The Use of New Technologies for the Collection of Statistical Information in Romania

Liliana Mihaela MOGA
Dunarea de Jos University of Galati, e-mail: liliana.moga@gmail.com

Alina Irina POPESCU
The Bucharest University of Economic Studies, alina.popescu@rei.ase.ro

Valentin Marian ANTOHI
Dunarea de Jos University of Galati, e-mail: valentin_antohi@yahoo.com

Abstract

Web based information collection becomes important for statistical analysis, because of the offered advantages and by the information technology progress. The paper analyzes web based information collection organization and implementation at national level and discusses the validity by using different data sources, such as survey online portal implemented by National Institute of Statistics description and functionality, legal framework existing web-surveys, security issues, and new technologies. The research will contribute to the theoretical foundation of web-based statistical information collection, identifies the stakeholders and each category advantages.

Keywords: online information collection, web-survey, statistical research, national framework

JEL Classification: B23, C81, C88

1. Introduction

Web-based information collection will increasingly be important for any statistical information collection, addressed to any reporting domain. The web-surveys are the most used application of web-based data collection. Web survey response rates in general public surveys are often dismal (Dillman, Reips, and Matzat, 2010). In addition, given a choice of responding by mail, a mode not held in high regard by many surveyors, most respondents choose mail (Smyth et al., 2010). The strengths and weaknesses of web-surveys are a debate subject the literature. Fricker and Schonlau (2002) argue that a mixed mode strategy is more economical than a mail-only survey if a sufficiently large number of responses are received over the web. Coverage error is a serious problem for many web-surveys. As not each person or even each company has Internet access, not everyone has the same probability of being included in the survey. This situation is more common in the rural areas. Internet penetration rates continue to increase and once this will minimize the differences the information collected online in demographic characteristics (Dever et al. 2008).

Together with problems related to sampling, coverage and non-response, web surveys also face problems of measurement errors: differences between the true answer and the answer recorded. These measurement errors can be different in web surveys compared to surveys with interviewers who, if properly trained, can explain whatever problems the

interviewee might have with the questions (Pedraza et al., 2010). The questionnaires designed for Internet surveys have to differ by the traditional way applied questionnaires. Gosling et al. (2004) made comparative analysis of six preconceptions about Internet questionnaires and of the way in which the design of the questionnaires affects the trust in online conducted statistical surveys.

A proper sampling frame allows obtaining a probability-based sample and using standard inference procedures, although methodological concerns continue, because not all sampled members are willing or able to complete the survey. Non-response error is not unique to web-surveys but as their response rates tend to be lower when compared to other modes, the problem is quite serious (Lynn 2008; Manfreda et al., 2008). Main reasons are: the inefficiency of response stimulating efforts (incentives, follow-up contacts); technical difficulties (slow unreliable connections, security, interaction of scripts with technologies, interaction of technologies with personality); personal problems in using a computer; personalization (Joinson and Reips, 2007); and privacy and confidentiality concerns (Wel and Royakkers, 2004; Galesic 2006; Kaczmirek, 2008). Research has shown that privacy policies tend to intensify privacy concerns rather than engender trust. One way to combat this dichotomy is to redesign their content, language, and presentation format (Pollach, 2007).

In order to minimize measurement errors, the web-survey instrument must be - due to its self-administration - easy to understand and complete, designed to keep respondents motivated to give optimal answers, and should reassure respondents about confidentiality (Couper and Miller, 2008). These issues are further complicated because an online questionnaire, unlike a paper survey, can vary among respondents due to different browser settings, user preferences, and hardware variations (Dillman, 2009).

Web-based data collection is mainly used by commercial firms and in academic fields, like psychology, sociology and health studies. Statistics institutes remain cautious about the implementation of web-based data collection, like web-surveys. Consequently, there is space for innovative and outreach research, particularly within the context of globalization and requests for cost efficiency, traditional slow and costly modes of data collection cannot be optimal for political advice. To increase confidence in web-based data collection, data quality is essential for academics and users (Gosling et al., 2004), while privacy is essential for respondents (Wel and Royakkers, 2004).

2. Methodology

In the first part of the paper a review of the relevant literature to identify the online tools used for collecting the information needed to carry out statistical surveys is performed. As a result of the literature analysis and empirical studies conducted so far, the factors preventing the widespread expansion of online surveys, as well as the advantages making that this questionnaire type be used increasingly more are identified. In the second part of the paper an analysis of the manner in which online questioning is implemented by the National Institute of Statistics of Romania is achieved. There are also analyzed both the legal framework which facilitated the introduction of online questioning, and the way in which software application developed for statistical data collection fulfils all user groups requirements. Most of the information about eSOP functionality is picked up from the

eSOP User Manual (2013) which is available on the web portal for the online collection of statistical information.

3. Legal Framework for the Development and Use of the Survey Online Portal

The electronic system implementation has been mainly carried out through the cooperation of the Ministry of Communications and Information Technology with the National Institute of Statistics. The general framework for electronic system operation of collecting statistical data, the responsible institutions as well as their responsibilities in this area are regulated in Romania by Ordinance no. 19 (2003) concerning the compulsoriness of using electronic system for statistical data collection. The Ordinance supplements the provisions of the Government Ordinance No. 9/1992 on the official statistics organization of, as amended by Law no. 11/1994, republished, with activities adjusted to the characteristics of information society. The electronic system for collecting statistical data is defined as the information system of public utility, accessible via Internet to a specified address, operated for the purpose of collecting statistical data through the on-line procedure, hereinafter referred to as Electronic System. The on-line procedure means the direct access use by Internet of the electronic statistical questionnaires, with a view to filling in the statistical data required by the National Institute of Statistics.

The Ordinance regulates the selection and obligations of the beneficiary institutions. Therefore, the list of reporting institutions (external users) will be established by Government Decision and will be periodically updated. Entering and extending the electronic system use by the reporting institutions will be performed in stages. The reporting institutions have the obligation to answer to the statistical questionnaires by using the electronic system, whenever this will be required. The list of statistical research included in the electronic system, as well as the rules for employing the electronic system have been drawn up according to the regulations, by the Ministry of Communications and Information Technology and the National Institute of Statistics. These have been approved by the Romanian Government. The National Institute of Statistics will establish a development program of the electronic system for including other categories of reporting units, other than public institutions.

4. The Collection of Statistical Information in Romania

The National Institute for Statistics is the institution carrying out statistical research in Romania, according to the Annual National Statistical Program, approved by Government Decision. The planning and design of statistical research is performed in accordance with the national legislation in force, regulations, recommendations and the European Union standards. The statistical research objectives, as well as the indicators followed are described in the reports of the Annual National Statistical Program. Statistical research is carried out based on information collected from enterprises and institutions selected according to scientific criteria, located throughout the country. The enterprise/institution

selection is based on the requests received from the internal beneficiaries of statistical information, which are ministries and other government institutions, trade unions, employers, research institutes, media etc., as well as from the external beneficiaries, which are the European Union institutions and other international organizations, a system that enabled to collect and validate online starting from the reference year 2010.

For a more efficient activity of information collection, the National Institute for Statistics has passed from the traditional system, which involved filling in the questionnaires and sending by normal mail or e-mail, to a new online system that allows the online collection of the information required for carrying on the infra-annual and annual statistical research. During the system implementation, it has been estimated that its employment will enable both reducing the paper costs, as well as the operationalization among the users and providers of statistical data. The time used by employees during the statistical data collection activity will be lowered by almost 60% (Mediafax, 2009).

The online takeover of statistical information is achieved through the web portal e-Survey Online Portal (eSOP). The online application allows the management of information on the reporting entity, the access to all information supplied in all statistical surveys through the eSOP system and the management of respondents for each entity inserted in the system and has the purpose of enhancing the collection of statistical information and their quality. The respondents are persons authorized by enterprises and institutions to fill in the information for statistical surveys.

The system users are direct beneficiaries of the facilities made available by the web portal of on-line statistical data collection and, depending on the location from where the system is accessed, users are split into domestic and external users. Domestic users are the operators at the County Directorates of Regional Statistics and will enter via Intranet / Internet the forms or statistical questionnaires sent on paper by the respondents. External users are the representatives of the units responding to statistical questionnaires. They may either enter the statistical questionnaires and/or forms into the on-line application, or send them printed to the County Directorates of Regional Statistics in order to be inserted by operators (internal users). For allowing access to users only to that information corresponding its role within the institution, for each field or group of fields of surveys and/or statistical research, security areas will be set up at the database level and the access to these areas will be allowed based on declaring the role of user who accesses the database.

4.1. Information Security and Privacy within the Survey Online Portal

Information privacy involves ensuring the internal security, external security and protecting the information against destruction. The external security implies protecting the information against foreign persons' access in the company/institution. The internal security is achieved through protecting information from the unauthorized access of insiders. In order to ensure privacy, the most employed method is the stringent organization within the information system by establishing strict levels of access into the system based on passwords. Information protection in case of hard elements destroying is performed through periodically saving the database, going until purchasing redundant equipment (servers fitted with multiple hard disks, through which is automatically saved the database within certain time intervals).

The information system ensures the definition of dynamic groups for each statistical research, updating being carried out based on entering each unit in a sample of statistical research or survey. By its centralized information system, the National Statistics Institute ensures that policies are evenly and consistently applied to all applications and systems, providing minimum the following functionalities:

- Authentication: the system allows multiple forms of authentication, from username and password, up to multi-level and security devices authentication;
- Authorization: the system restricts users access only to a certain type of resources, according to their role and responsibilities assigned;
- Single Sign-On: has been implemented SSO type mechanisms using multi-factor security mechanisms;
- Identity administration: the management of security policies is achieved in an automated and centralized manner.

The web portal for online downloading the statistical information from respondents has a series of security requirements for ensuring the information security, both when they are in motion, and also when the files are stored on disk. The information stored in the database is encrypted; the access of administrators (DBA) and privileged users to any data will be blocked. Moreover, each communication channel has to be of VPN type. It is essential that the data being stored in database be secured. For preventing the security issues deriving from the privileged users' access to information, the online solution provides mechanisms that limit the rights of privileged user access to the information stored in the database, and following installation, the privileged users will no longer have access but to areas strictly related to management, without being able to access the information stored in tables allocated to applications.

For the protection of information collected through the web portal has been taken into account the implementation of an information solution that would ensure the prevention of failures in the system running and a minimum time of recovery in case of hardware, software, human errors or natural disasters. The security architecture implemented is based on rules and concepts that enable adaptation to the security requirements, but at the same time ensure a high level of security for the information handled. The information stored in database is encrypted, and the security mechanism is a centralized one, which also controls the access to the component modules, ensuring the management of the information system users.

The privacy is achieved by passing their source under anonymity. The information provided will be aggregated and processed, losing any identifying element. No information on the respondent unit/institution will be transmitted to local authorities, policy makers at the central level, press or public. Also, the respondents are ensured that information will be used exclusively for statistical purposes by the National Institute of Statistics. The implemented security solution ensures the information confidentiality, which is an essential task in a system of data collection from respondents.

4.2. Implementation of Online Statistical Surveys

Statistical surveys are those surveys that collect, process and provide information of descriptive or analytical nature as concerns the society, economy, culture, etc.. Depending on methodology used, the surveys are conducted on certain time intervals. According to the survey periodicity, it is classified in:

- annual surveys: have a single period of information collecting;
- half-yearly surveys – statistical surveys that have two data collection periods, one for each semester;
- quarterly surveys - statistical surveys having four data collection periods, one for each quarter;
- monthly surveys - statistical surveys that have 12 periods of collection, one for each month.

The eSOP application facilitate the participation of selected units to take part in the surveys conducted and provide online the requested information. When publishing a survey in the online application, the observation units selected are notified through an email. For the surveys with more collection periods in a year (twice a year, quarterly or monthly surveys), the eSOP application sends to the observation units selected to participate in the respective survey a notification email at the commencement of each collection period.

To each collection period is allocated a time interval within which the observation unit can enter statistical information in the eSOP application. During this period the eSOP application forwards warning notifications concerning the end of the collection period. Each type of answer is performed so as to not allow entering values outside the defined range or of other type than defined. Also, for ensuring accurate data validation rules are set out. These rules are executed when navigating on the questionnaire pages and also to the final saving. Breaching these validation rules is announced by an error message shown at the top of the questionnaire page and contains a link to the view page of identified errors.

5. Conclusions

The new system implementation by the National Institute of Statistics has proved that it ensures, on the one hand the information collection from economic agents and institutions, at the same time lowering the costs of collecting and processing the statistical information. Also the ESOP system provides the operationalization among the users and providers of statistical information. The usage of ESOP application demonstrates that the online collection of statistical information is a viable alternative for carrying out statistical surveys by the national statistical institutes. Moreover, compared to the surveys conducted by academia or private companies, the statistical institutes benefit from the legislative support in relation to providers of statistical data, which enables them to ensure a representative sample for the surveys performed and a high credibility to the outcomes achieved.

Acknowledgement:

Part of the work reported in this article was supported by the European Science Foundation, COST Action IS1004 “WEBDATANET” (http://webdatanet.eu).

References:

Couper, M.P., Miller, P. (2008): Web survey methods – Introduction. *Public Opinion Quarterly* 72(5): 831– 835.

Dever, Jill A., Ann Rafferty and Richard Valliant (2008): Internet surveys: can statistical adjustments eliminate coverage bias?. *Survey Research Methods* 2(2): 47–62.

Dillman, D.A., Smyth, J., and Leah, C. (2009): Internet, Mail and Mixed-Mode Surveys: the tailored design method, Ed. John Wiley, Holboken, 3th Edition.

Dillman, D.A., Reips, U.D., Matzat, U. (2010): Advice in Surveying the General Public Over the Internet. *International Journal of Internet Science* 5(1): 1–4.

Fricker R, Schonlau M. (2002): Advantages and disadvantages of Internet research surveys: Evidence from the literature. *Field Methods* 14(4): 347–367.

Galesic, Mirta (2006): Dropouts on the web: effects of interest and burden experienced during an online survey. *Journal of Official Statistics* 22 (2): 313–328.

Gosling, S.D., Vazire, S., Srivastava, S., John, O.P. (2004). Should we trust Web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist* 59: 93–104.

Joinson, A.N., Woodley, A., Reips, U.D. (2007): Personalization, authentication and self-disclosure in self-administered Internet surveys. *Computers in Human Behavior* 23(1): 275–285.

Kaczmirek, L. (2008). Human-Survey Interaction: Usability and Nonresponse in Online Surveys, ed. Mannheim: MADOC.

Lynn, Peter (2008): The problem of nonresponse, in International Handbook of Survey Methodology, eds. Leeuw, E.D., Hox, J. and Dillman, D.A., New York: Lawrence Erlbaum Associates.

Manfreda, K.L., Bosnjak, M., Berzelak, J., Haasa, I. and Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. International. *Journal of Market Research* 50(1): 79– 104.

Pedraza, P., Tijdens, K., Muñoz de Bustillo, R., and Steinmetz, S. (2010). A Spanish Continuous Volunteer Web Survey: Sample Bias, Weighting and Efficiency / Una encuesta voluntaria y continua en la red en España: sesgo, ponderación y eficiencia. *Revista Española de Investigaciones Sociológicas (Reis)* 131: 109–130.

Pollach, I. (2007). What's wrong with online privacy policies? *Communications of the ACM* 50(9): 103-108, ACM New York, NY, USA.

Smyth, J.D., Dillman, D.A., Christian, L.M., and O'Neill, A. C. (2010). Using the Internet to survey small townsand communities: Limitations and possibilities in the early 21st century. *American Behavioral Scientist* 53: 1423–1448.

Wel, L., Royakkers, L. (2004). Ethical issues in web data mining. *Ethics and Information Technology* 6(2): 129-140.

NIS, eSOP web portal for the online collection of statistical data, User Manual (2013). http://www.insse.ro/cms/files/Site_eSOP_v2/files/eSOP%20-%20Manual%20pentru%20Utilizatorii%20Externi.pdf.

Government Ordinance No. 9 from 30/01/2003 concerning the compulsory use of electronic system for collecting statistical data, http://www.legi-internet.ro/legislatie-itc/guvernare-electronica/og-192003-colectare-electronica-a-datelor-statistice.html

Mediafax, INS will have a web portal to collect and process on-line the statistical data, november 2009, http://incomemagazine.ro/articles/ins-va-avea-portal-web-de-colectare-si-prelucrare-on-line-a-datelor-statistice.