



This module is part of the

Memobust Handbook

on Methodology of Modern Business Statistics

26 March 2014

Theme: Survey Frames for Business Surveys

Contents

General section.....	3
1. Summary	3
2. General description.....	3
2.1 The connection of registers and survey frames	3
2.2 Selection of the survey frame	6
2.3 The role of the master frame	7
2.4 Coordination of frames.....	9
3. Design issues	10
4. Available software tools.....	10
5. Decision tree of methods	10
6. Glossary.....	10
7. References	10
Interconnections with other modules.....	12
Administrative section.....	14

General section

1. Summary

The *survey frame* identifies and lists the units of the population altogether with their contact information, economic and geographic classifications and size categories. In the case of a sampling survey, it serves as a *sampling frame*. The survey frame supports accessing the data supplier and personalising and mailing the questionnaires. Furthermore, it has a role in the controlling and monitoring of the data collection phase, it helps to register and validate responses and to urge and evaluate non-responses. The survey frame gives information for the weighting, grossing, micro integration and analysing phases of the survey process as well.

For business statistical surveys, the main sources of the survey frame are the *business register* and the connecting *satellite registers* which record and maintain the statistical units and their characteristics.

The survey frame of a particular *survey instance* uses a *snapshot of the register*, a frozen register state for a given date. Since the business register serves as a base for different surveys, it is worth creating a *master frame* that is used as a common frame for these surveys. The master frame and the predefined *subpopulations* help to compose survey frames and support the integration of different surveys. Integrated survey frames improve the effectiveness of the data collection and the whole survey process and also help to reduce response burden.

2. General description

2.1 The connection of registers and survey frames

The *survey frame*, as it is discussed in the module “Statistical Registers and Frames – The Populations, Frames, and Units of Business Surveys”, is to describe *population* units and their attributes. In most cases not all population units are observed (sampling survey or representative survey); therefore the survey frame usually serves as a sampling frame to select the sample. It provides information to compose mailing lists for paper and electronic questionnaires, to personalise them and to monitor the data collection flow. As the survey frame describes the units and attributes of the population, it is the base for validation, weighting, grossing and analysis of the survey data.

For the survey frames the business register and the associated statistical registers have to provide for a reliable source where

- the coverage of the population is suited to the *target population*, the duplications and omissions in units are prevented
- the activity state and attributes of the units are suited to the reference time of the survey.

Statistical surveys use *snapshots of the register* rather than registers directly. The survey frames are selected from the register’s frozen state on a given date for the following reasons:

- the register keeps changing after the reference time of the survey, because of the systematic maintenance and ad hoc corrections,
- the survey frame (the population and their attributes) is expected to be stable, unchanged during the data collection and the whole period of the survey processing,

- the survey frame selection process and building on the survey frame have to be repeatable.

The survey frame has to reflect the state of the population in time in accordance with the reference time of the survey. Recently the countries of ESS don't apply a unified practice how they consider the changes of businesses during the year in the frames. There are two approaches:

- There are countries which use one frozen state of the register for the whole year, for the annual as well as for the short term statistics. In that case the population and the sample is unchanged during the year.
- There are others who use a separate frozen state for each short term survey instance, e.g., for each month. That takes into account the changes of the population during the year (birth, cease, etc.). The census type surveys and the "take all" type sample can use the up-to-date population. The selection of the "take some" sample can be once a year, but the weighting and estimation can refer to the changed population in computing "N". (See the topic "Weighting and Estimation")

The advantage of having one frame for the whole year is that the population and so the sample doesn't change during the year, thus the comparison is simple in the time series during the year as well as within the different statistical domains. Its disadvantage is that within-year changes are ignored. In certain cases this can cause bigger coverage error.

Having separate frames for the different *reference periods* brings the frame population closer to the target population in the different survey instances so the coverage error will be less than in the previous approach. It also has the advantage that the changes in the time series won't be so sharp at the beginning of the next year. On the other hand, this type of frame management can cause comparability problems between the reference periods of the year which have to be managed.

One solution to the problem of handling the changes of the attributes during the year is to use statistical attribute which is stable in whole year in contrast to the administrative attribute which follows the changes of the real world.

According to this distinction difference can be made between the maintenance and timeliness of the unit attributes:

- *administrative attributes* change according to the regulation of administrative sources because officially accepted values of these attributes have to be used in letters to the data suppliers, for example, the name, head quarter address, legal form, primary activity in the administrative source, the mode of book-keeping, composition of capital, etc.
- *contact attributes* – that support unit accessibility – change their values according to the changes in real life. The last known value is the best in order to reach the unit.
- values of *statistical attributes* – that support stratification and analysis of data – are stable during the whole reference year in order to assure the comparison of data, with the exception of changes causing significant distortion in the results of the survey. Statistical attributes are, e.g., the size category of the enterprise or its statistical principal activity.

Another aspect of the frame management is the role of the units ceased during the year. It is reasonable to keep the ceased units in the second and following periods when the same sample should be used as it was selected in the first period. In this case the estimation demands the unchanged rate of sample (n)

and population (N). But it is the fact that the response ability of a ceased unit is very low, response cannot be expected from them. It has to be managed in the data collection phase. If separate frames are used for the different periods of the year, then difference should be made between the ceased units with and without successors. The ceased units with successor can be substituted by their successor in the frame in order to avoid the duplication in the population.

Example: In Hungary, for annual survey with reference time of 31st December the copy of the *register units* according to their state on 31st December can be used. This snapshot contains the living units with their valid attributes on 31st December and the units that are ceased during the year with the last value of their attributes. In a monthly survey, separate survey frame is used for each survey instance. Each of them describes the state of the population for the given month. New units that come up in the course of the year have to appear in the frames of the following months while dead units can be omitted from the population or substituted with their successor. In the data collection phase the different response ability of the living and the ceased units are handled. In order to compare the different frames during the year statistical attributes for the principal activity, number of employees, size category of employees, turnover, and size category of turnover are applied. The value of these attributes is stable during the year, doesn't follow the changes of the indicators. The same statistical attributes belong to the units both in the short term frames and the annual frames.

A recommendation specifies for the business register what kind of attributes should be applied (EC, 2008, Eurostat, 2010). Most of these attributes are used in the survey frame as well as. The survey frame usually contains the following attributes of the register:

- the snapshot date
- type, identifier and name of the *statistical unit*
- identifier and name of the data supplier
- attributes supporting the connections to the *data suppliers* and *reporting units*, such as their name, seat and postal address
- attributes of the stratification, analyses of the units, like NACE, geographical location (NUTS, LAU), legal form, category of the number of employees and turnover value, activity state of the unit, etc.
- other topic-specific attributes

In the further phases of the survey process (sampling, data collection) the survey frame or the relating dataset is completed with newer attributes of the frame units like whether the unit is part of the sample, the intended and actual ways of response, etc.

Detailed contact information on the *collection units* (the name, address, e-mail address, phone number, etc. of the *data providers* and the *respondent* persons) is also necessary to carry out the survey. In practice, this is not part of the base register of the survey frame. This kind of contact information can change from survey to survey, from survey instance to survey instance, so it is maintained by the *surveying units* in a “contact register”. It is practical if all surveying units can see all contact

information, not only its own contact data but the data belonging to the other surveying units as well to use them for the common data suppliers.

2.2 *Selection of the survey frame*

In the GSBPM version 4.0 the creation of the frame belongs to the sub-process 4.1, sample selection. The units of the survey frame are selected from the register snapshot. The selection can be specified in different ways.

The best solution is to use an algorithm based on the *attributes of the register units* that delimit the units of the frame population. Built on the *business register* this kind of attributes can be, for example:

- *NACE* codes to define the *enterprises* with given *principal activities*
- the categories of the number of employees or turnover to define the given size of enterprises to be observed
- given *legal forms* of *legal units*, enterprises to involve into the population
- the activity state of the enterprises to select either only the active enterprises (with employees and/or turnover) or those as well that are in liquidation process, however, they can have some activity.

The selection by algorithm allows describing the algorithm only once and the result of the selection for the different survey instances will be different according to the attribute values of the register units in the different reference periods.

Sometimes the algorithm built on register attributes is not sufficient to define the population. This can be when:

- there are attributes in the definition of the population that are not part of the register, for example, the value of investment
- the subject of the survey is not the organisation or the enterprise but it is some kind of secondary activity that is not recorded by the register.

Example: one of the SBS surveys wants to observe the enterprises with industry as principal activity, that is, the NACE Rev.2 code is between 0500 and 3599. The size of the enterprises is defined by either the number of employees or the expected value of investment. The number of employees has to be over 20 persons or the value of investment over 1.5 million euro in a previous period.

In that case the principal activity and the employee number give an algorithm to select the enterprises from the business register snapshot and the list of enterprises to meet the requirement of the value of the investment is coming from another source (for example, from a survey in a previous period or the VAT data).

Example: Enterprises with foreign trade (Intrastat) activity have to be observed. To select the *frame population*, an external source is necessary to list the legal units that have a yearly intra-EU trade turnover above the exemption threshold. The list is created and calculated from data of VAT returns of the legal units.

The survey frame selection can be affected by feedbacks from results of the data collection phase of earlier survey instances of the given survey or other related surveys. The non-response causes and other survey paradata (information about the data supply) (Györki, 2012) inform us about the survey units. This information can be gathered on the questionnaire (to ask comment from data supplier) or in the data collection phase, e.g., during the urging step. They show whether a survey unit is not part of the target population or currently it is not active in the observed field. This information can be taken into account in repeated surveys. In the data collection phase of the next survey periods of the given year their different response ability can be taken into account. In the next year, however, the survey frame can be revised according to this information.

2.3 *The role of the master frame*

If the selection of more survey frames is built on a common snapshot of a register or registers, we talk about a *master frame* (Györki, 2012).

A master frame is quite useful in order to improve the efficiency and quality of the surveys, because it

- helps the coordination and comparability of the surveys,
- supports the effective coordination of samples,
- supports communication with the data suppliers,
- improves the efficiency of the data collections,
- utilises knowledge about the data supplier from other surveys,
- helps the common monitoring and evaluation of the data collections,
- supports the measure of the response burden.

The master frame may contain not only one statistical unit, but it can be created from more than one type of statistical units and from more than one register. For example, it can be composed from the union of the enterprises from the *business register* and farms from the farm register and non-profit institutions, covered by all the actors of the economy from the different registers. Similarly the different *LKAUs* from different *satellite registers* of the business registers like shops, accommodations, health care local units, etc. can be incorporated in the same master frame. On an international level the master frame can be based on the EGR (EuroGroup Register) in a similar way.

The master frame has to meet the following requirements:

- include register snapshots with same reference dates
- contain statistical unit with identified type
- contain units with unique identifiers
- be duplication-proof.

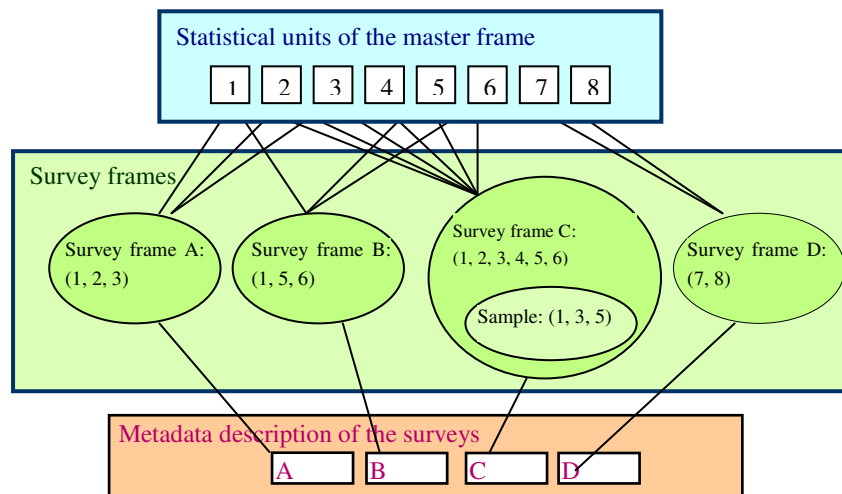


Figure 1. The connection between the master frame and the survey frames (Györki, 2012)

The master frame contains all common attributes (usually identification, contact, stratification attributes) as they are necessary for the survey frames built on them. The particular survey frames connected to the master frame can access and use these attributes.

Attributes characterising the connection between the frame unit and the given survey are part of the unique survey frames. The unique survey frame describes the *scope of data suppliers*; it makes distinction between the units of the population selected into the sample (the data suppliers) and those that are not. There are other attributes that can describe the unit role, type in the survey, in the data collection phase, for example:

- the way of mailing of questionnaires to the data supplier
- the intended and actual ways of response
- response willingness
- cause of non-response
- urging the response
- imputation methods
- etc.

This kind of paradata can be part of the unique survey frame or the connected paradata set in a unified structure. Most of them get values in the later sampling, data collection and processing phases. The unified attributes for the different surveys help the unified quality assessment and evaluation. For detailed description, see the module “Statistical Registers and Frames – The Quality of the Registers and Frames”.

2.4 Coordination of frames

The coordinated design of the target populations and survey frames can contribute to the integration of statistical data coming from different surveys. Statistical data integration means combining information from different administrative and/or survey sources to provide new datasets for statistical and research purposes. If this integration is taken into account during the development of questionnaires, definition and assignment of the survey frames and samples, data can be linked effectively. Otherwise success is not guaranteed.

Integration can be horizontal and vertical.

- *Horizontal integration* means linking statistical measures from different sources for a given population. For example, sales data from a retail trade data collection can be linked to data from a labour survey.
- In the case of *vertical integration*, a union of given statistical measures for different, separately collected subsets of a particular population is made. For example, unifying data on the land usage of agricultural organisations, collected by self-enumeration, with those of individual agricultural farms, collected by interviews, creates data for the whole national economy. A similar example is where the master frame is derived from the EGR (EuroGroups Register) and the populations by countries are given by the subpopulation of the EGR units.

Table 1 shows the possible directions of the integrations, vertically when a topic is built from more *subpopulations* and horizontally when a subpopulation is used by more topics.

Table 1. Features of data integration

Possibilities of data integration	Topics to be integrated horizontally				
	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
<u>Subpopulations to be integrated vertically</u>					
Subpopulation 1	X	X		X	
Subpopulation 2	X	X	X		
Subpopulation 3	X	X			X
Subpopulation 4	X				X
Subpopulation 5	X				

When the possible connections of surveys are reviewed during survey design, the definition of subpopulations can be prepared which the demanded new information can be composed from. At the assignment of the given survey frames from the master frame these predefined subpopulation descriptions can be applied. In this case, it is assured that the same subpopulation units are selected into the different survey frames, and later their statistical data can be joined together.

A survey frame can also be composed from disjoint subpopulations (populations without common units). If the subpopulation is observed by sampling, then the result for horizontal integration is better if the different surveys using the given subpopulation apply the same sample and weighting methods. It also helps the more effective validation of the topics.

Example: At survey design the following subpopulations are defined.

- Big enterprises of industry (employee number not less than 50) with full-scope observation (“take all” sampling)
- Small enterprises of industry (employee number between 5 and 50) with representative observation (“take some” sampling)
- Big enterprises of construction (employee number not less than 50) with full-scope observation (“take all” sampling)
- Small enterprises of construction (employee number between 5 and 50) with representative observation (“take some” sampling)
- Some other subpopulations for the other groups of principal activities ...
- Enterprises in financial sector
- Government, social security and non-profit institutions

For the SBS statistics different surveys observe the big and small industrial enterprises, the big and small construction enterprises, the financial sector, the government and social security sectors etc. The survey frames of these surveys are composed from the above predefined subpopulations.

Most of the questions in the questionnaire of the surveys are common for the topic of performance, labour force and investment. During the survey processing from the data of the different surveys the data of the whole economy can be united because the subpopulations describe the economy without overlap.

On the other hand the big enterprises are observed not only in the SBS questionnaires but also, for example, in the annual labour cost survey. Its survey frame is constructed from the union of the same subpopulations that are used in the SBS surveys for the big enterprises. So the labour cost data and the other labour force data from the SBS questionnaires can be easily linked together during the processing because their population is common.

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

European Parliament and the Council of the European Union (2008), Regulation (EC) No 177/2008 of the European Parliament and of the Council of 20 February 2008 establishing a common

framework for business registers for statistical purposes and repealing Council Regulation (EEC) No 2186/93. *Official Journal of the European Union*, L 61, 05/03/2008, 0006–0016.

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:061:0006:01: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

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 – The Design of Statistical Registers and Survey Frames
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
14. Weighting and Estimation – Main Module

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 Frames

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	26-01-2014	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 17:36