

CONFERENCE OF EUROPEAN STATISTICIANS

UNECE Work Session on Statistical Data Editing
(27 – 29 May 2002, Helsinki, Finland)

Topic (iii): Editing of administrative data

USE OF ADMINISTRATIVE DATA BY THE CZECH STATISTICAL OFFICE

Contributed paper

Submitted by the Czech Statistical Office, Czech Republic¹

Abstract: During the last decade Czech statistics has undergone a large transformation process to harmonize all activities of the Czech Statistical Office [CZSO] with international standards, concepts, method and best practices. The top priority was the compliance with the Acquis and other requirements of the European Statistical System as a basic condition for the accession of the Czech Republic to the European Union. This process is supposed to be completed before 2003. Administrative data and their use have been one of the basic priorities of this harmonization process. The activities in this area can be broken down into three main topics:

- 1) Evaluation of administrative data sources for purposes of the CZSO;
- 2) Use of administrative data sources for the Business Register and other registers;
- 3) Use of the administrative data sources as a substitution of statistical surveys.

I. INTRODUCTION

1. Since the beginning of the 1990s, a long-term transformation process has been in progress in the CZSO, which reflects the political, economic and social changes throughout society. Similar processes are under way in many countries of central and eastern Europe. All changes of statistical principles and tools should correspond not only to changes in society in the Czech Republic, but also to the tasks of harmonizing the state statistical service in the Czech Republic with regulations, standards and practice of the statistical service of the European Union. This is related to the process of preparing the Czech Republic for accession to the European Union. One of the important priorities of the transformation process in statistics is the use of administrative data sources. In preceding decades, the then Czechoslovak statistical service used and provided almost exclusively data from its own statistical surveys. The use, if any, of statistical data from other sources was of a quite different nature and reflected the needs and requirements of a centrally planned and controlled state economy. At that time the Czechoslovak statistical service had to observe strict regulations on the protection of state, economic and military secrecy, but the principle of statistical data confidentiality was non-existent. Owing to the assistance given by Eurostat and experts of many NSIs of the EU member states, the CZSO was successful in the implementation of methods and procedures of handling administrative data sources.

II. USE OF ADMINISTRATIVE DATA SOURCES IN PRACTICE OF THE CZSO

II.1. Evaluation of administrative data sources for purposes of the CZSO

2. Until now, the CZSO has established several criteria and views of evaluating administrative data sources with respect to their qualities and ways of use. This applies mainly to defining the source of administrative data, evaluating the qualities of both the source and data contained and, further, laying down the purpose to which a specific administrative data source can be used.

¹ Prepared by Stanislav Drapal (drapal@gw.czso.cz).

3. While defining the administrative data source, the CZSO principally distinguishes between:
 - data already produced for other than statistical purposes. They may be used by the CZSO and are available at reporting units. They are particularly accounting and taxation data, social security data, data intended for other central government authorities, etc. These data are located among enterprises, non-profit institutions, etc., and need not have been produced for the needs of the CZSO only. However, they cannot be taken over from another central government authority that has collected them, because e.g. legislation on tax administration does not allow this; this legislation in the Czech Republic prevents tax administrators from providing anyone with the contents of tax returns. In such a case the CZSO asks reporting units to provide data already produced for other purposes. Reporting units can put these data down in the CZSO questionnaire or use electronic data collection. To this end, reporting units receive from the CZSO a diskette with installation and data capture software, which takes data over from the reporting unit's records, stores them in electronic form of the CZSO questionnaire and makes basic plausibility checks. The data can be supplied to the CZSO on a diskette or sent by e-mail;
 - data from various state-owned organizations and their databases. These data have already been collected from reporting units and are stored in databases. The CZSO takes over either sets of individual data on individual reporting units or aggregate data structured as agreed in advance between the CZSO and the given database operator. This second type of administrative data source has precedence in the CZSO over the first one if it is allowed by legislation. It mainly applies to the use of some identification data on taxpayers (legislative restrictions) or complete data on social security payers (no legislation impeding the use by another central government institution).
4. In case of the first indent above, the CZSO can decide about the scope and character of the basic set and the sample of statistical units. In case of the second indent, the CZSO has available only sets already created by the administrative data source, but taking data over from databases is usually faster, less expensive and represents no burden on reporting units.
5. Every defined administrative data source is evaluated according to the following criteria:
 - legislative conditions of use of the data source;
 - technical conditions of use of the data source;
 - comparing definitions of the contents of indicators in the administrative data source with definitions of the CZSO;
 - judging dates of getting data from the administrative data source;
 - judging stability of the data source with respect to time horizon of the data use by the CZSO;
 - the level of coverage of required area of statistical evaluation;
 - quality of data of the administrative data source;
 - the scope and ways of using the administrative data source for the needs of the CZSO;
 - comparing the cost of using the administrative data source with the cost of a classic statistical survey;
 - considering other circumstances (transparency of the data source, its freedom from political interference, good image, history, etc.).
6. Having defined and evaluated the administrative data source, the CZSO lays down the purposes of its use. These purposes can be divided into the following two groups:
 - primary data source of producing statistical information, where the administrative data source substitutes a statistical survey or its data are combined with data from statistical surveys taken by the CZSO. For example, data on budgets of municipalities are not based on a CZSO statistical survey but taken over in the required factual structure from the Ministry of Finance;
 - data source for the creation and updating of statistical tools, chiefly registers (Business Register, Register of Agricultural Farms, Register of Buildings and Dwellings).

II.2. Use of administrative data sources for the Business Register and other registers

7. The Business Register (BR) of the CZSO is of an administrative and at the same time a statistical character. It embraces all businesses and non-profit institutions located in the Czech Republic. These units are included in the BR on the basis of an administrative registration at a commercial court of justice, the Trade Licensing Office, or other authority in compliance with the law. This administrative source notifies the CZSO of changes in registration or official suspension of activity or winding down. The number of units included in the BR went up from under 50,000 in 1990 to more than 2 million today. Until 2001, units in the BR were the types of 'Legal Unit' and at the same time 'Enterprise'. Since 2002 other types of units have been implemented. Also 'Local Units' (which have a different address than 'Enterprise' and at least one employee) are monitored for 'Enterprises' with 20 employees or more. Data on Kind-of-Activity-Units are observed for approx. 220 economically most important 'Enterprises'.

8. Basic data on every unit included in the BR are updated on a continual basis. A part of data is updated using the results of surveys taken by the CZSO. However, the CZSO carries out in the area of business statistics exhaustive surveys among units with 100 employees or more and sample surveys among units up to 99 employees. The sample ranges between 0.5 and 5 per cent. This means that most units up to 99 employees cannot be updated using statistical surveys. That is why administrative data sources are used. There were many obstacles in legislation and elsewhere for the CZSO to overcome in this field, particularly to be allowed using information from the Ministry of Finance. These data sources serve mainly the following major purposes:

- determining economic activity or inactivity of the unit;
- verifying the present address of the location of the unit;
- determining the number of employees to classify the unit into the correct size group.

9. In order to determine economic activity or inactivity of the unit, the CZSO uses two administrative sources: (i) a list of identification codes of units that showed a movement on the VAT account at the Ministry of Finance; the CZSO assumes that units with movement on this account are most likely economically active; (ii) the Social Security Office, which provides two basic items of information – a list of units that pay social security (for their employees or the self-employed for themselves) which corroborates economic activity, and data saying for how many persons social security was paid, which serves the classification of the unit into the size group according to the number of employees. There was a rule until 2001 saying that the number of employees in the BR was primarily taken over from CZSO surveys; where no survey was conducted, administrative data source was used. However, as definitions of the number of employees valid in the CZSO and in social security are somewhat different, both data were not fully comparable. Therefore, starting with 2002, the number of employees in the BR is determined in a standard manner according to this administrative data source only.

10. Most data for the Register of Agricultural Farms are acquired from regular statistical surveys and from Agrocensus (once in five years). As this Register is linked with the BR, most of its units use data from the BR to determine economic activity and the number of employees. This Register uses data of the Ministry of Agriculture (as an administrative data source) gained e.g. for the needs of providing grants and subsidies. Also an agreement was reached on data supply from the Agrarian Chamber and other voluntary professional associations of entrepreneurs in the areas of agriculture and forestry.

11. The Register of Buildings and Dwellings serves now predominantly the needs of the Population and Housing Census. Particularly the Central Population Register of the Ministry of the Interior, the registers of recipients of social benefits of the Ministry of Labour and Social Affairs, and the records of local and district councils on dwelling houses completed, on dwelling houses closed down and on election districts are used as the administrative data source for the needs of the Register of Buildings and Dwellings.

12. At present the CZSO works on linking all the three Registers, so that data in all of them is standardized. The Register of Buildings and Dwellings will serve the automated verification of the address of the location of the unit included in the BR. This process is demanding in view of achieving standardization of the methodology of all the CZSO registers, standardized technology and software.

However, all registers of individual authorities of central government in the Czech Republic are not yet linked, mainly for legislative and technical reasons.

II.3. Use of the administrative data sources as a substitute of statistical surveys

13. There are mainly two reasons why the CZSO strives for the use of administrative data sources as a substitute for statistical surveys:

- easing the administrative burden on respondents;
- increasing the effectiveness of resources from the state budget provided to the CZSO (accountability to the taxpayers).

14. The CZSO cooperates closely with most ministries, the central bank, as well as with the largest employers' associations. According to the Statistics Act, which has been in force in the Czech Republic since 1995, or on the basis of agreements, the CZSO uses the data sources of these institutions. This is usually in the form of mutual data exchange on terms agreed in advance. These terms mainly apply to strict limits on access to individual data and, further, technical and financial conditions of the data exchange. This conception aims at statistics being produced by ministries, the central bank and employers' associations, which should be guided by high demand for expert knowledge of a given area. For example, a specialized statistical unit of the Ministry of Health produces nearly all the health statistics. The same holds for education and culture. The Ministry of Industry and Trade and some of the employers' associations produce detailed statistical data concerning groups of manufacturing or products, technologies, etc.

15. There are two fundamental functions of the CZSO ensuring that this system of mutual exchange of statistical data is working:

- creating a standardized methodology (statistical classifications, definitions of indicators, sampling methods, registers) obligatory for all institutions involved;
- coordination of the division of labour into individual institutions, so that individual administrative data sources cover all areas of statistical information and duplication in data collection be eliminated.

16. These two functions are assigned to the CZSO by the Statistics Act and are respected by all involved institutions, which understand their necessity for the proper functioning of this system.

III. CONCLUSIONS

17. The use of administrative data sources has become an important part of the CZSO transformation in new political, economic and social conditions of the Czech Republic. More than ten years' work resulted in the establishment of the CZSO's own rules for handling administrative data sources. These rules and developed practice are similar to the practice of NSIs of the EU member states.

18. The defining of administrative data sources, their evaluation, and laying down the optimum aims and methods use is considered very important. In the course of the 1990s, the Czech Republic was successful in creating a working system of division of labour and exchange of statistical data between administrative data sources that are represented, on top of the CZSO, by most ministries, the central bank and large associations of entrepreneurs. In doing so, the CZSO plays the role of a single centre for methodology and coordination. Data from these administrative data sources serve both the direct statistical purposes (production of statistical information) and other purposes (creation and updates of registers).
