

**Session 5: Projects on the Improvement of Business Registers
(Wednesday morning, Session leader Eurostat)**

**The French Business Register :
From a quality approach to a statistical register**

1. INTRODUCTION

This document aims to show the approach to improve the quality of the French *Système national d'Identification du Répertoire des Entreprises et de leurs Etablissements* (SIRENE) business register, from the day-to-day management of this register to the construction of a specific statistical register.

After a brief presentation of the SIRENE register and its contents and purpose, we will discuss the users' expectations and give some examples of operations carried out to improve the quality, together with indications of volumes involved. Then we will see how this "quality approach" takes into consideration some criteria such as "relevance", "timeliness". Lastly, in a few words, we will show why it has become necessary to think about a new tool: a statistical business register.

2. DESCRIPTION OF SIRENE

All businesses and their local units in France are registered in a national business register called SIRENE. This register is used in the context of the statutory framework in which INSEE is responsible for identifying businesses (legal units) and their local units, on the basis of documents transmitted by *centres de formalités des entreprises* - CFEs (centres for business procedures), which were set up to deal with all administrative procedures falling within their remit. An act of 11 February 1994 established a unique identification number, which is now at the core of the inter-administrative system, since use of this identification number is mandatory for relations between businesses and administrative bodies.

The SIRENE register lists the civil status of businesses: the family name, first name (in the case of an independent business) or the company name (in the case of a legal person). Also listed are the legal form of the business, the address of its headquarters, the date it was set up and, where appropriate, when it ceased trading. The address of local units is indicated, as well as the date and purpose of their creation. In addition to these identification data, the register contains the APE (*activité principale exercée*) code assigned by INSEE identifying the principal business activity carried out, based on the NAF (*Nomenclature d'Activités Française*) French general economic classification, for each of the units listed.

This wealth of resources means that this inter-administrative register is also used in the area of statistics. For this reason, it was decided to record additional data for each unit registered in SIRENE, in accordance with a list drawn up after consulting statisticians using the system. Thus, in addition to identification variables, the business register includes economic classification variables (economically important units, units having employees, etc.); variables relating to the number of employees (number of employees when set up, number of employees at each year end); turnover; variables showing links between local units or legal units ; and management variables (particularly for management of dissemination of information and data confidentiality).

3. INVENTORY OF REQUIREMENTS

Before looking at quality in relation to SIRENE, it is essential to distinguish the different purposes of the register:

- Inter-administrative use;
- Dissemination of information contained in the register;
- Statistical use.

3.1. INTER-ADMINISTRATIVE USE

Requirements of administrative partners involved in managing the register

The term “administrative partner” is used to describe the statutory partners (centres for business procedures, social security bodies or tax administration). Their three main requirements are :

- Registration at the earliest possible opportunity;
- Unique identification number for each unit;
- Exhaustiveness of the registrations (legal units and units which are registered only for the needs of an administrative partner);
- Consistency between SIRENE and the legal and administrative registers for businesses (business-names register or trades register, social security and tax files), which is ensured by the mandatory unique identification number.

Requirements of businesses

The requirements of businesses dovetail to a large extent with those of the administrative partners: rapid registration; unique identification number; consistency with other administrative and legal files. Businesses also have other specific concerns, such as:

- The code describing their principal activity.
- The reference to termination in the register. If erroneous, this can penalize businesses, given the public nature of the register.

3.2. REQUIREMENTS CONCERNING DISSEMINATION OF INFORMATION CONTAINED IN THE REGISTER

One of the articles of the decree establishing SIRENE provides that INSEE shall give the public freedom of access to data contained in the register. The staff responsible are particularly concerned to ensure that the information contained in the “address” variable is correct. However, while they attach importance to the reliability of addresses, the services responsible also want to ensure that the register provides a correct activity code and a reliable indication of the size of a business.

3.3. REQUIREMENTS CONCERNING THE USE OF SIRENE AS A STATISTICAL REGISTER

SIRENE is, of course, widely used for statistical purposes.

Business demography

By definition, SIRENE is used to monitor business demographics: analysis of the economic climate (statistics for business start-ups are published monthly, and are closely followed); structural statistics relating to the demography of businesses and local units; description and monitoring of the characteristics of entrepreneurs and of new businesses.

Basis for sampling

Its exhaustiveness and content enable the register to be used as a reference database for sampling purposes. Identification data (name, address), economic termination, stratification data (activity, size, region of location, turnover) are therefore essential and must be of good quality. In addition, to meet the requirements of statisticians, they must be supplemented by metadata (date and source of latest update, quality indicators, etc.).

Survey managers

When utilizing data from administrative (tax or social-security) sources or from surveys, statisticians may sometimes query data in the register. They often have access to more recent data or are able to observe discrepancies, without being able to account for them (local unit address, for example, or economic data).

4. TAKING QUALITY INTO ACCOUNT

Quality, as we have seen, is the ability to meet the requirements of users. But if some requirements are the same for all users, some requirements are quite different, even come into conflict with others requirements. For example, in the requirement concerning the use of SIRENE as a statistical register, we meet the reference to economic termination in the register, while in the requirements of administrative partners, we meet the reference to a “legal” termination.

That is why, actually, to improve the quality of the SIRENE register is “to enhance the quality of SIRENE data by making up for the limitations of the inter-administrative system in terms of the requirements of user statisticians and those in charge of disseminating the register’s information to the public”. In a way, it is a quite “administrative” approach of the quality.

4.1. TAKING QUALITY INTO ACCOUNT IN DAY-TO-DAY MANAGEMENT

The quality of a register does not depend simply on regular operations to improve quality. It is at the level of day-to-day management of the register that quality control is required.

SIRENE receives standardized “administrative” information via “EDI”: the workstation has powerful automated processing capabilities. Some data sent by administrative partners can thus be processed entirely automatically while other data, rejected by the system, need to be processed manually. This advanced level of automation (with introduction of checks into the data-processing flow) guarantees uniform processing, which is in itself a factor in quality.

4.2. TAKING QUALITY INTO ACCOUNT AS AN “INDEPENDENT COMPONENT”

Alongside the day-to-day management of the register (registrations, for example), INSEE performs operations unrelated to administrative requirements. These operations to improve the register, referred to as “quality operations”, may take many forms but all have the same objective: to enhance the quality of SIRENE data by making up for the limitations of the inter-administrative system in terms of the requirements of user statisticians and those in charge of disseminating the register’s information to the public.

An operation to improve SIRENE may be carried out either at the request of one of the register’s users or at the request of the register managers themselves; they consist in lists of units to be checked (which are drawn up in order to resolve a specific problem concerning one of the variables) or consist in register improvement surveys (questionnaires sent to businesses in order to conduct an overall validation of their characteristics and those of their local units) ; they can be carried out on a periodical basis, or not.

4.3. OVERVIEW OF THE OPERATIONS CARRIED OUT AND OF THEIR VOLUMES

Each year some 100,000 legal units had their activity code updated or validated: the activity code of legal units is updated on an ongoing basis, as and when the results of the annual business surveys are processed.

About 15,000 units were checked at the request of statisticians (wrong address, probable termination of business); the user sends a file of units that he or she considers to contain errors, since the findings of a survey or other source differ from those of the register. The user waits for SIRENE to be checked and for an answer concerning each discrepancy to be provided.

Another example: if a questionnaire is sent to a legal unit, using the address indicated in SIRENE, and the questionnaire is returned undelivered, the service carrying out the survey returns this information to the register, and asks for the address to be checked.

About 15,000 possible duplicate entries were examined; the fact that names and dates of birth are contained in the register means that SIRENE teams can check that the same person is not registered twice.

Some 50,000 units possibly misidentified as operational are currently being examined. In this operation, SIRENE identifies units that, although recorded in the register as operational, are highly likely to have ceased trading (no recent movements concerning these units).

Mass updates

When the results of a survey provide information covering a wide range of a register's variable, they can be used to update the register. Currently, two sources are used for this type of update:

- one concerns the number of employees : every year, all units that have taken on paid employees in the course of the year have the numbers of their employees in the register updated, using a single comprehensive source (by spring 2008, 3.5 million units (legal units or local units) had had their employee statistics updated) ;
- the other concerns the turnover : by summer 2008, 2 million businesses will have had their turnover statistics updated.

“Quality operations” may also consist of surveys in the form of questionnaires sent to businesses in order to conduct an overall validation of their characteristics and those of their local units. This is, for example, one of the methods used to introduce into the register the new French general economic classification (NAF), Rev. 2, derived from the Statistical Classification of Economic Activities in the European Community (NACE), Rev. 2, which is in turn derived from the Standard International Trade Classification (SITC), Rev. 4. (200,000 surveys were sent last year).

Consistency with other registers

The SIRENE register must be consistent with the other administrative registers. An administrative body may therefore request a comparison to be carried out between those registers and SIRENE. These “consistency” operations enable the quality of the two registers compared.

4.4. HOW WILL THESE OPERATIONS GUARANTEE QUALITY IN THE SENSE OF THE EUROSTAT QUALITY REPORT ?

Here, we will try and see the links between these operations and the European quality report, which includes several components.

- A “survey” is relevant if it meets users’ needs. Apart from its inter-administrative use, the register belongs to the “relevant” component: the identification of statistical users and their expectations has been taken into account (addition of specific data for the statisticians for example). Furthermore, some “quality operations” are based on the request of statisticians (cf. 4.3). However, we will see (cf. 5) it is maybe no more sufficient.
- To talk about the “accuracy” component of the register is quite meaningless. But we can say that all the quality operations (units possibly misidentified, mass updates of the activity code or of the number of employees) will improve the “accuracy” of the sampling plans using the register.
- Two components are connected: timeliness and punctuality and accessibility and clarity of the information.

Timeliness and punctuality: because of its inter-administrative role, the register is daily updated thanks to the administrative partners. But this standardized information can be insufficient: for instance, while start-ups are recorded each day, the record of a cessation can arrive late (or never). What’s more, for mass updates, data are included as soon as available. Finally, the discrepancies are examined as quick as possible.

Accessibility and clarity of the information: statistical data have value when they are (easily) accessible by users. That’s why frozen versions of the business register are regularly produced. Some of them are available for all users, some are created for specific needs (short-term or infra-

annual surveys working of short-term administrative sources on job and on wages, working of structural administrative sources (wages, turn-over) some surveys (“tourism”, “employment”))

Frozen versions are produced quarterly (a file with all units - businesses and their local units - which have traded at least one day in the course of the past term and having an economic consistency or hired employees) and yearly (the same file is declined in an annual version). In order to take into account the actualizations of the register, several versions of these files are produced, in a well-defined schedule.

We can also mention that each month, an extract of the register is provided for the staff responsible of the dissemination of information contained in the register and that all statisticians can search in the whole register, thanks to a specific tool, the data of a list of units, at a given date.

Maybe we could say it is “completeness” ?

- Comparability of statistics :

Few statistics spring from the register, but we have to mention “business demography”. Indeed, SIRENE is used to monitor business demographics (statistics for business start-ups; structural statistics relating to the demography of businesses and local units; description of the characteristics of new businesses) : the day-to-day management of the register and the advanced level of automation provide comparability of statistics (in the time).

- Coherence

Maybe a lack of the register : cf. 5.

5. TOWARDS A NEW STATISTICAL REGISTER

Although its deep wealth, the use of SIRENE as a central tool for business statistics comes up against its twofold role (national identification system used in the context of a system of inter-administrative regulations, and central tool for business statistics) : you may be quite sure a firm ceased but, without information, you are not allowed to cease it officially in the SIRENE register. So the administrative register is essential, but has limitations for a statistical use.

That is why, we expect a lot from the future “statistical register”.

As part of the recent overhaul of the annual business statistics, it was decided to build a “statistical register”. As it is not yet developed, we cannot give many details but some lines have already been drafted about, especially about the aims of such a register.

5.1. NEW CONCEPTS, NEW DATA

Of course, this statistical register will be based on the “inter-administrative” one, but it will go further, will be closer to the requirements of the statisticians.

Among the “already-known” aims of the statistical register, we can mention:

- The definition of “statistical units”: the inter-administrative register manages legal units and local units. But more and more statistics are coherent in so far as the observed units are groups or parts of a group. In some French surveys anyway (such as the annual enterprise survey), the questionnaire is dedicated to the group and no more to its affiliates. The new register will include this level (and describe the links to define these new units from legal units).
- This register should include some new data, such as the “probability for a unit to be still operational”. The results of a survey could then be extrapolated to the whole population, taking into account this probability.
- Finally, thanks to these new data, the register will provide “reference populations” for the whole business statistics (exclusion of units only registered for the needs of an administrative partner).

5.2. A NEW “QUALITY MANAGEMENT”

- In this statistical register, statisticians will manage quality: they will decide which populations have to be checked, how they will correct them, independently from the SIRENE register. Of course, the administrative register will go on some “quality operations” such as the examination of duplicate entries, but the statistical register will develop its own ones.
- A “quality operation” would no more be “checking all mistakes in a population”: it would consist in targeting a population, sampling some units, checking them and applying the results to the entire population (using for instance the “probability for a unit to be still operational”).
- Data will be managed, updated continuously, in a coherent way.

Indeed, in the actual register, few surveys can update the register (especially there is no use of the infra-annual surveys). It has to be developed. For example, one of the register’s shortcomings is the existence in the register of businesses still operational when they are in fact defunct. When a survey detects a cessation, it would be possible to cease the unit in the statistical register but not in the administrative one.

- The register should be only based on automatized updates, since it guarantees coherent data. Furthermore, these updates have to be classified, structured, in order to take into account inconsistencies between several sources. Therefore, an entire procedure has to be described in order to get a federating and coherent register (classification of sources, dates, etc) .

- Finally, once this statistical register set up, we will have to define some indicators to measure quality of this statistical register.
- On the other hand, we have to develop an “information feedback”. Currently, when a data is updated in the administrative register, the whole administrative partners are informed (thanks to daily files). But, that is not the case of the statisticians: the majority of them only used the frozen versions (when they use them). With the new register, the wish is to develop such a system: to inform, without delay, all surveys managers that a data has been changed, an activity code updated, a unit ceased and the component “coherence” will be hugely improved. Such information should also return to the administrative register.

All these works will be managed by the French NSI in the coming months, in order to get a federating, accepted, statistical register, which will spread definitions, concepts and data in the whole business statistics system.