

**Distr.
GENERAL**

**CES/SEM.50/6
15 April 2003**

Original: ENGLISH

**STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR
EUROPE**

**STATISTICAL OFFICE OF THE
EUROPEAN COMMUNITIES (Eurostat)**

**CONFERENCE OF EUROPEAN
STATISTICIANS**

**Joint UNECE/Eurostat Seminar
on Business Registers
(Luxembourg, 25-26 June 2003)**

**THE USE OF ADMINISTRATIVE DATA SOURCES FOR THE UPDATING
OF THE POLISH STATISTICAL REGISTER**

Invited paper submitted by Central Statistical Office of Poland*

Summary

According to the Law on Official Statistics, the Central Statistical Office (GUS) runs the National Official Register of the National Economy Units (REGON). Tasks of this official register are: to provide service for units i.e. to allocate identity numbers to the REGON register on the basis of the applications that have been submitted, to deliver certificates on the allocated REGON identity numbers; to run the REGON register as the computerised database; to provide public administration bodies and others interested parties with data from the register.

* Prepared by Doroslawa Kurek, Deputy Director of Co-ordination of Statistical Surveys Division, Central Statistical Office of Poland.

For full survey service, in particular to make possible a wide information scope and in order to use all accessible data sources (in particular the surveys results), GUS also runs the statistical register (BJS).

The BJS statistical register has been created as a sampling frame for all surveys conducted by GUS on the population of Polish national economic units. It contains descriptions of enterprises (including enterprises which are not employers of hired persons), the public sector units and the non-profit units. The BJS statistical register contains information from the REGON register and supplementary information both from surveys and other sources, including administrative information systems.

The legal basis for the creation of the statistical register is the Law on Official Statistics. The data on economically active, legal units, which are accumulated in the BJS statistical register from sources other than the official REGON register, are subject to special protection as individual data, in accordance with the principle of statistical confidentiality from the Law on Official Statistics

The major source of information for the statistical register is the REGON register. Entering new legal units or deleting them is based only on the REGON register. Other sources of information are statistical surveys, administration systems of the Ministry of Finance and Social Security Service, other registers and lists, the official journal "Court and Commercial Gazette", information from the units, postal information, and press information.

The use of administrative sources for updating data from the statistical register is a priority for each statistic. The work on this task has been conducted by the Central Statistical Office on the basis of the development of projects creating central databases within the information systems of the Ministry of Finance and Social Security Service. These systems have been created taking into consideration their consistency with the REGON register.

The data obtained from the Ministry of Finance in 2002 made possible the following: the updating of the economic activity of 141911 units (4,2% of the active units) in the statistical register, a change of rules of the sample selection for the survey of enterprises that have the number of employees < 10, the beginning of work on the completion of non-responses in surveys which statistical services conduct on the basis of data from the tax system.

Increasing the use of administrative registers in statistical surveys also is one of the key priorities of Polish official statistics because it will make possible its ability to improve the quality of statistical surveys, to increase its statistical output scope, to better meet the needs of users and to gradually decrease survey costs.

I. INTRODUCTION

1. According to the Law on Official Statistics, the Central Statistical Office (GUS) runs the National Official Register of National Economy Units (REGON). The tasks of this official register are: to provide service for units i.e. to allocate identity numbers to the REGON register on the basis of the applications that have been submitted, to deliver certificates on the allocated REGON identity numbers; to run the REGON register as the computerised database; to provide public administration bodies and others who are interested with data from the register.
2. The rules on running the REGON register (sources of data updating, information scope) are defined in the Law on Official Statistics and in the 1999 Decree of the Cabinet on the way and methodology of running and updating the register of the national economy units. These also describe detailed terms and procedures of cooperation between services of official statistics and other units running official registers and information systems of the public administration. The REGON register is described in greater detail in Mr. Andrzej Bajor's paper "The role of REGON register in the Country Information System."
3. For full survey service, in particular to make possible full information scope and in order to use all of the accessible data sources (in particular the surveys results), the Central Statistical Office also runs a statistical register (BJS). The legal basis for the creation of the statistical register is Article 10 of the Law on Official Statistics, which authorises official statistics to use individual data accumulated in official statistics surveys to create a sampling frame for the purposes of the statistical surveys conducted by the services of official statistics. The statistical register has the following tasks: creating sampling frames for specific statistical surveys; creating lists for surveys; the description of units, which are subjects of statistical observation, for full and correct development of results.

II. GENERAL INFORMATION ON BJS STATISTICAL REGISTER

4. The BJS statistical register has been created as a sampling frame for all surveys conducted by the Central Statistical Office (GUS) on the population of Polish national economic units. It contains descriptions of enterprises (including enterprises which are not employers of hired persons), the public sector units and the non-profit units. The BJS statistical register contains information from the REGON register and supplementary information both from surveys and other sources, including administrative information systems.
5. As they are individual data, the data on economically active legal units which are accumulated in the BJS statistical register from sources other than the REGON register are subject to special protection, in accordance with the principle of statistical confidentiality from Article 10 of the Law on Official Statistics. Data protection is ensured by limited and controlled access to the BJS statistical register. Access to data is granted depending on the scope of responsibilities of the individuals using the BJS register.

6. The BJS register contains units which are subjects of statistical observation with the exception of individually owned farms. The statistical register has been developed on the basis of units registered in the REGON register. There are the following units in the BJS statistical register: legal units – 3776791 (including natural persons carrying out economic activity), their local units – 4233743 (registered in REGON), and in 2003, for selected domains, local kind of activity units will be entered into the BJS statistical register. As a rule, these units are not registered in the REGON register; however, they are subjects of surveys.

7. The enterprises are treated as an organised activity of legal units and the distinct identity numbers of the BJS are not given to them. In order that the legal units may have the enterprise, the following conditions have to be met: at least one local unit must exist and have > 0 employees (this means that the legal unit is economically active), appropriate kind of activity according to NACE, appropriate special legal form (units which have some special legal forms cannot create enterprises i.e. self-government organisational units financed by the government budget).

A comparison between subjective scopes of REGON official register
and BJS statistical register

REGON	BJS
Legal persons Organisational units without legal status Natural persons carrying out economic activity	Legal units including natural persons carrying out economic activity (excluding agricultural activity)
Local units of the legal units mentioned above	Institutional units (legal units which are classified according to the algorithm)
Farmers, which are agricultural farms owners	Enterprises (acc. to algorithm)
	Local units
	Local kind of activity units (LKAU) (in selected domains)

The subject units in the statistical register are described through:

- characteristics included in the REGON official register (some of these characteristics are updated with data from different sources);
- additional characteristics based on data from surveys or administrative systems – which are necessary for creating the lists of units for surveys and making possible data aggregation in accordance with the survey programme;
- additional classifications based on the characteristics mentioned before.

8. There are also attached the validity dates for characteristics and the sources of data (for characteristics updated on the basis of different sources).

9. The scope of characteristics of the BJS statistical register is as follow (characteristics, for which the official register (REGON) is the only source, are in italics):

identification characteristics

- *14-digit REGON identity number;*
- *unit name;*
- *abbreviated name.*

address characteristics

Address of registered office/seat of activity/ Mailing address: voivodship (NTS2); subregion (acc. to NTS3); district (powiat – NTS4); commune (gmina – NTS5); town/ village, street, street number, apartment/ office/ room number, post office, post code; telephone – area code, direct number, extension number; fax - area code, direct number, extension number; e-mail.

registration characteristics (which are important because of the many original registers)

- *registration authority;*
- *type of register/ record.*

legal characteristics

- *basic legal form;*
- *special legal form.*

characteristics of public sector units

- *form of financing;*
- *founding/ supervisory body.*

legal and economic status

ownership

type of activity

- types of activity acc. to Polish Classification of Activities (PKD) - all from REGON register and surveys (according to subclass);
- *main activity acc. to PKD, entered in REGON register (subclass);*
- main activity determined on basis of structural surveys.

size of the unit

- total number of working people – constant, as of 30 November each year in the BJS statistical register;
- size category for the purposes of determination of obligations (based on total number of working people which is constant);
- number of employees;
- turnover (obtained from surveys);
- revenues (obtained from tax system);
- number of local units.

connections with tax system

- Tax Identity Number – NIP from National Tax Payer Registration Database - KEP ;
- type of accounting records or carrying out economic activity by non-profit units.

other information on local unit

- *is the local unit the seat of management*, (if the local unit is the seat of management)
- total farm area and arable land area.

additional information

- definition of reporting obligations for a given year and participation in voluntary surveys;
- indication of participation in the survey or reasons for non-response.

III. UPDATING OF THE BJS STATISTICAL REGISTER

10. The major source of information for the statistical register is the REGON register. Entering new legal units or deleting them is based on the REGON register only. Other sources of information are statistical surveys, the administration systems of the Ministry of Finance and Social Security Service, other registers and lists, the official journal “Court and Commercial Gazette”, information from the units, postal information, and press information. These sources are classified and coded in the register.

11. The register is updated monthly and the data for the updating process are prepared systematically during a month. The main sources for this process are the changes entered into the REGON register. Other sources are survey results, which have been attached for the updating of the BJS, and information obtained by statistical offices (from 16 voivodships) and entered by Intranet application BJSKOR (130 persons have the access to the BJS statistical register through this application). During the entry of the changes, their sources and validity dates are given.

12. The possibility of collecting the data from the tax system of the Ministry of Finance and from the Social Security System is very important for official statistics. These sources

contain basic information on the total number of units. The statistical surveys based on data collected by means of questionnaires, in some cases are not able to observe the whole population. Considering this fact, there has been work on automating the data updated from the tax system and toward completing the statistical register over the last three years. The aim of this work is to improve the quality of the BJS statistical register and as a result of this process, the improvement of the quality of surveys. This includes, among others, working out the methods of imputation data in the case of non-response for complete surveys, the improvement of results precision in sample surveys and the development of estimation methods for small domains.

13. The Ministry of Finance can provide to the Central Statistical Office the data from the tax system under the following conditions:

- there is a legal base for the use of these data for statistical purposes;
- uniform computing system (the best would be the central system) for tax data exists;
- the information scope of the BJS statistical register and tax system are consistent.

IV. THE ANALYTICAL WORKS ON THE USE OF THE TAX SYSTEM AS A SOURCE OF DATA FOR THE UPDATING OF THE BJS STATISTICAL REGISTER

14. The Central Statistical Office (GUS) is provided with data from the National Tax Payer Registration Database (KEP) of the Ministry of Finance on the basis of the 1999 Decree of the Cabinet on the way and methodology of running and updating the register of national economic units. This includes as well detailed terms and procedures of cooperation between services of official statistics and other units running official registers and information systems of the public administration. Additionally, the other systems supply the Central Statistical Office with information on tax and cost on revenues, financial results and income tax (both for legal units and for natural persons carrying out business activity) on the basis of the Decree of the Cabinet on the Programme of Statistical Surveys of Official Statistics.

15. The following works are necessary in order to prepare the rules for supplying the statistical register with data from the tax system:

1. Comparison of the data sets of KEP database and the BJS statistical register concerning the subject units.
2. Equipping the BJS register with the NIP numbers which are identity numbers for the tax system.
3. Explaining differences in subject units between KEP and BJS records.
4. Preparations of adequate information (notes) in the Programme of Statistical Surveys of Official Statistics.

5. Comparative analyses of revenues in the tax system and statistical surveys together with a determination of reasons and range of differences.

16. **Ad.1.** The BJS statistical register is consistent with the REGON official register with reference to the legal units, therefore the differences in subjects units are analyzed between the KEP database and the REGON register. There has been work on the comparison of subject units since November 2000. The Central Statistical Office provided the Ministry of Finance with the target list of units, which were the sampling selection for surveys concerning enterprises that have the number of employees < 10 in 2000 (114 thousands of units). The aim was to check the activity of these units (if they have existed in KEP database). About 70 thousands of units were found in the KEP database. The analysis of results proved that 16 thousands of units (14 %) ceased their activity despite being marked as active in the REGON register. In December 2001, during the preparatory work for next year's survey of enterprises having the number of employees < 10 persons, the Central Statistical Office again provided the Ministry of Finance with the target list of units, which were the sampling selection. There were 114 630 units including 82 395 natural persons carrying out economic activity and 32 134 legal units. 394 natural persons and 9 203 legal units were not found.

Year	Total number of sample units	Number of sample units, which were not found in KEP database
2000	114 000	44 000
2001	114 630	9 597

17. As the above table of results shows, the methods of subject units comparison have been improved.

18. In 2002, the Ministry of Finance provided the Central Statistical Office with the entire scope of the KEP database - both the legal units, organisational units without legal status and natural persons carrying out economic activity. The following are the results of the units comparison:

Specification		Number of units	
		Total	Consistent
Legal units	KEP	642 486	531 514
	REGON	744 032	
Natural persons carrying out business activity	KEP	3 097 4782	2 301 673
	REGON	682 623	

19. **Ad.2** The KEP database contains a part of its units without REGON identity number (in spite of a rule of entering a REGON identity number to the tax declarations). Particularly, it concerns the natural persons carrying out economic activity, for whom the basic identity number is the PESEL number from General Electronic System of Population Registration. The work on comparison of subject units was finished together with creation of linked pairs of numbers REGON-NIP. This set of data was applied for providing the statistical register BJS

with a Tax Identity Number – NIP from the KEP database and afterwards was sent to Ministry of Finance in order to complete or verify REGON numbers in the KEP database. The number of units with the linkage REGON-NIP is as follow:

Specification	Number of consistent units	% of units in BJS equipped with NIP identity number
Legal units	531 514	71,4%
Natural persons carrying out business activity	2 301 673	85,8%

20. **Ad.3.** Deficiencies in links between the REGON and NIP identify numbers are caused by many reasons. Up to now, some of them have been identified:

- units which have registered themselves in the REGON register (including natural persons carrying out economic activity) but which haven't started their activity. They haven't submitted a tax return, and they don't exist in the KEP database;
- errors which occur in entries of the REGON and PESEL identity numbers. These make it difficult to create links between units from the KEP database and the REGON register automatically;
- units which have been moved out from the REGON register but which still submit tax returns (the basis for such a decision needs explanation);
- cases where the personal company has been registered in the KEP database while its partners have been registered as natural persons carrying out economic activity in the REGON register.

21. The voivodship statistical offices and tax offices take part in detailed explanation of mentioned differences. Each voivodship statistical office has divisions (country branches), that makes easier direct cooperation of official statistics and tax offices.

22. **Ad.4.** The Ministry of Finance runs centralised databases that contain the income tax returns for natural persons and legal units. Therefore, there is the possibility of use of all data sets of tax payers without technical drawbacks. According to the statistical law, the information on the use of the particular administrative source must be located in the Programme of Statistical Surveys of Official Statistics. In the Programme of Statistical Surveys 2002, an obligation to provide the data from Ministry of Finance tax system for Central Statistical Office was introduced. In mid-1999, the Central Statistical Office was provided with data from 1999 and 2000 for two selected voivodships in order to conduct methodological work. In the end of 2002, the Central Statistical Office was provided with the whole range of data from 2001.

23. **Ad.5.** The data from the tax system are derived from two databases:

- personal income tax database (PIT), which contains information taken from personal income tax return forms PIT;
- corporate income tax database (CIT), which contains information taken from corporate income tax return forms CIT and from income tax return forms for organisational units without legal status.

The kind of tax return form	The number of units		% of units consistent with BJS
	Total in particular files	Consistent with reference to BJS	
CIT	225 095	217 498	96,6%
PIT	1 258 231 ^{*)} among this, units, which are in KEP database marked as natural persons carrying out business activity: 1 022 855	951 345	75,6% 93,0%
<p>*) The differences between the number of units described by data from the PIT database and the number of taxpayers identified in the KEP database are the result of:</p> <ul style="list-style-type: none"> - settlements of personal companies partners according to the same rules as natural persons carrying out economic activity, - the obtained data set contained only the data of taxpayers who have settled according to general rules. Additionally, there should be information provided on persons who settle their revenues according to the rate of lump-sum tax. 			

24. A comparative analysis of units from the survey of enterprises which have the number of employees <10 and submitted reports in 2001 was conducted. A correlation of two variables was investigated: the revenues from this survey and the revenues from the Ministry of Finance databases (the common part of these sets consisted of 26414 records). The correlation between these variables was 79%. In most of cases, the analysis of distribution of relative error proves a very high quality of information on the revenues, taken from two accessible sources. The relative error of more than 10% (in relation to absolute value) occurs in about 16% of the analysed units; for about 79% of analysed units, this error is less than 3%. The maximum relative error, which is equal in relation to absolute value = 100%, occurs for about 0,3% units and it means that revenue according to information taken from questionnaire or from data of the Ministry of Finance, has zero value.

V. THE USE OF DATA FROM THE TAX SYSTEM IN 2002 AND 2003

25. The data from the tax system was applied to the analysis of the activity of units from the statistical register. The units have been divided into two groups: (1) units which were on

the basis of the BJS statistical register, the KEP database and report files described as non-active, and (2) units which were not active potentially and the fact of their activity must be verified during the surveys.

26. In order to define non-activity, the following rules were established:

- a) non-active units, for which a final cessation of activity is marked in the KEP database;
- b) potentially non-active units, which haven't been found in the KEP database and additionally which fulfil the following criteria:
 - the date of starting activity < 1997;
 - the date of any updating in the BJS statistical register < 2000;
 - the units didn't submit the reports in 2000 and 2001.

27. The number of units, which were marked as non-active on the base of:

- KEP database – 118502;
- tax system – 23409.

Total number is 141911 – 4,2% of active units

28. Data from the tax system was applied in order to change the rules of sampling selection in the survey of enterprises that have a number of employees < 10. On the basis of information obtained from the Ministry of Finance, a sampling frame was verified and divided into two parts: one part, which contains 2873601 units selected on the basis of the BJS statistical register and recognised as active according to information from the Ministry of Finance, and a second part, which contains 139869 non-active units according to Ministry of Finance information. The units that have non-standard, extremely high revenues were derived from the sampling frame. On the basis of analysis of “revenues” variable distribution, 9699 units with revenues comprising 39,2% in relation to the total amount of revenues (according to accessible information), were selected in order to be examined and attached to the rest of drawn units. The rest of active units was drawn according to algorithm which was created in order to ensure the same precision of “revenues” variable for different kinds of activity.

29. These efforts are expected to considerably improve the quality of this survey's results.

VI. CONCLUSIONS

30. The fact that the statistical services run the REGON register is very important for the quality of Polish official statistics surveys. This register has been designed in such a way that it meets a considerable part of statistics needs. The beginning of activity in the case of each legal unit follows a registration of this unit in the REGON register. It ensures the completeness of legal units in the statistical register, which is updated in this scope on the basis of the REGON register. For the improvement of surveys quality, while intensive

changes of economical conditions take place (information on cessation of units activity is not complete), there is the need to look for every possible source for updating the description of the units in the register.

31. The ideal to achieve this aim is to obtain the data from administrative sources, especially from the Ministry of Finance tax databases and from the Social Security Service database. It is in the interests of the units to declare their activity to these administrative systems and to update their characteristics there. The possibility of obtaining up-to-date information on units activity from these systems makes easier the elimination of overcoverage in the register (in full-scale surveys in order to decrease the costs of conducting the surveys and in sample surveys for increasing the precision of the survey results).

32. Current data on the number of employees and amount of revenues makes easier the creation of unit stratification by size and ensures better precision of surveys. Because of non-responses, the use of information from administrative systems in order to complete this lack of information is important. For this reason, the use of administrative data for updating data from statistical registers is a priority for the statistics.

33. The work on this task has been conducted in the Central Statistical Office on the basis of:

- statistical regulations which ensure that the official statistical services have the possibility to use administrative records for statistical purposes;
- the implementation of central databases creation projects within the information systems of the Ministry of Finance and the Social Security Service. These systems have been created taking into consideration their consistency with the REGON register;
- the implementation of a statistical register building project in the Central Statistical Office.

34. The data sets obtained from the Ministry of Finance in 2002 made possible:

- the updating of the economic activity of 141911 units (4,2% of the active units) in the statistical register;
- a change of rules of sample selection for the survey of enterprises that have the number of employees < 10;
- the beginning of work on the completion of non-responses in surveys, which statistical services conduct on the basis of data from the tax system.

34. Increasing the use of administrative registers in statistical surveys is one of the priorities of Polish official statistics.

35. As soon as the first large system, the General Electronic System of Population Registration (PESEL), had been implemented in Poland, the Central Statistical Office started work concerning the use of administrative data sources for official statistics. The data from this system were used for the National Population and Housing Census in 1988 and for surveys on migration. In 2002 the data from the PESEL system were used, among others, for the development of the National Population and Housing Census data 2002, and for

population surveys, vital statistics surveys, internal and international migration surveys and in demographic projections.

36. In 2002, 30% of statistical surveys were enriched with information resources that derived from administrative systems, mainly in domains such as: environmental protection, national accounts, finances of enterprises, labour market. Some surveys, especially the surveys concerning environmental protection are based primarily on administrative records.

37. Increasing the use of administrative data will make it possible to improve the quality of statistical surveys, increase the statistical output scope, better meet the needs of users and result in a gradual decrease of surveys costs.

GLOSSARY

GUS - Central Statistical Office

REGON - the National Official Register of the National Economy Units

BJS - the statistical register

KEP - National Tax Payer Registration Database KEP

NIP - Tax Identity Number – NIP from National Tax Payer Registration Database - KEP

PIT - Personal Income Tax Database,

CIT - Corporate Income Tax Database,

PESEL - General Electronic System of Population Registration.
