI' publications and issues (see Table 1). The NSI' information services are used by 38.4% of the respondents in different degree.

In the answer “Other” the respondents indicate that they find desired statistical information in specially publications, branch bulletins, from the NSI by phone, by information agencies and other.

Table 1

Distribution of the interviewed units’ answers of the question “Which of the sources listed below were used by you to find needed statistical information?”

Code	Sources	Responses		Often	Rarely	Don’t use
		Responded		Rate (%)	Rate (%)	Rate (%)
		N	Structure (%)			
A	Internet	2953	100.0	33.6	27.1	39.3
B	The NSI’ information services	2740	100.0	6.9	31.5	61.6
C	The NSI’ publications	2717	100.0	8.0	29.4	62.6
D	International statistical organizations’ publications	2613	100.0	2.6	12.3	85.1

The dominant part of the respondents (62.8%), which use Internet as source of information in their activity, are indicated that they receive (“Often” or “Rarely”) statistical data by the NSI’ web site (www.NSI.bg). Other Internet web sites like www.noi.bg (74.9%); www.minfin.government.bg (55.5%), www.portal.government.bg (44%), www.bnb.bg (43.0%) and www.Stat.bg (25.7%) are used for statistical data obtaining.

The level of knowledge (of the respondents) of the basis data and metadata included in the NSI web site is investigated by the question “Do you know that you can find in the NSI web site the following information?” (Table 2).

A little more $\frac{1}{4}$ of the respondents are indicated that they know about the National classifications and nomenclatures (30.7%) and “Data on statistical surveys” (26.0%) included in the NSI web site.

Table 2

Distribution of the interviewed units’ answers of the question “Do you know that you can find in the NSI web site the following information..... ?”

Code		Responses		Yes	No	Don’t use
		Responded		Rate (%)	Rate (%)	Rate (%)
		N	Structure (%)			
A	National Programme for Statistical Surveys	3892	100.0	19.2	20.0	60.8
B	Calendar of the statistical surveys	3850	100.0	14.1	22.5	63.4
C	List of the standard statistical indicators for which the data are disseminated	3855	100.0	19.1	21.3	59.6
D	Data on statistical surveys	3865	100.0	26.0	16.9	57.1
E	Statistical surveys’ methodologies	3840	100.0	13.7	22.9	63.4
F	National classifications and nomenclatures	3884	100.0	30.7	17.3	52.0
G	Catalogue of the NSI’ statistical surveys and publications	3828	100.0	18.4	21.4	60.2

It makes an impression that almost 1/2 of the respondents used the statistical information in their practice are declared that they didn't know before conduction of the survey (answer "No") that they can find in the NSI' web site the National Programme for Statistical Surveys, Calendar of the statistical surveys, Statistical surveys' methodologies, Catalogue of the NSI' publications and List of the standard statistical indicators for which the data are disseminated.

2. Concerning the time while the respondents are working with statistical information and the business circles' need of methodological explanations and instructions

2.1. In accordance with the respondents' opinion least part of the time is separated to "clarify the statistical methodology". More then 2/3 of the respondents indicate that they have no time to "clarify the statistical methodology". Only 3.3% of all respondents answer that they use "Large part" of their time for this purpose (Table. 3).

Table 3

Distribution of the interviewed units' answers of the question "Which part of the time during the work with statistical information is separated for the activities listed below in your firm?"

Code	Activities	Responses		Large part	Small part	No time
		Responded				
		N	Structure (%)	Rate (%)	Rate (%)	Rate (%)
A	Searching and collection of information	3108	100.0	11.2	52.2	36.6
B	Orientation in the information sources	3042	100.0	5.2	39.4	55.4
C	Information (Data) processing	3079	100.0	12.0	43.9	44.1
D	Clarification of the statistical methodology	3023	100.0	3.3	26.9	69.8
E	Specification of the statistical indicators' content	2927	100.0	5.6	34.3	60.1

36.6% of the respondents not separate time for "Searching and collection of information". At the same time 11.2% of interviewed units consider that they devote "Large part" of their attention to this activity in their enterprises and 52.2% - "Small part" of time.

As "Other" activities for which the time is devoted the respondents determine the time for comparative analysis and coordination of the information from different sources.

2.2. More then 1/3 of the respondents answer that during their work with statistical information they use in different degree ("Often" or "Rarely") methodological explanations and instructions in the areas listed below (Table 4).

Table 4

Distribution of the interviewed units' answers of the question "Is it necessary to use methodological explanations and instructions in the areas listed below during your work with statistical information?"

Code	Areas	Responses		Often	Rarely	No
		Responded				
		N	Structure (%)	Rate (%)	Rate (%)	Rate (%)
A	Statistical surveys	2921	100.0	5.3	24.9	69.8
B	Statistical indicators	2935	100.0	8.1	32.6	59.3
C	Statistical concepts and definitions	2912	100.0	7.2	27.2	65.6
D	Classifications and nomenclatures	2950	100.0	11.2	31.4	57.4

E	Data comparability	2913	100.0	8.7	26.3	65.0
F	Level of harmonization with EU legal acts	2880	100.0	4.0	15.9	80.1
G	Data timeliness and data periodicity	2886	100.0	9.0	24.3	66.7

2.3. Only 28% of the respondents are indicated negative answer (“No”) the question “Do you meet difficulties with statistical indicators interpretation?” They consider that it is not difficult for them.

45.1% of the respondents have problems during the statistical indicators explanation.. 6.5% of the interviewed units have problems “Always” and 38.6% - “Sometimes”. At the same time 26.9% of the respondents are indicated the answer “Can not assess”.

2.4. Investigation of the respondents’, users of statistical information, opinion on difficulties encountered during the statistical information usage, presents that the largest part of them are not met: “Discrepancies in the information” (59.2% of the respondents); “Very detailed information” (65.8%) and “Statistical data confidentiality” (69.9%). The respondents consider that they “Often” meet informative problems with: “Not illustrated statistical concepts and terms” (30.1% of the respondents); “Unclear contents and/or incompatibility of the indicators” (25.1%); “Lack / or incompleteness of methodological instructions” (28.2%); “Untimeliness obtaining of information” (19.9 %).

3. Business circles’ necessities of statistical metadata

The survey results present that:

1. Large part of the respondents (55.0%) would like to have “A List of available data bases and access to them” as 27.3% of the respondents consider that this information is “A much needed” for them.
2. “Glossary of statistical terminology” is necessary to work with statistical data in accordance with the opinion of 51.5% of respondents. 22.9% of the respondents consider that it is “A much needed” for them.
3. “A description of statistical indicators and time series “ is information, which is “A much needed” for 19.1% of the respondents. It is “Not so needed” for 30.8% of them. Approximately the same is the correlation of answers concerning the need of information on “The statistical surveys methodologies”, “Statistical questionnaires and instruction for their filling in”, “Data processing descriptions and descriptions of data presentation formats”.
4. The answers of the dominant part of the respondents present that the following quality characteristics as “Completeness (coverage)”, “Reliability”, “Comparability”, “Accessibility”, “Timeliness” are very important for them. More then 1/4 of the respondents are indicated that these characteristics are of “Large” importance for their activity. The “reliability” of the information is defined as “Large” importance by the biggest part of the respondents - (43.8%). The levels of importance of another surveys characteristics are defined by the respondents as follow: “Accessibility” (35.3%), “Timeliness” (31.1%), “Comparability” (30.2%).

4. Informative and technical means for business circles work with statistical data facilitation

4.1. “The glossary of statistical terms and concepts” is treated as mean that facilitate the work with statistical data by the largest part of the respondents (65.9%). 37.5% of the respondents are indicated the answer “Fully” and 28.4% of them – the answer “Partially” (Table 5).

Table 5

Distribution of the interviewed units’ answers of the question № 11 “Do you consider that availability of the following means and data bases will facilitate your work with statistical data?”

Code	Informative means	Responses				
		Responded		Fully	Partially	No
		N	Structure (%)	Rate (%)	Rate (%)	Rate (%)
A	Glossary of statistical terms and concepts	2774	100.0	37.5	28.4	34.1
B	Information system with statistical indicators and time series	2686	100.0	23.5	30.7	45.8
C	Information system for statistical classifications and nomenclatures	2708	100.0	30.3	29.6	40.1
D	Information system containing statistical instrumentarium (statistical questionnaires, instructions, information, inquires...)	2675	100.0	20.8	28.6	50.6
E	Information searching system (statistical data searching by key words, phrases ...) on the basis of common national standards for electronic data exchange	2692	100.0	35.7	21.5	42.8
F	On-line access to the NSI data bases	2680	100.0	37.9	21.4	40.7
G	On-line access to the NSI library	2630	100.0	32.4	21.5	46.1

Smaller part of the respondents (59.9%) in comparison with previous case considers that their work with statistical data can be facilitated by availability of “Information system for statistical classifications and nomenclatures“. For 30.3% of the interviewed units find that the level of facilitation is “Large”.

In accordance with the opinion of 57.2% of interviewed enterprises their work with statistical data will be facilitated “Fully” or “Partially” by availability of “Information searching system (statistical data searching by key words, phrases ...) on the basis of common national standards for electronic data exchange”. Approximately the same part of the respondents (54.2%) answer that their work with statistical data will be facilitated by availability of “Information system with statistical indicators and time series”. Smaller part of the respondents in comparison with previous case (49.4%) expresses the same opinion concerning availability of “Information system containing statistical instrumentarium (statistical questionnaires, instructions, information, inquires...)”.

About 3/5 of the respondents consider that their work with statistical data will be facilitated “Fully” or “Partially” by possibility of on-line access to the NSI data bases and to the NSI library.

4.2. As “Other” means of facilitation the respondents indicate the databases with regional data by regions and districts and permanent sending of statistical data by e-mails.

5. Business circles opinion on the priorities of introducing of the information and technical means of access to statistical data and of statistical information dissemination

5.1. In order to facilitate the access to the NSI statistical data ½ of the respondents find it necessary to introduce with first priority “Glossary of statistical terms and concepts”, “On-line

access to the NSI data bases”, “Information searching system (statistical data searching by key words, phrases ...) on the basis of used common national standards for electronic data exchange”, “Information system for statistical classifications and nomenclatures“ and “On-line access to the NSI library”.

5.2. The rate of the respondents, which consider that the NSI must introduce “Immediately” on-line access to the NSI databases, is large - (32.2%). The same is the respondents’ opinion of on-line access to the NSI library (27.1%) and of information searching system (31.1%).

More then 1/4 of the respondents (27.6%) find that with first priority and “Immediately” must be introduced the “Glossary of statistical terms and concepts” and 28.0% of them consider that it can be introduced “Later on”. (Fig. 2).

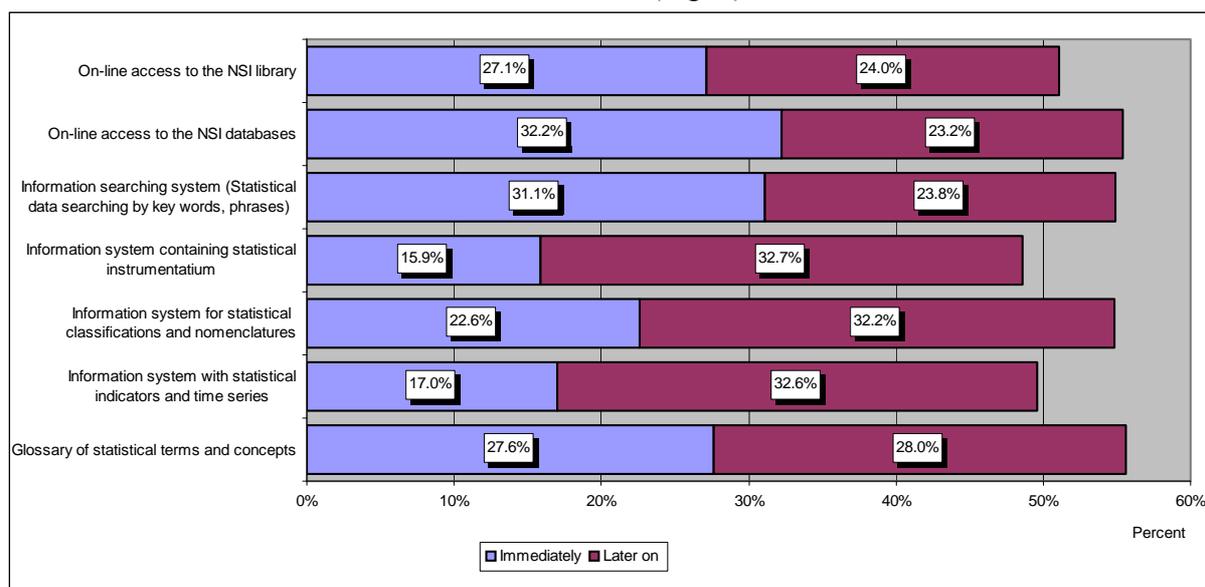


Fig.2. Rate of the firms answers “Immediately” or “Later on” the question “In your opinion what is the priority of introducing in the NSI the means and data bases listed bellow to facilitate your access to statistical information?”

5.3. The business circles respondents give preferences to the following technical means for obtaining the statistical information: printed documents, e-mail, diskettes, on-line and CD. The obtaining of statistical information as printed documents is more preference form in business circles in comparison with on-line services and diskettes or CD yet.

5.4. The respondents’ preferences for usage of modern means of official statistical data obtaining (by e-mail or on-line) are determined in considerable degree by the level of ICT development of the business circles in the country.

In accordance with the survey results in 79.6% of the firms exist working places equipped with PCs. 41.2% of the respondents indicate that local communication networks are built in their organization. 50.5% of them use e-mail and 59.2% - Internet.

The proportion of the respondents answers, distributed by forms of ownership, presents that in comparison with the public sector, working places equipped with PCs exist in the greater part of the private sector firms (80.2%). 60.0% of them have connection with Internet and 51.4% use e-mail. (Fig. 3).

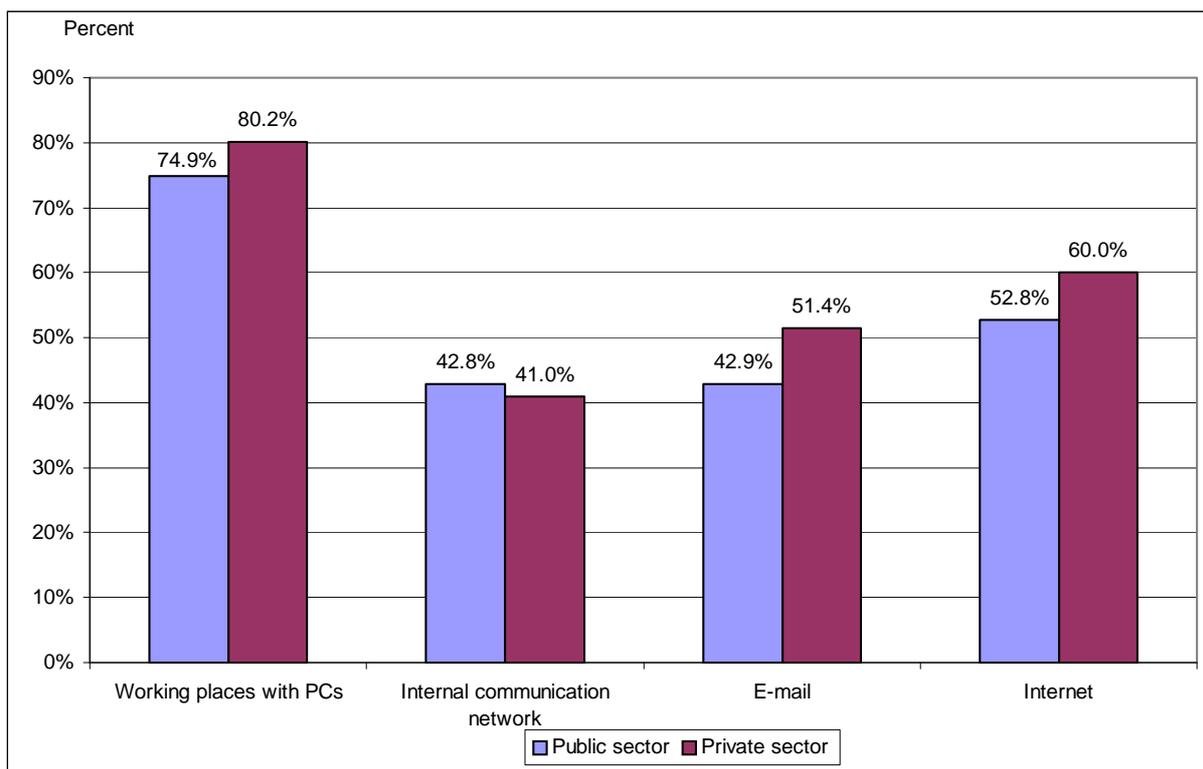


Fig.3. Rate of the firms, which as answers the question “ Which of the means for external and internal electronic data exchange, listed below, exist (you built) in your enterprise?” indicate that they have such means.

The proportion is changed in the field of communication networks usage. 42.8% of the public sector respondents and 41.0% of the firms in private sector indicate that their organizations (firms) are built such networks.

For large part of the respondents’ availability of local communication networks with e-mails and existence of electronic exchange systems for documents as well as the future perspectives for their development predetermine the choice of modern technical means of statistical data obtaining.

The received survey results will be used in further NSI activities regarding the problems of business circles needs of statistical data. The respondents’ opinions will be used in the process of formulating the NSI policy in the field of establishing of statistical meta-information systems, choice of metadata standards; new possibilities to the data submission and access to the data; introducing a new methods and technologies for data dissemination.