

GÉSA: The Tool for Survey Control, Quality Assessment and Data Integration

Ildikó Györki

Senior Advisor
Hungarian Central Statistical
Office
E-mail: Ildiko.Gyorki@ksh.hu

Supporting the flow of survey design and collection, the HCSO uses a standard metadata driven system, the so-called GÉSA system. This survey control system manages all economic and social statistical data collections of the office, observing the businesses and other institutions. The paper describes the aim and functions of the GÉSA system. It presents the unified tools of survey control for preparing the connection, integration of data collections through the way of the assignment of their survey frame. The paradata and other information connected to the survey frame and to the flow and result of the data collection are in the same structure and with the same code lists. It makes possible the unified monitoring, evaluation and quality assessment of the data collections.

KEYWORDS:

Statistical methodology.
Statistical survey.
Statistical system.

The transformation of the European Statistical System (ESS) is built on the Communication from the Commission to the European Parliament and the Council on the production methods of EU statistics (*Commission of the European Communities* [2009]). The so-called vision document (hereinafter referred to as vision) defines the objectives of reengineering to improve the coherence and comparability of data and to increase efficiency and cost effectiveness. Its two important elements are:

- replacement of the so-called “stove pipe model” in which statistics of different domains are produced independently, with an integrated model;
- standardization and integration of the formerly separated production processes.

The conception, methods and requirements of the vision fit well into the former developments and the strategy of HCSO.

The statistical office tried to create general systems for the phases of the statistical processing flow. These are built on metadata and their aim is to give a unified solution for as wide scope of data collections as possible. Besides, this unified design of the databases facilitates the integration of different statistical topics.

GÉSA (economic organisations and their data provision) is the survey control system for the observation of businesses and other institutions. It is the earliest metadata driven system in the HCSO which has been working since 1996. It is intended to give a general tool for supporting the tasks of the data collection phase of surveys, such as design, documentation, gathering of questionnaires, and evaluation of data collections. The solution is built on standard procedures.

At present, GÉSA manages and controls 132 data collections which account for 98 percent of all data collections where the data suppliers are institutions. In the case of more than 80, the questionnaires can be filled in and sent to HCSO via internet. Besides data collections, 50 administrative sources belong to GÉSA for the sake of unified processing. The number of modules, functions supporting the survey design, the data collection and its evaluation are more than 160. The system maintains a population with almost 2 million units, more than 380 000 data suppliers and 2-3 million pieces of questionnaires in one year.

As time went by, GÉSA has incorporated more and more functions and more and more data collections, its maintenance application has changed but its basic concept has remained stable.

Its main principles are the following.

- The terms and procedures connected to the data collections can be standardized. Giving a description of the unique characteristics of data collections in the metadatabase and using metadata, we can build metadata-driven procedures.
- The unique survey frames of data collections, their paradata¹ referring to the data suppliers and statistical units can be connected to a common master frame. Hereby,
 - the standard management, monitoring, quality assessment of all data collections maintained by the system;
 - common and unified communication with the data suppliers, and easy control of response burden become possible.

In the beginning, the GÉSA system dealt only with Business Register (GSZR)-based data collections, where data suppliers have a tax identification number. Later its scope was expanded to the data collections whose statistical units are institutions (dependent social institutions, non-profit organizations, etc.) without a tax identification number.

During this period, survey control became more precise as it was worked out for data collections with data suppliers which have to complete more than one questionnaire (for example by kind of activity units, local units, or settlements).

Concerning functionality, the most characteristic changes were brought by the widening scope of the proactive functions, thereby assisting data suppliers in responding in due time and in appropriate quality.

The support of work organization, the transparent task management and control are among the main functions of GÉSA. The system easily follows the changes in the organization of HCSO and in the responsibility of statisticians.

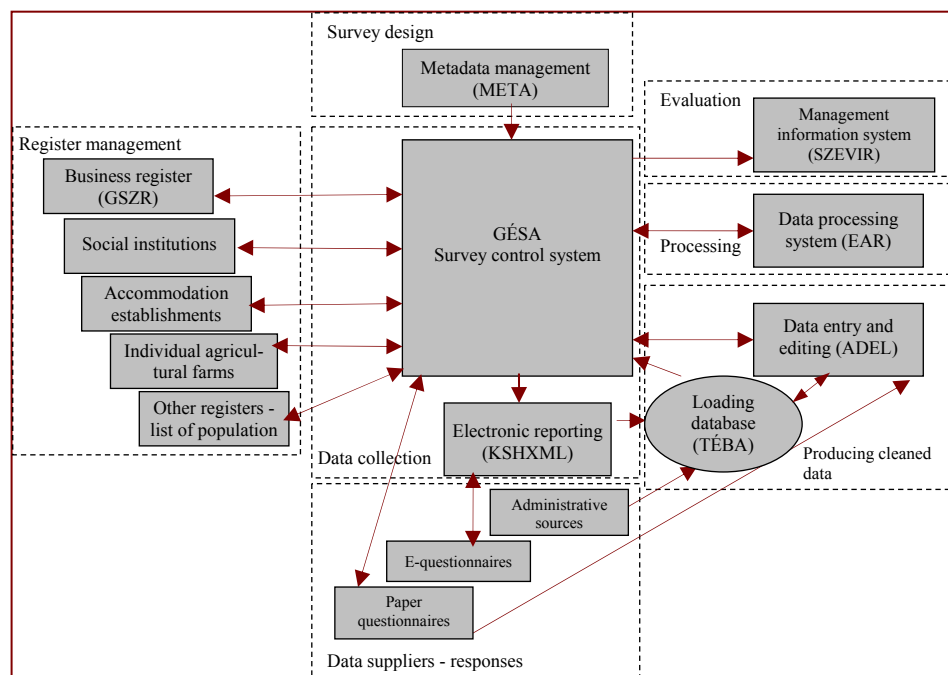
The main concepts, terms of the system and their relations are included in Appendix I.

1. The place and role of GÉSA in the statistical processing flow

GÉSA keeps track of data collections and questionnaires from survey design to loading and entering micro data into the database. Figure 1 shows the relation of GÉSA with the other phases and standard application systems of the processing flow.

¹ Paradata are data collected about the process of survey production to measure and improve quality and saving costs. Shortly, the difference between metadata and paradata is the following: metadata are data about data, paradata are data about process.

Figure 1. The environment of the GÉSA system – relation among the statistical processing elements



The wide scope of information, paradata which GÉSA gathers about data collections constitutes grounds for designing new data collections and for redesigning the existing ones after their evaluation. GÉSA provides a structured metadata description facility for documenting the information of the *survey design*.

Data collections involved in the system are built on several (business and other) registers. The snapshots of these registers connected to the reference time of data collections give input to the master frame. The *survey frame* and, in most cases, the sample frame of data collection are assigned from this common master frame, using the metadata description of the population and sampling.

GÉSA supports the personalized printing of the questionnaires and their mailing to the data suppliers, and forwards information to the electronic reporting system (KSHXML or the new system being under development) to inform the *electronic respondents* of their tasks. Several *proactive features* help data suppliers to send the questionnaires in time.

In the course of data collection, the work of statisticians is governed by the description of their tasks and that of the organizational units. All incoming questionnaires are registered. If some questionnaires are missing after expiry of the

deadline, *urging* letters of different degrees are sent automatically or individually to the data suppliers. *Response or non-response* is coded applying a unified nomenclature.

The progress made in data collection and in the processing of the questionnaires can be followed by both the statisticians and the management. Different statistics help the *evaluation* of the actual state and result of the process. From GÉSA information (on population, response, non-response, urging, etc., that is, from paradata), indicators on quality and response burden can be computed.

The GÉSA master frame, survey frames and paradata serve as a base for *other phases of the statistical production process*, for example, for loading micro data to the database (TÉBA – automatic registration and loading of questionnaire data), for entering, validating and correcting data (ADEL – frame system for data preparation), as well as for imputing, weighting, aggregating and making other procedures of statistical processing (EAR – unified statistical data processing system is under development).

The statistics about the result of the data collections are forwarded to the *management information system* every day where they can be analyzed in their tendency together with other information.

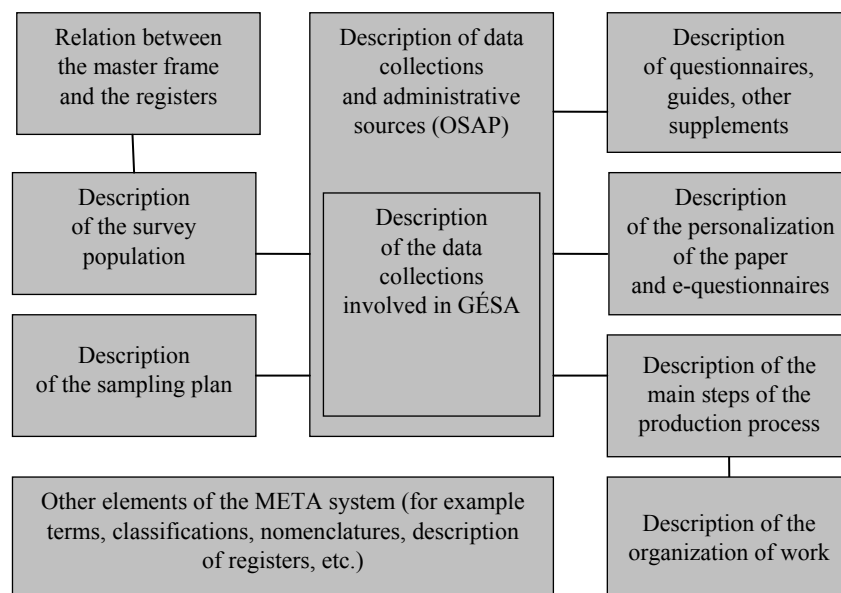
2. Description of data collections in the metadatabase

The metadata-driven operation of GÉSA is built on the description of the attributes of data collections, their supporting tools and steps of gathering data, as well as the division of work in the metadatabase. This metainformation is planned and entered in the survey design phase.

Figure 2 shows the overall diagram of the metadata necessary for the operation of GÉSA.

The primary aim of the *description of data collections* and administrative sources is to document the information needed to the yearly legislation about the National Data Collection Programme (OSAP). Besides legal commitments, it serves as a base for survey control, therefore this chapter as a part of the metadatabase contains all sources of statistics not only on mandatory but also on voluntary data collections and on the takeover of data from other (for example administrative) sources. The description renders account of the organizations that enacts and executes data collection, the topic, population, frequency, method (mail questionnaire, personal interview, etc.) and extension (full-scope, representative) of the observation, the size of the questionnaire, the response deadline, etc.

Figure 2. The main metadata groups describing data collections



As regards data collections included in GÉSA, these pieces of information are completed by more precise ones about the register such as the base of the survey frame, the type of data suppliers and statistical, reported units, the mode of printing and mailing the questionnaires, the possibility of electronic response, etc. The detailed information can be seen in Appendix II.

As it was already mentioned, assignment of a survey frame is built on such a snapshot of the given register that belongs to the reference time of data collection. To this function, the metadatabase has to describe the relation between the master frame (see Section 3) and the registers, as well as the exact *population*, rule and algorithm of creating a survey frame.

The base of sampling is the *sampling plan* for the representative data collections. It contains the rule and method of sampling, the description of strata and the attributes framing a stratum. The system selects the sample from the sampling frame in accordance with these rules and the sample allocation for the given period.

The description of data collections also includes *that of the questionnaire(s)*. If a subpopulation has to get another type of questionnaire that the others do, we have to identify and link it to that certain variant. Usually a guide and other supplements, code lists, a catalogue of terms, etc. belong to the forms, ordered to the data collections. Data suppliers get their questionnaires in a block whose composition (from different questionnaires, guides, etc.) is ruled also by the metadatabase.

Data suppliers get both paper and e-questionnaires in a personalized form. To standardize the operations related to, and the *personalization* of, the questionnaires, their design relies on *templates and rules*. The metadata-driven personalization is built on these templates and rules documented in the metadatabase.

A part of the metadatabase deals with the statistical processing flow. It describes the periods and deadlines of the various *phases of data collections*, for example, defines the deadline planned for response, or for the different urging types, according to the frequency of the given data collection.

To control and support the steps of data collections, it is necessary to describe the organization and staff of HCSO and the responsibilities connected to different functions and subpopulations. The task management built on this information defines the access to the data and functions of the survey control.

3. Creating a master frame

Most data collections are built on the Business Register (GSZR), in particular, on the description of legal units. Besides, there are other data collections where the survey frame comes from another register being only in loose connection with the Business Register and not integrated in it. These data collections include, for example, the personal interviews on individual agricultural farms that are listed in an independent register.

To provide a unified survey control system for different, independently maintained registers, GÉSA creates a unified survey frame, the so-called master frame. It is produced for each reference period from the given snapshots of the registers.

GÉSA differentiates two roles in the master frame. One of them is played by the data supplier units, for example, by legal units, institutions, individual farms, etc. HCSO has legal connection with them and expects their response. The other role is played by the statistical units which are either organizations as a whole or their parts defined by a given aspect, like activity, settlement, etc. Such a statistical unit can be:

- an economic unit, institution, nonprofit organization, individual farm, etc. (in their cases the data supplier and the statistical unit is the same);
- a part of the former category, engaged in an activity or in a group of activities (the so-called kind-of-activity unit – KAU);
- a settlement, a site of an organization (a local unit – LCU);

- an activity performed in a given site (a local kind-of-activity unit – LKAU), for example, accommodation, social activity, research activity, etc.;
- a special form of the previous category, the so-called specialized unit which aggregates the important activities at county level.

Both the data supplier and the statistical units are identified with eight characters. When they are legal units, the identifier is the first eight characters of their tax number. In other cases, the first digit of the identifier is a letter which characterizes the given type of units. Table 1 shows an example of data suppliers and statistical units in the GÉSA master frame.

The attributes characterizing the units include identification features, properties supporting the availability of units (name, seat address, postal address), demographic characteristics (date of establishment, beginning/ cessation of activities, operational status, etc.), economic, stratification aspects (principal activity, legal form, number of employees, county, settlement, composition of owners), links to other registers and to the owner, the maintenance organization.

They can be either administrative or statistical attributes. The first ones come from administrative sources, while the statistical attributes characterize the units according to their real activity. The value of the two attribute types may differ, for example, in the principal activity of the unit which can be either officially reported principal activity or statistical principal activity computed from value added. Similarly, the county where the headquarters is located can be other than the county where the unit performs its main activity. As regards operational status, the official status – legally ‘alive’ (active, under bankruptcy, liquidation or dissolution proceeding) or ceased (with or without successor) – does not always agree with the real status, as a lot of organizations do not wind up themselves.

For the purpose of statistical processing, the master frame makes difference between active and probably dead organizations, built on their official and non-official status. For the latter one the information is available from their announcements, data collections and tax returns. The population of active organizations is the so-called *base population* which consists of only a half (about 700 000) of the officially existing organizations.

Table 1

Data suppliers and statistical units in the GÉSA master frame

Data supplier identification number	Statistical unit identification number	Name of the statistical unit	Type of the statistical unit	Source
15329767	15329767	Szent István University	Economic unit	Business register (GSZR)
15329767	HA022187	Szent István University / Regional Knowledge Center	LKAU	Register of research and development units
15329767	HA025968	Szent István University / Pedagogic Faculty /	LKAU	Register of research and development units
15329767	K0139033	SZIE Dorm of Szent István University	LKAU	Register of accommodation establishments
15329767	K0169040	Zirzen Janka Dorm of Szent István University	LKAU	Register of accommodation establishments
16684801	16684801	Family Help Center Szentes	Economic unit	Business register (GSZR)
16684801	K0199087	Holiday Home of Szentes City	LKAU	Register of accommodation establishments
16684801	K0199088	Children's Camp of Szentes City	LKAU	Register of accommodation establishments
16684801	S0052898	Temporary Home of Families	LKAU	Register of social institutions
16684801	S0052902	Family Help Center – Temporary Home of Children	LKAU	Register of Social Institutions
12634048	12634048	Szeged Water Joint Stock Company	Economic unit	Business register (GSZR)
12634048	00019084	Szeged Water Joint Stock Company – Water collection, treatment and supply	KAU	Business register (GSZR)
12634048	00019091	Szeged Water Joint Stock Company – Waste collection and treatment	KAU	Business register (GSZR)
N0091559	N0091559	Golden Age 2004 Nursing Home Nonprofit Company	Non-profit institution	Register of non-profit organizations
M0010717	M0010717	Kiss Imre	Individual agricultural farm	Register of individual agricultural farms

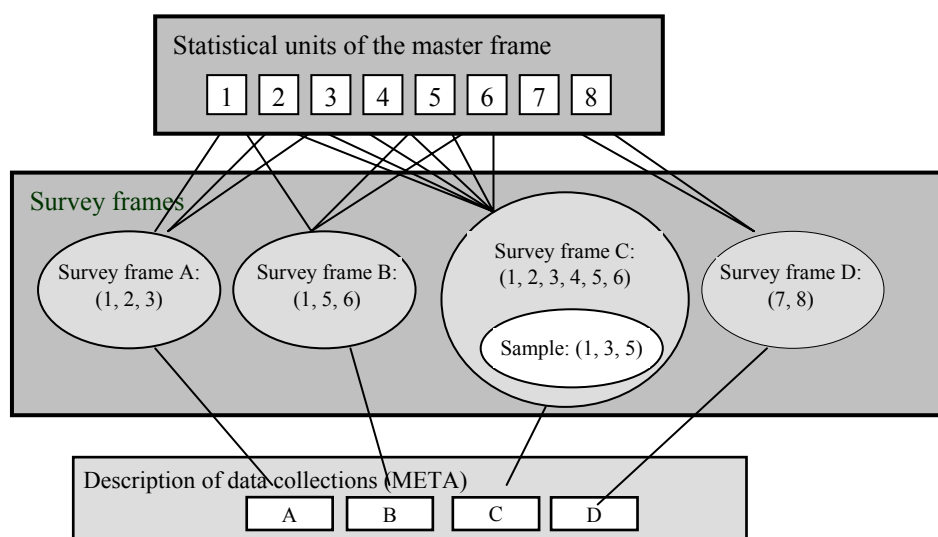
4. Assigning survey frames

The unified master frame serves as a base to select the survey frame of a given data collection by the algorithm, the definition of its population written in the meta-database. A unit of the master frame can be assigned to several individual survey frames.

To refine the coverage of the population, we can use the paradata of the given data collection for previous periods or those of other data collections for the same period available in the GÉSA system (such as the operating status of the data supplier, its response readiness, existence and operation of the observed activity, etc.)

Based on the former information, assignment of the survey frame is automatic. As regards annual data collections, it is built into the end-of-the-year snapshots of the registers. For data collections with short-term periodicity, survey frames are created for each reference period of the year to follow the changes in the organizations.

Figure 3. Assignment of the survey frames and samples in the GÉSA system



Concerning full-scope data collections, the survey frame, each unit of which has to supply data, defines the scope of the data suppliers. Representative data collections observe only a part of the survey frame: the units selected into the sample. In this case, the assignment of data suppliers depends on the method of sampling (which can be based on the survey frame) or another register may be used for this purpose. Where the master frame provides the sampling frame and the sampling plan describes random sampling, this process is a part of the GÉSA functions. In other

cases, sampling happens outside GÉSA, and the sample coming from other systems gives the scope of data suppliers of the representative data collection.

5. Conception of data integration in the GÉSA system

The provision of a base for the integration of statistical data begins with survey design. If we take integration into account during development of the questionnaires, definition and assignment of the survey frames and samples to data collections, data can be linked effectively. Otherwise success is not guaranteed.

Integration can be horizontal and vertical.

– The first means linking statistical measures from different sources for a given population. For example, sales data of a retail trade data collection can be linked to the data of a labour survey.

– In the case of vertical integration, we make a union of a given statistical measure for different, separately collected subsets of a particular population. For example, unifying data on the land usage of agricultural organizations, collected by self-enumeration with those of individual agricultural farms, collected by interviews creates data for the whole national economy.

Table 2

Features of data integration

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

GÉSA supports both types of integration as regards the description of the questionnaires and the specification of the survey frames.

5.1. Horizontal integration in the specification of the survey frames

A tool for realizing the horizontal integration is the specification, description of subpopulations in the master frame. The individual survey frames can be built up by several predefined and/or own subset definitions. The predefined, common subsets make possible that each data collection has the same data suppliers, statistical units as another, related one has. The data coming from data collections where the survey frames were created in this way can be directly matched to the common subsets. In GÉSA these predefined, identified subpopulations are called “segments”. Their description – in addition to their identifier and name – contains information helping the definition and assignment of the population, for example, the algorithm of the selection of the subpopulation from the master frame, the type of observation (full-scope or representative), the relation between the data supplier and the statistical units, the handling of organizations with different operating status, etc.

Examples for segments:

Code	Segment name	Segment short name
4194	Big industrial enterprises (B, C, D economic branch having more than 50 employees) observed in full scope from 2008 for annual and from 2009 for short-term data collections	Big enterprises of industry
4195	Industrial small enterprises observed representatively (B, C, D economic branch, between 5 and 49 employees) from 2009 for short-term data collections	Small enterprises of industry – short term
3789	Big enterprises of construction (F economic branch having more than 50 employees) observed in full scope from 2008 for annual and short-term data collections.	Big enterprises of the construction

The survey frame of data collections can be compiled from the segments included in this example as follows:

Data collection	Name of data collection	Segment	Name of segment	Type of segment
OSAP 1042	Monthly integrated statistical survey of industry	4194	Big enterprises of industry	Predefined segment
OSAP 1043	Monthly integrated simplified statistical survey of industry	4195	Small enterprises of industry – short term	Predefined segment
OSAP 1874	Quarterly integrated statistical survey of industrial, construction and financial enterprises	4195	Small enterprises of industry – short term	Predefined segment
		4194	Big enterprises of industry	Predefined segment
		3789	Big enterprises of construction	Predefined segment
		Predefined segment
		4409	Quarterly integrated statistical survey of industrial, construction and financial enterprises	Own segment of data collection

A predefined segment can be built not only on branches but also on any subsets of the master frame. In many data collections the municipalities are the data suppliers. Predefined segments belong to them as well.

Observation of municipalities:

Code	Segment name	Segment short name
3865	Municipalities from 2008 (without county and minority governments)	Municipalities

Some examples for data collections (among others) on municipalities:

Data collection	Name of data collection	Segment	Name of segment	Type of segment
OSAP 1206	Report on benefits provided in cash or in kind	3865	Municipalities	Predefined segment
OSAP 1832	Basic information on organizations providing social and child services	3865	Municipalities	Predefined segment
		1454	Child protection services	Predefined segment
OSAP 1761	Report on private accommodation establishments	3865	Municipalities	Predefined segment

5.2. Vertical integration in the specification of the survey frames

The formerly described segments contribute not only to horizontal but also to vertical integration. If the different subpopulations of a population are observed through different data collections (for example one data collection is conducted on industry, another one on construction and a last one on agriculture) which have common measures, indicators, and the subpopulations are specified with disjunctive segments of the population, measures for the total population can be computed from indicators characterizing the subpopulations.

If a significant proportion of the measures are identical, built on the same terms, and refer to the same period in the vertically connected data collections, we can connect them to a data collection group, to the so-called OSAP group. Within this, GÉSA supervises the disjunctive observation of subpopulations even when the characteristics of the units change.

Examples for the data collection groups are:

- the group of annual, integrated statistical surveys referring to different branches and group of branches, in the case of
 - big enterprises of industry;

- small enterprises of industry;
- construction;
- wholesale trade;
- retail trade;
- sale, maintenance and repair of motor vehicles and motorcycles;
- agriculture and services branches;
- financial intermediation;
- government, social security and nonprofit institutions;
- land area and sown area on 31 May, based on
 - self-enumeration on the agricultural enterprises of the business register;
 - personal interviews on individual agricultural farms.

6. Proactive support to the data suppliers

Following the assignment of the survey frame and data suppliers of the data collections, the GÉSA system helps, informs and supports the data suppliers with different tools so that they know their obligation of supplying data and the deadline of submission.

At the end of every year they get personalized information about their duties for the next year.

- A calendar lists those data collections to what the data suppliers are assigned as well as the deadlines of response, and gives an address where the questionnaire has to be sent.
- The questionnaires and their annexes (guide, classifications, terms, and other aids) are sent to the data suppliers in personalized blocks. If they respond via internet, they get only a sample questionnaire, but if they have previously mailed such a report, they get the same number of copies as that of the reference periods of the given year.

The calendar with response deadlines is not only sent to the data suppliers in a paper form, but it can also be accessed on the Hungarian site (http://portal.ksh.hu/portal/page?_pageid=36,1&_dad=portal&_schema=PORTAL) of HCSO by their identification number.

Figure 4. The calendar of data supply

The name of data supplier:

XXXXXXXXXXXX Trade and Service Ltd

Statistical identification number: XXXXXXXX 5610 113 15

**CALENDAR
of the sending deadlines of the questionnaires in 2011**

The data collection															
identi- fier	refer- ence year	name	sending address	sending deadline											
				January	February	March	April	May	June	July	August	September	October	November	December
1045	2011	Monthly Survey of Retail Sale	See here!	21	21	20	20	20	20	22	20	20	21	20	
1646	2010	Report on the sales of retail trade and catering by commodity groups	See here!	20											
1646	2011	Report on the sales of retail trade and catering by commodity group	See here!			20			20			20			
1872	2010	Monthly integrated statistical survey of agriculture, trade and services branches	See here!	12											
1872	2011	Monthly integrated statistical survey of agriculture, trade and services branches	See here!	HCSO Regional Directorate, Pécs Seat address: 7623 Pécs József Attila u. 10/A Mailing address: 7602 Pécs Pf.: 371											
1878	2010	Quarterly integrated simplified statistical survey of agriculture, trade and services branches	See here!	20											
1878	2011	Quarterly integrated simplified statistical survey of agriculture, trade and services branches	See here!			20			20			20			
2009	2010	Report on the number of job vacancies	See here!	12											
2009	2011	Report on the number of job vacancies	See here!			12			12			12			

The data suppliers, who already reported in the traditional way, get their questionnaires in a printed form. The personalized composition of the pack of questionnaires and the pre-printing of the known register data (identifier, addresses, name, some attributes) of data suppliers and statistical units facilitate the reduction of the data suppliers' response burden and the exact identification of the questionnaires.

In the questionnaire, the identification data (the OSAP number of the data collection, the reference period, the identifier of the data supplier and statistical unit) are printed not only with characters but also with a barcode. This, besides the formerly mentioned advantages, improves the effectiveness of the work of our colleagues responsible for data collection.

Personalization is built on the metadatabase chapter which deals with the description of printed materials (questionnaires, annexes, blocks, patterns, etc.). From the survey frames of various data collections, GÉSA packs and orders the questionnaires and their amendments belonging to the given data supplier. It selects the proper variant of the questionnaire and puts its copies into the pack based on the frequency of the particular data collection. It orders personal information to each page of the questionnaire. The unified personalization is helped by the templates. The printing office prints the blocks of questionnaires identified by a serial number, using the prepared information in the prescribed order. This method improves efficiency and quality and decreases the cost of mailing.

The data suppliers responding on the Web can see in the task list of the electronic reporting (data collection) system, which questionnaires they should complete in certain periods. The questionnaires of this system are also personalized just like the mailed ones. For the task list and personalization, information is provided by GÉSA.

Before the response deadline, the data suppliers and data providers (the agencies authorized to fill in and send the questionnaires in the name of the data suppliers) get an automatically generated reminder about the sending deadline by e-mail or – in absence of an e-mail address – by fax.

7. Supporting and evaluating data collection

The questionnaires sent to the office, the steps taken to ensure their submission, as well as other attributes and paradata of the data collection activity are stored, maintained by GÉSA in a unified data structure for all data collections and data suppliers, connected to the survey frames and the master frame. The collected information is detailed in Appendix III.

The data collection staff registers the questionnaires arriving by mail with the help of barcode readers or by entering their identifiers manually. The face of the questionnaires contains justification for a negative or a non-response if a blank questionnaire or a questionnaire containing a negative response is sent back. This information is managed by the so-called registration function of GÉSA.

Modification of the contact data of respondents, the time spent on filling in the questionnaires to weight response burden are maintained in a unified way in the phase of registration or entering the questionnaires.²

In the case of e-questionnaires, the registration data are the same as for the mailed ones: the arriving time, the reason for a negative answer, new, modified contact data, the time of questionnaire completion are all loaded to the survey frame in an automatic way.³

Submission of the missing questionnaires is urged. Similar to the formerly mentioned reminder sent after the deadline for arrival of the questionnaires, the persons responsible for completion and later their chiefs (response units) get an urging e-mail, fax, or letter. Urging has several degrees which is logged by GÉSA.

The reasons for non-responses or negative responses are coded according to a unified nomenclature. Their three types can be distinguished: reasons characterizing the problems/ errors of

- the units of the register describing the data suppliers and the statistical units:
 - false (107) or unknown (108) address;
 - not living or not active unit: ceased with successor (115), ceased without successor (101), under liquidation (102), under bankruptcy proceedings (103), not operating yet (104) or pausing activity (105);
 - active unit under liquidation (116) or under bankruptcy proceedings (117);
 - false classification: incorrect NACE (111), size by the number of employees (112)/ or settlement (county) codes (113);
- the register describing the observed activities of the units:
 - the data supplier has never performed the observed activity (201);
 - the unit has given up the observed activity (202);
 - the unit has paused the observed activity (203);
 - another reason that demands comment (204);
- readiness for data supplying
 - denial of response (801);
 - response is overdue (802) or will be overdue upon agreement (804);
 - connection with the data supplier has failed (803).

² See Data entry and validation of the ADEL system in Figure 1.

³ See Registration and loading electronic questionnaires of the TÉBA system in Figure 1.

The reason for non-response serves as a base to prepare the missing questionnaires for processing, marking the units whose data have to be imputed. Later, in the processing phase, the description of the statistical units is updated with the type of imputation.

8. Feedback, support of data processing

The attributes of the survey frames and paradata of the data collection are also used in the related systems and in the following steps of statistical processing.

The errors of the master frame (and registers), a part of which can be corrected, appear during data collection. The register (e.g. GSZR) gets a direct feedback on the correction and its result.

Such kind of errors can be:

- The address errors are reported by the post office in the mailing procedure when it sends back the questionnaires to HCSO. The register is corrected after making contact with the units having a false address.
- The under coverage of the population can turn out if the classification of the unit is false or the activity of the unit we want to observe isn't described in the register. By correcting the register and adding the missing unit to the survey frame, data collection can be improved.
- The reason for non-response or negative response can show the error in the operating status or other attributes of the unit. GÉSA passes these codes to the business register (GSZR). In the next data collection period, the selection of the survey frames is built not only on the official but also on the non-official operating status codes from data collections. These can make the survey frames more precise.

The processing phases following data collection use the survey frames of GÉSA.

- The data entry and validation phase approves only the questionnaires on statistical units of the survey frame. The validation procedures varying by subset identify the subsets by the attribute of the survey frame.
- The imputation procedure finds the missing units assigned for imputation in the survey frame. Subsequently, the survey frame of GÉSA gets feedback on the applied method.

- The weighting, grossing procedures build the estimation on the survey frame of GÉSA using the non-response codes and the attributes for partitioning the survey frame.
- The aggregation and other statistical procedures rely also on the attributes of the population units described in the survey frame.

The tools and methods applied in a well-prepared assignment of the survey frames – as it was detailed in Section 5 – promotes the record linkage and integration of data from different data collections, decreases the differences between the observed populations and the questionnaires describing them and improves the comparison of statistical data.

9. Monitoring, evaluation, response burden, quality assessment

During data collection, every step and information available from the automatic or manual registration of the questionnaires can be *monitored* in any moment. The following can be queried:

- the rate of responses sent in time or after urging;
- the distribution of the different media types used for submission of responses (electronic, paper mail, etc.);
- the proportion of the mailed questionnaires that were entered or loaded into the database;
- the rate of the missing or empty questionnaires in the case of which the reason for the deficiency was found out;
- the fact whether any arrangements shall be made either by the data suppliers or the staff carrying out data collection.

The subsets of the data suppliers belong to the responsibility of different statisticians. This is described in the metadatabase. Thus, the formerly mentioned monitoring features can be applied not only to the whole data collection but also to the subsets by statistician.

Data collection statistics can be *queried and evaluated* by both the supervisor responsible for and the staff carrying out data collection. This allows the latter to time and to organize its tasks.

Data collection statistics can be queried in time series as well, which makes it possible to follow up the tendencies of the changes in data supply, drawing attention to their possible setbacks.

The GÉSA system sends daily statistics, reports to the management information system.

Response burden indicators can be computed automatically from GÉSA information. Presently four indicators are composed:

- the number of the types of the questionnaires to be completed by the data supplier in a year;
- the number of the questionnaires to be sent by the data supplier in a year (taking their frequency and the number of the statistical units for one data supplier into account);
- the number of the fields in the questionnaires that shall be completed by a data supplier in a year;
- the average time of completion of the questionnaires by data collection. The data supplier may give the filling time voluntarily on the face of the questionnaire. GÉSA stores and processes this information by data supplier, data collection and year.

The indicators of response burden can be analysed on the basis of the most important attributes of the population. This tool makes an in-depth analysis of the indicators possible along the attributes. Among data suppliers, the critical groups and those institutions can be selected whose response burden is the highest.

The information collected during the data collection period through the mailing procedure of the questionnaires and registration of the arrived or missing forms, gives a picture of the accuracy of the registers and the survey frames based on them. It enables the automatic creation of various *product and process quality indicators* for the accuracy of the registers and data collections.

The so-called “expected” number of data suppliers (and statistical units) serves as a base for quality indicators. The difference between the total number of the data suppliers of the survey frame and the expected number of data suppliers is the number of units under liquidation or bankruptcy proceedings. They are “possible data suppliers” which are probably not active.

Two types of the coverage error of the survey frame can be computed.

- *Over-coverage*, the number of units in the survey frame not belonging to the population means the number of those units, who got non-response code 101, 102, 103, 104, 105, 118, 201, 202, 203, or 204 according to the reasons for non-response and negative response mentioned in Section 7. *The rate of over-coverage* is the quotient of over-coverage and the expected number of data suppliers.
- The measure of *under-coverage* is a bit more difficult. It can be inferred from the missing units of the survey frame that came to our

knowledge not from the base register but from another source. One part of these deficiencies can be corrected during data collection when we expand our survey frame with these new units, while the other part is cleared up only after that. The number of these units is called under-coverage whose rate is the quotient of under-coverage and the expected number of the units.

The errors of contact attributes and misclassification can be measured as well.

– *Error of the contact attributes*: The measure has two sources. One is the number of the questionnaires returned by the postal service in the mailing procedure and sent to the data suppliers again after correcting their address. The other is the number and rate of the address errors remaining in the survey frame, shown by the non-response codes (107,108).

– *Misclassification error*: The error of the attributes of the register (master frame) units (NACE, size categories by the number of employees, county, settlement, sector code, etc.) can be measured by the number of data suppliers in the case of which an attribute was corrected during the data collection phase or the non-response code was 112 / 113.

The main indicators of accuracy connect to unit response. GÉSA can provide automatic indicators for response, non-response or imputation of the units.⁴

– *Response rate*: the quotient of the number of the questionnaires (with response or negative response) and the expected statistical units.

– *Response rate with data*: the quotient of the number of the questionnaires with data response and the expected statistical units.

– *Unweighted non-response rate*: the number of the missing questionnaires in relation to the number of the expected questionnaires.

– The *weighted non-response rate* could be composed from the data of sample allocation and the non-response data of different strata. In practice it is not computed yet.

– *Imputation rate of units*: the number of missing statistical units where data were created by one of the imputation methods. It is related to the number of the expected statistical units.

⁴ Response, non-response or imputation indicators for the items are computed in the processing phase.

Some other indicators characterizing the data collections are the following.

– *The share of questionnaires according to the mode of submission* (mailed, electronic, e-mailed, etc.). 100 percent is the number of all arrived questionnaires.

– *The number and share of urges*: The number of the questionnaires sent prior urging or after 1, 2, 3, 4 urges (by phone, e-mail, letter, reminder, warning, etc.) correlating to the number of all arrived questionnaires.

10. Summary

The GÉSA system provides unified metadata-driven features for describing, maintaining all data collections observing institutions. The statisticians responsible for data collection can maintain all data collections belonging to them in one application system that makes their work easier. To collect a certain questionnaire, they can use paradata, reactions, remarks, contact information relating to other obligations of the given data supplier and the paradata of other data suppliers of the given data collection. These can help to improve the effectiveness of data collections. The standardized paradata, mailing modes, urging types, codes of response/ non-response and imputation provide the possibility of unified, efficient supervision and the evaluation of data collections.

The survey frames are managed in most of the statistical offices as part of single data collections. In some countries a need has also arisen to unify the data collection tasks and paradata, but we are unaware of a functioning system that gives such a standard solution that the GÉSA system does.

References

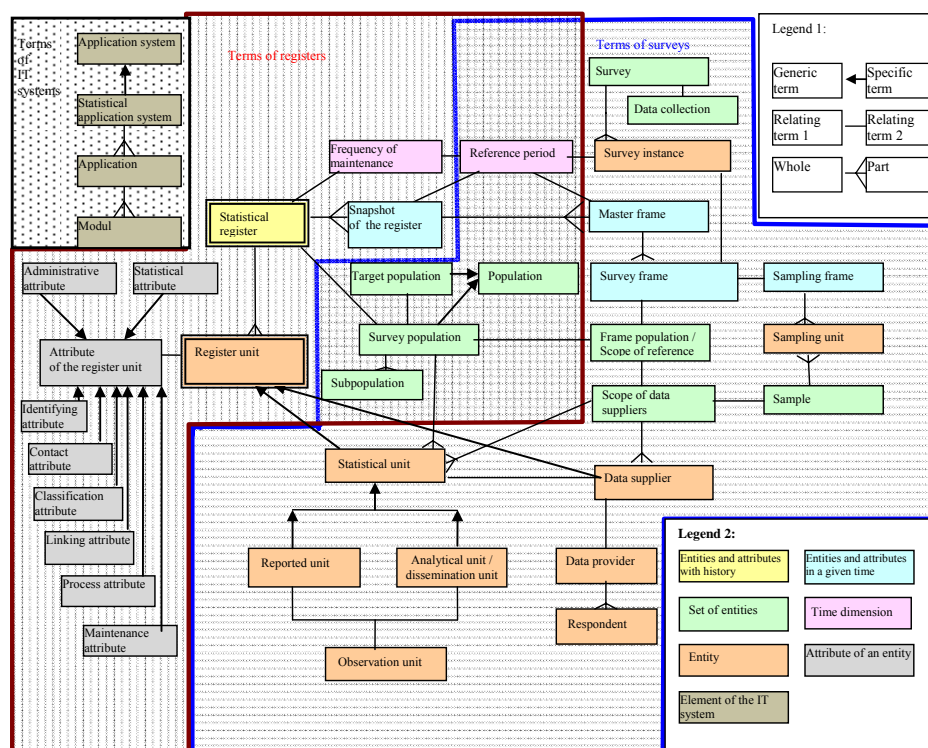
- BÉRARD, H. – PURSEY, S. – RANCOURT, E. [2005]: *Re-thinking Statistics Canada's Business Register*. http://www.fcsn.gov/05papers/Rancourt_Berard_Pursey_IVB.pdf
- CENTRAL STATISTICAL BUREAU OF LATVIA [2010]: *Metadata Case Studies*. <http://www1.unece.org/stat/platform/display/metis/Central+Statistical+Bureau+of+Latvia>
- COLLEGE, M. [2004]: *Assessing and Improving Data Collection Programmes and Survey Design Principles. OECD/UNESCAP/ADB Workshop on Assessing and Improving Statistical Quality "Measuring the Non-observed Economy"*. 11–14 May. Bangkok. http://www.unescap.org/stat/meet/wnoe/waisq_resource9p.pdf

- COMMISSION OF THE EUROPEAN COMMUNITIES [2009]: *Communication from the Commission to the European Parliament and the Council on the production method of EU statistics: A vision for the next decade*. COM /2009/0404. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0404:FIN:EN:HTML>
- 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*. Vol. 36. 30 March. pp. 1–11.
- 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. Vol 51. 5 March. pp. 6–16.
- EUROSTAT [2009]: *ESS Standard for Quality Reports*. Methodologies and Working papers. Luxembourg.
- GYÖRKI, I. – PAP, I. [2004]: *Metadata Driven Statistical Data Warehouse System at the Hungarian Statistical Office*. Work Session on Statistical Metadata. 9–11 February. Geneva. Working paper.
- GYÖRKI, I. – RÓNAI, M. [1999]: *Metadata Management*. Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata. 22–24 September. Geneva. Working paper.
- GYÖRKI, I. [1996]: *Survey Control as a Subsystem of the Statistical Information System*. Seminar on Integrated Statistical Information Systems and Related Matters. 21–24 May. Bratislava. Working Paper.
- GYÖRKI, I. [2000]: *Hungarian Solution for Metadata Model*. Conference of European Statisticians, UN/ECE Work Session on Statistical Metadata. 28–30 November. Washington, D.C. Working paper.
- KREUTER, F. – COUPER, M. – LYBERG, L. [2010]: *The Use of Paradata to Monitor and Manage Survey Data Collection*. *Joint Statistical Meeting of the American Statistical Association*. 31 July –5 August. Vancouver. Invited paper.
- MÉNARD, M. [2008]: *The Implementation of Tools to Support the Data Quality of the Business Register at Statistics Canada*. <http://www.oecd.org/dataoecd/13/12/41691288.pdf>
- MÉSZÁROS, Z. [2005]: *Szociális regiszterek és a KSH monitorozás néhány kérdése*. http://www.allamreform.hu/letoltheto/szocialis_ugyek/hazai/Meszáros-Zoltan_Szocialis_regiszterek_es_a_KSH_monitorozas.pdf
- STATISTICS CANADA [2011]: *Quality Guidelines – Survey Steps*. <http://www.statcan.gc.ca/pub/12-539-x/steps-etapes/4147793-eng.htm>

Appendix I

Terms of registers and survey control⁵

Relation of terms:



Administrative attribute: *Definition:* Administrative attribute of a register is a characteristic that can only be updated from administrative sources. Administrative attributes might change any time in accordance with the frequency of register maintenance. *Remark:* Administrative attributes cannot be modified, even if they are incorrect, however, reporting these errors toward the administrative sources is important. In some cases, correction of formal errors in the administrative data is allowed.

Analytical /dissemination unit: *Definition:* Analytical units represent real or artificially constructed units for which statistics are compiled. *Remark:* Analytical units are created by statisticians, often by splitting or combining observation units with the help of estimation and imputation.

⁵ The aim of the relation of terms is to create a consistent set of concepts for the statistical data collection phase. The sources of terms are the Hungarian metadatabase, the RAMON Eurostat's Concepts and Definitions Database, the OECD Glossary of Statistical Terms and UN/ECE Terminology on Statistical Metadata. Zoltán Vereczkei and Zolt Kóvári helped the author edit the definitions.

The goal is to compile as detailed and homogeneous statistics as possible using data on observation units.

Application system: *Definition:* Application system is a logically related group of applications designed to perform a particular task.

Application: *Definition:* Application is a coherent group of functions supporting the maintenance, process or inquiry of data of a given phase of statistical processing. It is called on-line application if the functions are performed in real time and the navigation among functions is supported by a menu. The term of batch application is used if the application performs the series of functions in the background.

Attribute of register unit: *Definition:* Attribute of a register unit is a regularly updated characteristic of a register unit. *Remark:* Attributes of statistical register units can be arranged in groups. Accordingly, attributes referring to identification, contact, classification, demographic characteristics, relation to other register units, attributes supporting register maintenance and statistical processes (for example organization of data collection, sampling, etc.) can be defined. In respect of maintainability and changes of attributes over time, administrative and statistical attributes are distinguished.

Classification attribute: *Definition:* Classification attribute is an attribute supporting grouping of units by a given characteristic of the population. *Remark:* Typical classification attributes are NACE, classification of units by legal forms, size categories by the number of employees, attributes used for settlement description (county, region, resort area, etc.).

Contact attribute: *Definition:* Contact attributes are attributes supporting localization and accessibility of register units. Such attributes are the name, address, telephone number, e-mail address, etc. of a unit.

Data collection: *Definition:* Operation of statistical processing aimed at gathering statistical data and producing the input object data of a statistical survey.

Data provider: *Definition:* Data provider is the organ (for example the bookkeeper) or person authorized to report data in the name of the data supplier.

Data supplier: *Definition:* Data supplier is the unit of the frame population from which the data about the reporting and observation unit can be retrieved. The organ carrying out the statistical data collection is in legal relation with the data supplier. The data supplier is asked /obliged for providing data. *Remark:* In the majority of the surveys, the data supplier reports about itself, therefore the data supplier and the statistical units are the same. In other cases, the two terms are different, one data supplier accounts for one or more statistical units (for example an enterprise reports about its settlements, a local authority reports about its institutions).

Frame population / reference scope: *Definition:* Frame population (reference scope) is the set of population units described in the survey frame. *Remark:* The frame population (reference scope) is usually the same as the survey population. In absence of a direct register on the survey population, the register of units that are able to report about the object of the survey, serves as a base for reference scope.

Frequency of register maintenance: *Definition:* Frequency of register maintenance is the time interval of the register content alterations. *Remark:* Registers can be maintained from different sources with different frequencies. In such cases, the most frequently used source determines the frequency of the register maintenance.

Identifying attribute: *Definition:* Identifying attribute is a synonym of the unique identifier.

Linking attribute: *Definition:* Linking attribute is an identifier of another register unit that is in a sort of relation with the given register unit. *Remark:* The type of relation can be under or upper dependency, source of maintenance, etc.

Maintenance attribute: *Definition:* Maintenance attribute is an attribute of the register unit describing the date of registration, update, cause and source of maintenance, validity, etc.

Master survey frame: *Definition:* Master survey frame is a snapshot of a register (union of registers) to assign the survey frames based on the given register (registers). *Remark:* An example of the master survey frame is the snapshot of the business register to define the survey frames of different economic statistical data collections. Another example can be the snapshot of the address register to make a common frame for population surveys. The common master survey frame, the common reference period helps the integration and linking of statistical data coming from different surveys.

Module: *Definition:* Module is a logical unit of the application created to perform a function, a given part of a task.

Observation unit: *Definition:* Observation units are the entities for which information is received. *Remark:* During data collection, this is the unit for which data is recorded. It should be noted that this may or may not be the same as the reported unit (the reported unit is, for example, the settlement, while the observation unit is the product being produced there).

Population: *Definition:* Population is the total membership or “universe” of a defined class of people, objects or events. *Remark:* Specific population definitions are target population and survey population. Target population is also known as the scope of the survey and survey population is also called as the coverage of the survey.

Process attribute: *Definition:* Process attribute is an attribute of the register unit describing a function, characteristic of a population unit in the statistical working process. *Remark:* Typical examples of process attributes are the survey related characteristics of institutions, such as their role in the population, their willingness to provide data, etc.

Reference period: *Definition:* Reference period is a time interval or a date to which the observed attribute (variable, indicator, measure) refers. *Remark:* Not only variables but data collections and their survey frames have reference period. The reference period of data collections are in accordance with the reference period of the observed variables (for example monthly data collection usually refers to indicators that reflect monthly or end-of-the-month situations of a phenomenon). Reference period of the data collection determines the reference period of its survey frame as well. The reference period of a survey frame is related to the date of the register snapshot (for example the first day of the month for monthly data collections).

Register unit: *Definition:* Register unit is the unit, entity of the register population with related descriptive information on identification, accessibility and other attributes. *Remark:* Register unit type – that is the collection of a given type of individual units – and register unit instance – that is a concrete, individual register unit – are distinguished. In the surveying process, data processing and dissemination phases, register units might function as data supplier, data provider or statistical (reporting, observation, analytical, dissemination) units.

Reported unit: *Definition:* Reported unit – or with other name, accounting unit – is the statistical unit about which information is sought. The data supplier accounts for as many reported units

as were assigned to it from the survey frame or belong to it on the base of the observed activity, phenomenon. If the data supplier has more than one reported unit, it has to fill more questionnaires. If an enterprise is the data supplier and reports about its settlements, then the settlement is the reported unit. *Remark:* In some other terminologies the reported unit means the organisation who submits the data in the name of the data supplier. In our terminology that term is called data provider.

Respondent. *Definition:* Respondent is the person who completes or approves the questionnaire what provides information about the reported unit.

Sample. *Definition:* Sample is the set of sampling units selected from the sampling frame. *Remark:* The aim of the observation based on a sample is to estimate the attributes for the whole population without the observation of the whole population.

Sampling frame. *Definition:* Sampling frame is an information-set for the survey population or the attributes of stratification being used as a basis for sample selection and in subsequent estimation procedures. *Remark:* Sampling frame often corresponds to the survey frame. In multistage sampling, the sampling frame of the primary and secondary stages may be different. In that case the sampling units of the secondary stage belong to the selected units of the primary sampling frame. For example the first step of sampling selects the settlements involved in the sample and the second one chooses the addresses in the selected settlements.

Sampling unit. *Definition:* Sampling unit is the entity or a group of entities of the sampling frame from what selection is made at a given stage of sampling.

Scope of data suppliers. *Definition:* Scope of the data supplier is the set of entities of the frame population assigned for data reporting from which data can be retrieved for the investigated population (statistical and observation units). *Remark:* In full scope data collection, the scope of data suppliers corresponds to the frame population. In representative or combined data collection, the scope of data suppliers is only a part of the frame population. It doesn't contain the statistical units of the frame population not selected into the sample.

Snapshot of register. *Definition:* Snapshot of a register is its frozen state on a given date. Instead of a register, snapshots are used for statistical processing because, unlike register units (that can be updated frequently), population units and their attributes must be constant during data collection and statistical processing.

Statistical application system. *Definition:* Statistical application system is an integrated group of applications supporting the given phase of statistical processing. It performs the maintenance, process and inquiry of data of the particular phase and gives interfaces to the relating applications, application systems.

Statistical attribute. *Definition:* Statistical attributes of a register are all non-administrative attributes of the register unit used in the statistical working process. Statistical attributes can be updated, corrected any time, irrespective of the administrative sources. For comparability reasons, the classification attributes remain unchanged within a year.

Statistical Register. *Definition:* Register is a continuously updated set of objects for a given population containing information on identification, accessibility of population units and other attributes, supporting the surveying process of the population. The register contains the current and historical statuses of the population and the causes, effects and sources of alterations in the population. Register data of population units are stored in a structured database. *Remark:* Statistical regis-

try is a less complex version of statistical register. Logging, change and history management are not mandatory in statistical registries.

Statistical unit: *Definition:* Statistical units are units of observation for which data are collected or derived. *Remark:* Statistical units observed during data collection are called reported units. In the analytical and dissemination phases, the terms of analytical or dissemination units are used.

Subpopulation: *Definition:* Subpopulation is a subset of a population. *Remark:* Subpopulation refers to populations that require different handling in the statistical working process. Subpopulations are usually specified to understand the distinguishing characteristics of these populations.

Survey frame: *Definition:* Survey frame is the set of survey population units together with their attributes referring to a given reference period. *Remark:* Survey frame describes the identification, contact, classification attributes of the survey population units for a given reference period.

Survey instance: *Definition:* Survey instance is a particular survey and reference period in which data are collected from respondents.

Survey population: *Definition:* Survey population is the population for which information during the survey process can be obtained. *Remark:* Concurrence or difference of survey and target populations is measured by coverage. In HCSO terminology, the survey population is called reference scope.

Survey: *Definition:* Survey is an investigation on the characteristics of a given population by means of collecting data (including censuses, full scope and sample surveys, usage of administrative sources) and estimating their characteristics through the systematic use of statistical methodology.

Target population: *Definition:* Target population is the set of units about which information is wanted and estimates are required. *Remark:* Synonym of survey scope.

Appendix II

The topics of the survey control metadata

Chapter A – General description

1. Identification of the data collection/ administrative source:
 - code of the program elements of the planning system;
 - OSAP identification number of the data collection and data source;
 - its name;
 - year of the reference period, frequency;
2. Identifier of a particular subpopulation, the reason for creating subpopulations;
3. Type of the data supplier unit;
4. Type of the statistical unit;
5. Mode of questionnaire printing;
6. Mode of personalization;
7. Guides, amendments to the questionnaire, and the rule of their maintenance;
8. The way of mailing the questionnaires.

Chapter B – Description of the population (survey frame)

The description of the population (or a certain subpopulation) is built on and specified from elementary subsets. The whole population is the union of the elementary subsets. There is no overlap between the subsets.

An elementary subset can be described by

1. referring to an existing, predefined subset or
2. specifying a new subset by:
 - the identifier of the subset;
 - the observation mode of the subset (full-scope, representative);
 - the way of maintaining the changes of organizations;
 - the operating status of the units of the subset;
 - the algorithm to assign the elements of the subset based on the attributes of the master frame (NACE, the size category of the number of employees, sector code);
 - the condition referring to the observed activity;
 - response/ non-response in the previous reference period;
 - definition of the frame (or a part of the frame) from external source, the way of maintaining the source list of statistical units.

Chapter E – Mailing, registering the questionnaires

1. Periods and time of mailing;
2. Characteristics of the composition of questionnaire sets and blocks;
3. Attributes of data collection and validation:
 - a) identifier and contact person of the organization unit responsible for data collection;
 - b) organization unit responsible for data entry;
 - c) organization units responsible for different validation and editing steps;
 - d) the fact whether the questionnaire was involved in electronic data collection;
4. Response deadlines.

Chapter L, F – Assigning the survey frame from external source

This chapter is filled in when the assignment of the survey frame of a data collection is built on other sources different from the master frame. Here the following have to be described:

1. The fact whether the information from another source aims at
 - a) giving new units of the survey frame or
 - b) disabling the assigned units of survey frame;
2. Possibility of common use of the list by other data collections;
3. Way of maintaining the changes of the organizations;
4. Priorities among the different sources of the survey frame;
5. Frequency of the external list;

6. Specification of selection if the list comes from the processing phase of another data collection.

Chapter R – Sampling in the case of representative data collection

The chapter is filled in when data collection has a not full-scope subpopulation. In this case, the chapter describes:

1. whether sampling forms part of GÉSA or the sample comes from outside the system;
2. the relation between the master frame and sampling frame, the way of sampling;
3. in the case of a common sample, the list of data collections using it;
4. the operating status of units;
5. the condition and frequency of sample refreshment.

Appendix III

Information on the survey frame and data supply (paradata)

1. Information on the survey frame

- a) Identifier of the data supplier;
- b) Identifier of the statistical unit;
- c) Identifier of the reference period of data collection;
- d) Identifier of the particular subpopulation;
- e) Identifier of the organization unit responsible for data collection;
- f) Way of observation (full-scope or representative observation, deletion from the observation);
- g) Strength of belonging to the population (voluntary or possible);
- h) Belonging to the scope of data suppliers (whether the unit is part of the sample or only an element of the survey population);
- i) Number of expected questionnaires (if the level of observation for the questionnaire is “deeper” than that for the statistical unit);
- j) Source of getting into the survey frame;
- k) Date of getting into the survey frame (in the case of follow-up recording);
- l) Date of deletion, modification (in the case of follow-up modification);
- m) Identifier of the user modifying the survey frame.

2. Information on the mailing procedure

- a) State of mailing of the questionnaire;
- b) Way of mailing (by post, web, e-mail, etc.).

3. Information on data supply

- a)* Arrival of the questionnaire (arrived, not arrived);
- b)* Way of data supply (post, web, e-mail, etc.);
- c)* Type of response (non-response, response with data, response without data (negative response), assignment for imputation, imputation by a given method);
- d)* Date of arrival;
- e)* Identifier of the user registering the arrival;
- f)* Reason of non-response or negative response;
- g)* Type of last urging (phone, reminder, warning, e-mail, fax, letter, urging e-mail, fax, letter, call-up letter, offence);
- h)* Date of last urging;
- i)* Identifier of the person urging data suppliers.

4. Comment on response or non-response

5. Filling time of the questionnaire

6. Contact data of the person

- a)* completing the questionnaire by chapters (name, phone, e-mail, identifier of the agency representing the respondent);
- b)* responsible for approval (name, phone, e-mail, post).

7. Log of the events of data collections

- a)* Identifier of the event (registry of the questionnaire, registry of response/non-response, different phases of urging, etc.);
- b)* Date of the event;
- c)* User of the event.

8. Demand for exemption from data supply