

Seminar

REEENGINEERING OF BUSINESS STATISTICS

Lisboa, 11 – 12 October 2007

2. IMPACT OF ADMINISTRATIVE DATA ON BUSINESS STATISTICS

2.3 INTEGRATION OF ADMINISTRATIVE AND SURVEY DATA

Presented by:

V. LAPENIENE

Statistics Lithuania

INTEGRATION OF ADMINISTRATIVE AND SURVEY DATA

Vilija Lapėnienė, Statistics Lithuania, e-mail: vilija.lapeniene@stat.gov.lt

Introduction

The Department of Statistics to the Government of the Republic of Lithuania (hereinafter referred to as „Statistics Lithuania”) is the central statistical office and the leading statistical institution in the country. The principal responsibilities and rights of Statistics Lithuania are laid down in the Law on Statistics of the Republic of Lithuania. The law defines official statistical data sources, gives Statistics Lithuania a right to access administrative sources for statistical purposes. The provision in the Law is very important, but in addition to that, Statistics Lithuania has built up a good cooperation with the national administrative bodies which collect administrative data. Regarding business statistics this concerns in particular the State Tax Inspectorate (hereinafter referred to as „STI”), State Social Insurance Fund Board (hereinafter referred to as „SSIFB”), Customs authority, State Enterprise Centre of Registers with which cooperation agreements are signed.

Constant increase in the demands for statistical information challenges national statistical institutions (hereinafter referred to as „NSIs”) to trade off between meeting user needs, ensuring quality of statistical information and simultaneously reducing the statistical response burden on respondents.

To solve this problem, Statistics Lithuania initiated concrete strategic actions at organizational and technological levels in order to deal with the existing and newly emerging requirements and reduce the response burden. The work is conducted in few basic directions: looking for possibilities to discontinue statistical surveys; revising indicators in statistical questionnaires in order to avoid any slightest appearance of duplication; improving statistical sample surveys; using administrative data sources more widely; introducing electronic tools for data collection; cooperating closely with business entities; elaborating a response burden measurement indicator and applying it for the burden assessment.

Statistics Lithuania among other measures which contribute to the reduction of the response burden has also taken on board the issue on the use of business accountancy data and, in relation to this, carried out a study in order to clarify whether it is possible to implement a project related to the automated statistical data extraction directly from the business accountancy systems. A target aimed at creating a system how to extract data directly from business accountancy systems for statistical purposes will become a top priority of Statistics Lithuania and it has to be considered on the Community level as well.

Statistics Lithuania initiated many projects and measures aimed at the reduction of the response burden. Hard work and efforts taken during the recent years let us understand that it will never work if the efforts are not joined together between NSIs, Eurostat, international organizations. More active approach should be also taken in terms of cooperation with policy makers. Success of efforts and measures will depend on all interested parties involved in balancing the response burden and users needs.

Business statistics is managed by conducting statistical surveys and using administrative data sources. The surveys are conducted on the basis of the Statistical Profile Business register (hereinafter referred to as the „SBR”). The SBR sources are as follows: administrative data sources

(Register of Legal Entities, Taxpayer Register of the SSIFB, VAT register (important element – unique identification code) and statistical surveys.

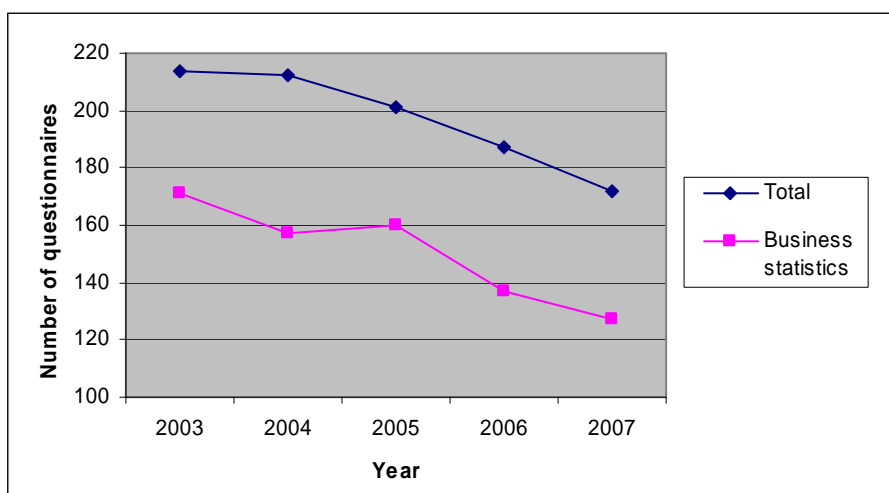
Statistical surveys are the main source of statistical data, respondents for which are natural and legal entities of the country. About 45 thousand or 60 percent enterprises are surveyed each year, which makes on average about 4 statistical surveys per enterprise.

Means for the reduction of the response burden

The response burden is a central goal of Statistics Lithuania aiming at the collection of adequate and accurate statistical information with the least possible burden on business. In order to reduce the burden the work is conducted in several basic directions: looking for possibilities to discontinue statistical surveys; revising indicators in statistical questionnaires in order to avoid any duplication; more wide use of administrative data sources. A number of measures were implemented. An analysis of the burden on respondents revealed that large enterprises receive too many statistical questionnaires, whereas small enterprises even if the number of the statistical questionnaires is not so high, have to participate in the surveys several years in a row. It is a problem of a small country. Statistics Lithuania tried to solve the issue by discontinuing surveys, or clarifying whether there is coherence between the indicators in different questionnaires. If cases of duplication are detected, i.e. if the same indicator is included in several statistical questionnaires, they are merged together, or the repeating indicators eliminated.

Discontinuation of a statistical survey, i.e. refusing sending statistical questionnaires to respondents without violation of users' interests, is very complicated. This might happen only if the required indicators are calculated on the basis of the available information or using alternative data sources. Statistics Lithuania cancels annual surveys only for those statistical domains, for which short-term statistics are collected and available. However, short-term statistics differ from the annual both in terms of coverage and the level of details. Nevertheless, some entire surveys (questionnaires) were retracted.

Statistical questionnaires, 2003 - 2007



Another measure to reduce the statistical response burden is related to more effective use of information which is already available at the statistical office. For instance, investment of enterprises might be measured using the annual financial accountability data of enterprises. From 2005 and onwards, 25 000 enterprises have been exempt from filling in the annual questionnaire

on investment. Annual investment indicators are to be estimated on the basis of the quarterly investment survey and statistical indicators of structural business statistics.

By the review of statistical questionnaires, Statistics Lithuania also evaluates the relevance of the surveys. Not relevant and recurrent indicators are identified. Up to now such a revision has been conducted within each statistical domain.

In order to better harmonize surveys and avoid the recurrence of indicators, a procedure and instructions for statistical questionnaires and reporting were prepared and approved. In order to reduce response burden on small enterprises which participate in sample surveys, planning of statistical surveys and links between them are considered of great importance.

In Lithuania's case, small enterprises are statistically significant since they make a major share of the Lithuanian enterprises. In some of the activities (e.g. services sector) they account for the absolute majority. The share of enterprises of the relevant size as compared with the total number of enterprises is presented in the table:

No.	Size of an enterprise	Share in the total number of enterprises
1.	Up to 5 employees	≈ 43%
2.	Up to 10 employees	≈ 65%
3.	Up to 20 employees	≈ 79%
4.	Up to 50 employees	≈ 91%
5.	Up to 100 employees	≈ 96%
6.	Up to 250 employees	≈ 99%

Enterprises with 10 employees make up more than half of all Lithuanian enterprises. Every year while making samples, a rotation principle for enterprises is applied. In short-term statistical surveys, one fourth of enterprises is rotated. Thus, theoretically every concrete enterprise should participate in the same survey for 4 years. Unfortunately, it is not always possible to implement this in practice. The problem is related to the fact that Lithuania being a small country accommodates a relatively small number of enterprises. In many economic activities the number of enterprises accounts for few tenths or only a few enterprises. This is very characteristic for industrial activities. For instance, only 6 enterprises are engaged in manufacture of paper and paper products; 15 - in publishing of books, brochures and other publications, 5 - in manufacture of fertilizers and nitrogen compounds. The same can be said about many activities in service and trade sectors. In such cases, rotation is not possible and the same enterprises, even if they are small, are obliged to participate in the surveys and provide information for the statistical office for many years in a row. Besides, in many sample surveys stratification is applied. If, in order to achieve accuracy of an estimate, the number of small enterprises, even if it is not too small, is divided into two or three strata, the critical limit is again reached when a sample covers the majority of the population.

An analysis of the sampling process was conducted by breaking down the surveys according to their importance and focusing on the selection of enterprises. At Statistics Lithuania, samples for all surveys are made simultaneously at the end of the year using the latest information available in the SBR. After completing the sample selection procedure, a list of sampled enterprises is compiled, which clearly identifies how many surveys fall per enterprise. All cases where small enterprises happened to be included into more than 3 sample surveys are separately analyzed, and in case a possibility exists they are replaced with other enterprises.

Wider use of administrative data

No doubts, the biggest reserve for the reduction of the burden on respondents is definitely the use of administrative data sources. The use of administrative data has many other advantages: non response rate is very low, two administrations involved give a positive effect on the data quality, data collection system has a cost-saving effect.

Disadvantages have to be mentioned as well. First of all, it is the dependence on institutions managing administrative data an additional need for the coordination of inter-institutional activities. The possibility to use administrative data sources does not solve the problem itself. Coherence of indicators remains an important question when using administrative sources. It is true that other institutions do not have exactly the same indicators that statisticians need. Statisticians have to calculate indicators they need from the information available. In order to simplify the work, definitions of the basic indicators available in the administrative databases are clarified. Based on these clarifications, a comparative analysis of definitions of indicators used for statistical purposes and those available in administrative databases was conducted, which provided a possibility to identify the indicators that could be used for statistical purposes. The second step in this direction is to carry out calculations and try to analyze differences between statistical information obtained from statistical surveys and analogous information obtained using administrative data sources. All the proposals concerning the relevance of administrative data to be used for statistical purposes are analyzed.

As it was already mentioned, it is not easy to use administrative databases directly due to differences in definitions of indicators, different deadlines for reporting, technical problems, legal and organizational aspects as well as difference between the goals of institutions for which data are collected. In any case, Statistics Lithuania considers the highest priority to find a solution to reduce the response burden, and at the same time to obtain relevant information. Currently, Statistics Lithuania uses 110 administrative data sources from different institutions and organizations. Most of the data are received in an aggregated form (from 71 sources). Data are taken directly from the SSIFB and other institutions databases. 30% of statistical indicators are produced using administrative data sources.

Currently, administrative data are mainly applied in demographic, labour, structural business statistics, short term statistics, foreign trade statistics. Also, data from administrative sources are used:

- for the estimation of non-response;
- as auxiliary information for the calculation of estimates;
- as additional information for the preparation of sample plans;
- for updating of the SBR;
- for the production of flash estimates.

Use of administrative data in Structural Business Statistics

For the compilation of Structural Business Statistics (hereinafter referred to as „SBS’’) several data sources are used. Large enterprises have to fill in a statistical questionnaire (6%). For the rest of enterprises of limited liability, annual financial reports provided to the State Enterprise Centre of Registers are used (20%). For legal persons of unlimited liability (individual enterprises), profit declarations collected by the STI are used. The SBS data for natural persons carrying out economic activities (more than 50% of the total number of entities) are produced using the annual income declarations that are also received from the STI.

Economic entities which have to fill in a statistical questionnaire employ about 60 % of the total number of employees and have about 80% of the total turnover. This share varies from activity to activity.

SBS 2005 Share of administrative sources

Activity	Number of enterprises, %				Number of persons employed, %				Turnover, %			
	Fin. statements	Statistical questionnaire	Profit declarations of ind. enterpr.	Income declarations of physical persons	Fin. statements	Statistical questionnaire	Profit declarations of ind. enterpr.	Income declarations of physical persons	Fin. statements	Statistical questionnaire	Profit declarations of ind. enterpr.	Income declarations of physical persons
Total	20	6	23	51	20	60	13	7	15	78	6	1,1
Industry	21	12	21	45	15	73	10	3	6	90	3	0,3
Energy	32	49	19	0	2	97	0	0	2	98	0	0,0
Construction	19	7	10	65	23	63	7	7	16	78	4	1,5
Trade	18	5	25	51	21	51	19	10	18	72	9	1,0
Hotels&restaurants	34	8	45	13	32	43	24	1	25	58	15	1,3
Transport	35	12	46	7	21	63	16	1	19	72	9	0,1
Telecommunication	55	18	27	0	8	90	2	0	4	95	1	0,0
Business services	24	5	17	55	29	46	12	13	35	52	6	7,1
Social services	9	3	21	67	21	43	16	20	24	62	6	7,2

Preliminary SBS data are produced only on the basis of data collected through statistical questionnaire. The rest are estimated using weights of the missing enterprises (turnover) from the SBR. Preliminary results are available after 8 months of the reference year. The final SBS data are the census data, for which instead of estimations the data from administrative sources are added to the data received from statistical questionnaire. Statistics Lithuania gets all administrative data at the individual level. A single identification number of enterprises allows combining different data sources. As far as administrative sources do not have all data needed for SBS, various imputation methods are applied. Final SBS data are available after 18 months of the reference year.

Advantages of using administrative data in SBS: full coverage of active enterprises, including the small ones, possibility to make all kinds of aggregations (by activity, ownership, size class, regions), better quality of the survey results, reduced response burden on small enterprises, updating of the SBR.

Use of administrative data in Annual Survey of Earnings

Another example is an annual survey on earnings which is complicated and time consuming for respondents. However, this survey is very important for both internal and external users. A methodological solution was searched for finding the ways to calculate annual indicators based on the short-term statistics and information from administrative data sources. For the quarterly earnings survey, the sample was increased in order to increase the representativeness at the regional level. After analyzing information in the database of the SSIFB, a methodology for its use was prepared.

Introduction of administrative sources for estimation of data of Labour Statistics is surely a useful experience. Burden on enterprises as well as for the staff of Labour Statistics Division was significantly reduced. Approximately 40000 of enterprises do not need to fill in the annual questionnaire on earnings, the staff of Statistics Lithuania do not need to enter and check those questionnaires and users are able to get information sooner than they used to. Useful experience

allowed starting an analysis of possibilities to introduce administrative sources in other surveys on earnings.

Use of administrative data in Short term statistics

A comprehensive comparative analysis of Labor indicators in Short term statistics (Wages and salaries and Number of employees) received from statistical surveys and relevant indicators available in the SSIFB database was conducted. The exercise resulted in a decision that the indicators of short-term Labor indicators can be based on administrative sources. Since 2007 collection of these indicators from enterprises was discontinued; it led to the simplification of relevant statistical questionnaires. Despite some problems caused by differences in definitions (which were solved methodologically), the quality of statistical indicators is good. It was proved by many calculations and data analysis.

The positive result of the action was that more than 16000 of enterprises received shortened questionnaires. But on the other hand, some “negative effect” was noticed. The enterprises used to fill in data on 3 labor indicators: number of employees, wages and salaries and hours worked. When the first two were deleted, more erroneous data occurred on the third indicator – hours worked. It requires conducting the data editing procedure more carefully.

Interinstitutional data warehouse (DW)

Statistics Lithuania participates in establishing an Inter-institutional Tax Data Warehouse (hereinafter referred to as „the DW”). The DW is a state information system established for the tax data collection and preparation of relevant and quality information based on the collected data and timely provision of information to the parties of the agreement. The joint order on the foundation of the DW was signed by six institutions (STI, SSIFB, Customs, Statistics Lithuania, Ministry of Finance, Ministry of Economy, Information Society Development Committee).

Presently, the STI DW workplaces are established at Statistics Lithuania and other institutions enabling their DW users to access data. All information systems ensure data protection from possible external damage by software, organizational and physical measures. Users are provided with information following the laws and legal acts regulating the provision of such data and disclosure of confidential information. The project resulted in practical useful outcome giving a possibility to all interested parties to exchange information. It also contributes to the alleviation of the burden on respondents.

Electronic data collection

Creation of a favourable, user-friendly, time-saving, cost-efficient and secure environment for respondents is one of the major projects initiated at Statistics Lithuania. Certain technological solutions were proposed. First of all, it is related to the introduction of electronic questionnaires.

Statistics Lithuania started the work on the introduction of such questionnaires in 2001. Starting from 2003, the system started to function in real conditions gradually introducing new statistical surveys in an electronic form. Vigorous changes in information technologies stimulated virtual reorganization of this system in 2006, thus resorting to the *ABBY eFormFiller* software means for filling in statistical forms. The software had been used for filling in tax declarations, thus it was already familiar to respondents.

A practical and convenient possibility was opened up in a specialized website *e-formos.stat.gov.lt* to find prepared electronic forms, schedule for their submission and information on data transfer via Internet.

Up to now, 63 electronic statistical questionnaires (40 % of the total number of statistical questionnaires) have been introduced, respondents could submit statistical forms directly from their working places. In 2007, about 10 thousand (18 % of economic entities sampled for the surveys) respondents expressed a wish to provide statistical reports electronically.

The advantage of the system is that not only the procedure for the submission of statistical forms was simplified, but also the statistical reporting burden on respondents was reduced in general. The number of electronic forms will be gradually increased.

A survey of respondents, potential providers of statistical data using e-forms, was conducted. It was targeted at the assessment of advantages and shortcomings of the system in use, as well as - detecting the reasons why the system had not been used. Based on the survey results, more efforts were directed towards further development and advancement of the system and promotion of transmission of data by electronic means.

Use of business accounting data

A number of statistical tools and methods which would facilitate meeting statistical obligations have been promoted in cooperation with Eurostat and NSIs recently. In the Communication from the Commission to the European Parliament and the Council on reduction of the response Burden, Simplification and priority-setting in the field of Community statistics (2006) the Commission refers to a number of areas suggested for the simplification in the short run, such as switching to a single flow reporting system in Intrastat, promoting use of administrative data, reengineering of business statistics, agricultural statistics, etc. It also speaks about the simplification tools, one of them being the use of business accounting data. It says the following: "Direct use of standardized accounting data using electronic tools is another promising way forward. The data are to be received from the enterprise accounts without a special questionnaire and avoiding the burden on respondents. The accounting data can be supplemented by small surveys as in the case of administrative data. The Commission will support the MSs by aligning the definitions in the legal acts on those in accounting standards, by developing the necessary electronic tools, and by deciding methods of targeted complementary surveys and corresponding estimation techniques". The latter subject has been tackled in a number of cases during the recent years in international forums, at the Community level as well as in the Members States (hereinafter referred to as „the MS“). It was discussed in the session on Improved Data Reporting organised by the Conference of European Statisticians in 2005 in Geneva, UNECE/Eurostat work session on Electronic Data reporting, in February 2002, etc. NSIs have been also working in this direction, i.e. undertaking projects on data extraction from business accountancy systems.

Statistics Lithuania among other measures which contribute to the reduction of the response burden has also taken on board the issue on the use of business accountancy data and, in relation to this, carried out a study financed from the Structural Funds in order to clarify whether it is feasible to implement a project related to the automated statistical data extraction directly from the business accountancy systems. The project should be aimed at creating a public e-service for business entities which would give a possibility to extract automatically data accumulated in different financial, wages and salaries, goods and other business accountancy systems using electronic ways. It would also contribute to increased data reliability; shortened data transmission time; increased

satisfaction of business entities; decreased costs for data transmission; reduced burden on respondents. Within the feasibility study an analysis of technological possibilities was conducted which showed, that appropriate technological solutions have to be adopted on both sides, i.e. in business as well as at the statistical office. Besides, the focus was given on the revision of statistical questionnaires.

It should be noted that a number of different issues was pointed out, which are to be dealt with before the start of the project. Based on the analysis made, it became clear that only less than 30% of the data can be extracted directly from the business accountancy systems and could be used for statistical purposes. In some cases they should be slightly modified to meet the statistical requirements. Whereas, over 40% of data cannot be automatically transmitted from the accountancy systems for statistical use due to differences in definitions, classifications, etc. Currently, over 20% of the data are produced by business enterprises specially for statistical purposes. Granted that 30% is not a sufficient and significant figure, it might be a good opportunity and achievement to start using accountancy data directly. On the other hand, the percentage could be increased by extensive cooperation between statisticians and international accounting authorities pursuing the same goal of reducing the burden on respondents and harmonising accountancy and statistical standards, especially taking into account that the MSs use the same international accounting as well as statistical standards. With this in view, NSIs efforts are not sufficient, and more systematic approach and measures should be targeted at the Community level as well.

Results achieved and next steps

To implement measures aimed at the increase in the efficiency of official statistics management, the activities have been focused on the reduction of the statistical reporting burden and strengthening the partnership with respondents. Following the annual practice, Statistics Lithuania released information to the respondents, indicating statistical surveys in which a certain enterprise will have to participate during the coming year, which statistical reports it will have to fill in, which novelties were introduced. All statistical questionnaires and instructions for filling them in were placed on the website of Statistics Lithuania. This makes it easier for the enterprises to produce the required statistical data, plan the time to be spent for providing filled-in statistical questionnaires, and get acquainted with the requirements set for electronic reporting. In 2006, the response rate accounted for 94.6 per cent (against the planned 93%). The indicator is an integral part of quality statistical information.

Taking into consideration requests of respondents, Statistics Lithuania actively co-operates with business entities providing them with statistical information on the key performance indicators and changes in a relevant sector as well as the place of a particular enterprise within the economic activity. Activities related to the improvement of statistical sample surveys and using more administrative data sources were targeted to avoid the increasing statistical reporting burden on economic entities. The average number of statistical surveys per respondent (economic entity) made up 3.6 surveys in 2006, whereas the plans stood at 3.7 surveys.

Participation of economic entities in statistical surveys

	2005	2006	2007
Active economic entities	72330	73344	76516
Selected for statistical surveys	46717	45025	45384
Share of total active economic entities, %	65	61	59
Number of statistical forms approved	187	175	157
Average number of statistical surveys per respondent	3,7	3,6	3,5
Response-rate of economic entities, %	93	94,6	95 (plan)

In order to distribute statistical response burden evenly, reduce it, or even influence decisions concerning the provision of certain statistics, the statistical response burden has to be assessed. Therefore, Statistics Lithuania requested the respondents to indicate how much time they spent on the preparation of statistical data and filling in of the statistical questionnaires. Based on the information obtained and guided by the response burden calculation methodology prepared by methodology specialists of Statistics Lithuania, a response burden indicator was assessed for the first time. One respondent spent on average 13.3 hours for the needs of statistical accountancy.

The use of administrative data continues to remain an important factor for the reduction of the response burden. Therefore, the implementing measures of the Programme of the Government of the Republic of Lithuania for 2006–2008 included the establishment of an Inter-Institutional Commission which would analyse and coordinate activities related to possibilities of using administrative data sources for data collection, coherence, use and exchange them between institutions and in case of necessity, provide the Government with the relevant proposals. The creation of such a Commission was initiated by Statistics Lithuania.

The use of administrative data sources for statistical purpose will bear continuous importance as a factor reducing the response burden on respondents. Taking into account the statistical response burden indicator (calculation of which based on enterprises time costs has been started in 2006), procedures for excessively time-consuming statistical reports will be improved, aiming at their simplification.

EU programmes

The Action Programme for Reducing Administrative Burdens in the European Union set in the Communication from the Commission to the Council and the European Parliament, the European Economic Committee and the Committee of the Regions hold out the prospect of making a significant contribution to improving the business climate in the EU. In the explanatory memorandum of the draft Decision for the MEETS programme the Commission describes the programme as follows: “Business and trade statistics are faced with major challenges in the coming years: the strong integration of world markets is changing the structure of the economy and also European integration is progressing further. Statistics have to react to these trends and have to describe new phenomena. At the same time, the administrative burden on businesses must be reduced. These two challenges are not contradictory to each other. The need for new types of indicators can go hand in hand with efforts to modernise the production system. New types of indicators providing the needed information could be obtained by using new sources of information and by linking already existing types of business statistics without the necessity to increase the reporting burden on the enterprises. But an investment is needed to exploit the potential synergies efficiently, this is the main reason for this programme.” We do believe that coordination of different

burden reduction measures on the Community level will allow achieving tangible results in the MSs.

Conclusions

Statistics Lithuania initiated many projects and measures to reduce the burden on respondents, a number of new projects are in the pipeline. However, not every measure resulted in tangible outcome and the results not always met the expectations. Hard work and efforts taken during the recent years made us understand that it will never work if the efforts are not joined together between NSIs, Eurostat, international organizations. A more active approach should be also taken in terms of cooperation with policy makers.