

# The hidden side of a successful story – implication of wide use of administrative data sources at national statistical institutes

Metka Zaletel<sup>1</sup>, Irena Križman<sup>2</sup>  
Statistical Office of the Republic of Slovenia

## Abstract

Trends at majority of national statistical institutes show that administrative data sources are more and more widely used in different areas of statistics and for different reasons. The most important reasons are of course reduction of the response burden, ensuring timely coherent data to the users and decreasing the amount of routine work at national statistical institutes. This is the bright side of the story of success, but there is also the dark side which is not exposed so often. There are a number of questions which will be addressed in the paper: what are the implications of wide usage of administrative sources within the national statistical institutes, what kind of changes does it cause for the data processing, what are the new challenges for the human resources at the national statistical institutes, and are the costs maybe even higher? A few case studies will be shown with the comparison of the input of human and financial resources.

**Key words:** administrative data, human resources, data processing, financial resources, human resource management

## 1. Introduction

Together with authorised producers in the national statistical system the Statistical Office of the Republic of Slovenia (hereinafter SORS) in the previous Medium-Term Programme of Statistical Surveys 2003-2007 already adopted as one of its most important long-term

---

<sup>1</sup> Metka Zaletel, Director for General Methodology and Standards, Statistical Office of the Republic of Slovenia, Vozarski pot 12, SI-1000 Ljubljana, Slovenia, [metka.zaletel@gov.si](mailto:metka.zaletel@gov.si)

<sup>2</sup> Irena Križman, Director-General, Statistical Office of the Republic of Slovenia, Vozarski pot 12, SI-1000 Ljubljana, Slovenia, [irena.krizman@gov.si](mailto:irena.krizman@gov.si)

objectives the plan of activities that enable as wide as possible use of existing data from various administrative registers and records and the use of innovative methodological and IT approaches. This practice has continued with the next Medium-Term Programme (2008-2012). With continuous activities in this field, producers of national statistics reduce administrative burdens of data providers (enterprises, family farms and persons). SORS and authorised producers of official statistics study the reduction of the burden of reporting units within authorised statistical bodies in Slovenia and abroad and present the results in annual reports on implementing statistical surveys. It is very important to keep the balance between the demands of data users and the requirements for reducing the burden of reporting units. This balance reflects in the awareness that the primary mission of national statistics is to provide quality, comparable, timely and relevant data with the minimum possible costs for everybody involved. Reduction of burdens in statistics can be achieved by minimising demands of users (number of variables, frequency of data collection, excluding some of the reporting units) and by using already collected data, either exchanging them within the statistical system or using administrative data with modern technology and organisation. All these new activities cause new challenges, new tasks and new demands at the national statistical institutes. These new challenges are described and clarified in sections 3 and 4 and the comparison of cost and burden is shown in section 5.

## **2. Wide usage of administrative sources**

The pre-conditions for using administrative sources are high quality and linked records. Slovenia has a great advantage with its good register infrastructure, improving records, systematic approach to eliminating administrative barriers at national level and good statistical legislation that enables obtaining all individual data together with identifiers.

Firstly, the legal framework of the Slovenian national statistical system should be pointed out. The National Statistics Act in its Articles 28 and 32 clearly defines the role of SORS as the coordinator of the national statistical system while using administrative data sources for different purposes. Two issues might be raised here as being very important for the efforts to reduce the response burden:

- The possibility to influence the content and structure of administrative sources which enable SORS and other authorised producers to promote standardisation and use of national classification systems on one hand, and on the other hand to introduce

necessary variables into the administrative data source and to take care of regular updating of the source where it is feasible to cooperate with other governmental bodies.

- The possibility to use all different kinds of public and private data sources (registers, databases, etc.) free of charge and link them for statistical and research purposes.

However, as one might expect, the National Statistics Act is not enough to realize the strategies mentioned in the previous chapter. Very pro-active involvement of representatives of the national statistical system is necessary to establish a stable system at state level which offers data of good quality to the users.

SORS's policy for co-operation with other important players at the national level has been very fruitful during the last few years. Agreements signed with different institutions made possible for SORS and other players within the statistical system to use the data in a transparent and efficient way. First, it is necessary to distinguish three different types of institutions in Slovenia regarding the role they play in the statistical system.

1. The role of other institutions within the statistical system (i.e. authorised producers) is similar to the role of SORS except for its coordination role. It means that these institutions are entitled to use administrative data for statistical purposes on one hand, and to exchange the data with identifications with other authorised producers on the other hand (again, for statistical purposes only). Of course, the relations within the system do need to be described in special agreements of understanding stating clear definitions and division of tasks. The possibility to exchange data with identifications is of crucial importance to minimize the burden, since data are collected only once for multiple purposes.
2. The role of institutions – owners of different administrative data sources and registers: SORS and other authorised producers – as described above – have the possibility to use different administrative data sources free of charge for statistical purposes. SORS has signed agreements with a vast majority of these institutions to clarify the purposes of data use and to clarify the way of receiving the data. In general, one can conclude that most of the agreements consist of two parts – one is more strategic, the other is more operational. Here, it is necessary to point out that a data source is received and

edited once for multiple purposes within the national statistical system. This kind of organisation ensures stability within the statistical system.

3. The role of institutions collecting data for multiple purposes: in some cases there is no possibility to establish administrative data sources (e.g. waste statistics). Since many ministries and/or governmental agencies need the data for administrative purposes, we have established a system to collect data at a single point. The agency collects data for its purposes and statistical purposes at the same time (and clarifies all purposes in the questionnaire), and afterwards SORS uses the collected data. Again, all these relations are described in Annual programme of statistical surveys and in the agreements which also consist of the strategic part and the operational part (including the definition of the questionnaire).

### **3. Implication of the use of administrative sources to data processing at national statistical institutes**

Data processing at national statistical institutes is a complex issue which has been described and tried to standardize many times in the past. Almost none of these attempts proved to be very successful, especially in cases when different countries tried to use the same templates. One of the important reasons for this is the changing nature of data processing and the development of new processes which occur virtually all the time.

Here we would like to compare the complexity of the processes of data collection based on questionnaires versus the processes of data collection from administrative sources in combination with data collection by using a short questionnaire. Our points of interest are the processes which are caused by the use of administrative records. What really needs to be done when administrative sources are used? There are three levels of tasks to be performed at national statistical institutes:

1. Legal – administrative tasks are important when an administrative source enters the NSI: agreements to receive the data need to be prepared and signed, basic controls of received data are to be performed (to verify that the data received are compatible to what has been requested and agreed). These tasks are infrastructural and are done once for all different uses of the same administrative data source.

2. Centralised control and editing of an administrative data source: for each data source, there must be an expert with the profound knowledge of the content available from the source. Besides this expert, there must also be several IT experts dealing with the tasks of creating a central repository of the data sources and the task of preparing applications for data cleaning and data editing. It is necessary to emphasize that the administrative data sources almost resemble living organisms and that the changes in legislation are very rapidly reflected in their structure. Nowadays, many of these changes occur also due to the reduction of the response burdens. These tasks are also infrastructural and are performed once for all different uses of the same administrative data source. Again, it is important to point out that more than half of the administrative data sources at the Statistical Office of the Republic of Slovenia are used only once because they are quite “limited” in their contents (e.g. register of cattle). In such cases, the tasks described above are not infrastructural.
3. Application of the administrative source in a survey: when combining data from different data sources, it is again necessary to clean and edit the data to eliminate some irregularities that arise due to different methodologies or the discrepancy between the reference periods of different administrative sources. Very detailed prioritisation of data sources (including the questionnaire) is very important, sensitive and also influenced by amendments to the valid legislation. The survey manager who is responsible for this part of the process must have wide knowledge of statistical methods and of the legislation which determines the shape of administrative sources (s)he is using. Besides, the vast majority of administrative data sources are not collected for statistical purposes, so the survey manager needs knowledge on how to adequately adjust and adapt the methodology of the source to get the data that suit the statistical purpose.

On the basis of the above description, we can summarize the new needs of national statistical institutes as follows:

	New needs	Decreased needs
Infrastructural tasks	- Better support of services for legal matters (preparation of agreements, follow-up, etc.)	- Less operational work for interviewers' management - Less operational work in data

	<ul style="list-style-type: none"> <li>- Statistical experts on the content of the data source</li> <li>- Influence to standardisation and development of administration environment</li> <li>- IT experts (data matching, E&amp;I)</li> </ul>	<ul style="list-style-type: none"> <li>cleaning</li> <li>- Less financial resources for interviewers' payment</li> </ul>
Subject-matter tasks	<ul style="list-style-type: none"> <li>- Better knowledge of data sources</li> <li>- Better knowledge of statistical methods (E&amp;I, prioritisation, etc.)</li> <li>- Knowledge about the necessary adjustments of different methodologies</li> <li>- Education of users</li> </ul>	/

The brief overview shows that less clerical work is needed and the workload of different experts (legal, IT, subject-matter specialists) is getting higher or – in other words – more experts are needed.

#### **4. New challenges for human resource management and training at national statistical institutes**

As we described in the previous section, there are new tasks emerging at NSIs and there are also some tasks which are being terminated. Human resource management is exposed to two new challenges: on one hand there are employees who at present perform more or less routine clerical work and their number should decrease quite rapidly and on the other hand, there are survey managers whose present knowledge of statistical methodology does not cover all the issues concerning the use of administrative sources. Quick response to these challenges is urgently required, otherwise NSIs cannot successfully start (or continue) to use an ever increasing number of administrative sources. In other words, the increasing use of administrative sources to reduce the response burden has caused new needs also in the field of human resource management. The issues that need to be solved in the near future by virtually the majority of national statistical institutes are:

- How to solve the problem of having employees who have been performing routine work for the last 20 years, but are not needed any more? Some of them will retire, a minority of them will find a new job elsewhere, some of them will still be needed for the continuation of routine work, but by rough estimation about 20-25% of them will not be needed any more. These people should be re-trained to get a new job within the NSI, so new training programmes will have to be developed for them.
- New training programmes should be developed also for the survey managers and their assistants. Topics such as statistical techniques in cases of combining different sources, the content and the nature of the administrative source are the most important. One has to bear in mind that there is no simple recipe how to use the administrative source in a particular survey, but the process differs from case to case. That's why it is not easy to organize this kind of training; it should be based on examples and good practices. At the Statistical Office of the Republic of Slovenia we are trying to organize it in the form of dedicated workshops lead by experienced statisticians. About half of the audience has acquired some experience with using administrative data sources and the other part of the audience is starting to use administrative data in their surveys. There are several purposes of such workshops: employees with experience can exchange the good practices, employees without experience can gain new knowledge, and technical support experts can get to understand the new problems, issues and challenges that arise from the use of administrative sources.
- At last, but not least, the number of employees assigned to a certain statistical survey might be raised in case of some surveys because of the complexity of the new tasks. However, bearing in mind the limitations regarding employment in the public sector, this problem can by no means be solved easily.

## **5. Cost and burden – case study on two surveys at the Statistical Office of the Republic of Slovenia**

Let us introduce two surveys which will serve as a case study:

1. The Survey on Travels of Domestic Population is a case of simple data collection: the sample size is about 6,000 households, the sample design is stratified systematic sample, the interviewing mode is CATI and this is a quarterly survey.

2. EU-SILC is a case of complex data collection: the sample size is about 12,000 households (9,000 responses), the sample design is a rotating panel, stratified two-stage sample, 40% of variables are collected by the CAPI or CATI questionnaire (dependent on the availability of phone numbers and on the wave of interviewing), 60% of variables are collected by different administrative sources and this is an annual survey.

During the exercise of estimating the response burden and costs arising at national statistical institutes, SORS estimated the costs and burden also for these two surveys:

*Table 1: Costs and burden - comparison of two case studies, 2007*

	<i>Survey on Travels of Domestic Population</i>	<i>EU-SILC</i>
<i>Burden of households</i>	368.40 hours EUR 3,294.40	4,354.00 hours EUR 38,924.80
<i>Costs at SORS</i>		
<i>Interviewers</i>	EUR 23,889.20	EUR 144,059.43
<i>Data processing, preparation, dissemination – EUR</i>	<b>2,966 hours</b> <b>EUR 34,998.80</b>	<b>7,551 hours</b> <b>EUR 89,101.80</b>
<i>Total costs and burden</i>	EUR 62,182.40	EUR 272,086.03

These two surveys are unfortunately not completely comparable because of different sample sizes, partly different interviewing modes and a slight difference in the number of observed variables. But on the other hand, it is quite clear that data processing should be comparable if administrative sources didn't play an important role in EU-SILC processing. One can observe that more than 7,500 hours were spent on EU-SILC in 2007 and almost 3,000 hours were spent on the survey on travel. The difference is caused partly by a larger number of variables, but mostly by the extensive use of administrative variables and – speaking frankly – numerous problems while combining 15 different administrative sources by the field questionnaire. Our estimate is that about 1,500–2,000 hours were spent just for the data matching, cleaning and prioritizing, which means around EUR 20,000 extra costs within the

office. At this point, it is necessary to emphasize what would have happened (during the fieldwork) if the administrative data were not used at all: the burden of households would have increased by 100% or by almost additionally EUR 40,000 per year.

## **5. Conclusions**

The usage of administrative sources at national statistical offices is on the increase, but implications of it are more than serious and unfortunately also neglected by the majority of EU Member States' governments. It is necessary to clarify that reduction of the response burden by using administrative sources in fact generates new demands and needs within national statistical offices. In the case studies, we revealed that use of secondary sources is costly for the offices, but the response burden on the part of households or enterprises is even more expensive. One has to bear in mind that higher response burden does not only mean "more time and money" for households and/or enterprises, but it increases also the perceived burden of the data providers - even at a higher rate. To summarize, we would like to point out the final recommendation, which is use as many administrative data sources as possible, but handle them with caution within the office and put the emphasis on the human resource management and training.

Furthermore, the questions treated in the paper are opening new areas of interest: what are the real inputs and outputs of national statistical institutes, what does "an average user" get from the national statistical institute and how to ensure that the users will get the maximum of available statistical outputs.

## **6. References**

- [1] National Statistics Act: [http://www.stat.si/eng/drz\\_stat\\_zakonski.asp](http://www.stat.si/eng/drz_stat_zakonski.asp)
- [2] Medium Term Programme of Statistical Surveys 2008-2012, Statistical Office of the Republic of Slovenia: [http://www.stat.si/eng/drz\\_stat\\_zakonski.asp](http://www.stat.si/eng/drz_stat_zakonski.asp)
- [3] Better Regulation and Reduction of Administrative Burden (Ministry of Public Administration, Slovenia): [http://www.mju.gov.si/en/better\\_regulation/](http://www.mju.gov.si/en/better_regulation/)
- [4] Križman, I. (2007): Further Decrease of Administrative Burden - an important Objective of Slovene National Statistics, presented at the conference Statistical Days, Radenci, 2007: <http://www.statisticni-dnevi.si/content/view/32/37/lang,en/>