

LIFI : the French groups register

Insee (French national statistical institute)
Henri Mariotte

Abstract

The French enterprise group register (Lifi) is produced by the National Statistical Institute (Insee). It is a part of the French system of business registers, with the national statistical business register SIRUS. Administrative data (from French tax authority and national bank for bank rating), commercial data and business survey data (for enterprises controlled by French Government) are used as input for its production. In addition, Lifi will be connected to the European Group Register (EGR) for data on multinational enterprise groups operating in France with data flows going into two directions. For those multinational enterprise groups data available in Lifi is sent as input to the EGR and data from the EGR will be used to complete Lifi on foreign affiliates.

Input data sources used by Lifi provide information on financial relationships between legal units (shareholders and affiliates) and information on legal units involved in these relationships. An algorithm determines the Global Group Head (GGH) and the enterprise groups perimeters. For a given enterprise group two perimeters are available in Lifi depending on the percentage of control : strict perimeter for legal units controlled or owned over 50 % and enlarged perimeter for legal units owned by less than 50 % .

The French enterprise group register was created in 1979 and has evolved since then. In 2015 a new application has been put in production. The new application treats automatically the input data and runs the algorithm that delineates the GGH and the group perimeters. The application also generates a list of potential errors to be analysed and corrected by experts manually if needed.

Each year data from Lifi on enterprise groups (perimeter and basic business register characteristics) is sent as input to the French statistical business register (Sirus).

This paper presents the major developements of the application launched in 2015 :

- on input data : administrative source is used as input data. A business survey could then be cancelled and business respondent burden reduced
- on concepts : the enterprise group is considered as a statistical unit for which the following variables are available in the register :
 - a stable identification number to follow the enterprise group from one year to another in case of continuity
 -
 - nace code
 - global decision center (GDC) and nationality
 - other characteristics as employment, number of affiliates...

Having the enterprise group considered as a statistical unit in the register enables to detect and correct potential errors at the enterprise group level if major changes about these characteristics are detected between two years.

- on connection with the EGR : the application will be used for two mains tasks : 1/ to receive and treat EGR data on foreign affiliates of multinational enterprise groups operating in France. 2/ to identify foreign legal units by using the EGR legal entity identifier (LEID)

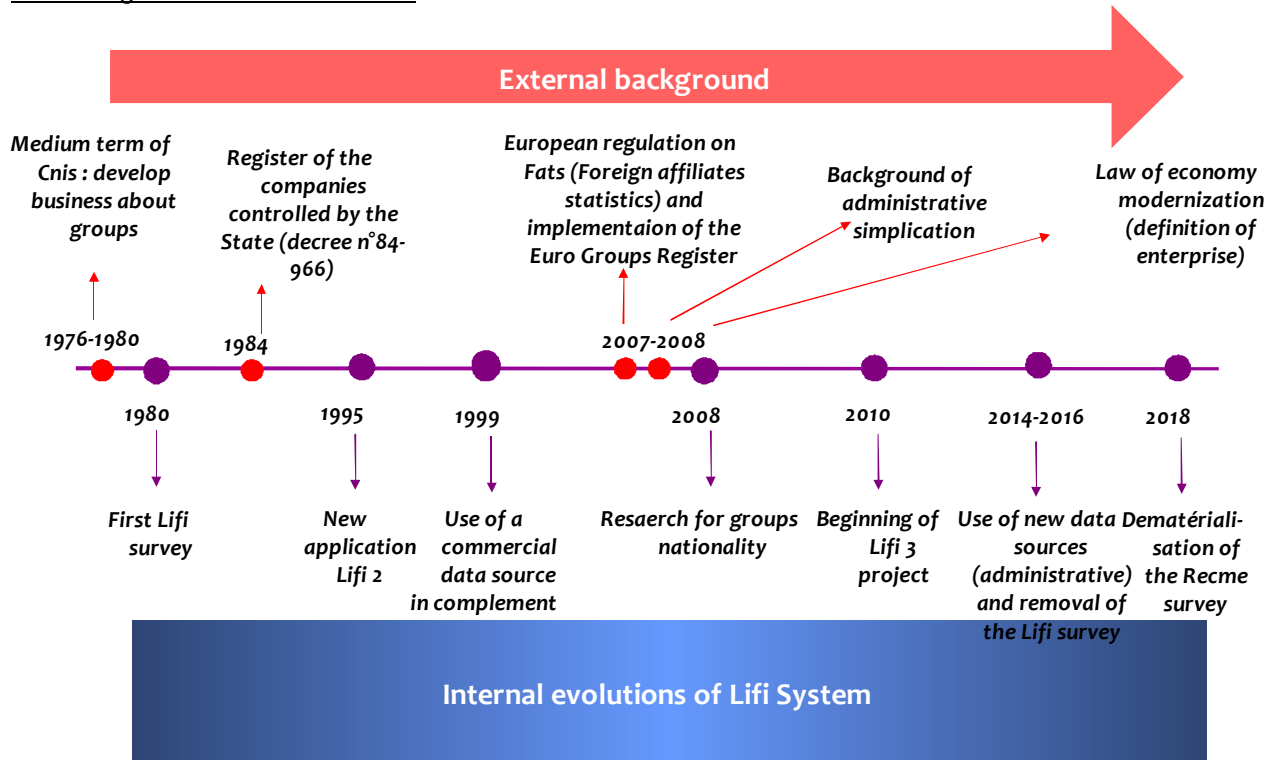
Thanks to the recent improvements of the French enterprise group register the French statistical system is now better equipped for measuring the globalisation of the French economy.

LIFI overview

LIFI ('Liaisons Financières', i.e. Financial links) is both the French group register and also the application managing and feeding this register.

LIFI relies on the knowledge as thorough as possible about the financial relationships between companies (i.e. legal entities): a parent company holds all or part of the shares (voting rights i.e. control) of a subsidiary. Using these relationships, LIFI identifies the group heads and the set of subsidiary directly or indirectly owned by the group heads, which is the group perimeter (core).

LIFI background and data sources



LIFI was set up in 1980 due to a recommendation of the CNIS (national comity for statistical information) done to respond to an evolution of economy where creation of subsidiaries and buying and selling companies expanded in France, whether for organisation or tax optimisation.

Originally and until 1999, LIFI was based on a survey of company shareholders and their holding companies. The companies surveyed were about 40 000. The sampled companies were beyond either an employment or turnover threshold or significant shareholders. Shareholding data were available in tax returns provided by tax authority to Insee.

A more ergonomic application taking better account of expert needs, using a new algorithm for building groups (still used) and integrating the updated French activity nomenclature (NAF rev1, implemented in 1993) was set up in 1995.

Since 1999, Insee noticed that the sample did not ensure an optimal coverage of the scope. As Insee did not have available resources to handle a bigger sample and did not want to increase the burden of respondents it was decided to complete the sample by commercial data. Bureau van Dijk was chosen and has won the tenders ever since.

Methodological improvements were gradually implemented in data processing. Since 2007-2008, following the implementation of FATS regulation, nationality of few groups were checked. Checks especially focused on French group even though the group head was located in foreign country.

Since 2009-2010, following elements contributed to the launch of a new project to lead the implementation of a new application :

- Insee was encouraged to reduce the burden on companies by using administrative sources providing the same information as the survey ;
- The integration of improvements made after the 1995 implementation of the application ;
- Eurostat set up the European Group Register (EGR) and Insee wished to fully be involved in this project and set up tools enabling smooth exchanges.

Tax returns data from tax administration was favoured to take over from the survey. But, it turned out that data were only available in November-December N+1 for the reference year N. It was too late to feed the group register. Contact was made with 'Banque de France' (i.e. national central bank) as it had tax returns duplicate from the companies, on a voluntary basis, in the framework of the rating banking system. Even though this data source was less complete than the tax returns it was available sooner in April N+1 for the reference year N.

The project took place from 2011 to 2015. The new application was gradually put into production between 2014 and 2016. Improvements are still implemented when they are validated. A dozen-expert team, led by two officers, have been using this application. The team has handled manual corrections input (e.g. units and relationships) or output data set up through algorithms, e.g. groups. The 2014 group data were the first to be processed by the new application.

The data sources for the LIFI application are :

- Banque de France (French central bank) data collected as part of the banking rating process ;
- Commercial data of Bureau van Dijk (ORBIS data source) ;
- Tax returns from 2015 ;
- A survey of 700 legal units to manage the RECME (State entity register) ;
- Annual report and website of major groups as release data is mandatory (CAC 40 and SBF 120).

A specific register, the RECME

In 1984, following a massive wave of nationalisation occurred in 1982, the government wished to have a state of play of companies it controlled directly or indirectly. Therefore, it was decided to set up a register of the companies controlled by the State, so called RECME, and Insee would manage the RECME. RECME is a part of the groups information system and managed like this one except for the data source input. Actually the RECME is fed with a survey, as State entity is not under the scope of the other data sources.

Features about administration data sources as LIFI inputs

Data provided by the Banque de France and then from 2015 by the tax administration enabled Insee to remove the survey of companies about their relationships between parent companies and subsidiaries. Moreover, it allowed a significant drop in respondents burden but also clearly led to higher data quality and completeness.

This part deals with background and features on the impact of the new LIFI data source. Until 2011, LIFI main data source was the survey, LIFI recorded 45,000 groups implanted in France including around 200,000 legal units, and among them 4,000 French groups were multinational groups.

Thanks to the Banque de France data source, the number of the group drastically rose to 90,000 groups including 350,000 legal units. And among them 6,000 French groups were multinational.

The new data source effect was spectacular. However, the additional groups recorded were rather small as the economic weight of the group slightly increased.

Due to the 2015 data provided later on by the tax administration, the number of recorded groups increased to 120 000. Thus, the legal units within groups increased steeply to 400,000. However, the additional groups were very small. The impacts on the economic weight of the groups were insignificant. The additional groups mainly consisted on two legal units: one productive unit and real estate company for managing the assets.

Thus, administrative data source improved the coverage and therefore the exhaustiveness of the French group register, however it did not significantly impacted the economic weight of the groups in France.

Data clearing on legal units and relationships

As already mentioned, the group information system is based on financial relationships between companies. Using different data sources, Insee seeks the extensive knowledge of the relationships between a shareholder, i.e. parent company, and its subsidiary, regardless the percentage holding. This percentage holding is a good proxy of the voting right on the board, and therefore of its effective control.

Using different data sources enables to ensure the completeness of basic data (legal units belonging to a group and relationships between the legal units). However, it may lead to inconsistent data and in order to remove these inconsistencies set up priority rules between data sources. Thus, a couple of checks are carried out with a thorough review :

- The percentage holding choice ;
- The processing of legal units held at over 100 %.

A same relationship can be recorded in several data sources with different percentage holding. In case of automatic processing, priority is given to a data source. In case of manual checking, experts control all the inconsistent relationships, that is to say when one source records a relationship within a percentage holding higher than 50 % (majority held) and another a percentage holding less than 50 % (minority held).

Due to multiple data sources, a legal unit may be owned as to more than 100 %. Depending on data source priority, processed either automatically or after expert checking, a relationship can be deleted or corrected to remove the inconsistency.

The relationships between a subsidiary and the State, a natural person (or family), a hedge fund or private equity are deleted. But, the percentage holding is kept at legal unit level.

Moreover, the relationships between a hedge fund and the managing financial institution are deleted as well.

After this process, for each legal unit, one gets the shareholding distribution between French legal units, foreign legal units, natural person (or family), the State and hedge fund.

Algorithm for building groups

Once the data have been cleared, then the step of group building and real relationships takes place.

The algorithm is building the group perimeter, so-called core, step by step. The algorithm builds in parallel "real relationships" linking the group head to all the legal units belonging to the group perimeter.

Setting up of potential group heads

The algorithm starts by setting up potential group heads. That is to say, legal units whose capital is not majority owned by a legal unit and which own majority of capital of at least one legal unit. T0, as it shown on the graph (yellow phase).

Setting up of groups within real relationships processing

For each potential group head, building group and real relationships processing is triggered. This process relates on retained relationships.

Rank 1

- All the relationships involving the potential group head as shareholder are selected
- All the legal units mainly held (percentage holding > 50 %) by this potential group head are integrated in the potential group head perimeter
- Thus, one gets the rank 1 the subsidiaries. Using relationships T0-> S1 and T0-> S2 and adding legal units S1 and S2 as the graph shown (red phase)

Rank 2

- All the relationships involving potential group head and all subsidiaries of rank 1 already integrated in group perimeter are selected
- For each subsidiary with relationships not selected in Rank 1 one sums the percentage holding
- All the subsidiaries for which the sum of percentage holding is higher than 50 % are included in the group perimeter
- Thus, one gets the subsidiaries of rank 2. Using the relationships T0 → S4, S1-> S3 and S2-> S4 and adding the legal units S3 and S4 as the graph shown (blue phase).

Rank n (n>2)

- Same selection procedure as Rank 2
- All the relationships involving potential group head and all subsidiaries of rank less than n ($\leq n-1$) already integrated in group perimeter are selected
- For each subsidiary with relationships not selected and not already integrated into the group perimeter, one sums the percentage holding
- All the subsidiaries for which the sum of percentage holding is higher than 50 % are included in the group perimeter
- Thus, one gets the subsidiaries of rank n.

The algorithm stops when no subsidiary among the selected relationships can be integrated to the group perimeter.

The real relationships are identified by the relationship between the group head and all the legal units belonging to the group perimeter (core).

The legal units outside group perimeters but minority held by one or several legal units of a group are in the expanded group perimeter. Using the relationships S3->S5 and S2->S6 and adding into the expanded group perimeter as the graph shown (black phase).

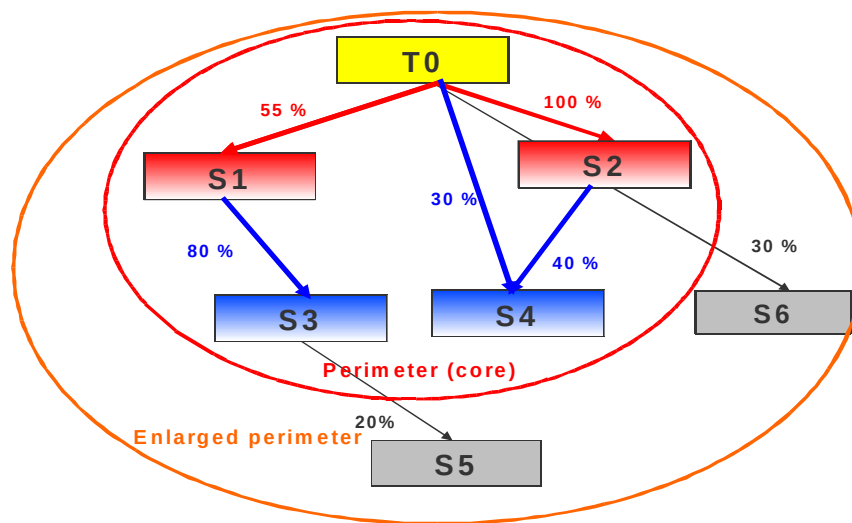
One legal unit can belong to several expanded group perimeters or can belong to on group perimeter and to expanded group perimeter of another group.

The group is identified using a provisional ID number GAAAnnnnnnP where AA is the current processing year, nnnnnn a 6-digit order number (000001, 000002...) and P stands for the provisional character of this ID. Definitive ID will be seen later.

The processing for groups

Once the group perimeter is set, legal unit indirectly controlled by the group head may be added. It may happen the direct shareholder of the legal unit was unknown. The indirect relationship is often known through the group annual report.

Following the step of building group, the step of continuity group management and its identification



takes place.

Group continuity management: identification

Process specification of the continuity rules for groups are very close to those about enterprises' continuity rules used under profiling. Actually, they have been jointly set.

For the legal units within a group perimeter (last output of the algorithm) the set of legal unit intersects with the previous group perimeter is selected as both group Ids (group ID from year n and n-1).

If the set of legal unit intersects accounted for more than 50 % of the total group workforce as compare as the current and previous group states, there is group continuity. For these groups the provisional ID number of the group is updated by the previous ID.

If the group is not continuous, the provisional identifier is retained as final ID.

The first year processed by the application is 2014. The first year for group identification was therefore 2013 (i.e. previous year for 2014). Most of group IDs were number as G13xxxxxx and other IDs were G14xxxxxx for new groups. The released group ID, GAAxxxxxx consists on 9 digits: G for Group, AA the first year of group processing, xxxxxx an order number.

Additional processing

After comparing group information from the previous year (n-1) on : group head, perimeter, size and main activity, discrepancies popped up. Group experts individually check all group with discrepancies and large new groups.

Foreign groups are checked using ORBIS. This ORBIS database allows tracking group head and Global Decision Centre (GDC).

Data groups and perimeter are compared at the end of the process to: OFATS data (Outward FATS), profiling of largest groups process and data on consolidated perimeter from tax administration.

For sizeable groups (i.e. more than 250 employees) with a group head located in a foreign country, experts check group nationality.

The full process leads to the all set of groups. For each group, following information is available:

- Group perimeter (set of controlled legal units), full French group perimeter
- The set of minority controlled legal units
- Global Group Head (GGH)
- Global Decision Center (GDC)
- Nationality of the group, i.e. GDC nationality
- Main activity, using Nace code rev2 and French activity nomenclature (NAF)
- The group size in term of workforce, number of French or foreign legal units

Calendar of LIFI cycle

For a given reference year N, LIFI cycle starts in May-June N+1 and ends in April-May N+2.

At the beginning of the cycle (May to August), experts process the identification of misidentified legal units and data collection from larger groups Annual Reports.

Then (September), after corrected and completed data loading relationships are processed (percentage holding, legal units held over 100 %)

The algorithm runs to set up groups and groups process step occurred between October N+1 and April N+2.

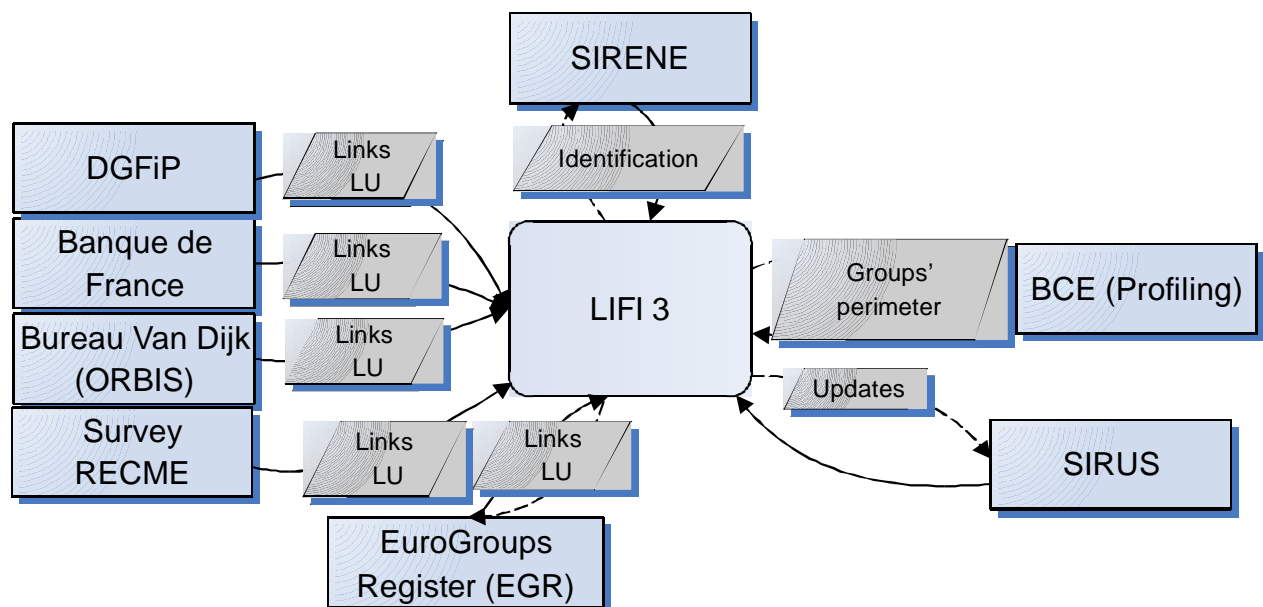
LIFI a part of French Business Register network

LIFI is one of the elements of the French Business Register network with :

- Sirene, the French administrative register for the legal unit and their local units ;
- SIRUS the French Statistical Business Register for enterprises, their legal units, Groups, profiled enterprises, enterprises' network ;
- BCE (i.e. enterprise setting up database) the French statistical register of profiled enterprises, i.e. enterprises with more than one legal unit.

Each year at the end of LIFI cycle, groups' data, i.e. perimeter, expanded perimeter, GGH, GDC and nationality, feeds SIRUS.

Shortly, LIFI will feed SIRUS, as its become available, with provisional groups data as soon in



September-October N+1.

Dataflow between LIFI and European Group Register (EGR)

The dataflow implementation from LIFI to EGR and vice-versa was one of the very aim of LIFI renewal.

However due to LIFI delay and EGR implementation difficulties, integrated dataflow between LIFI and EGR will soon occur.

LIFI is deeply involved in EGR setting up. Actually, every years LIFI provides data (i.e. legal units, relationships and group perimeter) about multinational groups implanted in France. LIFI must be the reference for EGR of the French part of multinational groups, i.e. in term of groups' legal units and relationships.

Vice versa, EGR data will feed LIFI. Actually, EGR data on foreign legal units included in EGR but not in LIFI yet will feed LIFI, for all groups implanted in France. GGH, GDC and nationality of foreign groups implanted in France will be checked using EGR data. At least, their Legal Entity Identifier (LEID) provided by EGR IS (EGR Identification Service) will identify all foreign legal units belonging to the groups implanted in France. EGR IS is one of EGR service available to Member States. This will improve LIFI in three fields :

- Routinely identified foreign legal units with LEIDs provided by EGR. Thus, one could get reliable and durable foreign legal units IDs ;
- Improve the coverage, in term of foreign legal units, of groups implanted in France ;
- Improve the data quality in term of GGH, GDC and nationality of groups implanted in France.

Furthermore, under LIFI EGR exchanges' framework, Insee could offer its expertise about French Legal units and their national ID (Siren number) to improve EGR IS data quality and allow other Member States to get reliable ID for the French legal units belonging to their groups

LIFI input for profiling at national level

The groups set up by LIFI are basic information for profiling implemented by Insee.

Starting from LIFI data, Insee profilers, with group collaboration, delineate profiled enterprises within around 50 large groups. Profiler set up enterprise characteristics e.g. size, main activity, and perimeter. In a second step, profilers will prepare economic information about these units. Later on, LIFI will use as additional data source profiler data collection, especially group control perimeter.

For smaller groups, automatic profiling is also based on LIFI data. The enterprise perimeter consists on the French part of the group perimeter without legal units belonging to financial or agricultural sectors and non-market units.

References :

[1] The calculation of the automatically profiled enterprises characteristics in the statistical business registers, G. Chanteloup, OECD Meeting of the Group of Experts on Business Registers (2017)

[2] Links between trade statistics and profiling in the French business register, O. Haag, OECD Meeting of the Group of Experts on Business Registers (2015)

[3] Construction of a statistical database linked to the interadministrative business directory of the National Enterprise and Establishment Register Database: registering samples to measure the response burden, Hussein-Skalitz A., 12th Meeting of the Group of Experts on Business Registers, Paris 2011