



This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: The Design of Statistical Registers and Survey Frames

Contents

General section.....	3
1. Summary	3
2. General description.....	3
2.1 Survey frame design for integrated business statistics	3
2.2 The design of collection units.....	5
2.3 The flow and control questions of the survey frame design.....	6
2.4 The survey document	9
3. Design issues	9
4. Available software tools.....	9
5. Decision tree of methods	9
6. Glossary.....	9
7. References	9
Interconnections with other modules.....	11
Administrative section.....	13

General section

1. Summary

The specification of the *survey frame* is one of the most important parts of the design phase in the business process of surveys. In most cases, the *business register* (BR) meets the requirements to assign the survey frame of a business survey. Sometimes other information or another register is needed to supplement the BR in order to observe a given type of activity of the *enterprises*. Sometimes the unit of the *target population* is not able to supply data; in these cases another unit should be chosen as *data supplier*. Therefore the survey design phase has to define the unit and extent of target population, the *collection units* and the register as a base of the survey frame. The reference time of the frozen state of the register and the conditions in selecting and assigning the units of the survey frame also should be specified. The survey frame has to contain the attributes (contact, size, classification, other stratifying attributes, etc.) of the units demanded by the sampling, data collection and processing phases.

If no suitable register is available for the creation of the survey frame, then it's the task of the design phase to determine the way of its substitution or the sources to either correct the BR or to create a new register for the survey.

Survey design has to prepare for the integration of surveys with suitable assignment of survey frames, during the frame coordination. It has to assign the building blocks of the *populations* and the common classifications that might help to integrate data coming from different surveys.

2. General description

2.1 Survey frame design for integrated business statistics

The creation of an integrated economic statistical system should begin at the design of the statistical surveys. The output of the different statistics can be connected if

- their *populations* are harmonised by scope, units and reference time of the surveys and
- they use the same terms and classifications

The harmonisation can be *horizontal* when data of two surveys are available for the same population (or for the same *subpopulation*). For example, if the investment data and the output data come from two separate surveys then they can be linked with each other if they are observed for the same population or the surveys have common subpopulation.

It can be *vertical* when more subpopulations of the economy can be integrated into a bigger population without overlapping. For example, vertical integration is when the union of the same indicators that are collected separately for the industry and the construction are created. Another example is the creation of a bigger population by the union of the data about the big, medium, small and micro enterprises collected in separate surveys.

The modular approach of the system of business surveys - in contrast with the practice of surveys specified separately - has proven to reduce response burden and costs, in terms of human and financial resources.

For this purpose the best tool is to have a central *business register* (BR) that covers the most complete scope of economic units and contains all necessary information on *statistical units* demanded by

different surveys. The *register* identifies the units and describes their contact, demographic, economic and other attributes like economic activity, size, geographical location, legal form and activity status, etc. (See: “Statistical Registers and Frames – Building and Maintaining Statistical Registers to Support Business Surveys”). This BR provides a common base for defining the *survey frames* and *sampling frames* of the different business surveys.

The integration of the surveys is conducted if the population of the surveys is built from subpopulations as building blocks (Györki, 2012). The building blocks together can cover the whole economy in a given *reference period* without overlap. To describe these building blocks, the so called *master frame* can help us. The master frame comes from the *snapshot* of the BR and contains the statistical units for a given reference time. The building blocks are the subpopulations of the master frame. The population of the different surveys can be compiled from these building blocks. The same building blocks applied in the different survey frames provide for the comparability of the units of the surveys. On the base of the above principles the data can be integrated easily (See: “Statistical Registers and Frames – Survey Frames for Business Surveys”).

Usually not all statistical unit types are recorded in the national BRs as described in the Statistical Units Regulation (EEC, 1993). Sometimes the survey demands a statistical unit type that cannot be provided by the central BR or the completeness and quality aspects are not adequate enough to meet the expectations of the survey.

In that case there are other possibilities:

- To build on the available register, but to seek another statistical unit type that is able to report about the originally planned units. For example, if the need is the observation of the *kind of activity units* (KAU) but the register is not extended to that unit, then the enterprises can be asked about their activities. First choice is when the unit of the target population remain the KAU but the *data supplier* is the enterprise and the enterprise fills in as many questionnaires as many activities it has, that is, the KAU is the *reporting unit* of the questionnaire. The second one when the enterprise becomes the unit of the target population instead of the KAU and the KAU becomes only a variable within the questionnaire.
- If the aim of observation is a given type of activity that is not part of the BR but there is a *satellite register* for the demanded statistical unit then this satellite register can be the base of the survey frame. This satellite register has to be harmonised with the BR as a central register in the identification and classification attributes of the units, the rules of maintenance and the reference time. The harmonised registers create a common source for the survey frames and make the integration of different surveys built on the harmonised registers possible.

Example: the survey tends to observe the retail trade activity and a satellite register is available with information on the local units with retail trade activity (*local kind of activity unit*, LKAU). We have the following alternatives:

- Survey frame is built on the retail trade register and the unit of the population is the LKAU. If the timeliness and completeness of the satellite register are satisfactory, then we can choose this solution.
- Survey frame is built on the BR at enterprise level, but the condition of the population comes from the satellite register assigned only to enterprises with retail trade activity. In that case the

unit of the population is the enterprise, and the variable or the reporting unit of the questionnaire is the LKAU. For the update of the LKAUs with survey information then this alternative will provide higher quality for the observation.

- If there is no information for the assignment of the target population from the BR and their satellite registers then secondary, *administrative sources should be found* (UN ECE, 2012)
 - to create and maintain a register for the given population or
 - to compile a list of population units.

In the first case the task is the design of a new register harmonised with the BR (Eurostat, 2010). It should facilitate the connection with the BR through an identification number, the maintenance should take the activity status of the connecting units into account, and common classifications should be used.

Example: The observation refers to enterprises, institutions, universities with research and development (R&D) activity. A list of the *legal units* and their *local units* with living R&D activities is needed from administrative sources.

If these units can be connected to the units of the BR (e.g., to enterprises or local units) and there is no special attribute of the units for the R&D activity that it should be recorded, then the list serves only for selection of the population from the BR. So the survey frame is built on the BR and it contains enterprises (or local units) which have R&D activity.

If the aim is to observe the activity on, e.g., local unit level and these local units are not part of the BR or it is necessary to know other additional attributes to the units or the activity, then it will be reasonable to create a register to maintain the units with R&D activity. This register has to meet the criteria of the definition of the satellite register as far as possible. In this case the survey frame can be built on the new register.

2.2 The design of collection units

At the design of the population and frame of the survey the ability of population unit for data supply has to be considered. When the demanded information is not available at the given level of the enterprise another statistical unit type within the enterprise should be chosen.

Example: Financial data are not generally available at local unit level, only at enterprise level. In that case the enterprise becomes the unit of the target population. The frame should be built on enterprise level and the financial data can be asked only about the enterprises. However other (e.g., labour, production) data can be observed at local unit level.

At indirect surveys the data supplier serves information about the observed statistical units, reporting units. These units are usually the same, but sometimes they are different. The survey frame has to contain information on the connection of the data supplier unit and the observation unit. The reporting unit fills in as many questionnaires as many observation units belong to it.

Example: Social institutions are sometimes only partly independent; they are controlled by the municipalities, therefore financial data are available only at municipalities. In that case the data supplier, the *collection unit* will be the municipality and the observed reporting unit will be the social institutions. Of course if the social institution is an independent legal unit, enterprise, then it can report about itself, meaning that the data supplier and the reporting unit will be the same.

2.3 *The flow and control questions of the survey frame design*

To control the register and survey frame design for a survey/surveys the flowcharts below summarise the necessary steps and the main decision points.

a) The *first phase* of the design is the selection or creation of a register for the target population. The flowchart (Figure 1) describes the three main possibilities in this phase:

- To use an available register
- To modify the available register or create a new one
- To modify the observation with building on a new reporting or collection unit

The selection of the adequate solution can be supported by the following control questions:

- Do we have a register that can be the base of the target population?
 - Does the *register unit* match the target population unit?
 - Does the register cover the scope of the target population?
 - Does the register contain suitable attributes to access the data suppliers, to assign and stratify the population?
 - Does the *frequency of register maintenance* suit to the periodicity and reference time of the survey?
- If we don't have an existing adequate register then the following aspects should be considered:
 - Can we modify the maintenance of the existing register to meet the requirements of the survey?
 - Are there any sources that create and maintain a new register with the purpose to be the base of the target population of the survey? Do the units, attributes, periodicity of maintenance, quality of the sources meet the requirements?
- If we cannot create a register for the ideal target population, can we choose another unit type of the BR that can be the data supplier to report about the originally planned units?

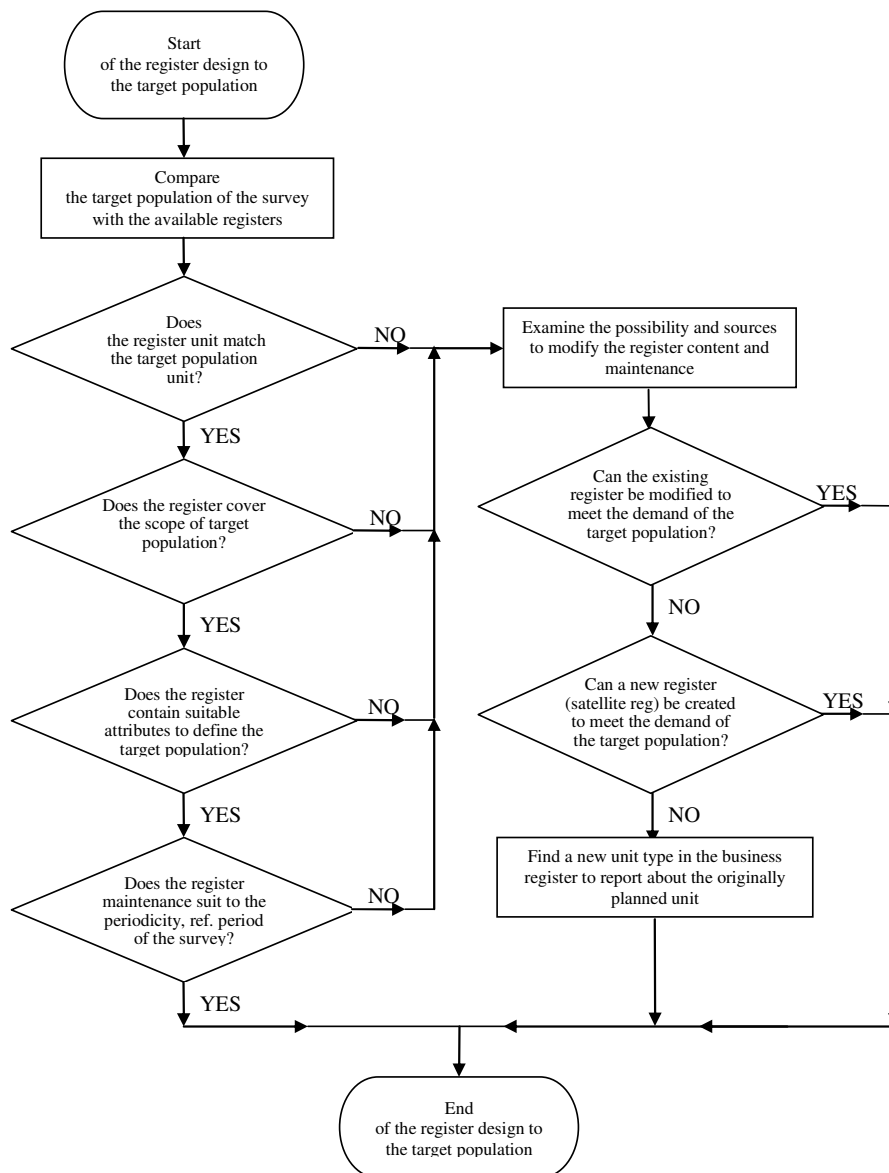


Figure 1. Register design to the target population

b) The *second phase* is the design of the survey frame taking into account the integration of the survey data with other surveys (Figure 2). In this phase we have to consider

- the connection of the given survey with other surveys (units, scope of population, reference time, variables) – to support the integration with well-defined population, subpopulations,
- the possibility to create a *master frame*, if more surveys use the selected register,
- the way and source of the assignment of the survey frame built on the frozen state of the register (or the master frame).

The selection of the adequate solution can be supported by the following control questions:

- What are the relations among the surveys if more than one is built on the BR?

- Are the levels of observation the same?
 - Do they refer to the same time, period?
 - Can the populations of the surveys supplement each other to a new wider population? Are there common variables of the surveys to serve statistics for the union of the populations?
 - Do the surveys have common subpopulations to integrate the statistics from the common parts of the surveys?
 - If the answers to the above questions are all yes, then which building blocks of the population can help the integration of the surveys?
- What are the reference times and units worth creating master frames to support the assignment of survey frames if more than one survey is built on the BR?

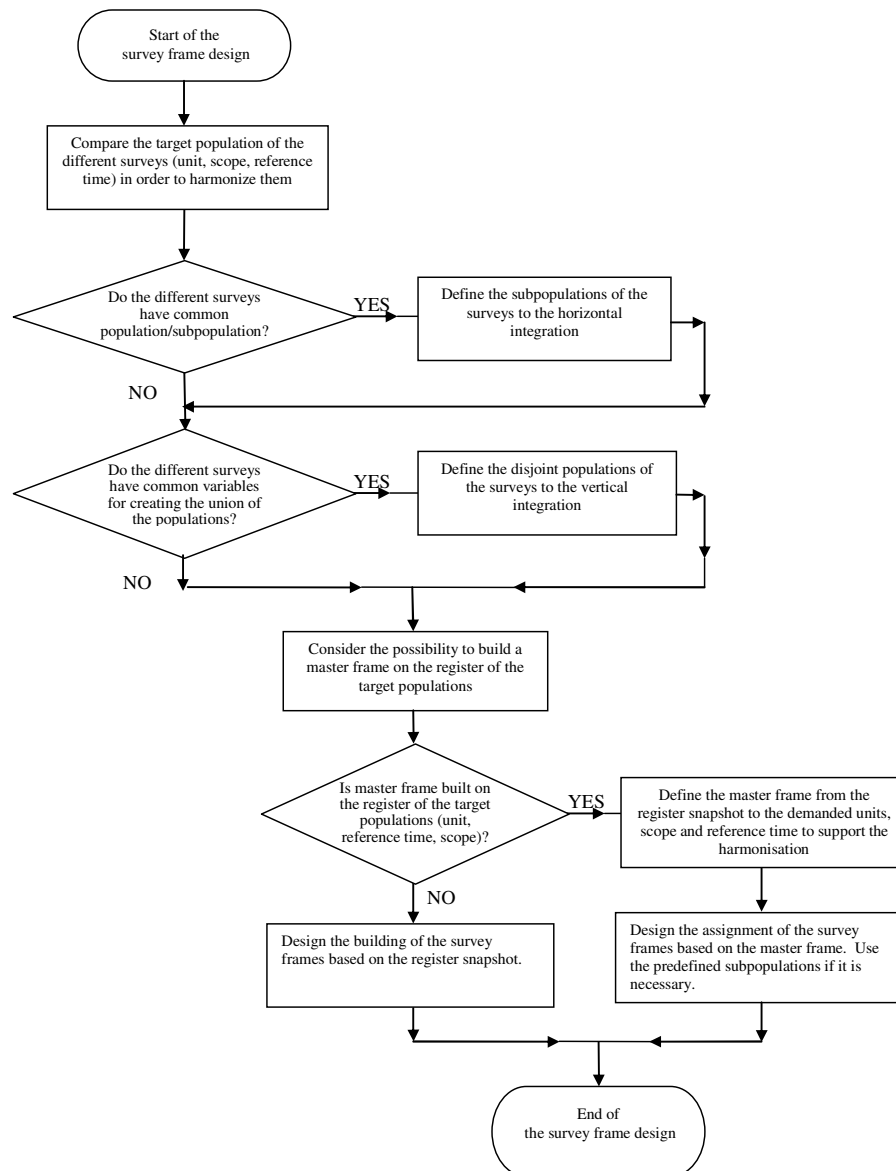


Figure 2. Survey frame design

2.4 The survey document

The aspects of the design and building of the survey frame have to be described in the survey document and the metadata of the survey. The survey document has to provide answers for the following questions:

- What is the scope, unit and reference time of the target population? What was the population of interest according to the user needs? Why do we differ from that?
- What is the source of the survey frame? What are the register and the register snapshot on which the survey frame is built on? Is there a master frame to select the population?
- What are the possible deviations between the target and survey populations? What characteristics show that an element of the frame population is not part of the survey (target) population? What kind of activity can help to correct the frame, to decrease the under coverage of the target population?
- What is the statistical unit, the data supplier and the possible reporting unit of the survey?
- What are the definitions of the important subpopulations the population is composed from?
- What connections are considered to other surveys to prepare the integration with them?
- Without the register serving as a base for the survey frame, is there any other register that can be used for statistical processing (for sampling, analysis)?

3. Design issues

4. Available software tools

5. Decision tree of methods

6. Glossary

For definitions of terms used in this module, please refer to the separate “Glossary” provided as part of the handbook.

7. References

Council of the European Communities (1993), Council Regulation (EEC) No 696/93 of 15 March 1993 on the statistical units for the observation and analysis of the production system in the Community. *Official Journal of the European Union*, L 076, 30/03/1993, 0001–0011.

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31993R0696:EN:HTML>

Eurostat (2010), *Business Registers - Recommendations Manual, 2010 edition*. Eurostat Methodologies and Working papers. http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-32-10-216/EN/KS-32-10-216-EN.PDF

Györki, I. (2012), GÉSA: The Tool for Survey Control, Quality Assessment and Data Integration. *Hungarian Statistical Review*, Special number **15**, 48–78.

http://www.ksh.hu/statszemle_archive/2012/2012_K15/2012_K15_048.pdf

UN DESA (2007), *Statistical Units*. United Nations, New York.

<http://unstats.un.org/unsd/isdts/docs/StatisticalUnits.pdf>

United Nations Economic Commission for Europe (2012), *Using Administrative and Secondary Sources for Official Statistics. A Handbook of Principles and Practices*.

http://www.unece.org/fileadmin/DAM/stats/publications/Using_Administrative_Sources_Final_for_web.pdf

Interconnections with other modules

8. Related themes described in other modules

1. User Needs – Specification of User Needs for Business Statistics
2. Overall Design – Overall Design
3. Repeated Surveys – Repeated Surveys
4. Statistical Registers and Frames – Main Module
5. Statistical Registers and Frames – The Populations, Frames, and Units of Business Surveys
6. Statistical Registers and Frames – Building and Maintaining Statistical Registers to Support Business Surveys
7. Statistical Registers and Frames – Survey Frames for Business Surveys
8. Statistical Registers and Frames – The Statistical Units and the Business Register
9. Statistical Registers and Frames – Quality of Statistical Registers and Frames
10. Sample Selection – Main Module
11. Data Collection – Main Module
12. Response – Response Process
13. Micro-Fusion – Data Fusion at Micro Level

9. Methods explicitly referred to in this module

- 1.

10. Mathematical techniques explicitly referred to in this module

- 1.

11. GSBPM phases explicitly referred to in this module

1. “2.4. Design frame and sample methodology” for frame design
2. “4.1. Select sample” for establishing the frame
3. All phases for register maintenance

12. Tools explicitly referred to in this module

- 1.

13. Process steps explicitly referred to in this module

1. “2.4. Design frame and sample methodology”
2. “4.1. Select sample”

3. All processes for the register maintenance

Administrative section

14. Module code

Statistical Registers and Frames-T-Survey Frame Design

15. Version history

Version	Date	Description of changes	Author	Institute
0.2	29-05-2012	first version as a separate module	Ildikó Györki	HCSO
0.3	16-07-2013	revised version after EB review	Ildikó Györki	HCSO
0.4	17-11-2013	revised version after EB second review	Ildikó Györki	HCSO
0.4.1	11-12-2013	preliminary release		
0.5	29-01-2014	revised version after EB third review	Ildikó Györki	HCSO
1.0	26-03-2014	final version within the Memobust project		

16. Template version and print date

Template version used	1.0 p 4 d.d. 22-11-2012
Print date	21-3-2014 17:37