The French business registers system: How to improve the quality of the statistics by combining different statistical units

Haag Olivier[1]

Abstract

The French business statistic register (SIRUS) contains different statistical units and the links between them. These information are needed to calculate the most relevant statistics on enterprises. In France, the most useful statistical units are:

- The legal unit, which is the best data collection unit and the backbone of the system. Its identifier is shared by all the administrations.
- The enterprise, which is the best statistical unit for the economic analyses. The enterprise is defined by the European regulation as the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making. The use of this statistical unit as reporting unit has become compulsory due to economy globalization.
- The enterprise groups are used for financial issues.

The French business registers system is also composed of four business registers:

- Three authentic sources; one for each type of statistical unit mentioned above: SIRENE is the administrative register for the legal unit, LIFI for the enterprise groups and BCE for the enterprise.
- One with all the links between them: the French business statistic register called SIRUS.

SIRENE and SIRUS are deployed; reengineering LIFI is ongoing; the BCE already exists for 110 large enterprises, defined from manual profiling process, and is being built for smaller enterprises through an automatic process.

First, this article presents some examples showing the interest of these different types of statistical units for economic analyses. It then presents the different business registers essential to manage this different statistical units and focus on the links between them. In a third part, this article shows how we combine different statistical units to obtain the best statistics possible. Finally it presents new statistical characteristics, calculated for each type of statistical unit and stored in SIRUS, which improve the quality of the survey's frame and the management of the statistical burden.

Keywords: Business register, statistical unit, survey design

[1]Haag, Insee, 18, boulevard Adolphe Pinard 75675 PARIS,
email : olivier.haag@insee.fr

1. The French statistical unit Legal Framework

The French national legislative framework distinguishes three main units:

- The legal unit;
- The local unit;
- The enterprise groups.

The next points are going to focus on each type of units and a fourth statistical unit: the enterprise. This last statistical unit has no administrative existence but it is essential to calculate statistics as relevant as possible.

1.1 The legal unit

The legal units the backbone of our system of data collection.
First because these units registered in a national business register called SIRENE (see below) and its identifier is shared by all the French administrations (tax, employment, customs...).
Secondly because the legal unit has to declare a lot of economic information to the tax administration (turnover, total assets, balance sheets, amount of investment). Besides, according to the French law, the tax administration must make available all this information to the French National Statistical Office (NSO) in order to reduce the administrative burden for the companies.
Third because the French statistical law of 1951, forces the legal unit to answer to the mandatory surveys. For instance, the French structural business survey, the short terms statistics, are thus collected at the legal unit level.
In conclusion, a lot of economic and statistic information are available at the legal unit level. And it is easy for the French NSO to merge all these data through micro data linking process, because the legal unit answers to all the administration with the same ID number.
That is why, until 2014, the legal unit was used for the dissemination of business statistics. The legal unit was regarded as an enterprise for the structural statistics and as a Kind Activity Unit for the short-term statistics.
But this situation is now changing and the point 1.4 below will explain why and how.

1.2 The local unit[2]

The situation is about the same for the local units. The legal framework is the same than the legal unit one. The only difference relies on the nature of the data collected.
The local unit is used as data collection and dissemination unit for:

- Employment, because the French administration collects information at this level. An example of map that localises the employment of the French territory can be found on the following link. It is built from the localisation of the local units.

(http://www.insee.fr/fr/themes/tableau.asp?reg_id=99&ref_id=TCRD_030).

- Statistical survey, about the energy usage, the waste generations, the expenditure to protect the environment...

2 According to the European regulation No 696/93 of 15 March 1993 on the statistical units: The local unit is an enterprise or part thereof situated in a geographically identified place.

1.3 The enterprise group[3]

On the opposite, there is no obligation for the enterprise groups to respond neither to administrative needs nor to statistical surveys. Only the large groups have to draw up consolidated annual accounts and consolidated annual reports. But these data are consolidated at the world level and cannot be used to calculate national statistics.
That is why, for the moment, the French statistics do not use the group as a dissemination unit. But the enterprise group is useful to create a new classification of legal units.

- Legal units that belongs to a **Foreign multinational groups** (GET-MNE): groups having at least one subsidiary in France but whose global decision centre (GDC) is abroad (16,000 groups);
- Legal units that belongs to a **French multinational groups** (GFR-MNE): groups having at least one subsidiary in France and GDC is in France (about 5,000 groups)
- Legal units that belong to a **Franco-French group** (GFR-FRA): groups with French subsidiaries only (about 65,500 groups).
- **French legal units** (IND-FR)**:** legal units not belonging to a group and with their registered office in France (more than 3.7 million legal units).

Table 1 *Breakdown of exports by type of enterprise*

Type of enterprises	Number of enterprises	Number of exporting enterprises	Share of number of exporting enterprises (%)	Export turnover (in €K)	Export turnover (%)	Average export turnover per exporting enterprise
French multinationals	5 637	4350	77.2	366 804 945	57.8	84 323
Foreign multinationals	15 969	8 408	52.7	188 704 708	29.7	22 443
Franco-French groups	66 500	21975	33.0	37 112 849	5.8	1 689
French independent	3 751 193	161299	4.3	42 062 147	6.6	261
Total	3 839 299	196032	5.1	634 684 650	100	3 238

3 According to the European regulation No 696/93 of 15 March 1993 on the statistical units: An enterprise group is an association of enterprises bound together by legal and/or financial links.

This table shows the importance of French multinational enterprises when it comes to exports. They make over half of exports despite the fact that there are less than 5,000 of them that export. Conversely, what is more surprising is that about one-quarter of French multinationals do not export at all! However, these non-exporting multinationals are the smallest ones, as they represent less than 5% of the value-added and 7% of the workforce of French multinationals as a whole. [1].

1.4 The enterprise[4]

The enterprise is a statistical concept and not a legal concept. In France, before 2012, we used to equate the enterprise with the legal unit, and the whole system of business statistics relied on the legal unit. But assimilating the legal unit to the enterprise is not relevant anymore for group's affiliates and subsidiaries. Indeed, they lose their autonomy in decision-making and do not respect the European definition of an enterprise.
INSEE has therefore decided to move from a definition based on the legal unit towards a more appropriate statistical definition.

In order to define the perimeter of the French enterprises, two methods of profiling are used to identify these enterprises inside each group [2]:

- The **largest groups** present in France, or the most complex ones (those that have a large number of subsidiaries and multiple activities) are profiled "**manually**"(meaning further to meetings with representatives of the groups) by the members of a specialised division of INSEE. There are about 55 groups of this kind; most of these groups have more than 10,000 employees in France.
- **The other groups are profiled "automatically"**. They are approximately 80,000.

This new concept is now used to calculate the business statistics. The next figure compares two vision of the French Economy: one calculated from the legal units and the other from the enterprises.
The four enterprise categories are defined as follows:

- **micro-enterprises** employ fewer than 10 people and neither their annual turnover nor balance sheet total exceeds €2 million: a little under 3,000,700 enterprises;
- **SMEs**(small and medium-sized enterprises) employ fewer than 250 people, and [their annual turnover is less than €50 million or their balance sheet total is less than €43 million]:slightly over 135,000 enterprises;
- **ETIs**(intermediate-sized enterprises) employ fewer than 5,000 people, and [their annual turnover is less than €1.5 billion or their balance sheet total is less than €2 billion]: about5,000 enterprises;
- **Large enterprises** are those that do not fall into the previous categories: 245 enterprises.

4 According to the European regulation No 696/93 of 15 March 1993 on the statistical units: the enterprise is the smallest combination of legal units that is an organizational unit producing goods or services, which benefits from a certain degree of autonomy in decision-making, especially for the allocation of its current resources. An enterprise carries out one or more activities at one or more locations. An enterprise may be a sole legal unit.

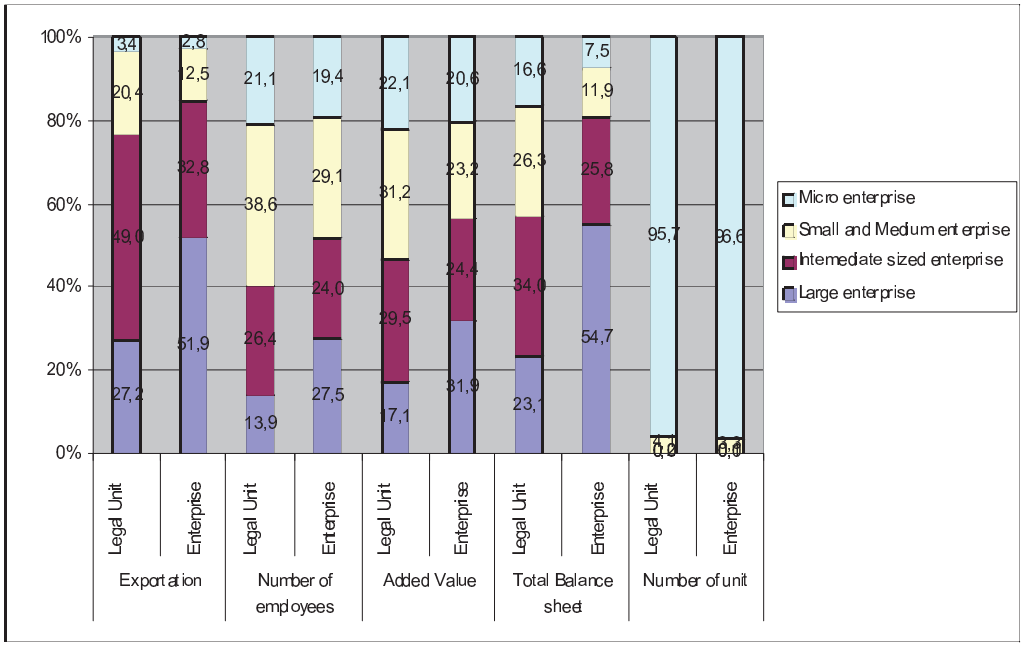


Figure 1: Share of legal units and enterprises in French economy by enterprise size in 2014.

How to read this chart: 95.7% of legal units are considered as micro enterprise. Their exportations represent 3,4%. 96.6% of enterprises/groups (uncontrolled legal units and groups) are micro enterprises. Their exportations represent 2,8%.

We notice that the enterprise view is more concentrated than the legal unit view. The weight of the micro enterprises and SME's has reduced to the benefit of the large enterprises. This result is partly due to the fact that the enterprise group can be organized in small legal units that are specialized.
For example one large enterprise group can create micro legal unit to:

- export its production
- register its fixed assets etc.

In the legal unit view, the data of these legal units are accounted among the micro enterprises but in the enterprise view they are accounted among the large enterprises.
According to us the enterprise view gives the true panorama of the French economy that allows the French government to set up the best economic policy possible.

In conclusion, we have underlined in this first part, that we need different kind of units to produce the most relevant statistics.

But one difficulty is that data are not systematically available for the unit of interest. For example we do not have economic variables at the enterprise group level.

To address this issue we need to consider differently the data collection unit and the reporting unit. Therefore, we need to know the links between these two types of units and to be able to create different frames. The statistic business register is the backbone of this new process as described in detail in the following chapter.

2. The French business register network

2.1 Four business registers

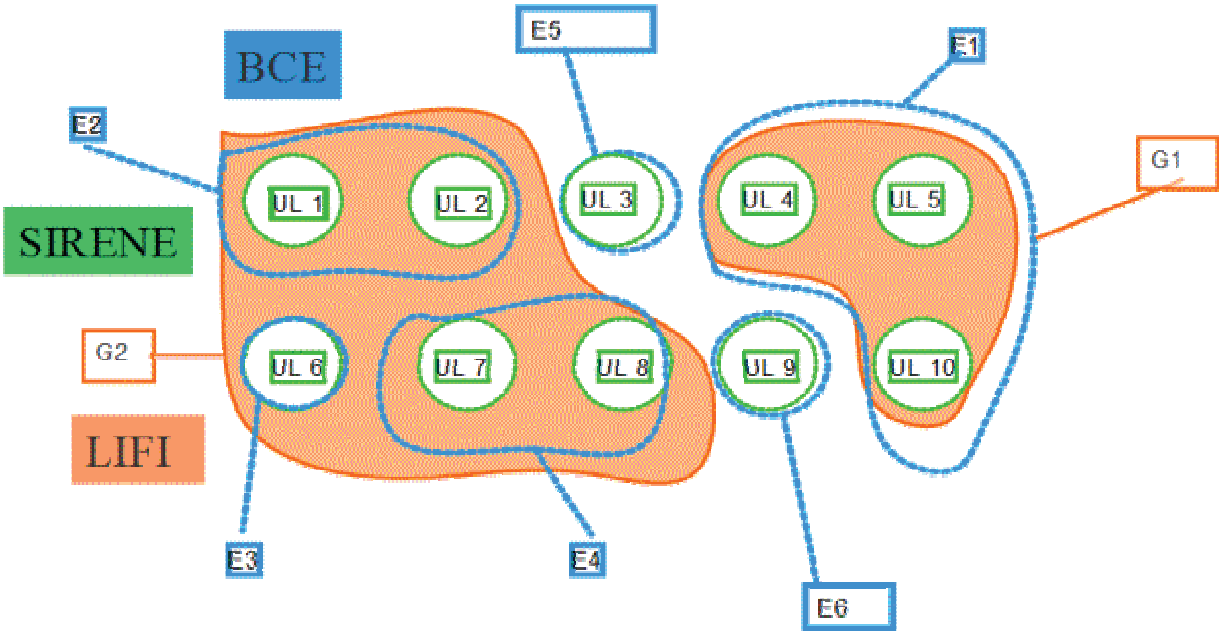
This chapter is going to present briefly the 4 French business registers and their links, as previously developed in more details [2].
The French system is based on 3 so-called "authentic source" business registers, each dealing with one type of statistical unit, and a so-called "statistical" business register (called SIRUS), concatenating all the different information and serving as the sole basis for all statistical operations requiring a reference sampling frame.

The three "authentic source" business registers are:

- SIRENE: the administrative register of legal units (UL 1 to 10 in the figure 2);
- LIFI: the statistical register of enterprise groups (G1 and G2 in the figure 2);
- BCE: the statistical register of enterprises (E1 to 10 in the figure 2).

The diagram below shows the different statistical units managed by the business registers and the links between them.

Figure 2: The different statistical units, their links and theirs business registers



2.1.1 SIRENE

SIRENE is an inter-administrative business register created in 1973. It is an exhaustive business register of legal units serving for exchanges between administrations. It contains a single identifier shared by all the registers of French government (taxes, customs, central bank, etc.). Thanks to this identifier, Insee can perform a lot of micro-data linking.

2.1.2 LIFI

The LIFI business register identifies the enterprises groups and contains the links between the legal units within these groups (core and extended perimeter). The French statistical definition currently in force takes the absolute majority of voting rights as control criterion for defining the contours of groups.

2.1.3 The Enterprise Creation Database (BCE): the register of enterprise

The BCE business register identifies the enterprises and contains the links between the legal units within these enterprises.
A "French" enterprise (the basis for "French" statistics), is therefore:

- Either an independent (non-group), French legal unit;
- Or the French footprint of a whole group of legal units;
- Or the autonomous part of the French footprint of a group of legal units5. These autonomous parts are obtained by a manual profiling process

The latter two types of enterprises mentioned above are more commonly referred to as "profiled enterprises" and are managed by a specific business register: the Enterprise Creation Database (BCE).

2.1.4 SIRUS

SIRUS stands for "system of identification in the business register of statistical units" and is a statistical business register of statistical units, in particular of enterprises within the statistical meaning of the term (see definitions below).

The main objectives of SIRUS are the following:

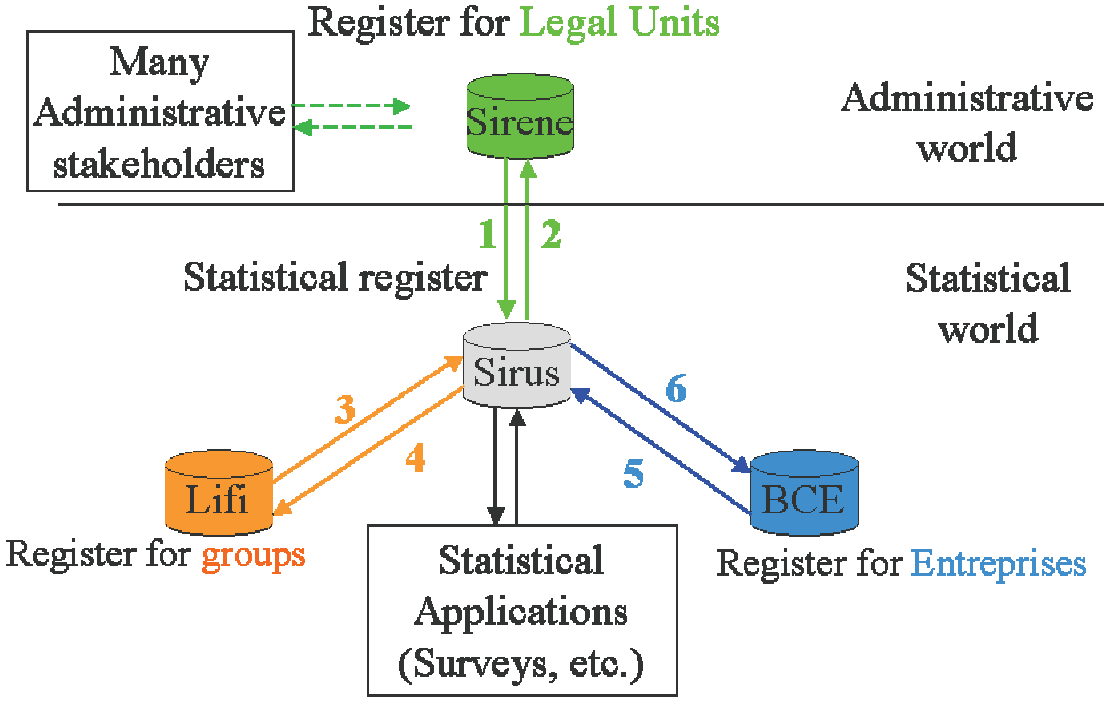
- **To list groups and enterprises** (within the statistical meaning of the term), **and the legal units and establishments that comprise these enterprises**. SIRUS records the links between the various statistical units. For all these units, SIRUS also records characteristics that are useful for the creation of frame as turnover, sector classification (APE) and salaried headcount, thanks to updates from a multitude of sources (the other business registers, but also from statistical surveys). If SIRUS needs other characteristics, there are webservices between SIRUS and the other business registers that allows SIRUS to obtain it. But they are not stored in the SIRUS database.
- **To provide** business statisticians with **reference populations**. In this way, at a given moment in time and for a given reference period, an enterprise will be allocated to the same reference population in all surveys and will have the same characteristics in all of them.
- **To provide new statistical information**, in particular **the classification of enterprises in four categories** (micro-enterprises, small and medium-sized enterprises, intermediate-sized enterprises and large enterprises).

5 Which does not necessarily mean that it will be composed of a whole number of legal units, as in certain rare cases, legal units may be split between several enterprises, as indicated in Paragraph 1 of the definition.

- **To manage the "statistical cessation" of units**. This will make it possible to distinguish between a unit that has an economic activity and a unit that is legally active but has no economic substance. See chapter 4 for further details
- **To record the response burden** of statistical surveys, meaning the time spent by enterprises filling out statistical survey questionnaires. See chapter 4 for further details

2.2 Their links

Figure 3: The French business register network



The main rules of the French business registers network are:

- SIRUS is the core of the system
- SIRENE makes the links between the statistical and the administrative Worlds
- There is no direct link between the business registers. All the flow goes from a register to SIRUS and from SIRUS to an other register
- SIRUS makes the links between the register and the statistical Worlds. For instance, the entire frame is constituted by SIRUS.
- SIRUS is always up to date for the creation of legal units because it is daily updated by SIRENE

3. The combination of different statistical units

As previously mentioned, the legal unit is the best data collection unit but cannot always be considered as the best reporting unit. However, the profiled enterprise is the best reporting unit but cannot be used directly as a data collection unit.
That is why, for a business survey, the survey designer must now ask two questions:
- What is the unit of interest (or reporting unit) of my survey?
- What is the best data collection unit of my survey?

The next part will explain how, thanks to the statistical business register, the links between these two different units will be possible and what are the new issues the statistician will face.

3.1 A top down approach for the delineation of the enterprises

The French profiling process used to delineate the enterprise within the enterprise group is based on a top down approach.
- For the manual profiling, the profiler asks the group to define the enterprise composition of legal units within the group.
- For the automatic profiling, the entire group is considered as an enterprise.

In these two cases the enterprise group is the starting point to delineate the enterprise explaining why we can talk about a top down approach.

3.2 A bottom up approach for the consolidation of the data collected

- For the manual profiling, there are different types of consolidation
    - A bottom up approach. In this case the profiler obtains the internal flow between the legal unit within the enterprise
    - A top down approach. In this case the profiler obtains the account for the enterprise directly from the group
    - A mixed method. The profiler obtains the value of the non-additives variables from the group. This information generally comes from the IFRS concept. And the values of the additives variables are calculated by adding the legal units data.
- For the automatic profiling. The method used is a bottom up approach. The intra-flows for the additives variables are calculated by algorithm. [3]

In most cases, as the enterprise group cannot be directly used to obtain tax information for instance, the legal unit stays the data collection unit. In this way, to obtain the data at the enterprises level, it is necessary to consolidate the legal units data. Therefore we can talk about a bottom up approach.

3.3 The impact on the frames and the survey process [4]

3.3.1 Two frames for one survey

Since the reporting units are now different from the data collection unit for the profiled enterprises, the survey design can be seen as a two-stage cluster sampling. As a cluster, an enterprise is randomly selected and then all the legal units within this enterprise are

included in the sample. But the cost constraint is still based on the number of legal units surveyed, which is now random.
The new challenge is therefore to optimise the survey design in order to have the best precision on estimators possible at the enterprise level under the constraint of a limited number of collection units.
For this new methodology of the survey sampling design two frames are necessary:

- One frame composed of enterprise which allows to define the sample
- One frame composed of legal units and the links between the legal units and the enterprises. This frame allows defining the legal units that have to receive a questionnaire.

These two frames are now delivered by SIRUS.

3.3.2 A new definition of the take-all strata[6]

It is not possible to keep the same criteria for the definition of the take all strata for an enterprise sample as the ones used for a legal unit state. Indeed the largest enterprise contains a lot of legal units. This large enterprise often belongs to the take-all strata with the legal units criteria. And if all the legal units of the large enterprise received a questionnaire the sample size exceeded the expected one.
For instance, the size of the French SBS survey is 120 000 units. If we applied the legal unit definition of the take-all strata at the enterprise level, 130 000 legal units should be collected.

In order to reduce the legal units number in the take-all strata, two procedures have been implemented:

- Increase the thresholds of the take-all strata in terms of turnover and number of employees. The impact of this action is to reduce the number of enterprises that are exhaustively interrogated;
- Define rules in order not to collect all the legal units of an enterprise. For instance, a cut off of legal units achieving 95% of the turnover within each exhaustive enterprise is applied. With this criteria the legal units that represent the lowest 5% are not surveyed but imputed (from the tax data). In the previous example, using this method allows to diminish the size of the exhaustive part for 30 000 legal units.

3.3.3 A new method of data editing

As the unit of interest, the controlled unit in the data editing process, is different from the data collection unit, forces us reconsidering the data editing process of our survey.
Indeed, the monitoring process checks the quality of the enterprise data. In case of inconsistency, the clerks have to check the legal units data and if necessary contact the legal unit with an apparently wrong answer. This new process makes the clerk's job more complex, and new tools need to be developed in order to make it easier.

[6] A take all strata is a strata where all the units are surveyed and have a sampling weight of 1.

4. New characteristics useful to improve the quality of the statistics calculated for each statistical unit and linked together

The business register SIRUS contains new variables that improve the quality of our statistics. These variables are calculated for each type of statistical units and are in some cases linked together.
This chapter presents three variables:

- The statistical cessation that allows to improve the quality of the frames by reducing the over-coverage;
- The registration of the statistical burden. This information is useful for the sampling coordination;
- The definition of the continuity of the enterprises. This information is crucial for the comparison of statistics from year to year.

4.1 The statistical cessation

4.1.1 The definition

We can identify different phases during the life of an enterprise.

1. The birth;
2. The period of economic activity;
3. The end of the economic activity (t1 in the figure below);
4. The end of employment (t2 in the figure below, t1 is often equal to t2);
5. The legal cessation of activity (t3).

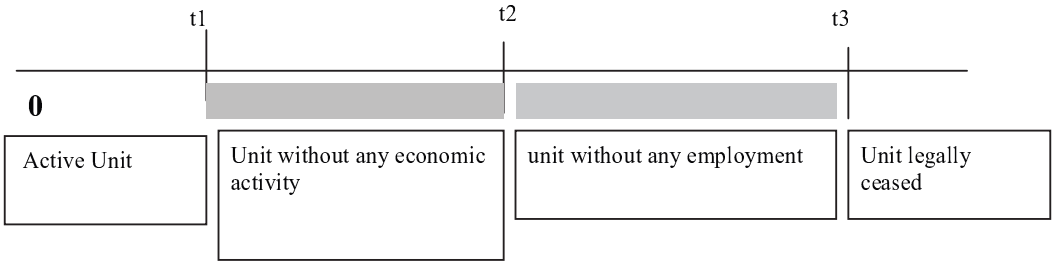


Figure 4: The real life of an enterprise

In France, the procedure of liquidation is long and costly, explaining why some legal units do not go through it or needs a long period to achieve it.
In fact, the date t3 is often far from the date t1. This is a problem for the quality of our statistics because it generates an over-coverage of our frame. Indeed, enterprises without any economic activities are wrongly considered as active unit by SIRENE as illustrated on the figure below.

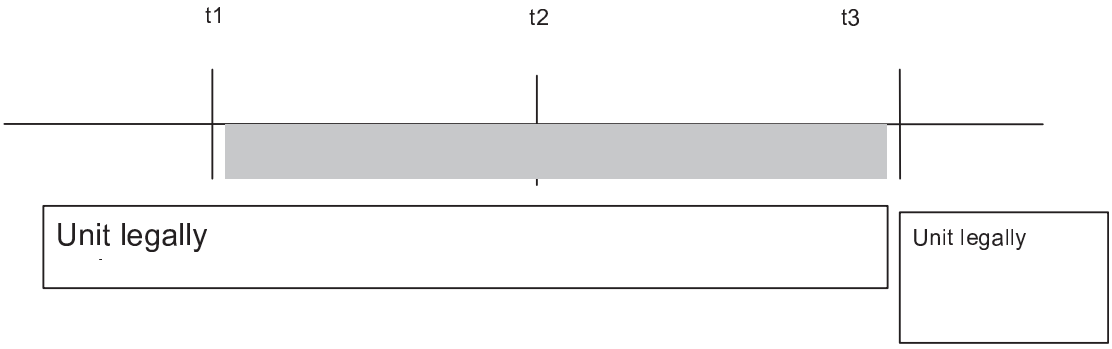
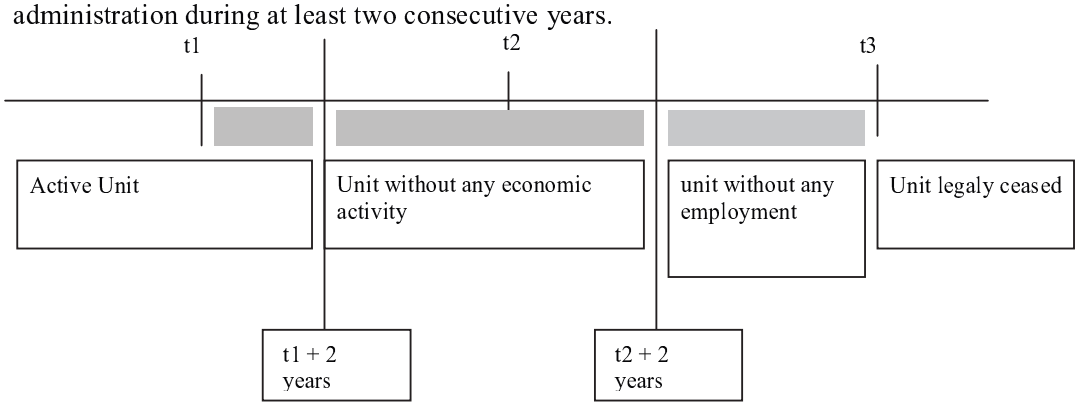


Figure 5: The life of an enterprise according to SIRENE

To improve this situation, new modalities of cessation are calculated in SIRUS:
- Unit without economic activity;
- The statistical cessation;

These modalities are calculated in SIRUS thanks to a micro data linking with Administrative sources.
- A legal unit is considered without any economic activity if it declares no turnover to the tax administration during at least two consecutive tax periods;
- A legal unit is statically ceased if it declares no employment to the social administration during at least two consecutive years.



0

Figure 6: The life of an enterprise according to SIRUS

For instance for the French results of 2014, there were:
- More than 4 millions active units
- 293 675 units without any economic activity
- 899 336 statically ceased (mainly micro enterprise, of which some had never begun any economic activity).

4.1.2 The link between the different statistical unit

The statistical cessation is first calculated at the legal unit level but is then reported to the other type of unit as follows.
- If a legal unit is statistically ceased or considered without any economic activity, all its active local units obtain the same status.
- If an enterprise group or an enterprise contains less than one active legal unit, it is statistically ceased or considered without any economic activity according to the status of the legal units of its perimeter.

4.2 The measure of the statistical burden

4.2.1 The definition

In France, a question about the time needed by the enterprise to answer to the survey appears on all questionnaires.
We therefore have three information about a survey burden;

- The sample size
- The number of respondents
- The time needed by the enterprises to answer. This information is known for all types of data collection units and is used for our sampling coordination process [5]

4.2.2 The link between the different statistical unit

The last information permits to calculate the burden for each type of statistical units by adding the burden of all the data collection units of its perimeter. The table above gives an example.

Table 2: Burden for the enterprise E1

	Burden (in minutes)					
	Survey 1	Survey 2	Survey 3	Survey 4	Total for the data collection unit	Total for the unit
Enterprise E1					0	75+15+81+32+12 =215
LeU 1	30	45			75	75+15+81 =171
LoU11			15		15	15
LoU12			25	56	81	81
LeU 2		32			32	32+12 =44
LoU21			12		12	12

The enterprise E1 contains two legal units LeU 1 and 2. The LeU 1 contains 2 local units LoU11 and 2 and the LeU12 has a single location LoU21.The LeU1 needs30 minutes to answer to the survey 1. The enterprise is never a data collection unit but we consider than its burden corresponds the amount of the burden of its legal units.

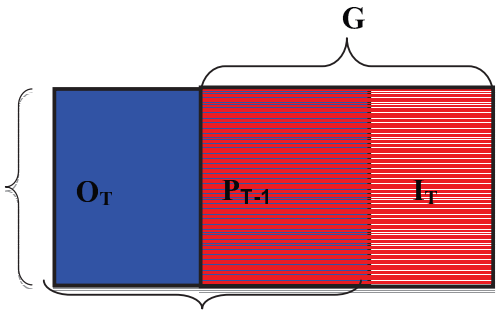
4.3 The continuity

The delineation of the perimeter of enterprise groups or enterprises in depends on a reference year.
For the reference year T, a draft version of the perimeter is available in SIRUS in October T+1. This version is useful for the creation of the frame of the Structural Business Survey for instance. The definitive version is available in April T+2.

For each data supply, the continuity of the enterprise (group) is calculated. If the enterprise (group) is the continuation of a previous one, it will keep its identifier. If not, a new identifier will be assigned. The links between the enterprises predecessors and successors are stored.

A group G, for year T, is the continuation of a preliminary group PG known in year T-1, if and only if the perennial part of these two groups is bigger (in terms of number of persons employed) than 50% of the number of persons employed in each one (50% is an adjustable threshold).
The same definition applies of course for enterprises.

We can summarize the situation with the following scheme:



$PG = O_T + P_{T-1}$
$G = P_{T-1} + I_T$

G is the continuation of PG if:
- For the year T, the employment of P_{T-1}>50% of the employment of G and
- For the year T-1, the employment of P_{T-1}>50% of the employment of PG_T

In the other cases we can only say than G is a successor of PG and PG is a predecessor of G.
This continuity is crucial for the identification of the enterprise group and the enterprises and to the dissemination of consistent longitudinal statistics.

Conclusion

With the globalisation of the economies and the increasing importance of the enterprise group in the French economy, it is not possible today to use only the legal units to calculate all the business statistics.
We have to take into account the enterprise group to calculate consistent and relevant business statistics.
But with the important issue in terms of reduction of statistical burden, it is crucial to gather administrative data via micro data linking for instance. This information are large for the legal unit but not for the enterprise group. We therefore have to consider two different types of statistical units for conducting a survey: the legal unit as the data collection unit, and the enterprise as the reporting unit.
In these circumstances, the role of the statistic business register is crucial, since it is the only one able link these different kinds of units, and allowing to follow them up over time.
For the French NSI, if the business register permits to take into account this new situation, further efforts will still have to be made to achieve the best quality possible. For instance, we are going to continue our work on profiling by:
- Increasing the number of groups that are manually profiled
- Increasing the quality of our algorithm of consolidation
- Setting up a new survey to obtain the intra flow of the large groups that cannot be manually profiled, due to a lack of resources.

References

- [1] Olivier Haag (2016), Profiling: a new and better way to apprehend the globalization, Q2016, Madrid
- [2] Jean-Marc Béguin and Vincent Hecquet (2014), Profiling in France: Implementation and results, 24th Meeting of the Wiesbaden Group on Business Registers, Vienna
- [3] Jean-Marc Béguin (2013) Calculation of the main SBS characteristics for the enterprises equal to the enterprise groups, Lot 1- TASK 2
- [4] Ronan Le Gleut (2016), The unit problem: a first assessment of the impact of profiling on sampling, ICES V, Geneva
- [5] Emmanuel Gros (2016), The procedure of sampling coordination for business survey implemented at INSEE; ICES V, Geneva