



Economic and Social Council

Distr.: General
5 July 2011
English
Original: French

Economic Commission for Europe

Conference of European Statisticians

Group of Experts on Business Registers

Twelfth session

Paris, 14–15 September 2011

Item 5 of the provisional agenda

Reducing respondent burden – the role of business registers

Construction of a statistical database linked to the inter-administrative business directory of the National Enterprise and Establishment Register Database: registering samples to measure the response burden

Document prepared by the National Institute of Statistics and Economic Studies (INSEE) of France

Summary

For many years, France has maintained an administrative directory using the legal entity as its base unit. Since the notions of “enterprise” and “legal entity” have until recently been regarded as one and the same in the French business statistics system, the directory, although basically administrative, has been widely used for statistical purposes.

In the mid-2000s, a major programme was launched to re-engineer structural business statistics and, in particular, to introduce a true definition of “enterprise”, which would no longer be equated with a legal entity. The establishment of a statistical business database has formed part of the programme from the outset.

The database now under construction will make it possible to bring new statistical units (in particular “enterprises”) into play, to identify their constituent legal entities and to ensure homogeneity of statistical characteristics in the business statistics system.

In time, the database will allow the response burden of businesses to be measured – a frequent request that cannot easily be met at the moment (see “French methods and practices regarding the statistical burden”).

I. From SIRENE an administrative directory...

1. Up until 2010, France had a comprehensive administrative directory, which is described in detail in “Improving the quality of business registers” (see references below).
2. All legal entities and all their local establishments (branches) in France are registered in a national business directory called SIRENE.¹ This forms part of the statutory framework in which INSEE is responsible for identifying legal entities and their local establishments on the basis of documents transmitted by the business procedures centres (*centres de formalités des entreprises* – CFE), which were set up to deal with all the administrative procedures to which legal entities are subject. An Act of 11 February 1994 established a unique identification number for businesses, which is now a linchpin of the inter-administrative system, since its use is mandatory in all relations between administrative bodies and businesses.
3. SIRENE records the civil status of businesses, their legal form, the address of their headquarters, the date they were established and, where appropriate, the date they ceased operating. The addresses of local establishments are indicated, as well as the date of their creation. In addition to these details, for each entity listed, the register contains the number of employees on the payroll and the code assigned by INSEE to identify the principal business activity carried out (*code d'activité principale exercée* – APE), which is based on the French general economic classification system (*Nomenclature d'Activités Française* – NAF).
4. Because of this wealth of information, in addition to its statutory inter-administrative functions SIRENE also used to be used for statistical and trade purposes and, over time, non-statutory “statistical” data such as indicators of economic importance, variables relating to the number of employees at each year end and annual turnover figures have been added to it.
5. Using SIRENE to produce statistics, however, poses two major problems. First, it only covers legal entities (companies and their local establishments). As increasing numbers of corporations go global, the assimilation of a “legal entity” to an “enterprise”² ceases to be appropriate in some cases. Those cases are, unfortunately, economically significant, and the assimilation of legal entities to enterprises can, through the double counting of turnover figures, for example, give a distorted picture of the economy.
6. Second, one cannot “do what one wants” with an administrative directory: entities cannot be added or removed without notifying the associated administrative bodies or the registered entities themselves, which means that SIRENE lacks the flexibility that a more statistics-oriented database would have.

¹ Established by amended decree No. 73-314 of 14 March 1973.

² I.e. “The smallest combination of legal entities constituting an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making on matters including the allocation of its current resources.”

II. ... to a statistical database, SIRUS

A. Structural overhaul of annual business statistics

7. A major overhaul of annual business statistics was launched at INSEE in early 2006. Known as RESANE (Refonte des Statistiques Annuelles d'Entreprises), the programme has been being phased in since 2009. It has three objectives:

- To reduce the response burden on businesses by using administrative sources as systematically as possible
- To improve productivity generally by completely re-engineering the processes used to date to produce structural business statistics
- To improve the quality of the statistics generated in terms of timeliness (productivity gains) and relevance and, hence, to adhere more closely to the exact definition of "enterprise" established in European Council Regulation (EEC) No. 696/93 by taking the "corporate group" dimension into account in the preparation of business statistics³

8. The point of the third objective, "taking the 'corporate group' dimension into account", is to monitor two additional types of entity which the administrative directory had not previously covered: enterprises and corporate groups.

9. To that end, two further projects have started in parallel:⁴

(a) The first, called "Profilage" (profiling), consists of identifying "enterprises", within the meaning of Regulation No. 696/93, among groups of legal entities. The project will advance in stages:

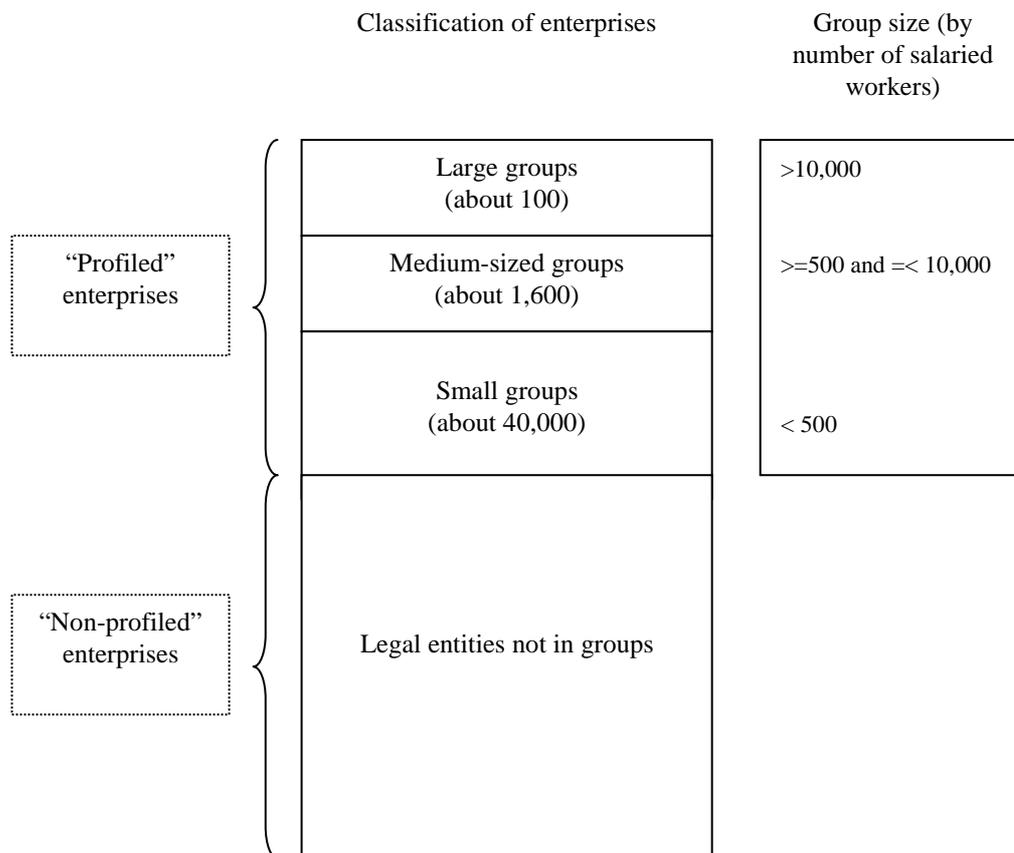
- For about 100 of the largest and/or most complex and/or most "demanding" groups (representing 22% of salaried workers and 26% of value added), enterprises will be defined by agreement with the groups from which they have been assembled. Profiling will thus be done manually, and the enterprises will be defined at a yearly meeting with the group;
- For almost 38,000 "small" groups (those with less than 500 workers on the payroll or with no more than two French subsidiaries), the enterprise will be assimilated to the group as a whole, and its characteristics (sector of activity, for example) will be obtained through an algorithm. Such groups represent 19% of salaried workers and 22% of value added;
- Approximately 1,600 intermediate groups (representing 17% of salaried workers and 17% of value added) will be profiled using a semi-manual, semi-automatic method, less refined than the one used for the individual profiling of large groups. The method might be based on the results of a specific survey designed to determine the degree of consolidation within the group.

All enterprises defined in this way will be called "profiled enterprises". Legal entities that do not belong to a group will still be classified as enterprises.

³ A group is a set of juridically independent legal entities linked by capital ownership, one of which wields decision-making authority over the others. In France, corporate groups have been monitored through surveys since 1980.

⁴ The same teams are working on both projects!

Figure 1
Definition of enterprises within groups



(b) The second project consists of establishing a statistical database whose main unit is the "enterprise" as broadly defined by the project outlined in (a) above. This project started up in 2009 but was conceived in 2005 during the discussions leading to the RESANE programme.

B. The SIRUS database

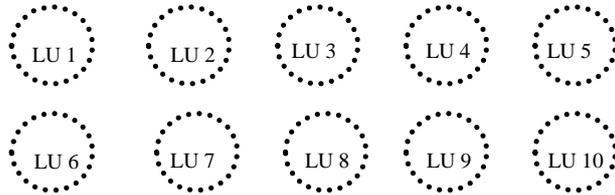
1. The processing of new statistical units and the links among them

10. How is this new database actually constructed? SIRUS registers enterprises identified by the "Profilage" application and then recovers, from the SIRENE administrative directory, the legal entities that do not fall within those profiled enterprises: these entities, which are therefore independent (i.e. not linked to a group), must therefore be "enterprises".

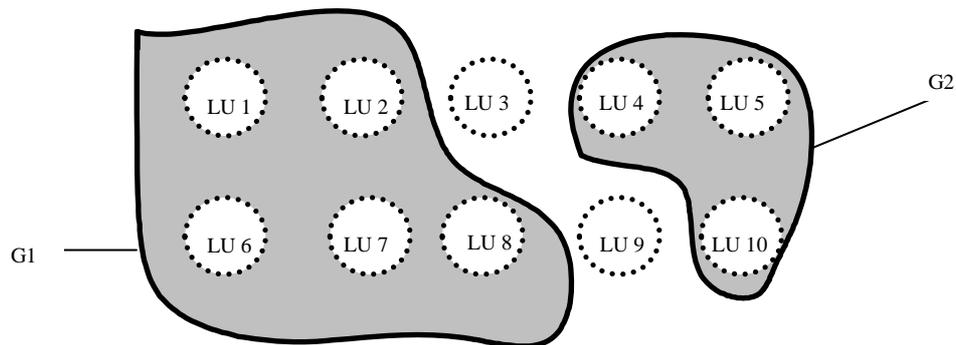
11. SIRUS is not limited to "enterprises", since it will record other statistical units of interest, as its name (identification system in the statistical units database) implies: corporate groups (which form the basis for profiling enterprises); the local establishments

of each enterprise; the legal entities they comprise; and the networks⁵ and links among them all.

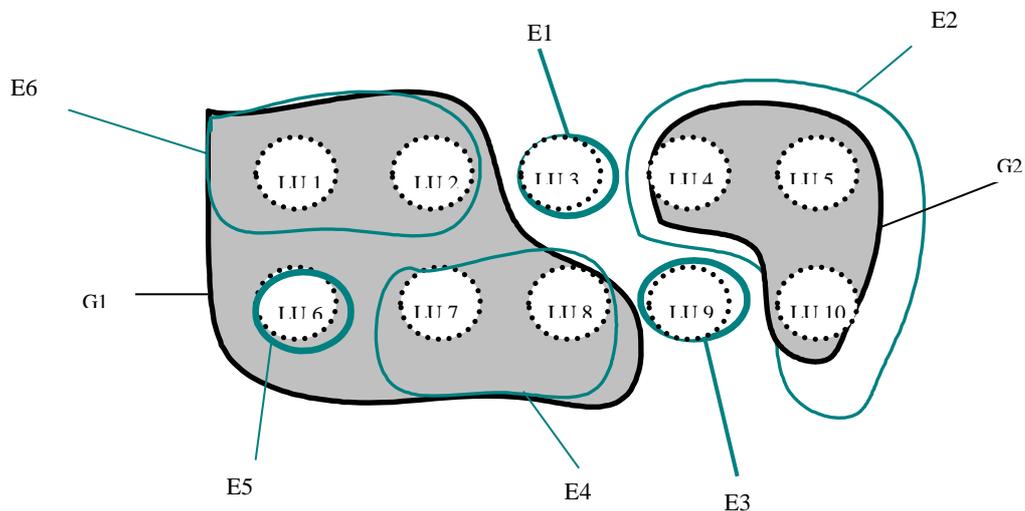
By means of example, consider a population of 10 legal entities (LU) from SIRENE:



Several are owned by a large corporate group (G1) and others by a small corporate group (G2), as shown below:



Profiling the groups converts the population of 10 legal entities into a population of six enterprises, as shown in the next figure: two independent legal entities; three enterprises derived from the manual profiling of the large group; and the small group, which has been classified as a single enterprise.



⁵ A network refers to a set of points of sale (legal entities or establishments) that are connected to a head point and form lasting relationships by creating common interests.

2. A new data registration method

12. SIRUS will, of course, register not only enterprises (profiled and non-profiled), but also certain characteristics of those enterprises, such as size, principal business activity, number of paid employees and turnover. To do so it will centralize update requests from “partner” surveys and applications, accepting or refusing them according to rules governing priority among the sources of updates. For example, if a survey finds that a unit appears to be “inactive”, that information will be transmitted to SIRUS, which will drop the unit unless a contradictory signal is received from another survey. The database will thus operate as a kind of “arbitrator”.

13. Any change to a value in SIRUS will be logged, and SIRUS will notify “partner” applications of changes to units in their areas of interest. Thanks to this updating, logging and transmission of data,⁶ one can be sure that each enterprise will, at any given moment and for any given reference period, be assigned to the same reference population and display the same characteristics.⁷ Furthermore, having a single reference value for certain statistics will make for greater overall consistency in the system. When results are published, for example, the principal business activity classification (APE) will be the same regardless of the source of the publication.

14. All the characteristics recorded by SIRUS will be dated (the value of a variable on a given date will thus be known), and all processing operations will be dated as well (the date on which a variable assumed its current value will thus also be known). These two pieces of data will make it possible to construct various types of sampling bases.

15. By means of example, consider the case of an entity classified under the industrial sector in 2009 that shifts to the transport sector in December 2010. The change is reported to SIRUS only in July 2011. The information kept in the register will be as follows:

<i>Sector</i>	<i>Date occurred</i>	<i>Date information processed in SIRUS</i>
Industry	1 March 2009	1 March 2009
Transport	31 December 2010	15 July 2011

If SIRUS is asked to create a sampling base for 31 December 2010:

- The enterprise will be classified under “industry” if the base is created on 1 January 2011
- The enterprise will be classified under “transport” if the base is created on 1 August 2011
- The enterprise will be classified under “industry” if the base is created on 1 August 2011 and SIRUS is asked for the information it had in March 2011

3. New variables

16. New information accessible only through SIRUS will be made available to all. For example, in December 2008, a decree implementing the French law on the modernization of the economy divided enterprises, using the notion of “enterprise” as defined in the

⁶ On matters such as termination of activity, change of sector and additions to the payroll.

⁷ Some applications nowadays use “snapshots” of the SIRENE directory produced on date “t” without taking into account changes recorded by SIRENE post-t even if they involve a pre-t event.

European regulations on statistical units, into the following categories: microenterprises, small and medium-sized enterprises, intermediate enterprises and large enterprises.⁸ SIRUS calculates the category to which each enterprise belongs.

17. Another new feature: the name of an individual INSEE consulting official assigned to facilitate communication with enterprises, particularly profiled enterprises, will be recorded in SIRUS. This is the person whom the enterprise should first contact if a problem involving several surveys arises or if the enterprise needs to be put in touch with the INSEE official assigned to a specific survey.

18. In short, SIRUS will be the ideal tool for:

- Defining reference populations for statistical surveys
- Constructing sampling frames or bases (according to the parameters set by the survey designer), whether the unit of interest be the enterprise, the group or the local unit
- Describing statistical units on the basis of their constituent units

19. SIRUS will be the only tool that will make it possible to apply simultaneously both Regulation No. 696/93, on statistical units, and Regulation No. 177/2008, on business registers, the two regulations that define the statistical units that must be used pursuant to current European legislation on the matter (structural business statistics, short-term statistics, labour costs, etc.).

20. Everything to do with the construction of SIRUS and its updating by “partner” applications is currently under development. The database is due to go into production, and the first reference populations to be created, in September 2011. Work on survey registration will begin then as well.

III. Registering the survey burden

21. Having presented SIRUS, let us turn to its role in reducing the statistical burden on survey respondents. Before any attempt is made to reduce it, that burden must first be registered and measured.

22. In France, businesses can be asked to complete surveys for several different INSEE departments (wage surveys and economic data surveys), and specialized ministerial statistics services (e.g. surveys on employment issues, transport or construction). Since information on the survey samples used by INSEE departments and ministerial services is not centralized, there is no quick and cheap means of ascertaining how many questionnaires a given legal entity or enterprise has been sent, how many surveys it has been asked to participate in or how many it has actually completed. Quantifying the time spent on answering surveys is next to impossible.

23. That will change when SIRUS is launched because it will register the samples used in all business surveys conducted by the public statistics system. To do that, it will work in liaison with:

- The Statistical Methodology Unit for Enterprises. This specialized unit of INSEE performs much of the sampling work for business surveys conducted by INSEE and ministerial statistics services. SIRUS will feed into the sampling process upstream by drawing up survey bases for the Unit according to the specifications it is sent

⁸ Decree implementing Act No. 2008-776 on the modernization of the economy, of 4 August 2008.

(field and unit type). Then, after the sampling, the Unit will return data on the survey base to SIRUS, including data on the sample used, such as the sampling stratum of each unit and each unit's weight in the sample. SIRUS will thus know whether a unit has been surveyed or was in the survey base, and will have the stratification data.

- Other services. SIRUS will be able to provide survey bases on demand to other services as well. It will also receive and register all the sample data from other surveys conducted by the business statistics system.

24. It will thus be possible to know whether a unit has been sampled, when, by which service, for which surveys and in which years, and in which survey base or bases it has been included. This information will be available for all types of units, whether they be profiled enterprises, corporate groups, legal entities or local establishments. Every effort will also be made to recover the samples from previous surveys so as to build up a historical record of the survey burden.

25. In the future, in order to ensure that SIRUS records all samples, it will keep track of every survey conducted. For that purpose it will enlist the support of the National Council for Statistical Information, which serves as a forum for dialogue between statisticians and representatives of civil society (professional associations, employee associations, researchers, etc.). INSEE and the ministerial statistics services submit their work programmes to the Council annually, and it is the Council that gives consent for statistical surveys to be conducted. The Council therefore has information on all the surveys conducted by the French public statistics system in any given year.

IV. Measuring the response burden

26. SIRUS aims to enhance the business statistics system even further by recording the survey response burden, i.e., the time spent by enterprises on answering questionnaires. Some surveys already routinely ask about the response burden (see CES/2009/35). What is more, the Government has made it compulsory for a question on response time to be included in all statistical surveys as of January 2012 as one of the recent measures introduced to simplify administrative procedures for companies. The answers will be collected in SIRUS. If a business does not answer the question, note will be made of whether it responded to the survey or not.

27. Here again, all survey services will have to cooperate with SIRUS so that post-survey data (lists of respondents and response times) are collected even when samples were established without assistance from SIRUS. Ideally, surveys of the latter type would be minimized over time so that the survey burden on businesses can be optimized.

28. The structure of SIRUS allows the burden to be recorded for both legal entities and enterprises. The most important measurement, however, will be of the burden at the enterprise level (if surveys are done at the legal entity level, the data will be totalled to obtain the enterprise-level figure).

29. Ideally, it will one day be possible to call up two pieces of information for each enterprise in the database:

- How many times it has been surveyed over the last year (together with the surveys in which it has participated and hence the number of questionnaires it has received)
- The number of hours spent on answering statistical surveys over the last year

V. Reducing the survey burden

30. Reducing the survey burden on businesses is even more important than measuring it. How can SIRUS help?

A. Surveying businesses, whether they constitute a group or a subgroup, will de facto reduce the survey burden on (profiled) enterprises

31. Once the number of survey questionnaires sent to all legal entities in general and to those belonging to a profiled enterprise in particular is known, it becomes possible to measure, inter alia, the returns to be made by that enterprise and, thanks to profiling, it will henceforth be sent just one questionnaire. This will be one of the arguments that those now entrusted with profiling large groups can use to persuade them that it is in their interests to participate in the profiling process.

32. Consider, for example, the returns from the corporate groups profiled first: 30, mostly of the largest groups, together representing 8,900 companies around the world, of which 3,400 are legal entities operating in France. After consultation with them, 95 enterprises are likely to be selected (between 1 and 7 enterprises per group, 3 on average). That means that possibly only 95 questionnaires will need to be sent out to obtain enterprise-level information that would today come from sending out 3,400 questionnaires.

33. One problem still needs to be overcome, however: recent tests with groups show that enterprises cannot always be asked to provide certain information.

34. The legal provisions governing data collection (and the legal obligation to respond) currently apply to legal entities but it remains to be seen how they can be adapted to apply to enterprises.

35. Some surveys will probably therefore continue to be conducted at the level of the legal entity and, in that case (as mentioned previously), the burden on the enterprise will be calculated by adding up the burden on each of its constituent legal entities.

B. Negative coordination will also be possible with SIRUS

36. In early discussions on SIRUS, the idea was to create a statistical database able not only to establish survey bases on the basis of field specifications furnished by the survey designer, but also to establish survey samples using a special tool. That idea has been dropped, but the fact that SIRUS contains data on the constitution of survey bases and on the response burden, in addition to sample data, still has its advantages.

37. One method currently used by INSEE to reduce the response burden on businesses is to negatively coordinate samples and, when establishing a sample, to favour enterprises that have not been recently selected, while preserving an unbiased sample. This method, described in “The two-by-two coordination of samples” below, is based on the random allocation of a number to each entity. Unfortunately, it only allows samples of enterprises to be negatively coordinated two by two.

38. Research carried out at INSEE a few years ago, however, has led to the development of a different method, now being tested, which should make it possible to coordinate several surveys negatively.

39. The method is based on the survey burden. This changes, of course, with each sample selected since it depends on participation in previous surveys; the sampling process consists in selecting, in each stratum, the units with the smallest accumulated burden.

40. The method is described by Christian Hesse in “*Généralisation des tirages à numéros aléatoires permanents, ou la méthode JALES+*” and is currently being tested at INSEE with a view to developing it into a software package. The SIRUS statistical database will supply the software with data on each unit’s survey burden since it is the only central repository of sample information.

41. As reported above, SIRUS will also centralize information on response burdens. That information could be incorporated into the method now under development to distinguish between surveys and weight them accordingly. In other words, an enterprise that responds to a 10-page questionnaire with an average completion time of three hours would be less likely to be selected for future surveys than one that has responded only to a half-page questionnaire with an estimated average completion time of 10 minutes.

42. In closing, we present a few notes on the timetable for the project: the database established in the autumn of 2011 will be gradually expanded with samples from previous years, 2011 and possibly 2010 in particular. Hence first indications of the survey and response burden should become available in the summer of 2012.

43. The SIRUS statistical database will, thanks to its incorporation of the “enterprise” dimension and calculation of the statistical burden on enterprises, allow the sampling process in the French business statistics system to be coordinated in new ways, which should reduce the statistical reporting burden on enterprises.

The two-by-two coordination of samples

44. In the case of stratified simple random sampling plans, the sample selection algorithm generally used at INSEE consists of attributing a random number between 0 and 1 to each unit of the sampling base in an initial survey and then selecting the units in each stratum that have the smallest random numbers. After the sampling, the random numbers are permuted in such a way that the units sampled because they had had the smallest random numbers end up grouped with those that have the largest numbers and vice versa (i.e. the non-sampled units end up with smaller random numbers), as shown below.

45. In a sample that has two strata (industry with a survey rate of 50% and transport with a survey rate of 10%), for example, the units are classified in their stratum by random number in ascending order. The shaded units are the ones selected.

A	B	C	D	E	F	G	H	I	J
---	---	---	---	---	---	---	---	---	---

Industrial sector (survey rate: 50%)

K	L	M	N	O	P	Q	R	S	T
---	---	---	---	---	---	---	---	---	---

Transport sector (survey rate: 10%)

Once the sample has been drawn, the permutation of the random numbers would generate the following arrangement:

F	G	H	I	J	A	B	C	D	E
---	---	---	---	---	---	---	---	---	---

Industrial sector (survey rate: 50%)

L	M	N	O	P	Q	R	S	T	K
---	---	---	---	---	---	---	---	---	---

Transport sector (survey rate: 10%)

46. During the next sampling, after assigning random numbers to the units not included in the first sample base (units U and V in the scheme below), the units with the smallest

random number in each new stratum of the second sampling base would be selected, as shown below.

47. Imagine that a second survey is conducted after the first and that, in the second survey, the survey rate was to be 60% and there would be only one stratum. The result would be as follows:

F	L	G	M	N	H	I	U	O	J	A	P	B	Q	C	R	V	S	D	T	E	K
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Transport and industry sector (rate: 60%)

Only A and B receive two questionnaires. There is some overlap between the two surveys, but the number of units that receive two questionnaires is kept to a minimum.

VI. References

“French methods and practices regarding the statistical burden”, paper presented by France, DGINS 2008 Conference, Vilnius.

Depoutot, R.: “Reengineering French structural business statistics: an overview”, Conference Q10, Helsinki.

INSEE (2009): “Redesigning French structural business statistics: how can the response burden on companies be lessened”, UN ECE, paper CES/2009/35.

INSEE (2007): “Improving the quality of business registers”, UN ECE, paper ECE/CES/GE 42/2007/3.

Hesse, C.: “Généralisation des tirages à numéros aléatoires permanents, ou la méthode JALES+”, INSEE, Direction des Statistiques d’Entreprises.

Laflamme, G., Rivière, P. (1997): “OCEAN, toward a general system for sample collection and coordination”, *Courrier des Statistiques*, English version No. 3.

Hecquet, V. (2010): “Le profilage et son impact sur la représentation de l’appareil productif, papier présenté au colloque de l’Association de comptabilité nationale.