



**Statistics Netherlands**

Division of Process Development, IT, and Methodology  
Methodology Department



**Statistics Norway**

Department of Data Collection

---

**MODERNISATION AND QUALITY OF BUSINESS STATISTICS:  
THE NSI PERSPECTIVE**

**Paper presented at the  
2013 European Establishment Statistics Workshop (EESW),  
9-11 September 2013, Nuremberg, Germany**

**Gustav Haraldsen<sup>1</sup> and Ger Snijkers<sup>2</sup>**

*Summary: This paper is based on two previous papers discussing the Modernisation and quality of business statistics: Erikson, Haraldsen & Snijkers, 2012; Snijkers & Haraldsen, 2013a. In these papers three basic challenges have been identified: the emergence of the information society, globalisation, and thirdly a shift in the balance of power between NSIs and businesses. In the current paper we try to take this discussion one step further and point at different possible paths to follow in the future*

<sup>1</sup> Statistics Norway: [gustav.haraldsen@ssb.no](mailto:gustav.haraldsen@ssb.no)

<sup>2</sup> Statistics Netherlands: [g.snijkers@cbs.nl](mailto:g.snijkers@cbs.nl)

The views expressed in this paper are those of the authors, and not necessarily reflect the policies of neither Statistics Netherlands nor Statistics Norway.

---

## 1. Four kinds of challenges

By tradition, data collections from businesses are basically built on the same survey design as social surveys. Still, surveying businesses are quite different from surveying households or private persons. While each household or person in social surveys normally have the same weight, larger businesses count more than the smaller ones in businesses surveys. Consequently most business samples are stratified by size. In most people's life being sampled to a social survey is a rare event. In contrast businesses often take part in several surveys at the same time, and several of them will be panels. Because most business surveys are mandatory and because the business respondents complete the questionnaires within working hours, response burden tend to be a more pressing issue than response rates. Making contact is a two-step process where the business management decides on survey participation, and secondly on the response process, i.e. who should complete the questionnaires and what priority the task should be given. Furthermore, those who respond generally act as informants and are not asked personal questions (Willimack and Snijkers, 2013; Snijkers and Jones, 2013). Also information retrieval is rather about searching for and judging available information than about memorizing (Haraldsen, 2013).

Surveys were established as a cost efficient method to collect information at a time when counting all was the alternative. Surveys also have strengths in that they are designed for statistical or analytical purposes. Bob Groves list the most important attributes of surveys to be research controlled data collection, uniform, multivariate measurements and a tight inferential paradigm (Groves, 2013). However, one may both question the quality of business surveys in general and more specifically ask to what extent the aforementioned differences between business surveys and social surveys is reflected in the business survey designs. It is this kind of challenges we address in the newly published book "Designing and Conducting Business Surveys" (Snijkers et al., 2013).

Utilizing information from administrative registers is about to become an important alternative to business surveys. As reporting obligations to the Tax-office and other authorities have been digitalized and computer power has exploded, technically register data have become much easier to access than before. In Northern Europe in particular, legal restrictions are also manageable. In some of these countries national statistical institutes are now obliged to check administrative sources to statistics before they are allowed to launch surveys. They have a so-called data collection strategy in place, resulting in multi-source/mixed-mode data collections (Snijkers, 2009). Until now, however, quality considerations have played a minor part in these investigations. To a large extent it has been taken for granted that registers hold a higher quality than samples, and that this does not need to be documented. We think that quality challenges both residing in administrative registers and which arise when administrative data are transformed to statistical data have been underestimated (See Zhang, 2011).

With what we call the information society, drawing samples because it is cheaper than counting have become a weak argument for conducting surveys. Rather, managing an abundance of data has become a more important challenge. This includes looking for appropriate data sources, extracting the needed data from it, and checking its quality. The information abundance of our time also affects NSIs position as information processing institutions. We meet a growing competition from other information processing actors. This is not only a competition between institutions, but also between qualities. The main dimension in quality considerations is that between professional evaluations and evaluations made by users of statistics. What is currently called "big data" or "organic data" is superior on timeliness, which is a highly appreciated quality by users of statistics, but weak on accuracy, which is the quality indicator survey methodologists first of all strive for. As with all wars the battle between timeliness and accuracy may be won by one of the parties or be solved by a compromise; what Robert Groves calls a blending of designed and organic data (Groves, 2013).

Finally in this overview we will point at a kind of challenge which is specific and fundamental to economic statistics. Because of globalization the structure and behavior of businesses are changing. Businesses no longer stick to national borders. The internationalisation of businesses and economies are making the traditional units and concepts of business statistics harder to use. Hence, concepts that found national statistics and national accounts are more and more difficult to uphold. In this way globalization undermines the national perspective which has traditionally been the main focus in business surveys and official economic statistics.

From a methodological point of view this may either be considered to be a sampling or a measurement problem. From a sampling perspective, it is a coverage problem. From a measurement perspective it may rather be a question of construct validity. The fundamental questions are these: Are our traditional indicators still the best way of reflecting the value and development of our economy? And if so, are results from business surveys the best way of measuring business behaviors and transactions? When considering this last question, bear in mind that business behavior and transaction today more and more take part and leave traces on Internet. It is these digital traces which we call organic data (Groves, 2011), and which Google and other competitors on the statistical market pick up.

Above, we have identified four challenges in modernizing business statistics: modernizing current business surveys; quality issues in register based statistics; blending designed and organic data; and producing statistics in a globalized world. These four challenges point at different paths of modernizing business statistics, and various paths a national statistical institute can go. In the next sections we will briefly discuss these challenges in some more details.

## **2. Tailoring Computerized Business Surveys**

Self-administered surveys are the dominant mode in business surveys. In discussions about how self administered household surveys should be modernized, mixing web and paper questionnaires is a heavy discussed issue. In business surveys the perspective is different. Internet penetration is much higher in businesses than in households, much of the data held in business is already in a digital form and many of the tasks businesses perform are already online. This applies particularly to larger size businesses which have a higher weight in business statistics than the smaller ones. Because of this, mixing modes in business surveys should not be a discussion about web versus paper, but a discussion about different kinds of computerized data collection modes.

The business population is extremely heterogeneous, and this leads to a high variety in the number of surveys different businesses should take part in and how much information they should report.

- When the information asked for is available in an accountancy or personnel system, it seems to be a cumbersome detour for respondents who report in a computerized questionnaire, first to log into the companies administrative system, then look up figures there, copy them and finally paste them into the relevant response box. A more efficient procedure, particularly in large deliveries from large companies and in recurrent surveys, would be to link the response boxes to the relevant posts in the administrative system so that it will be automatically updated each time figures are to be reported. This kind of automatic data capture has already been realised (under different names) in several countries, e.g. using XBRL. The matching process, i.e. the process of extracting the needed data from business registers, and linking it to variables in a statistical system, is however not straight forward, as different definitions may be used and the two systems may be based on different units. When this kind of inconsistency is a problem, it calls for meta-questionnaires, investigating the content of business registers.

- Contrary to what is the case in social surveys, businesses typically take part in several surveys at the same time, many of them running monthly or quarterly. Often accountancy firms also serve as respondents for several middle range or smaller companies. Hence, when web questionnaires are offered as a response option for businesses, there is a need for an administrative tool that keeps track of the reporting obligations. This is what we commonly call a web portal for business surveys

Looking at present administrative survey systems, they are usually still quite primitive. Questionnaire specific user ids and passwords are common, while a common login procedure to all relevant questionnaires would be more practical. By “relevant” we mean that only those surveys where the business belongs to the sample should be listed. The administrative page should also offer search functions, status information and different ways of ordering the list (e.g. according to deadlines or completions status). Finally the administrative page should be designed according to the same visual principles as those used in the questionnaires. Until now next to no usability research has been carried out in order to develop user friendly administrative pages in web questionnaire portals.

In addition to improving the administrative tools offered in business web surveys, there are several ways in which the questionnaires themselves could be improved. Looking at the content of business questionnaires, different questions often call for different competences. This is because the content is tailored to different kinds of statistics and not to different kinds of business respondents. A more user friendly approach would be to collect questions from different statistical divisions, but aimed at the same kind of respondents and put them together in the same questionnaire.

It is also worth noting that the structure of business questionnaires often is quite demanding and hence lend itself to advanced web designs. Looped designs combined with preloads and complicated branching is quite common e.g. in price index surveys.

- Smaller companies are not sampled as often as the bigger ones and generally have less to report than bigger firms. For businesses asked to respond to only one or a small number of surveys, the need for a questionnaire administration system is low, and may rather cause frustrations. Smaller businesses may also have simple administrative systems which do not lend themselves to automated data capture software. In these cases a quick, easy at hand reporting method, like reporting by smart phone or iPad, should be offered.

This kind of technology should also be an alternative when the questionnaire only contains a small number of questions, or when screening questions reveal that the company has nothing to report.

Even if the main mode of data collection in business surveys are and will be self administered, we think interviewers could play an important part in several steps during designing and conducting business surveys. High competence in personal communication is needed when feasibility studies and qualitative interviews are used during survey development and testing. Interviewers could play an important role in motivation campaigns directed towards respondents or managers who in turn pick respondents and decide what priority the questionnaire completion should have. And finally by manning help desks with interviewers, one kind of help could be to collect the data as an interview with some of those who call in.

### **3. Making response tasks easier**

As already mentioned, many NSI’s have adopted a data collection strategy where data already available from other surveys or administrative registers are investigated before the institutes are allowed to launch new business surveys. As a consequence of this trend new methods and procedures for evaluating the validity and reliability of

secondary sources are developed. In an article recently published in *Statistica Neerlandica*, Li Chun Zhang (2011) develops a typology of quality challenges when administrative registers are used for statistical purposes. The typology is based the well known double barrelled survey error model developed by Groves and his colleagues in 2004. One of Zhang's main points is that utilizing administrative registers for statistical purposes both involves quality challenges similar to those known from surveys and quality challenges that have to do with the transformation of administrative data to statistical data, and with integrating information from different administrative sources. In this way Zhang ends up with a longer list of quality challenges than that used in surveys.

An effort to translate these and other quality considerations into a practical, robust and efficient evaluation tool is found in the work package "Improve the use of administrative sources" of the Blue Ets project (Daas et al., 2012). In this project five quality dimensions and 28 indicators measuring these were identified. The five dimensions were Technical checks, Accuracy, Completeness, Integrability and Time-related issues. An initial design for a Quality Report Card was proposed and is presently tested in several countries. The results should be evaluated with Zhang's two-step model in mind.

Zhang's description of the different steps and quality challenges when statistics is produced from administrative registers coincides to a large extent to the tasks business respondents are faced with when they collect information from their company's administration system and feed the results into traditional business questionnaires. Mismatch between the survey questions and available information is the most common source of response burden mentioned by business survey participants (Haraldsen et al., 2013). Even if traditional business questionnaires can be improved, this may be a more fundamental problem which calls for an alternative approach. Instead of asking business respondents to adjust information collected from administrative sources to our information needs, an alternative method would be to collect the available data as they are and combine this with a kind of metadata questionnaire which map definitions and other characteristics that are necessary to adjust the received data for statistical purposes.

#### **4. Blending designed data with other data**

If we could combine the quality standards and richness of traditional surveys with the speed by which actions and transactions are recorded on Internet and other sources of organic data, this would obviously be bringing the best from two worlds together. The basic model seems to be statistical surveys which establish a sound benchmark combined with organic data which are used to measure trends of change. However, in his keynote speech at the latest NTTS conference Bob Groves (2013) pointed at several obstacles on the way towards this goal, the most important perhaps being that commercial interest may hinder cooperation between producers of different kinds of data. He did not see any quick-fix to this problem. He envisages a public-private partnership built between non-commercial, traditional survey institutes and commercial collectors of organic data which obviously are not built in one day.

Furthermore, Groves also called for new model-based methods in order to solve integration of data from various sources (survey data, register data, and organic data), and of different quality. This brings us to multi-source/mixed-mode data collection designs (Snijkers, 2009) to get our data, and its implementation. We also need data integration theories (Zhang, 2011) that replace traditional inference theories based on sampling. Zhang (ibid, p.61) concludes his paper by stating: "The 20<sup>th</sup> century witnessed the birth of and maturing of sample surveys; the 21<sup>st</sup> century will be the age of data integration." To make this happen also legal, ethical and privacy issues need to be addressed, as well technical issues like data matching, and data storage. Again we talk about methods that may take considerable time to develop. While we are waiting for these problems to be solved, we would like to see some simpler and less ambitious combinations of designed and organic business data.

## 5. Meeting globalization

The fourth challenge NSIs are facing is globalisation. Because of globalisation the structure and behaviour of businesses is changing, as well as the behaviour of individual people. Businesses no longer stick to national borders. As pointed out earlier, this may either be considered as a coverage problem or a more fundamental question about the concepts measured in business statistics.

Viewed as a coverage problem, globalisation complicates the defining of business units in a business register, as well as during the data collection process. Multinational groups may not keep track of all cross-border transactions with a level of detail necessary for providing data to statistical offices (data for tax purposes is of course still kept). With administrative offices more and more centered to one specific place within the group, there might not even be somebody within the country that can be contacted for information or data, and concepts or questionnaires in native languages may not be understandable to those who can actually provide the information. (The international Roundtable on Business Registers studies the effects of globalization on defining and identifying business units: <http://unstats.un.org/unsd/methods/citygroup/wiesbaden.htm>). In terms of quality, globalisation has effects on coverage as well as non-response and measurement errors.

Viewed as a conceptual problem, the question is if our traditional indicators still reflect developments in our economy, and are actually based on the behaviour of establishments. Multinational enterprise groups no longer operate as a set of individual national enterprises held together as a group by an owner, but rather as a set of business operations that can be located and relocated between countries as is called for to ensure an efficient production. Functions are more and more specialised and concentrated to specific enterprises/countries within the group. With this in mind, the concepts that found national statistics and national accounts are more and more difficult to uphold. These are fundamental issues in a statistical system, which relates to the relevance of statistics.

## 6. The future path of National Statistical Institutes

Two pairs of concepts seem to come up several times in this paper. That is the distinction between sample and measurement issues and the distinction between accuracy and time-related quality issues. Whatever kind of modernization we are talking about it seems that the discussion can be linked to these pairwise concepts. They point at strengths and weaknesses of different data collection methods, different kinds of evaluation criteria, and at different modernization strategies that can be followed.

Making decisions with regard to these concepts has consequences for the path NSIs will be going in the future. NSIs are no longer a monopolist in the information market, but there are many businesses and organisations that also produce and disseminate statistics. Because of the information society, the relationship with our customers has changed. The question we need to pose ourselves is: what will be our role in the future? What will be our added value, our unique selling point?

We need to rethink our position in the information market. It is our view that we should not fight our competitors on their strength: we cannot beat Google with regard to timeliness, but we should rethink our added value, and our unique selling points. We can e.g. be a benchmark organisation: official statistics is a public good that is indispensable for benchmarking other information, also information produced and used by the government. In addition to traditional statistical indicators, our strength lies in giving complete pictures of phenomena by combining all data sources we have. We have access to a broad spectrum of data sources while individual big data sources and registers are narrow focussed in topic. And because we have a huge number of data, so we could be the National Data Archive and National Data Service Centre, responsible for public data archiving. This position,

however, can only be established if we can be trusted as a sincere and independent public organisation, and meet quality standards.

A second conclusion we can draw is that the challenges are too big for each NSI individually to deal with. NSIs need to globalise, and collaborate on international levels with regard to output, the organisation and innovation of production processes, and research. On European level initiatives in these directions have been taken. Integration and streamlining of business statistics in general Eurostat regulations is discussed in the FRIBS initiative (Framework Regulation Integrating Business Statistics), thus moving away from stove-pipe surveys. Solutions with regard to the production processes are being discussed on high, strategic levels in the High Level Group on Business Architecture of Statistics (HLG-BAS), and include industrialisation, and implementing standardised plug-and-play modules, taking the GSBPM model as a framework. As for research, we have ESSnet projects and the MEETS program, and next we have the Horizon2020 framework stimulating collaboration with universities, to develop new methodologies, and develop courses aimed at training staff to be fit for modernising official statistics. And we have the EMOS initiative, to set-up a European Master in Official Statistics.

Finally, and to quote Groves (2013) again, we can conclude that “we live in interesting times”; we are at the beginning of new developments, a new constellation for producing statistics. It is our job to make a sustainable fundament for the future.

## 6. References

- Daas, P., S. Ossen, and M. Tennekens, 2012, Quality Report Card for Administrative data sources including guidelines and prototype of an automated version. BLUE-ETS report 4.3.
- Erikson, J., G. Haraldsen, and G. Snijkers (2012), The Future of Statistical Data Collection? Challenges and Opportunities. Paper presented at the UNECE Seminar on New Frontiers for Statistical Data Collection, 31 October-2 November 2012, Geneva, Switzerland. ([www.unece.org/stats/documents/2012.10.coll.html](http://www.unece.org/stats/documents/2012.10.coll.html))
- Groves, R.M. (2011), Three Eras of Survey Research. *Public Opinion Quarterly*, Vol. 75, pp. 861-871.
- Goves, R.M. (2013), Official Statistics and Big Data. Keynote address at the 2013 European Conference on New Techniques and Technologies in Statistics (NTTS), 5-7 March 2013, Brussels.
- Haraldsen, G. (2013), Quality Issues in Business Surveys. Chapter 3 in: Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack, *Designing and Conducting Business Surveys*, pp. 83-125. Wiley, Hoboken.
- Haraldsen, G., J. Jones, D. Giesen, and L-Ch. Zhang (2013), Understanding and Coping with Response Burden. Chapter 6 in: Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack, *Designing and Conducting Business Surveys*, pp. 219-253. Wiley, Hoboken.
- Snijkers, G. (2009). Getting Data for (Business) Statistics: What's new? What's next? Paper presented at the 2009 NTTS Conference (New Techniques and Technologies for Statistics). Brussels, Belgium.
- Snijkers, G., and G. Haraldsen (2013a), Modernisation and Quality of Business Statistics: The NSI Perspective. Paper presented at the BLUE-ETS final conference, 8 March 2013, Brussels. ([www.blue-ets.istat.it](http://www.blue-ets.istat.it))
- Snijkers, G., and G. Haraldsen (2013b), Managing the Data Collection. Chapter 10 in: Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack, *Designing and Conducting Business Surveys*, pp. 431-457. Wiley, Hoboken.
- Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack (2013), *Designing and Conducting Business Surveys*. Wiley, Hoboken.
- Snijkers, G., and J. Jones (2013), Business Survey Communication. Chapter 9 in: Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack, *Designing and Conducting Business Surveys*, pp. 359-430. Wiley, Hoboken.
- Willimack, D., and G. Snijkers (2013), The Business Context and its Implications for the Survey Response process. Chapter 2 in: Snijkers, G., H. Haraldsen, J. Jones, and D. Willimack, *Designing and Conducting Business Surveys*, pp. 39-82. Wiley, Hoboken
- Zhang, L-Ch. (2011), Topics of Statistical Theory for Register-based Statistics and Data Integration. *Statistica Neerlandica*, Vol. 66, No. 1, pp. 41-63.