
**BANK OF FINLAND
STATISTICS DEPARTMENT
WORKING PAPERS**

22.5.2000

2/2000

Heikki Hella and Hannu Viertola

**A review of sampling frames in business surveys:
structure, use and maintenance problems**

PREFACE

This review paper considers business survey frames mainly from the theoretical perspective. The theoretical part of this paper considers different aspects of business frames, highlighting the problematic aspects of business frames. Since it is not possible in a brief paper to discuss all the key issues concerning business frames, we focus on the most important ones. The latter part of this paper is a brief review of the reconstruction of a business survey frame in practice. Our special appreciation is extended to Jorma Hilpinen and Harri Kuussaari, who helped us in the preparation of this paper with their insightful comments.

TABLE OF CONTENTS

1. Introduction.....	4
2. Concepts and definitions.....	4
3. Registers and frames: structure, errors and maintenance.....	6
3.1. Properties of frames	7
3.2. Definition of frame errors	7
3.3. Business frame: maintenance and quality.....	8
4. Bank of Finland's BoP survey frame update.....	9
5. Conclusions.....	11
6. References and bibliography.....	12
7. Appendix.....	14

1. INTRODUCTION

Business survey sampling is becoming a more proper statistical procedure for collecting data for official statistics and scientific research. In designing a business survey, one must first define the target population and then the frame population according to the goals of our sampling investigation. That part of the target population that can realistically be sampled is called the frame population.

The quality of the frame is a critical aspect of the overall design of a survey. Controlling the quality is a continuous and challenging task. As is well known, births and deaths of frame units cause continuous additions and deletions to the frame over time. To improve and/or maintain the level of quality, one must incorporate various procedures for eliminating duplication and for updating for births, deaths, out-of-scope units and changes in characteristics of frame units (eg classification, contact information, size).

In this paper, we consider briefly the main challenges and problems of registers and frames encountered in practice. We focus particularly on problems and errors in respect of business frames. *First*, we present the main concepts and definitions; *second*, we consider the properties and quality of registers and frames and the types of frame errors; *third*, the special problems of business frames (single, complex, small, large enterprises) are considered; and *finally*, we discuss briefly the Bank of Finland's project on reconstruction of BoP business surveys.

2. CONCEPTS AND DEFINITIONS

In order to get a clear picture of survey frames, explicit definitions of commonly used terms are needed. However, because of differences in national practices, there are differences in definitions and understanding of concepts.

Business registers and frames play a significant role in conducting regularly repeated business surveys. Business registers are databases that meet all the requirements of survey frames and business surveys. It should be relatively easy to construct a frame or multiple frames from a business register.

The first term we need to define is *business*. This term generally refers to an economic unit involved in the production of goods and services. This is the broadest definition and it encompasses a variety of institutions (enterprises, farms etc). The term also has a more restricted definition. It may refer to commercial and industrial activities only. This more restricted definition is closest to the needs of constructing frames in a financial context. The term *survey* refers to the collection of data to be used for statistical purposes.

Many surveys are conducted within the context of a program. A single organization can conduct several business surveys simultaneously. This requires an efficient procedure for providing survey frames for studying individual businesses. Figure 2.1 illustrates a *statistical program* with six regularly conducted business surveys.

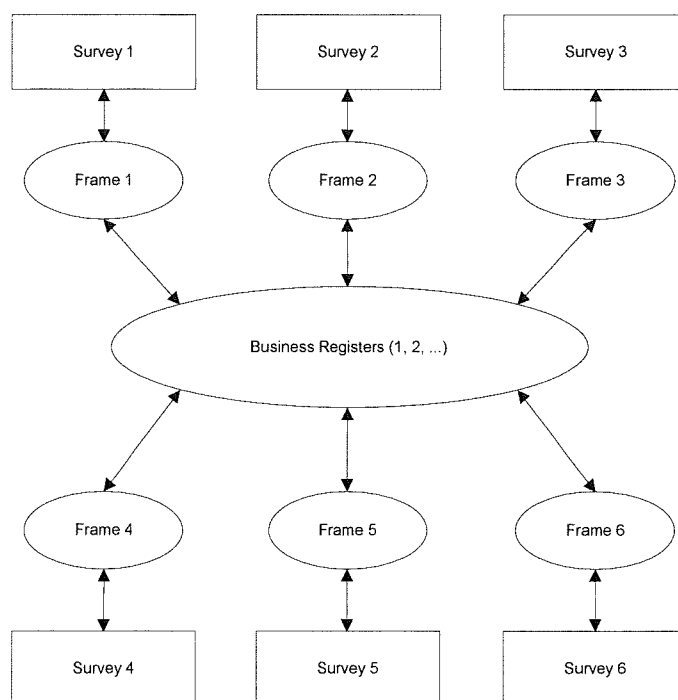


Figure 2.1

As can be seen in Figure 2.1, several different business registers may be used to construct a survey frame. Each business survey could have its own frame constructed from the registers in the middle.

Every survey is designed to satisfy all data requirements. A survey is most commonly a *sample* that is collected from a survey frame using some sampling method. If the whole population is examined, the collection procedure is called a *census*. A *business survey* is a certain type of survey in which the units of interest are businesses or parts of them. The *target population* is the set of units on which information is needed. The units in this set often cannot be assembled in full, mainly for operational reasons. This leads to the concept of a *sampled population*, which is a set of units from which the survey sample is actually selected. It is now possible to define the term *survey frame (sampling frame)*. A survey frame is the set of units comprising the sampled population (sampled population = sampling frame). Optimally, the target population and the sampled population would coincide. In practice, this ideal is hardly ever achieved, so the two populations differ more or less. The less they differ, the better.

A *statistical unit* is a unit defined for statistical purposes. A unit on which data are required is called a *target statistical unit* and a unit on which the data are actually collected is referred to a *collection unit*. Thus, the target statistical unit and the collection unit are not necessarily the same. In order to identify each unit in a survey frame, the frame must contain some relevant data items, which are described in the next paragraph.

Frame data are data items included in survey frame such as identification, classification, contact, maintenance and linkage data (Colledge 1995). *Identification data* in survey frames are data items that uniquely identify each unit, ie name, address, and certain numeric or alphanumeric identifiers. *Classification data* are items that are used to classify the units. Classification data are used to stratify the population for selecting the sample. The most important classification data contains some information on unit size. Naturally, there will often be several variables that measure the size of a unit (turnover, employees etc). Other important classification items in business surveys are industry and region. In some cases, classifications are needed for institutional sector and legal form of the firm. *Contact data* may include certain administrative data such as contact person, mailing address, telephone number, email address of a contact person, and possible previous survey response history. These are used to locate units in the sample. Finally, *maintenance and linkage data* are needed to update the survey frame as regards eg dates of additions and changes to the frame.

In addition to previously defined concepts, two different types of frames can be defined. First, a *list frame* is a list of units that contains relevant (frame) data. In business surveys, the starting point for constructing a list frame is usually some kind of business register or administrative data list that contains information on the set of businesses that are registered as active. Secondly, an *area frame* is a set of geographic areas from which areas are selected and the associated units enumerated.

A survey frame is thus a set (list, register) that typically contains some important data items and information in the form of *auxiliary variables*, x . These auxiliary variables do not usually contain the *output variable*, y , itself but are (or should be) closely related to it. The closer the auxiliary variables x and output variable y , the better and more accurate the results. This of course is directly related to the quality of the business surveys and statistics. However, there are not always relevant auxiliary variables available and thus there may not be any relevant information that could be used in planning the construction, the frame or the sampling. In addition, it is often a very complex process to get relevant auxiliary information that will help in the construction of a proper survey frame. Sometimes, however, when the target population is sufficiently small, the needed auxiliary information can be obtained via a business survey.

3. REGISTERS AND FRAMES: STRUCTURE, ERRORS AND MAINTENANCE

A typical business register is built using administrative sources (eg tax information from the national board of taxes) provided by organizations that produce administrative data as a by-product of their main functions. For instance, Statistics Finland has the comprehensive updated business register system of over 200,000 single enterprises and the new register of circa 720 enterprise groups (complex businesses). Also a private service enterprise can have a register of published business data (eg accounting data).

It is highly desirable to have an accurate and up-to-date business register, because of the direct impact on quality of survey data. Samples are collected from a frame based on this business register. The most important activities in maintaining a business register are updating the list and size information of the enterprises. Archer (1995) has viewed the requirements, maintenance, use and quality of business registers. Another recent review paper is that of Woodhouse (1999). Because registers and frames have many common requirements, problems and maintenance aspects, we consider only frames in the following.

3.1. Properties of frames

In the following, we list the properties that a frame should have, followed by the main types of frame errors in practice. Särndal et al, (1992, pp 10-11) presents eight properties that a frame should have:

1. The units in the frame are identified
2. All units can be found, if selected in sample
3. The frame is organized in a systematic fashion, eg ordered by size
4. The frame contains a vector of auxiliary information for each unit
5. The frame specifies the domain to which each belongs
6. Every element in the population of interest is present in the frame
7. No element not in the population of interest is present in the frame
8. Every element in the population of interest is present in the frame.

Properties 1 and 2 above are the minimum requirements. In practice, a frame is never perfect¹. It should be noted that a frame can contain several types of statistical units, but in using them we must apply care and logic in proper survey applications and analyses. As Colledge (1995) mentions, units of different sizes cause problems for register (frame) development and maintenance for small businesses and large businesses. In the case of large businesses, there are often difficulties in defining appropriate statistical units for complex organizations, especially when an organization has changed.

3.2. Definition of frame errors

There are three main types of frame error:

- a) Coverage (missing, extraneous, duplicate units)
- b) Classification (size, industry, geography)
- c) Contact (name, address, telephone number, email address)

In addition, there may be errors in stratification variables (eg incorrect revenue and employment class). Laniel et al (1996) show that it is not a trivial task to measure the level of these various errors, especially for complex units and missing units.

Errors in coverage and classification has negative effects on sampling estimation results. Especially as regards small frames and samples, missing units result in a downward bias; extraneous units may introduce an increase in variance; duplicates may introduce upward bias etc Errors in classification may be particularly dangerous in small frames because misclassified units may result both in bias and increased variance of estimates. These kinds of deficiencies are common with frames in different countries (Laniel et al 1996).

Laniel and Hidiroglou (1999) present a long list of ways to improve the quality of frames (coverage) and sampling estimation. One of the most efficient procedures is to combine several frames (*multiple frame methodology*) (see also Cox and Chinnappa 1995). Rotation and two-phase sampling reduce a misclassification problem.

¹ It is noteworthy that there are many differences between household (social) surveys and business surveys (Colledge 1995). One of these differences is in the frame and its updating (Laaksonen 1995).

3.3. Business frame: maintenance and quality

According to Hidioglou (1994) frame and sample maintenance consist of the following:

- (i) updating with new business (births)
- (ii) identifying enterprises on the frame that are no longer in operation (deaths)
- (iii) keeping track of classification changes on the frame
- (iv) reducing the response burden via sample rotation or partial replacement of the sample.

Business frames contain both administrative data and commercial data, which are linked and matched with each other². The practical problems (coverage, statistical unit, classification etc) using these data sources and updating links can be very different for small vs large businesses (Colledge 1995, Laaksonen 1995). Further, since distributions in business frames usually are highly skewed, there will be the problem of handling outliers (=outlying observations). Especially with small business frames, the shortcomings in maintenance points (i)-(iii) may be problematic because they may lead to biased estimation results. Moreover, in protecting against outliers, the use of robust statistical tools is very important for quality control. In the context of large sampling surveys, integrated control systems have been developed for survey processing (Bethlehem 1997). These are also applicable to small scale surveys.

Estimation is a powerful tool for remedying deficiencies in a frame. Laniel and Hidioglou (1999) emphasize the role of auxiliary information. They present three ways (forms) to use this information:

- (i) direct calibration³
- (ii) imputation
- (iii) combining administrative and survey data via modelling procedures.

The cyclical dynamics of an economy may have essential influences on the handling of frame errors. When the economy is on a strong upward trend, checking the change of size of the enterprises and the number of births of enterprises may be problematic. In the context of a downward trend, the elimination of deaths can cause difficulties.

Business populations and frames can also reflect various interbusiness relationships, eg when two businesses are owned by the same legal entity, or their activities are combined in the same set of financial records (Sigman and Monsour 1995).

The main factors affecting the quality of a business frame are accuracy, coverage and timeliness⁴. A well managed business register and frame should have quality standards (a system of quality indicators) that are defined in consultations with the various users. Maintenance and improvement of quality has become a real challenge. The 13th International Roundtable on Business Survey Frames included a session on the quality of the register (see Appendix).

² In general, there are differences of meaning between economic, accounting and statistical concepts.

³ By calibration we mean a class of estimation methods, ie multivariate regression techniques are applied to dependent variables (target variables of a frame/sample) and to independent variables (here: auxiliary variables). The auxiliary variables can be used in calculating sample weights and in evaluating the quality of the estimates.

⁴ Särndal et al (1992, pp 545-546) give Wright & Tsao's (1983) list of the five main steps for a program of frame construction and maintenance.

4. BANK OF FINLAND'S BOP SURVEY FRAME UPDATE

The business survey system has been used for nearly ten years to collect data on foreign assets and liabilities for compilation of the Finnish Balance of Payments (BoP) Statistics. During that period, the Finnish economy, especially the business sector and the financial institutions, has experienced remarkable structural changes. New financial institutions have emerged and old ones disappeared, and new financial instruments have been introduced in the markets. These developments have had many effects on the survey frames, and maintenance of the frame has become a most challenging task as we conduct business surveys in a rapidly changing economic environment. The frame for the Bank of Finland's monthly BoP survey was updated in spring 2000, employing an extensive frame survey.

In a small country such as Finland, business surveys have two main features: 1. The amount of large enterprises is restricted, and 2. The distribution of enterprises by size is highly skewed. However, experience gained by the Bank of Finland suggests that the distributions of BoP financial variables and current account variables are different, and both are skewed. The sensitivity to bias is a major risk in small business surveys. Missing data on a few large enterprises may cause highly disturbing effects on survey results. This problem can often be related to outliers occurring in survey studies (Hidiroglou and Srinath 1993).

The statistical unit is difficult to define consistently throughout the BoP surveys, since the frames contain both single and consolidated (complex) enterprises. The statistical survey analysis for the above types of enterprises was carried out both jointly and separately in the frame update project. The amount of BoP survey-relevant consolidated enterprises is limited to 30 – 40 in Finland. The problem became increasingly complicated by the fact that some financial instruments are used mainly by consolidated enterprises and not by affiliated enterprises within conglomerates.

The use of auxiliary information in defining the sampled population becomes particularly important when there is the combination of a small frame and financial variables. In the BoP frame survey project, the respondents were selected employing balance sheet data as auxiliary information for foreign assets and liabilities. The auxiliary variables were regressed against the target variables in the current sample. The financial variables are not available in the official enterprise register in Finland.

The deficiencies of the frame can be attenuated via several procedures. One effective procedure is the *multiple frame technique* (see Laniel & Hidiroglou 1999). For instance, Statistics Finland has recently applied this technique in designing a survey on foreign trade in services in 1999.

Figure 4.1 below describes the Bank of Finland's BoP survey frame reconstruction process.

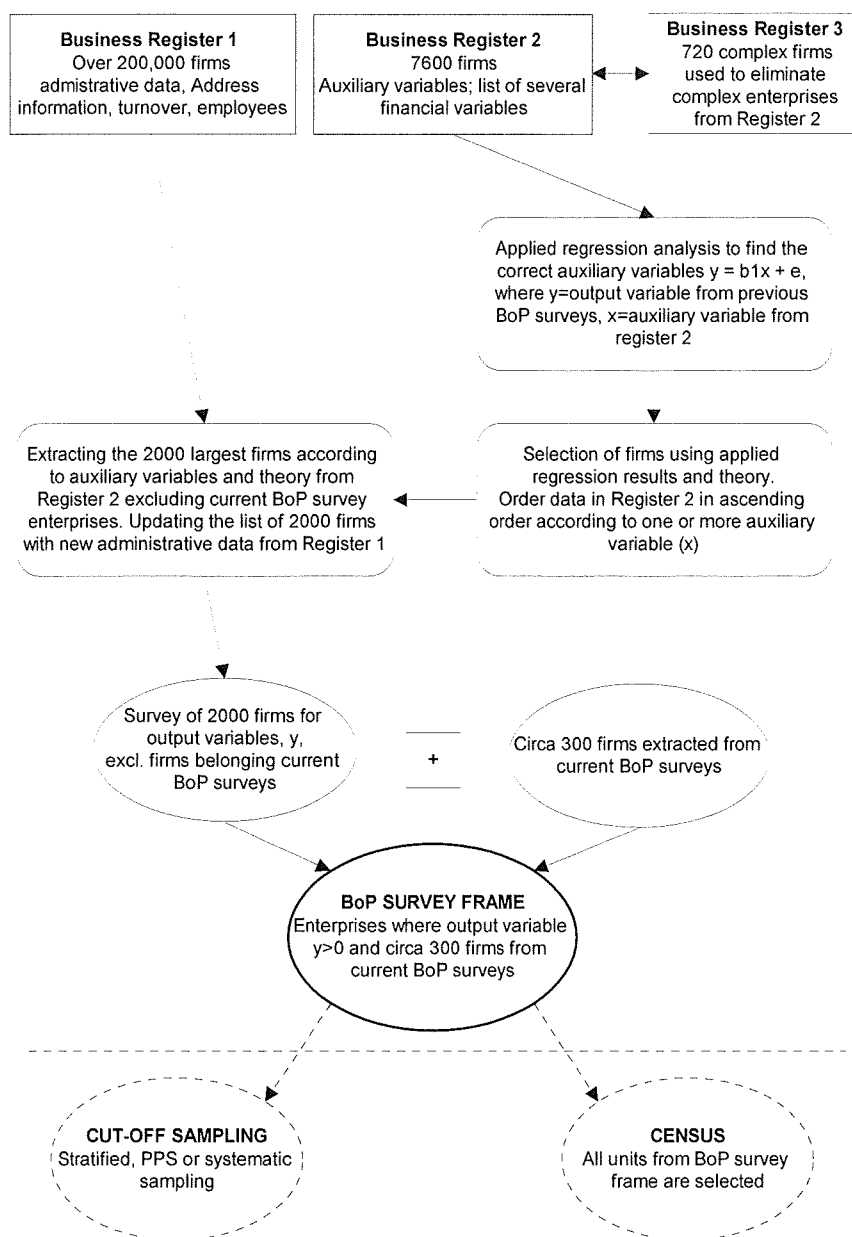


Figure 4.1

In practice, the maintenance of our BoP frame consists of annual basic checks of the list of enterprises, as well as the continual task of controlling and updating contact information etc.

With a small and restricted population, one might consider taking censuses rather than samples. A census might be preferable if moreover the data collection process is largely computer assisted and integrated, and the interviewers are not used for collecting data. The cut-off sampling technique is preferable with a relatively large frame population, especially if the distributions of target variables are highly skewed with many small enterprises and very few large ones.

5. CONCLUSIONS

The maintenance of statistical registers and frames is a dynamic and challenging task. In practice, register and frames are never perfect. The key problem is how to keep the quality high in changing conditions while meeting users' various needs. Source data for registers and frames, input and classification processes, and output should be continually evaluated and matched against the objectives. This whole system should include efficient and modern monitoring and quality control mechanisms, including robust statistical methods and tools.

Moreover, the rapid development of new financial instruments that provide and intermediate international financing presuppose continual careful study of their effects on the structure of the BoP frame. In summary, we should have well-defined quality standards (*quality profile* system) to support our survey work and management and to guide our improvements to these. The overall objective is that the costs and other frame maintenance resources should be allocated in an optimal manner.

6. REFERENCES AND BIBLIOGRAPHY

- Archer, D. (1995) Maintenance of Business Registers. In *Business Survey Methods*, edited by Cox et al. John Wiley & Sons, Inc.
- Bethlehem, J. (1997) Integrated Control Systems for Survey Processing. In *Survey Measurement and Process Quality*, edited by Lyberg et al. John Wiley & Sons, Inc.
- Biemer, P. P. and Fecso, R. S. (1995) Evaluating and Controlling Measurement Error in Business Surveys. In *Business Survey Methods*, edited by Cox et al. John Wiley & Sons, Inc.
- Colledge, M.J. (1995) Frames and Business Registers: an Overview. In *Business Survey Methods*, edited by Cox et al. John Wiley & Sons, Inc.
- Cox, B. G. and Chinnappa B. N. (1995) Unique Features of Business Surveys. In *Business Survey Methods*, edited by Cox et al. John Wiley & Sons, Inc.
- Hella, H., Sorsa, M. ja Viertola, H. (2000) BOP:n YSV-kyselyn otantakehikon uudistaminen ja päivitys: Apumuuttujien etsintäharjoitus yritysten tasetietojen avulla. Työraportti 15.3.2000, Suomen Pankki, tilasto-osasto.
- Hidiroglou, M.A. (1994) Sampling and estimation for establishment surveys: stumbling blocks and progress. ASA 1994 Proceedings of the Section on Survey Research Methods, 153-162.
- Hidiroglou, M.A and Srinath K.P. (1993) Problems Associated With Designing Sub-annual Business Surveys. *Journal of Business & Economic Statistics*, 11, 4, 397- 405.
- International Roundtable on Business Survey Frames, (annually) 1986-1998 (all presentations in cd-rom). Statistics Finland.
- Konschnik, C.A., Monsour, N.J. and Detlefsen, R.E. (1985) Constructing and maintaining frames and samples for business surveys. American Statistical Association, 1985 Proceedings of the Section on Survey Research Methods, 113-122.
- Laaksonen, S. (1992) Johdantoa aihepiiriin käsitteisiin julkaisussa: Laaksonen, S., Byckling, T. ja Rosenqvist, G. (toim.) *Yrityssurveys nyt ja tulevaisuudessa*. Tilastokeskus, raportti n:o 1992:5.
- Laaksonen, S. (1995) International comparability of business surveys: Some key factors based on European experiences. ISI'95 Session, Beijing, China.
- Laaksonen, S. (1997) Weighting and Auxiliary Variables in Sample Surveys. Suomen Tilastoseuran vuosikirja 1996. Helsinki.
- Laaksonen, S. Byckling, T. ja Rosenqvist, G. (toim.) *Yrityssurveys nyt ja tulevaisuudessa*. Seminaariraportti Suomen Tilastoseuran ja tilastokeskuksen 27.8.1992 järjestämästä tilaisuudesta. Tilastokeskus raportti n:o 1992:5
- Laniel, N. and Hidiroglou, M.A. (1999) Sampling and estimation issues for annual and sub-annual Canadian business surveys. ISI'99 Session, Helsinki, Finland.

Laniel, N., Mach, H., Finlay, S. and Dionne, S. (1996) Measuring errors on the business register. Proceedings of Statistics Canada Symposium 96.

Niemi, H. & Tourunen, K., (toim., 1996) Tilastoista tiedoiksi. Tilastokeskus.

Sigman, R.S. and Monsour, N.J. (1995) Selecting samples from list frames of businesses. In *Business Survey Methods*, edited by Cox et al. John Wiley & Sons, Inc.

Statistics Canada, Quality Guidelines, October 1998.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992) Model Assisted Survey Sampling. Springer-Verlag, New York.

Teikari, I. (1997) Special Features in Sampling Designs of Business Surveys. In *The Evolution of Firms and Industries, International Perspectives* edited by S. Laaksonen. Statistics Finland, Research Reports 223.

Thomsen, I. and Holmoy, A.M.K. (1998) Combining Data from Surveys and Administrative Record Systems. The Norwegian Experience. *International Statistical Review*, 66, 2, 201-221.

Woodhouse, T. (1999) Establishing Industrial Classification, Life Cycle and Coverage Quality Measurements In Statistics New Zealand's Business Register. The 13th International Roundtable on Business Survey Frames, Paris.

Wright, T. and Tsao, H.J. (1983) A frame on frames: an annotated bibliography. In T. Wright (ed.), *Statistical Methods and the Improvement of Data Quality*, pp. 25-72. Academic Press, New York.

7. APPENDIX

An Example of the Program of the Annual Roundtable Meeting on Business Survey Frames

The 13th International Roundtable on Business Survey Frames, Paris 1999

Contributed papers:

Session 1.	Progress report ⁵
Session 2.	Electronic business
Session 3.	Business demography
Session 4.	Technology
Session 5.	Globalization
Session 6.	Quality of the Register
Session 7.	France's day

For example, Session 6: Quality of the Register

The presentations:

No 1:	Quality aspects of the use of administrative data sources for the statistical business register in Germany
No 2:	Estimating the number of falsely active legal units in the French Business Register
No 3:	Establishing Industrial Classification, Life Cycle and Coverage Quality Measurements In Statistics New Zealand's Business Register
No 4:	Quality assessment and improvement in the Danish Business Register
No 5:	Quality assurance at the source – A strategic alliance between the Australian Bureau of Statistics and the Australian Taxation Office
No 6:	Product and process quality declarations in Statistics Sweden Business Register
No 7:	The development of a Single Business Register as an authentic register (Netherlands).

⁵ Separate report by each country.