



# Are Publicly Available Online Businesses Lists Appropriate to be used as Sampling Frames in Croatian Business Surveys?

*Berislav Žmuk*

*Faculty of Economics and Business, University of Zagreb, Croatia*

## Abstract

**Background:** In order to conduct a probability business survey, a high quality sampling frame of enterprises is needed. **Objectives:** The goal of the paper is to investigate publicly available online businesses lists which can be used as sampling frames in business surveys in Croatia and to find out whether they have a satisfactory quality level. **Methods/Approach:** Publicly available businesses lists in Croatia are examined and their appropriateness for use in different modes of data collection as business sampling frames is inspected. The advantages and disadvantages of businesses lists are discussed and compared. **Results:** Overall 11 online businesses lists are considered as business sampling frames in the paper. The comparison analysis has shown that businesses lists from the government institutions are the best choice when business surveys are conducted face-to-face, by mail and/or by telephone. However, none of the observed businesses lists is good enough to be used in a business web survey. **Conclusions:** The research has shown that the publicly available online businesses lists are of a satisfactory level of quality only if traditional data collection modes are used. Unfortunately, they are not appropriate in business web surveys. Therefore, the development of a business register in Croatia is a prerequisite for conducting probability and representative business web surveys in the future.

**Keywords:** business register; business survey; Croatia; mode of data collection; population coverage; sampling frame

**JEL classification:** C83

**Paper type:** Research article

**Received:** Apr 25, 2017

**Accepted:** Jun 24, 2017

**Citation:** Žmuk, B. (2017), "Are Publicly Available Online Businesses Lists Appropriate to be used as Sampling Frames in Croatian Business Surveys?", Business Systems Research, Vol. 8, No. 2, pp. 26-39.

**DOI:** 10.1515/bsrj-2017-0014

**Acknowledgments:** This work has been supported by the Croatian Science Foundation under the project STRENGTHS no. IP-2013-9402.

## Introduction

The survey process consists of many different steps (Bethlehem, 2009). The earlier steps in the survey can be considered as more important ones. The reason for that

can be found in the fact that if a researcher makes a mistake in a very early step that mistake will have a certain (negative) impact on all further steps. Consequently, such a mistake will be very hard to correct unless it is spotted at the right time.

At the very beginning of the survey process, a researcher is dealing with a problem of choosing an appropriate sampling frame (Ross, 2005). A sampling frame is a list or a procedure which is used for identification of all elements of a target population (Groves et al., 2004). Target population consists of elements which are an object of researcher's interest. Elements can be individuals, households, enterprises, schools, and similar (Martirosyan et al, 2010). A researcher uses the sampling frame to select a certain number of elements which are going to be invited to participate in the survey. The answers and characteristics of elements which participated in the survey are used for making an inference about the whole population.

Researchers have two options: they can use already prepared sampling frames or they can make their own sampling frames. There are some advantages and disadvantages of both options. Generally speaking, own sampling frames are time and costs demanding whereas already prepared sampling frames could be inaccurate, not up to date, or the definition of elements could be different from the definition in the research. In this paper, focus will be given to publicly available online businesses lists that can be used as sampling frames in business surveys.

Business surveys are specific due to the fact that a researcher selects enterprises in the sample, but the answers are provided by an employee or more of them (Cox et al., 1995). Enterprises, without any doubt, have a very important role in each economy. Consequently, it is important to make research about businesses because certain weak spots and possible difficulties can be detected that way. Furthermore, business surveys can result in recommendations which enterprises can use for improving their business further and achieving better business results.

However, in order to be able to make a valid inference about enterprises the probability sampling approach must be used. In probability sampling, each element needs to have a known probability which is higher than zero to be selected in the sample. Consequently, the sampling frame has the main role in probability sampling. A sampling frame which cannot ensure that each element of interest has a known probability to be chosen in the sample is considered not to be of high quality and should not be used to make an inference about all elements. If an inference was made by using low quality sampling frames, the possibility of making misleading conclusions is fairly high.

Although Croatia is the European Union member state, some reports show that it is falling behind in economic development in comparison to the other European Union member states (see World Bank, 2017). Because of that significantly more business surveys should be conducted first to identify the problems with which they must cope. After that some recommendations could be brought to increase the competitiveness level of Croatian enterprises. Such survey research should be conducted promptly because otherwise Croatia could lag behind too much. Consequently, already prepared sampling frames should be used in conducting business surveys. However, the main research question is whether publicly available online businesses lists, which can be observed as already prepared sampling frames, of Croatian enterprises have a satisfactory level of quality or not. In order to start business operations, an enterprise must be registered and get permissions for work from different government institutions. Consequently, government institutions should have a lot of data about enterprises. Because of that, the research hypothesis of the paper is that business sampling frames based on publicly available online businesses lists which are provided by government institutions are of higher quality than those

which are based on publicly available online businesses lists provided by other institutions or persons. Unfortunately, in the present literature the possibility of using publicly available online businesses lists as sampling frames in business surveys has not been considered yet. Consequently, this original paper is of great importance for further development of business surveys.

The paper is organized as follows. After the brief introduction, in the second chapter sampling frames are observed in detail from the theoretical perspective. In the third chapter, the way of selection of publicly available online businesses lists which are going to be observed in the paper is described. In the fourth chapter, the selected publicly available online businesses lists having the role of business sampling frames are inspected and their differences are discussed. Chapter five presents final conclusions and recommendations for further research.

## Sampling frames characteristics

If a sampling frame provides some probability of selection for each element in the population, a probability sample can be drawn from the population (Warnecke, 2005). However, regardless of whether researchers use an already prepared sampling frame or they make their own sampling frame it is very hard to get a totally perfect sampling frame. A sampling frame should include only the elements of the target population but it is problematic or, in some cases, impossible, to get a complete list only of elements which are of interest. Consequently, some differences between the target population, elements which are of interest, and the survey population, elements which are included in the sampling frame, are expected (Stoop, Harrison, 2012). The smaller the difference between the target and the survey population is, the higher the quality of the sampling frame is. At the same time the probability of making a valid inference about the population is increasing.

The following reasons can be observed as the basic reasons which lead to imperfect frames: missing elements, blanks or foreign elements, duplicate elements, clusters of elements, many-to-many matching elements (Kish, 1995). The problem of missing elements or the non-coverage problem is present when some target population elements are not listed in the sampling frame. This problem can be solved by using supplemental and additional sampling frames that cover non-covered elements. Furthermore, if the effect of non-coverage is known, non-coverage weighting adjustments can be used to take into account the missing elements problem. The problem of blanks or foreign elements is present when elements in the sampling frame do not have corresponding population elements. In other words, the sampling frame can contain elements which are not a part of the target population. Consequently, some screening is necessary to omit and delete such elements from the sampling frame. The problem of duplicate elements is present when the target population element is linked to two or more sampling frame elements. In that way, an element has a higher probability to be chosen in the sample and it can be sampled more than just once, also. To solve this problem, the screening process needs to be performed. The problem of clusters of elements is opposite to the duplicate elements problem. The problem involves the fact that more than just one target population element can be selected by a sampling frame element. The final problem, the many-to-many matching elements problem, incorporates the duplicate elements and the clusters of elements problem.

All sampling frame problems caused by previous reasons can be solved. However, the real question is whether it is reasonable to do that taking into account the required time and money. Some problems can be solved very easily, e.g. duplicate elements can be found very fast and deleted from the sampling frame by using

modern computers. On the other hand, in order to include some missing elements in the sampling frame a lot of resources could be needed, i.e. field work should be conducted to find out contact addresses of some enterprises. No matter which problem is present at the used sampling frame, a researcher should strive to use the best sampling frame. However, that does not mean that a researcher should use just one sampling frame per survey. They can combine more of them but the advantages and disadvantages of all the used sampling frames must be taken into account. A researcher must keep in mind that with the use of more sampling frames the complexity of analysis rises and the problems can get bigger than in case of using just one sampling frame.

Kish (1995) recommends three general ways of avoiding sampling frame problems. The first recommendation is just to ignore and disregard the problem. However, this can be done only if the consequences of ignoring a problem are insignificant and if the costs of solving the problem are too high. The second recommendation is to redefine the target population to fit the sampling frame. So, instead of fitting a sampling frame to the target population, the reverse is done. However, the aims of the research should be kept in mind because too much target population redefinition could result in an inability to reach research goals. In order to avoid sampling frame problems, sometimes the best approach is to sit down and manually correct the entire sampling frame by deleting all blanks, splitting each cluster, and so on. This third recommendation requires more time and money than the first two recommendations but it is preferred over the previous ones.

Business sampling frames have certain special characteristics which make them different from other sampling frames. First of all, contact information about enterprises can be found in the business sampling frames because they are the target population. However, respondents are not enterprises but people who are connected with enterprises in a certain way. So respondents can be owners, managers and other employees. Depending on research aims, it can be decided that sales managers, accountants, or other specific employees will be contacted and asked to participate in the survey. This ensures that competent employees provide answers about an enterprise and its processes. However, it is very difficult to know what each person in an enterprise is competent for and whether they should be contacted or not. Furthermore, the list of all employees and the organisation of business processes in enterprises is unknown to a researcher. Because of that, in most cases, a general invitation to an enterprise is sent and the recipient is asked to forward survey information and the survey invitation letter to the most appropriate person in the enterprise.

The same as individuals, enterprises can also decide whether to participate in the survey or not. However, the situation with enterprises is somewhat different than with individuals. Namely, enterprises have to deal with competition on the market. Because of that enterprises unwillingly share information about their business processes. Consequently, some enterprises have a strict policy on participation in surveys. Furthermore, because of sensitive topics, enterprises are more likely to participate in surveys conducted by private institutions than by government agencies. The reason for that lies in the fact that a higher level of confidentiality can be reached in surveys conducted by private institutions (World Bank, 2013).

## Data and methods

In order to investigate the quality level of publicly available online businesses lists in the role of sampling frames in business surveys in Croatia, a selection of publicly available online businesses lists was made. Because we live in the digital age, the

search for businesses lists or sampling frames which can be used in business surveys was conducted by using the Internet sources. In other words, only digital businesses lists were considered. The reason for observing only digitalized businesses lists is their easiness of use. Furthermore, if needed, corrections in a digitalized sampling frame can be made easily and very fast. The same is valid for selecting enterprises which will be contacted and asked to participate in the survey.

The publicly available online businesses lists which could serve as sampling frames were found by using Google search and keywords such as "sampling frame", "business sampling frame", "list of enterprises", "Croatian enterprises" and similar. This procedure simulates the searching process that would be done by a researcher whose intention is to conduct a business survey by using already prepared publicly available online businesses lists or business sampling frames. The search of publicly available online businesses lists was conducted in December 2016.

As expected, the search gave (too) many results. Unfortunately, only a small number of results led to businesses lists. Furthermore, only businesses lists that include Croatian enterprises and are in the Croatian language are observed. An additional requirement is that the access to the businesses lists is free of charge and without any restrictions. Finally, overall 11 businesses lists as business sampling frames were taken into account and analysed in the paper. The list of the observed businesses lists as sampling frames and their hyperlinks are provided in Table 1.

*Table 1*  
Selected businesses lists

Businesses list	Hyperlink
Bisnode	<a href="http://www.bisnode.hr/">http://www.bisnode.hr/</a>
Creditreform	<a href="http://secure.creditreform.hr/">http://secure.creditreform.hr/</a>
Croatian Bureau of Statistics	<a href="http://www.dzs.hr/Hrv/important/Roj/roj.asp">http://www.dzs.hr/Hrv/important/Roj/roj.asp</a>
Croatian Chamber of Economy	<a href="http://www1.biznet.hr/HgkWeb/do/extlogon">http://www1.biznet.hr/HgkWeb/do/extlogon</a>
Department of Justice	<a href="https://sudreg.pravosudje.hr/registar/f?p=150:1">https://sudreg.pravosudje.hr/registar/f?p=150:1</a>
Energetika-net.com	<a href="http://www.energetika-net.com/korisno/baze-podataka/tvrtke">http://www.energetika-net.com/korisno/baze-podataka/tvrtke</a>
Fininfo	<a href="http://www.fininfo.hr/">http://www.fininfo.hr/</a>
IRENA	<a href="http://www.irena-istra.hr/index.php?id=3416">http://www.irena-istra.hr/index.php?id=3416</a>
Kompass	<a href="http://hr.kompass.com/">http://hr.kompass.com/</a>
Tvrtke.com	<a href="https://www.tvrtke.com/">https://www.tvrtke.com/</a>
Yellow Pages	<a href="http://www.zutestranice.com/">http://www.zutestranice.com/</a>

Source: Author's work

The characteristics of the selected businesses lists will be inspected in detail and tested. The advantages and disadvantages of each businesses list as a business sampling frame for conducting mail surveys, telephone surveys, personal interviews and web surveys will be emphasized. The observed businesses lists will be compared and their overall quality level will be estimated.

## Results and discussion

Before a business sampling frame can be used in a survey, a researcher should check whether the observed business sampling frame meets all expected requirements. In the paper, the characteristics of the observed businesses lists are going to be inspected. It will also be checked whether they can be used in a

business survey as sampling frames. Firstly, business sampling frames are expected to include a complete list of enterprises. Furthermore, it is expected that the business sampling frame is updated. That is especially important when small enterprises are under the research because they appear, disappear and change basic information almost on a daily basis. A researcher expects that a business sampling frame has a user friendly interface and that enterprises can be easily selected. Not only is it assumed that enterprises can be easily sampled, but it is assumed that necessary data about enterprises are also available and can be easily accessed. Depending on the survey method of data collection, additional information about enterprises might be needed. For example, if a web survey is conducted, e-mails of enterprises are needed, and if a mail survey is conducted, postal addresses of enterprises are necessary. So, in addition to inspecting main characteristics of the observed businesses lists, their appropriateness for the use as sampling frames in different survey data collection methods will also be observed.

The observed businesses lists are going to be inspected in alphabetic order. Therefore, *Bisnode* businesses list is studied first. After the web page of *Bisnode* is loaded, the option of searching enterprises from Croatia and from other countries is given. Unfortunately, at the same time the main disadvantages of this businesses list can be noticed. Namely, there are no additional options for searching and filtering enterprises according to their characteristics such as their size or legal form. Furthermore, the complete list of enterprises is not available. However, once an enterprise is chosen, basic information about it is shown. The following information is shown: full and short name of the enterprise, address, county, VAT number, identification number, legal form, date of establishment, enterprise's representatives, and the main activity. There is an additional option to get a credit report for the observed enterprise but this service must be paid. Finally, it can be concluded that this businesses list can be used as an additional source of information about enterprises. So, this businesses list cannot be used as a sampling frame on its own.

*Creditreform* is the following businesses list under inspection. *Creditreform* has the same disadvantage as *Bisnode*. Namely, there is a search bar in which a name of an enterprise must be written first. So, the names of enterprises must be known in advance. However, unlike *Bisnode*, after typing an enterprise name *Creditreform* gives enterprises which have the typed name included in their names and offers some options to search these enterprises according to their VAT number, address or the main activity. Unfortunately, the complete list of enterprises is not presented to a researcher. After an enterprise is selected the following information is available: date of establishment, VAT number, identification number, registered activities, enterprise's representatives, address, telephone number, e-mail, web page, and the number of employees. However, in order to get more information, a researcher should pay. The same as *Bisnode*, *Creditreform* could be used as a supportive business sampling frame only.

The Croatian Bureau of Statistics has developed its own businesses list. This businesses list can be considered as an administrative one because the Croatian Bureau of Statistics collects data about enterprises according to law, decisions and regulations on the National Classification of Activities (see Official Gazette, 1994, 2007a-c). Consequently, enterprises are obligated to send information about their registered activities to the Croatian Bureau of Statistics and, naturally, pay for this service. Due to that this businesses list can be considered as up to date. Furthermore, enterprises also send the following information: name, address and legal form. The Croatian Bureau of Statistics can provide other available information about an enterprise but upon a request and after the payment was made. However, in order

to get any information about an enterprise, the correct identification number must be known. The next drawback of this businesses list is the fact that a complete list of enterprises is not available and visible to a researcher (without any charge). So, this businesses list cannot be used as a business sampling frame without additional business sampling frames.

The *Croatian Chamber of Economy* has developed its businesses list by using data from the Commercial Court. Consequently, it can be observed as an administrative one because enterprises are obligated to give certain information to the Commercial Court. This list provides a lot of information to a researcher: identification number, short and full name, county, municipality, address, telephone number, e-mail, the size of the enterprise, the main activity, the legal form, the responsible persons, the number of employees, bank account. Unfortunately, it has to be emphasized that not all those data are known for all enterprises (in most cases e-mail is missing). However, enterprises can be filtered according to different characteristics (main activity, size, and similar). What is more, after selecting certain characteristics, in the next step, a full list of enterprises which have these characteristics is shown. There are some limitations when a list of enterprises is observed. The first limitation is that only 15 enterprises can be listed on the screen. The second limitation is that the process of enterprises filtering according to certain characteristics will result in 500 enterprises at most. In other words, nevertheless the fact that more than 500 enterprises have given characteristics, the business sampling frame will provide only the list of the first 500 enterprises with those characteristics. Despite these drawbacks, this businesses list can be used as a sampling frame alone without other additional sampling frames.

Unlike the *Croatian Chamber of Economy*, the *Department of Justice* uses data from the Commercial Court directly. However, in order to get any information about an enterprise, a researcher should know its identification number, VAT number or the name of an enterprise in advance. Because of that this business sampling frame cannot be used alone. After an enterprise is selected, the following information can be obtained: identification number, VAT number, the information whether the enterprise is in the dispute at the Commercial Court, name, address, location on the map, the legal form, registered activities, the list of persons in the supervisory board and persons who represent the enterprise, selected information about the legal relations, a hyperlink to the latest financial statements. Various legal documents which were given to the Commercial Court are also available.

*Energetika-net.com* is a businesses list that is oriented towards enterprises from the energetics sector only. Consequently, a full list of Croatian enterprises is not provided but the list of energetic enterprises is. So, if a researcher observes only enterprises from the energetics sector, they can consider using this businesses list as a sampling frame. However, if other enterprises are also a subject of researcher's interest, other businesses lists and business sampling frames must be used as well. *Energetika-net.com* is not an administrative businesses list, which additionally raises the question whether the information about enterprises is up to date. After an enterprise is selected, the following information is available: address, telephone number, e-mail, web page, activity.

In order to start using *Fininfo* as a business sampling frame, a researcher should know the identification number, the VAT number or the name of an enterprise. After selection of an enterprise, the following information is given: address (but not full!), identification number, VAT number, registered activities, size, web page, telephone number, the lists of owners, persons in the supervisory board and persons who represent the enterprise. The location of the observed enterprise is shown on a map

also. If researchers would like to know something more about the observed enterprise, they should pay a certain amount. After such a payment, a researcher gets different accounting indicators and financial statements.

*IRENA* stands for Istrian Regional ENergy Agency and it was founded by the County of Istria. Consequently, this list focuses on enterprises from Istria whose activity is related to energy. Because of that *IRENA* includes only 40 enterprises. The short list of enterprises is the main disadvantage of this businesses list. However, the list of all 40 enterprises is available and visible to a researcher at the very beginning. The following information about enterprises is available: name, activity, address, web page, name of a contact person, telephone number, e-mail. However, not all those data are available for each listed enterprise.

In order to get a list of enterprises, in *Kompass* a researcher must search enterprises according to their activity. After an activity is typed in, a preliminary list of enterprises which are somehow connected with that activity is shown. After that the list of enterprises can be reduced by adding some additional characteristics such as the number of employees. The following information is usually given for an enterprise: address, telephone number, VAT number, web page, bank account numbers, memberships in associations, number of employees, initials of the names of owners, persons in the supervisory board and persons who represent the enterprise, the main and secondary activities. In case a researcher wants to have a complete list of enterprises with different business indicators included, an additional payment is required.

*Tvrtke.com* also requires some information about enterprises before it can provide a list of them. An advantage is that in addition to the search based on enterprise's name, a researcher can search enterprises according to products and services, activity, county and place. The main disadvantage of this businesses list is the user unfriendly design. Namely, *Tvrtke.com* includes too many ads. Furthermore, some privacy issues could be also present (i.e. *Tvrtke.com* for some reason wants to know the location of its user). The provided list of enterprises is hard to export because of many figures, also. The following information is available after an enterprise is selected: address, telephone number, VAT number, identification number, main activity, contact person, number of employees. This businesses list is free of charge to use by researchers, but enterprises can buy a "service package" which grants them some advantages over other enterprises. For example, enterprises which have paid for additional services appear on the top of the list. This fact must be taken into account when enterprises are sampled.

*Yellow Pages* offers an option to search enterprises according to their name, activity or address. There are also some ads but not so many as at *Tvrtke.com*. *Yellow Pages* does not include all enterprises but only those enterprises which users have included. Because of that it is questionable whether the information about enterprises is regularly updated. However, the following information about an enterprise is usually given: address, telephone number, contact person, web page, activities. However, if an enterprise pays an additional charge it can include more information about itself, such as an e-mail address.

Table 2  
Main characteristics of the observed businesses lists

Businesses list	Admin. source	Free access	Limited access	Visible full list of enter.	All types of enter. included	All Croatian enter. included	User friendly interface
Bisnode	-	+	+	-	+/-	+/-	+
Creditreform	-	+	+	-	+/-	+/-	+
Croatian Bureau of Statistics	+	+	-	-	+	+	+
Croatian Chamber of Economy	+	+	-	+	+	+	+
Department of Justice	+	+	-	-	+	+	+
Energetika-net.com	-	+	-	+	-	-	+/-
Fininfo	-	+	-	-	+/-	+/-	+
IRENA	-	+	-	+	-	-	+
Kompass	-	+	+	+	-	-	+
Tvrtke.com	-	+	-	-	+/-	+/-	-
Yellow Pages	-	+	-	+	-	-	-

Source: Author's work

Table 2 briefly presents main characteristics of the observed businesses lists. Only three out of eleven observed businesses lists clearly and undoubtedly use an administrative source for their database. Consequently, such businesses lists are more likely to be up to date, to include all enterprises and to have more information about enterprises than the other businesses lists. All the observed businesses lists can be used free of charge, which is important taking costs into consideration. However, some businesses lists require additional payment for additional information and services. From the sampling point, the best solution would be that all enterprises are listed on one page which would make the process of sampling much easier. Unfortunately, less than half of the observed businesses lists provide a fully visible list of enterprises. It has to be emphasized that none of the observed businesses lists had an option of automatic export of the list of enterprises. The quality of the observed businesses lists as business sampling frames differs very much. There are businesses lists for which we can be sure that they include all types of enterprises and all Croatian enterprises but there are also businesses lists for which we can be sure that they do not include all enterprises. However, for some businesses lists it is questionable whether they include all enterprises or not (they are marked with "+/-" sign in Table 2). Generally speaking, all the observed businesses lists are user-friendly and easy to use. Still, some of the observed businesses lists have overextended use of ads which has a negative impact on users' experience.

Table 3

Information about enterprises available in the observed businesses lists

Businesses list	Name	Address	VAT number	Telephone number	E-mail	Web page	Additional contact information
Bisnode	+	+	+	-	-	-	+
Creditreform	+	+	+	+	+	+	+
Croatian Bureau of Statistics	+	+	-	-	-	-	-
Croatian Chamber of Economy	+	+	+	+	+/-	+/-	+/-
Department of Justice	+	+	+	-	-	-	+
Energetika-net.com	+	+	-	+	+	+	-
Fininfo	+	+/-	+	+	-	+	+
IRENA	+	+	-	+	+/-	+/-	+
Kompass	+	+	+	+	-	+	-
Tvrtke.com	+	+	+	+	-	-	+
Yellow Pages	+	+	-	+	+/-	+/-	+/-

Source: Author's work

Table 3 summarizes main information about enterprises which are available in the observed businesses lists and which can be useful in the research. The enterprises' name is the obligatory part of each business sampling frame and all the observed businesses lists have it. However, one or more enterprises might have a very similar name. Because of that it is useful to know the VAT number which is unique for each enterprise. However, some observed businesses lists do not have this information. All the observed businesses lists have provided information about addresses of enterprises. However, *Fininfo* does not provide the full address because it omits the postal code. A majority of the observed businesses lists have information about official telephone numbers of enterprises. However, the availability of information about e-mails and web pages is not so widespread. The enterprises themselves could hold responsibility for that as well. Namely, there are a lot of enterprises which do not use an e-mail address and a web page as a way of business communication. Furthermore, there are enterprises which have an e-mail address but they do not use it. Consequently, some of the observed businesses lists have information about e-mails and web pages for only a certain number of enterprises. Researchers would appreciate if a business sampling frame had additional information which would help them in the research. For example, it would be great if businesses lists would have lists of managers who are responsible for a certain part of business, direct e-mail address of the main manager, and similar. Some of the observed businesses lists give such information but their accuracy is questionable.

Table 4

Applicability of the observed businesses lists as business sampling frames in different survey data collection methods

Businesses list	Face-to-face survey	Telephone survey	Mail survey	Web survey
Bisnode	+/-	-	+/-	-
Creditreform	+/-	+/-	+/-	+/-
Croatian Bureau of Statistics	+/-	-	+/-	-
Croatian Chamber of Economy	+	+	+	+/-
Department of Justice	+	-	+	-
Energetika-net.com	+	+	+	+
Fininfo	+	+	+/-	-
IRENA	+	+/-	+	+/-
Kompass	+/-	+	+	-
Tvrtke.com	+	+	-	-
Yellow Pages	+	+	+	+/-

Source: Author's work

The applicability of the observed businesses lists as business sampling frames in different survey data collection methods is inspected and final conclusions are shown in Table 4. Almost all the observed businesses lists provide enough information which enables a researcher to conduct a face-to-face or a mail survey. There are some businesses lists which do not include information about official telephone numbers of enterprises. Consequently, some businesses lists cannot be used in a telephone survey. Unfortunately, many of the observed businesses lists cannot be used in web surveys as a business sampling frame.

If the observed businesses lists are observed individually it can be concluded that the *Bisnode* businesses list can be used as a business sampling frame in face-to-face and mail surveys. However, its use as a business sampling frame is questionable because *Bisnode* does not include all types of enterprises and all Croatian enterprises. Furthermore, the full list of all enterprises in this businesses list is not available. *Creditreform* could be used as a business sampling frame in face-to-face, telephone, mail and web surveys. However, *Creditreform* has the same disadvantages as *Bisnode*. Because of that it is questionable whether the *Creditreform* businesses list can be used as a sampling frame in probability sampling.

The same is true for the use of the *Fininfo* businesses list as a sampling frame. Its use in probability sampling is arguable and it cannot be used in web surveys. The *Tvrtke.com* businesses list has even worse properties than *Fininfo*, namely, *Tvrtke.com* cannot be used as a sampling frame in mail surveys because all the disadvantages mentioned above are present there also.

The main disadvantage of *Energetika-net.com* and *IRENA* businesses lists is their orientation towards enterprises with certain characteristics only. Nevertheless, it is arguable whether all enterprises with emphasized characteristics are included in the first place. Because of that their usefulness in face-to-face, telephone, mail and web survey should be considered very limited.

On the other hand *Kompass* and *Yellow Pages* businesses lists do hide the fact that they do not include all enterprises. However, an advantage is that a list of all

listed enterprises is available at both businesses lists. Still, that does not change the fact that their use in probability sampling is very limited.

All the above mentioned businesses lists are considered to be non-administrative. In other words, they do not use data from one or more administrative sources. The *Croatian Bureau of Statistics*, the *Croatian Chamber of Economy* and the *Department of Justice* businesses lists are considered to be administrative ones. Their main advantage over the other observed businesses lists is the fact that they really include all types of enterprises and all Croatian enterprises. This advantage is so huge that it leads to the acceptance of the research hypothesis of the paper. So, it can be concluded that business sampling frames based on publicly available online businesses lists which are provided by government institutions are of higher quality than those which are based on publicly available online businesses lists provided by other institutions or persons. However, if these three administrative businesses lists are observed, some disadvantages can be found also. The main disadvantages of the *Croatian Bureau of Statistics* businesses list are a lack of all enterprises and additional contact information. Consequently, the *Croatian Bureau of Statistics* businesses list cannot be used in a survey on its own. The *Department of Justice* businesses list also suffers from a lack of all enterprises but it ensures more information about enterprises than the *Croatian Bureau of Statistics* sampling frame. However, the *Department of Justice* businesses list can be used as a sampling frame only in face-to-face and mail surveys.

From all the observed businesses lists, the *Croatian Chamber of Economy* businesses list has the best properties and it can be used as a sampling frame at all survey data collection methods. However, the *Croatian Chamber of Economy* businesses list has certain disadvantages when a web survey is conducted but only enterprises can be blamed for that. Namely, not at all enterprises use e-mails or the Internet.

An unambiguous answer to the research question cannot be given. If the capability of the observed businesses lists to be used as business sampling frames in different survey data collection methods is considered, it can be concluded that the existing businesses lists could be used in face-to-face, telephone and mail surveys. Those survey data collection methods can be considered as traditional ones. However, when modern data collection methods are observed, the situation is not so good. Namely, the observed businesses lists, with the exception of the *Croatian Chamber of Economy* sampling frame, offer very weak support for conducting web surveys. It is expected that with the increase of computer literacy, appropriate businesses lists will be developed also. However, in order to speed up this process, the government should also be involved. The potential positive effects of increased computer literacy and availability of direct and instant access to information about the situation in enterprises should not be neglected.

## Conclusion

Survey results can indicate certain serious problems in enterprises. In that way, certain solutions can be recommended and, consequently, enterprises can achieve better business results. However, survey conclusions have a greater significance if a probability business survey is conducted. One of the main requirements for conducting such as survey is having an appropriate business sampling frame.

In the paper overall 11 publicly available businesses lists and their adequacy for use as sampling frames were inspected. There are far more available businesses lists but it is assumed that those 11 can very well describe the present situation and point out the most important problems with widely available businesses lists. Furthermore,

those 11 businesses lists have some basic characteristics which make them different from other lists. For example, certain businesses lists are oriented towards a specific group of enterprises, other businesses lists give additional information if a researcher pays for additional service, while some businesses lists have user friendly interface, and so on.

The characteristics of all the 11 businesses lists have been discussed and compared. The comparison has shown that businesses lists which use administrative sources as a source of information are more likely to be appropriate to be used as sampling frames in probability sampling.

However, there are some problems when a web survey is intended to be conducted because not all enterprises use e-mail and the Internet. In order to improve this situation, an increase in computer literacy in enterprises is expected.

The main limitation of the paper is that only 11 publicly available online businesses lists were considered. In the future research online businesses list available at specialized organizations should also be taken into account.

It has to be emphasized that nowadays the nonresponse problem in surveys is becoming increasingly important (Dale et al., 2007). In the further research it would be interesting to investigate to which extent imperfect sampling frames have an impact on the response rates. If it were shown that perfect and imperfect sampling frames result in statistically similar response rates, that would result in a completely new view on investing additional efforts in improving imperfect sampling frames. However, the level of sampling frames imperfectness also has an important role here.

## References

1. Bethlehem, J. (2009). *Applied Survey Methods: A Statistical Perspective*, Hoboken, John Wiley & Sons.
2. Cox, B. G., Binder, D. A., Chinnappa, B. N., Christianson, A., Colledge, M. J., Kott, P. S. (eds) (1995), *Business Survey Methods*, Hoboken, John Wiley & Sons.
3. Dale, T., Erikson, J., Fosen, J., Haraldsen, G., Jones, J., Kleven, Ø. (2007), "Handbook for Monitoring and Evaluating Business Survey Response Burdens", available at: <http://ec.europa.eu/eurostat/documents/64157/4374310/12-HANDBOOK-FOR-MONITORING-AND-EVALUATING-BUSINESS-SURVEY-RESPONSE-BURDEN.pdf> (15 December 2016).
4. Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. K., Singer, E., Tourangeau, R. (2004). *Survey Methodology*, Hoboken, John Wiley & Sons.
5. Kish, L. (1995), *Survey Sampling*, New York, John Wiley & Sons.
6. Martirosyan, L., Arah, O. A., Haaijer-Ruskamp, F. M., Braspenning, J., Denig, P. (2010), "Methods to identify the target population: implications for prescribing quality indicators", *BMC Health Services Research*, Vol. 10, No. 137, pp. 1-8.
7. Official Gazette (1994), Law on the National Classification of Activities, No. 98.
8. Official Gazette (2007a), Decision on the National Classification of Activities 2007 - NKD 2007., No. 58.
9. Official Gazette (2007b), Decision on the National Classification of Activities 2007 - NKD 2007., No. 72.
10. Official Gazette (2007c), Regulation on the classification of business entities by National Classification of Activities 2007 - NKD 2007., No. 80.
11. Ross, K. N. (2005), "Sample design for educational survey research", available at: [http://www.unesco.org/iiep/PDF/TR\\_Mods/Qu\\_Mod3.pdf](http://www.unesco.org/iiep/PDF/TR_Mods/Qu_Mod3.pdf) (15 December 2016).
12. Stoop, I., Harrison, E. (2012), Classification of Surveys. In *Handbook of Survey Methodology for the Social Sciences* (Gideon, L. ed.), New York, Springer.

13. Warnecke, R. B. (2005). Sampling Frames. In Encyclopedia of Biostatistics. John Wiley & Sons.
14. World Bank (2013), "G20 Financial Inclusion Indicators: Methodology", available at:  
<http://databank.worldbank.org/data/download/g20fidata/G20%20Set%20Methodology.pdf> (9 December 2016).
15. World Bank (2017), Doing Business 2017: Equal Opportunity for All, Washington, World Bank.

## About the author

Berislav Žmuk, PhD, graduated with a major in Accounting, post-graduated in Statistical Methods for Economic Analysis and Forecasting, and gained his PhD degree in Business Economics at the Faculty of Economics and Business, University of Zagreb. Currently he is an Assistant Professor at the Department of Statistics, Faculty of Economics and Business, University of Zagreb where he teaches the following subjects: Statistics, Business statistics and Business forecasting. In 2013, he successfully completed Sampling Program for Survey Statisticians (SPSS) at the Survey Research Center (SRC), Institute for Social Research (ISR), University of Michigan in Ann Arbor, Michigan, USA. In 2015, he completed several survey methodology courses (Introduction to Web Surveys, Introduction to Questionnaire Design, Mixed-Mode and Mixed-Device Surveys) at Gesis - Leibniz Institute for Social Research in Cologne, Germany. In 2016, he completed two survey methodology courses (Introduction to Survey Data Analysis, Advanced Survey Data Analysis and Survey Experiments) at Essex Summer School in Social Science Data Analysis in Colchester, United Kingdom. In 2017, he completed Survey Research: Statistical Analysis and Estimation course at Utrecht Summer School in Utrecht, the Netherlands. His main research fields include applications of statistics in business and economy, survey methodology and statistical quality control. The author can be contacted at [bzmuk@efzg.hr](mailto:bzmuk@efzg.hr).