

QUALITY OF THE STATISTICAL INFORMATION – IMPORTANT PRECONDITION FOR THE RELIABILITY OF GLOBAL ECONOMIC ANALYSIS

UDC 519.257

Veselka Pavlova

University of National and World Economy – Sofia, 1700 Sofia,
Studentski grad "Hristo Botev"

Abstract. *In accordance with the tasks assigned by the Phare Project, a survey on the users' satisfaction with statistical data was realized in August-September 2005 in Bulgaria. It was based on a summary of the subjective opinions, assessments and recommendations of 1502 users. They were distributed in several basic consumers' segments – business enterprises, central public administration, district administration, municipal administration, non-government organizations and educational institutions. The components referring to the quality, as recommended by Eurostat are: accessibility; relevance; data juxtaposition; data reliability; terms for information provision; accompanying analyses and assessments (metadata) and clarity of the indicators.*

Key Words: *quality of statistical information, components of quality, assessment of quality components*

In the current process of Euro integration and globalization, the main task for every statistical institution is to provide sufficient information with good quality as a reliable base for national and international analysis. Concerning users' satisfaction with the quality of the data produced and provided by the National Statistical Institute (NSI) of Bulgaria, a specialized study was carried out in August and September 2005¹.

1. HOW TO DETERMINE THE QUALITY OF STATISTICAL DATA AND ITS COMPONENTS

According to Eurostat's definition, statistical quality consists of "existing features and characteristics of a product or service for satisfying existing or suggested needs on the

Received October 05, 2006

¹ The author of this paper was a regional expert in Phare Project "Regional Statistics Improvement and Assessment of the Quality of the Statistical Data", EuropeAid/ 116362/D/SV/BG

part of statistical information's users"². Eurostat has been developing an approach for the assessment of the quality of statistical data. This approach is based on general documents as follows: Definition on quality in statistics; Glossary of the main terms connected with quality; Standard Quality Report; Variance Estimation Methods in EU; Standard Quality Indicators³.

According to the definition⁴, quality is characterized by **seven components**: **relevance**, appropriateness, importance of the statistical data for the user; **accuracy** (measured by the error size as sum total of its components – non-stochastic, stochastic, arbitrary); **timeliness and punctuality** of the statistical data and observing the respective terms for conducting the statistical survey; **accessibility and clarity** of the distributed statistical data; data **comparability** in territorial and temporal terms; **coherence** regarding the content between preliminary and final data, between short-term and annual indicators, between indicators in the same social and economic area as well as with the indicators in the national accounts; **data completeness**, lack of available data and clarifying the reasons for this.

2. BRIEF PROFILE OF THE USERS OF STATISTICAL INFORMATION AND THEIR SATISFACTION WITH ITS QUALITY

It is interesting that the distribution of users among **organizations different in size** – small (up to 49 employees), medium (from 50 up to 249 employees) and big (over 250 employees) organizations, is relatively even with a slight majority of small and medium ones (Table 1).

Table 1. Distribution of the interviewed persons by type of organization(%)

Type of Survey Organization per the Number of Employees	User Groups					
	Business Enterprises	Central Public Administration	District Administration	Municipal Administration	NGOs and Educational Institutions	Total
up to 49	45,3	4,8	85,2	15,8	21,2	35,4
from 50 up to 249	32,0	11,8	14,8	75,3	41,1	35,6
more than 250	22,7	83,4	0,0	8,9	37,7	29,0

At the same time, from the user groups' point of view, this distribution is absolutely uneven. For the business enterprises sample, for example, users engaged in small organizations are predominant – 45,3% out of all users engaged in business enterprises belong to small sized ones. Only 22,7% of them belong to large enterprises. As it might be ex-

² Quality Declaration of the European Statistical System, Eurostat, 2002.

³ According to: Assessment of Quality in Statistics, Luxembourg, 4-5/04/2000 и Quality in Statistics, Luxembourg, 23-24 May 2005 and Svien Nordbotten, Strategies for Improving Statistical Quality, Statistical Commission and Economic Commission for Europe, UN/ECE, Work Session on Statistical Data Editing, Rome, Italy, 2-4 June, 1999.

⁴ Quality in Statistics, Initial Workshop, p.p. 4-16, Sinaia,2004.

pected, more than 83% of the users within public administration belong to large structures with over 250 employees whereas the highest percentage of respondents within municipal administration (75,3%) and NGOs and educational institutions (41,1%) belongs to middle sized organizations.

An addition, thorough analysis (Table 2) gives the possibility to find out that, unfortunately, the highest shares of persons who do not use statistical information belong to the business enterprises (42, 2%). On the contrary, about 1/3 of the NGOs and educational institutions are regular statistics users. The regional and municipal administrations use statistics mostly on occasional basis (53% of the cases) and in certain periods when they prepare business profiles, report annual activity or apply for financing under a new program.

Table 2. Frequency of statistics use and user type (%)

How often does your job require use of statistics?	User Groups					Total
	Business units	Central public administration	District administration	Municipal administration	NGOs and educational institutions	
Permanently	5,8	14,1	16,7	15,3	31,8	11,1
Twice, three times per week	2,3	7,0	7,4	7,9	9,9	4,6
Several times per month	11,0	14,1	18,5	22,3	17,9	13,9
More intensively in particular time periods	38,7	33,4	55,5	49,1	29,8	39,2
I don't use statistics	42,2	31,4	1,9	5,4	10,6	31,2

Upon combining the branch profiles of statistics users with the frequency of the use of statistics, it should be concluded that health care and social service jobs require a permanent referring to statistics (41,2% of the respondents); in education this percentage is 23,8% and in public governance – 15,6%. Statistics are being used on weekly basis or 2-3 times per month by employees in the fields of: health care – 29,4%, public governance – 26,3% and financial brokerage - 25,0%. Therefore, a conclusion can be made that **regular statistics users are mainly public administrations within public governance, health care and social services, as well as education.**

When referred to the respective **user groups**, the highest percentage belongs to the persons from the district administration (35,3%) who are fully satisfied with NSI's information. (Fig. 1). At the same time, among this very group are the users whose expectations are not satisfied (13,7%). Apparently, there is a certain polarity in the users' opinions, which probably linked to the degree of competence of the respondents and the frequency of the use of the statistical data.

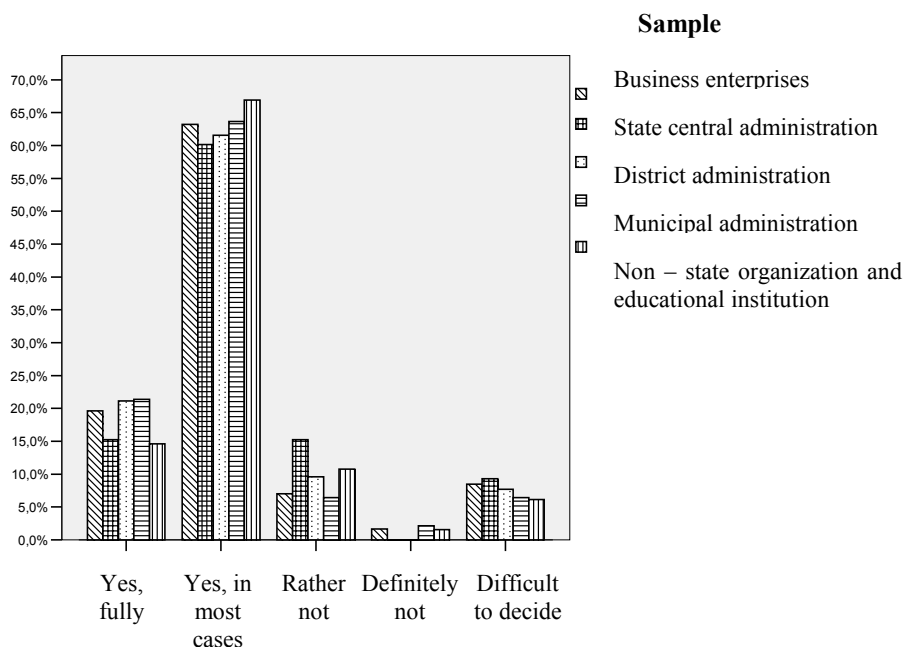


Fig. 1. Users' satisfaction with the statistical information

Depending on their **organization's size**, between 26,0% and 29,0% of the interviewed persons are not fully satisfied with the statistical information provided by the NSI. Between 60,0% and 65,0% of the respondents are partly satisfied regarding expectations and the relative share of those whose expectations are not met varies from 9,0% to 13,0%, where the latter percentage is again for the small-sized organizations.

Fully met expectations regarding NSI's information have 46,2% of those **working** in mining and quarrying, 42,9% of those employed in hotels and restaurants and 37,5% of financial brokers. Fully disappointed remain 22,2% of those dealing with real estate renting and business activities and 18,8% of the financial brokers.

The analysis connected with the **assessment of certain statements** on the very quality of the statistical information of NSI demonstrates that the respondents most agree with the reasonable price (42,8%), followed by the high data reliability (36,1%). The statement on the very high quality of the provided information is supported by slightly more than 1/4 of the interviewed. What is positive is that only 6,7% of the respondents fully reject this statement. At the same time, this is also the lowest percentage regarding disagreement among the compared statements. The highest relative share regarding disagreement belongs to the statement on the satisfaction of users' expectations (10,2%).

When considered by **user groups**, the assessments of the interviewed users regarding the separate statements accumulate around the partial agreement with them. The highest relative share belongs to the surveyed persons from the central state administration followed by those from the municipal administration and the non-government organizations

and educational institutions that have partly confirmed the statements on information quality. Most frequently, complete agreement was expressed by the employed in the district administrations followed by those from the municipal ones. There is a certain variation in the negative responses regarding the different statements. For example, the highest relative share of those who disagree with the statement on the statistical data's high quality is among the interviewed from the district administration (7,7%). Least satisfied are the expectations of the same group of respondents (13,7%). The statement on data reliability is rejected by the biggest number of respondents from the municipal administration and the one on the reasonable price of data from the persons in the non-government organizations and educational institutions (almost 20,0% of the interviewed in this group).

The **organization's size** does not greatly influence the users' assessments. The highest percentage of those who fully supported the statements on the high quality of information is among the interviewed from the medium-sized organizations from (29,0%) and the fully met expectations to 45,6% for the statement on the statistical price of information. We also observe that the highest percentages of those who fully reject all statements are among the interviewed from the small-sized organizations. The explanation of this phenomenon is probably in the fact that the small-sized organizations use statistical data less frequently and belong mainly to the group of business enterprises. We can assume that this very group is in a process of structuring its needs and requirements regarding the statistical information provided to it.

The comparative analysis by the **economic profile** shows that the full agreement with the statements of the quality of statistical data is expressed by the employed mainly in hotels and restaurants, by those working in health care and, in some cases, in education, mining and quarrying, transport and communications. Those who completely reject the statements of real quality in the highest degree are those dealing with real estates, renting and business activities and the financial brokers.

If we compare the **resulting grades** in the different statements we will find out that the statement which was valued highest was that regarding the reasonable price of information (4,98), followed by the statement on its high reliability (4,79), followed by the statement on the high quality of information (4,67) and finally by the statement regarding the users' degree of satisfaction (4,57). At the same time, the highest variation in the opinions of the surveyed users is seen in the assessments related to reliability and quality.

On the whole, the **resulting average grades** are high and hinting at an acceptable level of the quality of the offered statistical data. Particularly interesting in this study is the binding of the statistical information users' satisfaction with their assessment of the importance of the quality components⁵ of the statistical information. Based on the conclusions that the assessments of the components' importance are predetermined and highly reflect the interviewed' personal impressions from the quality of the statistical information used, then the analysis of this interconnection can characterize two main things. On one hand, it can demonstrate to what degree the resulting assessments are objective and consciously expressed, and, on the other, it can help us assess the actual degree of satisfaction of the users' needs regarding the statistical information.

⁵ LEG on Quality Implementation Group (2005), "2004 LEG Implementation Status Report", CIRCA Doc. ESTAT/02/Quality/2005/13.b/2004, Seventh Meeting of Working Group "Quality in Statistics", Luxembourg, May 2005.

The careful consideration of the obtained results demonstrates that, in most cases, there is a strong connection between the statements. For example, over 93,0% of the interviewed who evaluate the separate quality components as "high" agree (fully or partly) with the defined statements. In some cases these percentages are more than 96,0%, e.g. as this is with the statements on the high quality and reliability of the statistical information. Therefore, we could assume that the resulting assessments regarding the statistical quality of information and the respondents' degree of satisfaction with it reflect the objective situation as regards information. At the same time, the majority of the respondents have clear requirements towards the NSI's data as well as criteria for quality assessment of the statistical data and the degree of their satisfaction.

3. GENERAL DESIGN AND SUMMARIZED INFERENCES FROM THE PROJECT

The general scheme of the realization of the Phare project "Regional Statistics Improvement and Assessment of the Quality of the Statistical Data", EuropeAid/116362/D/SV/BG is presented on Fig.2.

A detailed review of the existing methods that are applied in EU statistical practice for assessment of data quality (including quality indicators produced from them) is made, as well an analysis of the differences between the varieties of methods used. General statistical programs like SPSS or STATA, that offer very good tools for data analysis, starting from simple descriptive statistics (one-way distributions, frequency tables, number of missing/non-missing values etc.) were used. SPSS (v. 14) contains two separate modules: Data Validation and Missing Value Analysis. STATA has user-written add-ons for imputation.

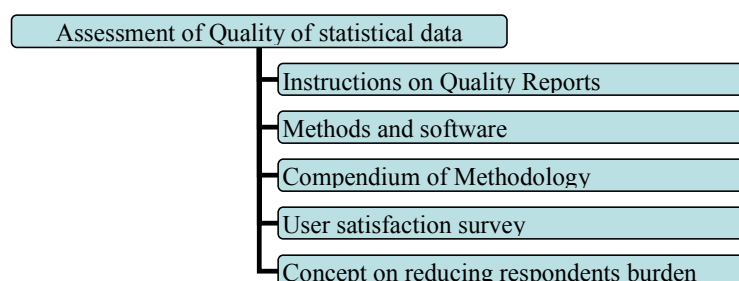


Fig. 2. National Phare Program Results

On the basis of the survey's results and their analysis the **following conclusions and generalizations** can be drawn:

1. Concerning the type of user group, the majority (over 75%) of the users of statistical information produced and provided by the NSI, is concentrated in the big residential areas and the country's capital;

2. Regular statistical information users are mainly state structures related to the state administration, health care and social work and education. Those who use very rarely or do not use at all statistical information from the NSI are primarily the users from business enterprises (42,2%);

3. Over 1/3 of the interviewed persons use statistical information produced and provided by the NSI, where 11,1% use it constantly and 13,9% use it often (2-3 times per week). The prevailing percentage belongs to the users who utilize statistical information more widely only in certain periods;

4. There are two main reasons for an organization's not using NSI's data: first the interviewed think that such information is not necessary for their organization (48,1%) and second is that it does not provide the information it needs (36,2%) itself. The first opinion prevails for the users from the business enterprises and the second is popular mainly among the users from the state and municipal administrations;

5. Users' satisfaction with the provided statistical information is evaluated by way of two prisms: users' satisfaction with the information servicing provided by the NSI and assessment of separate statements connected with the statistical data's quality. The analysis demonstrates that more than 82% of the interviewed persons are completely or in most cases satisfied with the quality of the information servicing and only 1,4% are entirely dissatisfied;

6. The majority of users (64,9%) think that the statistical information provided by the NSI is sufficient. At the same time, 5,4% of the interviewed persons had to require additional data very often. In more than half of the cases these are persons who considered the statistical information as very important;

7. The assessed average grades for the quality components of the statistical data vary from 4,40 to 5,28 and support the conclusion already made that the quality components are evaluated as comparatively high by the users;

8. The assessments regarding the importance of the separate quality components of the statistical information show that they are predominantly estimated as "average". The highest assessment is given for the "accessibility" component (47,3%), followed by the "data reliability" (41,1%) and "clarity of the indicators used" (38,1%). The lowest evaluation receives the component "accompanying analyses and assessments" (14,9% of the interviewed give a unsatisfactory evaluation);

9. If determining the importance of the separate quality components is based on the personal experience, we can take them also as assessments of the quality components of the statistical information provided by the NSI. On the whole, the evaluation of the quality components is above the average level and can be assumed to be completely satisfactory. The lowest evaluation of almost all quality components of the statistical data seems to be given by the group of those dealing with real estates, renting and business activities.

10. The overall evaluation of the quality of the NSI's information is reducing to the affirmation or rejection of certain quality-related statements. The main are: quality of the statistical data, degree of coinciding between the users' expectations and reality, reliability and accessibility of the price of the provided statistical information:

11. The comparison between the users' level of assessment of a lot of statements shows that the highest evaluation is given to the accessibility of the price of provided information followed by the one regarding the high degree of data reliability. The users put in the third position the statement on the high quality of the information and in the last, fourth place - the statement on the users' satisfaction regarding their expectations. At the same time, the highest variation in the users' opinions can be seen in the assessments regarding reliability and quality;

12. The main directions in which the users would like to make changes in the NSI's information are: a) the terms for preparation and provision of the statistical information, b) the punctuality and reliability of information, c) the coverage of the indicators used and the c) juxtaposition of the statistical data.

The review of LEG implementation status in NSI can be considered as very relevant about evaluation the progress achieved on quality in statistics. A clear picture of the stages of recommendations fulfillment is outlined. Taking into account that quality improvement is a continuous process, further development could be expected even for the recommendations with status "Met" as regards the recent attainments about "Quality".

KVALITET STATISTIČKIH INFORMACIJA - BITAN PREDUSLOV ZA POUZDANOST GLOBALNE EKONOMSKE ANALIZE

Veselka Pavlova

Istraživanje o zadovoljavanju potrošača statističkim podacima realizovano je u Bugarskoj u periodu avgust-septembar 2005. Bazirano je na pregledu subjektivnih mišljenja, opozivanja i preporuka 1502 korisnika. Oni su bili podeljeni u nekoliko osnovnih potrošačkih grupa - poslovna preduzeća, centralna javna administracija, okružna administracija, opštinska administracija, nevladine organizacije i obrazovne institucije. Komponente koje se odnose na kvalitet, a predložene su od strane Eurostat-a su: dostupnost, relevantnost, pouzdanost podataka, uslovi za pribavljanje informacija, prpratna analiza i jasnost korišćenih indikatora.

Ključne reči: kvalitet statističkih informacija, komponente kvaliteta, analiza.