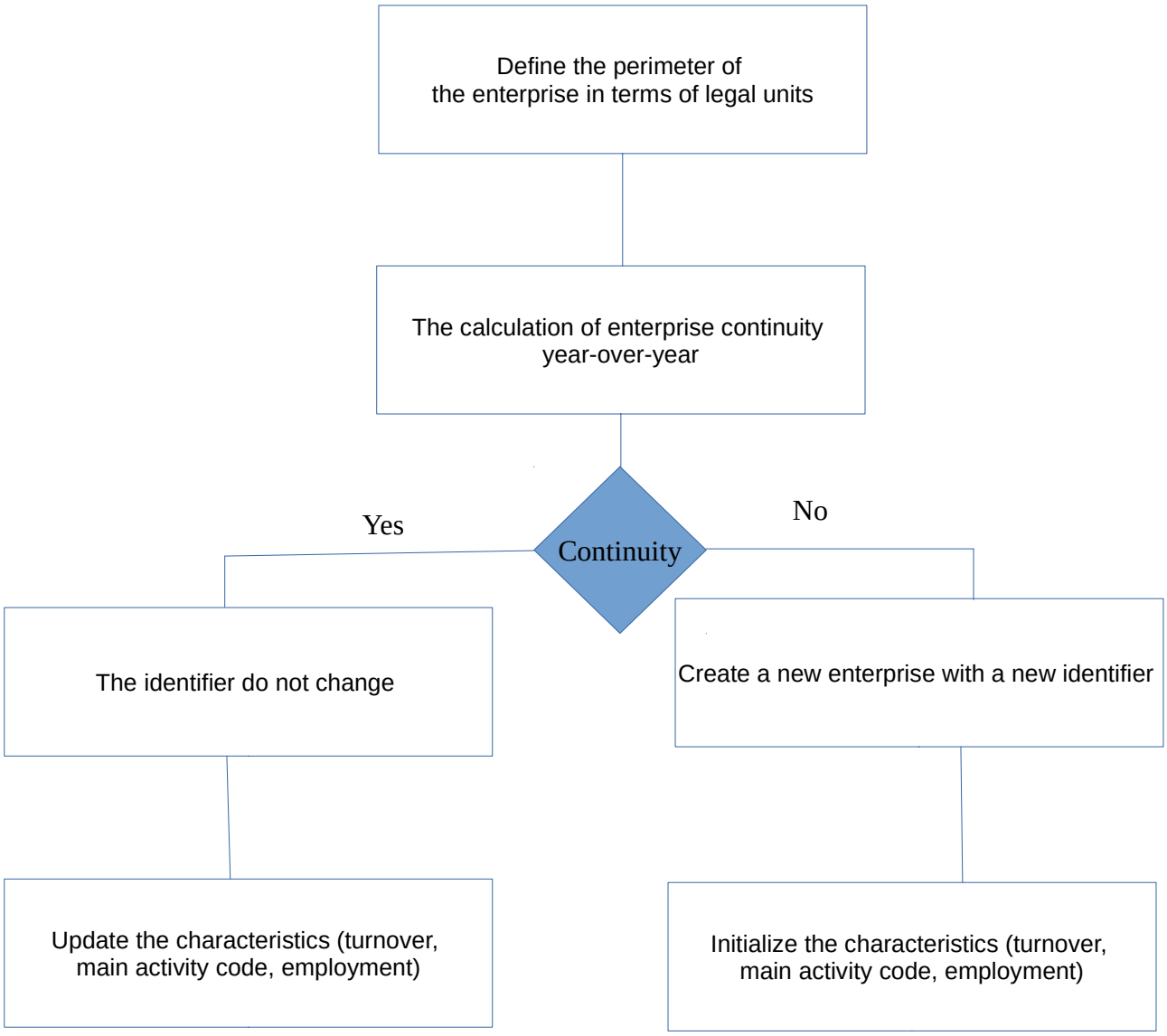
The calculation of the automatically profiled enterprises characteristics in the statistical business register

In order to comply with the 1993 regulation, INSEE set up a profiling process and therefore takes into account the concept of “enterprise group” to calculate the best possible structural business statistics.
Thus, INSEE started with manual profiling. Today, about fifty of the largest groups are profiled and give rise to a hundred enterprises. The Business Register characteristics of this enterprises are directly defined in face to face (manually profiled) with these groups. But this face to face method is costly and could not be applied to all the groups. Then, to treat the less largest groups (not manually profiled), INSEE has set up an automatic profiling system since the year 2016. About 70 000 enterprises have been created in the statistical business register and they bring together about 230 000 legal units. The French Structural Business Survey will take into account these new enterprises. Thus, to set up the SBS frame, it’s also essential to initialize the characteristics of these enterprises that are necessary for the definition of the scope and the stratification of the sampling frame.
This paper will focus on the different stages essential to create, each year, these enterprises in our business register. These stages are represented in the pattern below.



1 – The definition of the perimeter of the automatically profiled enterprises

Insee will set up an automatic profiling process for all the groups with French affiliates not manually profiled. All the groups perimeters are stored in the French Group Business Register called LIFI [1]. This register complete the French Statistical Business register called SIRUS. Starting of groups perimeters, the SBR creates at least one ENT for each group. Actually, an IT application called 'Profilapp' will automatically create ENTs. However Profilapp may be consider as part of the network of French SBR [2].

The ENTs' perimeters are delineated as follow :

The LEUs of the perimeter of each group are split in two parts :

- 1st part : the LEUs under SBS scope excluding financial sector. The subset of these LEUs form the ENT's perimeter
- 2nd part : the other LEUs if any (Foreign or belonging to agriculture or financial sectors) are each considered as an ENT consisting of single LEU

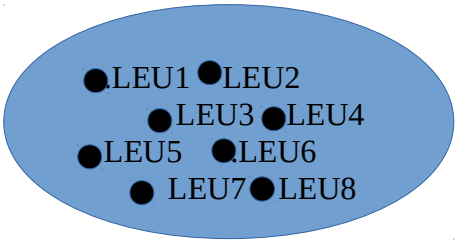
Thus, each group will include at least one ENT. This ENT may be reduced to a single legal unit and eventually independent LEUs which will in fact be out of SBS scope.

Exemple : The pattern below present different scenario.

Type 1 : all the LEUs belonging to the group are under SBS scope

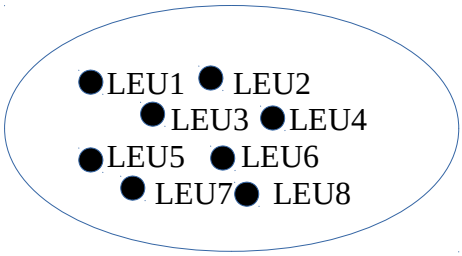
One ENT will be created.

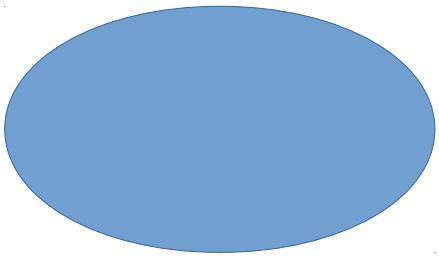
ENT perimeter=Group perimeter={LEU1,LEU2,LEU3,LEU4,LEU5,LEU6,LEU7,LEU8}



Type 2 : none of the LEUs belonging to the group ae under SBS scope

Eight ENTs equal to single LEU will be created. For these ENTs no further steps are necessary.



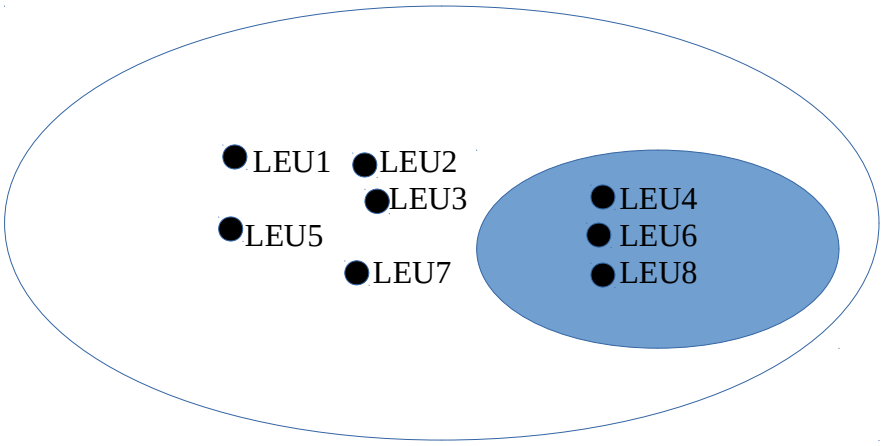


Type 3 : a portion of the LEUs belonging to the group are under SBS scope

One ENT gathering the LEUs under SBS scope will be created and 5 ENTs equal to single LEU outside the SBS scope will be created.

ENT1 perimeter={LEU4,LEU6,LEU8}

ENT2={LEU1} ENT3={LEU2} ENT4={LEU3} ENT5={LEU5} ENT6={LEU7}



This decision has been taken to facilitate the future inclusion of the profiled enterprises by the national accounts. In France, the financial sector is treated by the French national central bank. That is why, to avoid the double counting of the financial unit Insee must eliminate the financial units of the perimeter of its profiled enterprises.

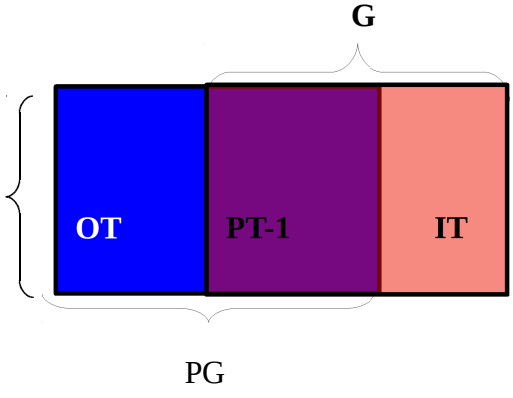
An ID is given to each ENT before the continuity calculation. For each data supply, the continuity of ENT is calculated. If the ENT is the continuation of a previous one, it will keep its ID. If not, a new ID will be assigned. The links between the enterprises predecessors and successors are stored.

The next chapter explains in details this algorithm of continuity.

2 – Le calculation of the continuity and the identification of the unit

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= OT + PT-1
G=PT-1 + IT

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.

When the profiled enterprise is created (new identifier), the characteristics of the SBR are initialized. The main characteristics for the French survey frame are the main activity code, the turnover and the employment. These variables are calculated from a bottom-up algorithms explained in the following chapters.

When the profiled enterprise is continuous, these characteristics can be updated.

- if the profiled enterprise answered to the Structural business survey last year, only the employment is updated. We consider that the main activity and turnover of the survey are better even if the perimeter change a little.
- If not, all the characteristics are updated from the bottom-up algorithms explained in the following chapters.

3 – The calculation of the main activity code

The ENT activity is one of the key variable for SBS and for all the SBR's units. In France the activity is recorded using the French Activity Classification (NAF) derived from the Nace. Actually NAF match perfectly with the Nace but has an additional five-digit level.

Each ENT stored in the SBR should have a Nace code. That's why Insee made a study and set up an algorithm to automatically initialize the Nace code of the ENTs. The principle and the results of the study and the algorithm are detailed below.

The Nace code of the LEU is:

- Defined by the breakdown of turnover into activities according to a top down method for units that were once interviewed in SBS (an annual structural survey ESA in French);
- Declarative when the legal unit is created, or by request from the unit for updating its activity.

The Nace code for ENTs automatically created is determined from an ascending algorithm[1] as used to calculate Nace code for legal unit in the French SBR. It is based on the Nace code of the LEUs of the perimeter. The Nace code is calculated on the basis of turnover without taking into account of the breakdown into activities, simulating the fact that the LEUs are all mono-active, the activity of their Nace code being their activity. For this method, it is therefore the turnover which is used as a proxy of value added. However, for this proxy to be as credible as possible, the turnover of the trade was divided by 3 or 9 in some cases. In fact, for commercial activities it is more the margin than the turnover which must be considered as the proxy, hence this division of turnover without which the trade sector would be overrepresented. Finally, some units, because of their nace code, are considered as ancillary. In this case, their turnover is considered as 0. The LEUs identified as "ancillary" are not taken into account for the calculation of APE at the ENT level. To define the ancillary unit, the ratio "LU workforce/ENT workforce" is taken into account. When these ratios are low for certain sectors, it can be considered to be an ancillary activity and the LU in question is removed from the calculation of APE at the ENT level.

The ratios and activity sectors concerned are defined, according to the situation, at the detailed APE level or at the 2, 3, or 5-digit APE level.

1[1] Link illustrating this method: http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_GENINFO_CLASS_2

Ratio	Sectors
<10%	"4778C" "4941B" "6430Z" "8559A"
<20%	"7810Z"
<20%	"691" "692" "642" "701" "711"
<10%	"66" "68" "77" "62"
<33%	"46"

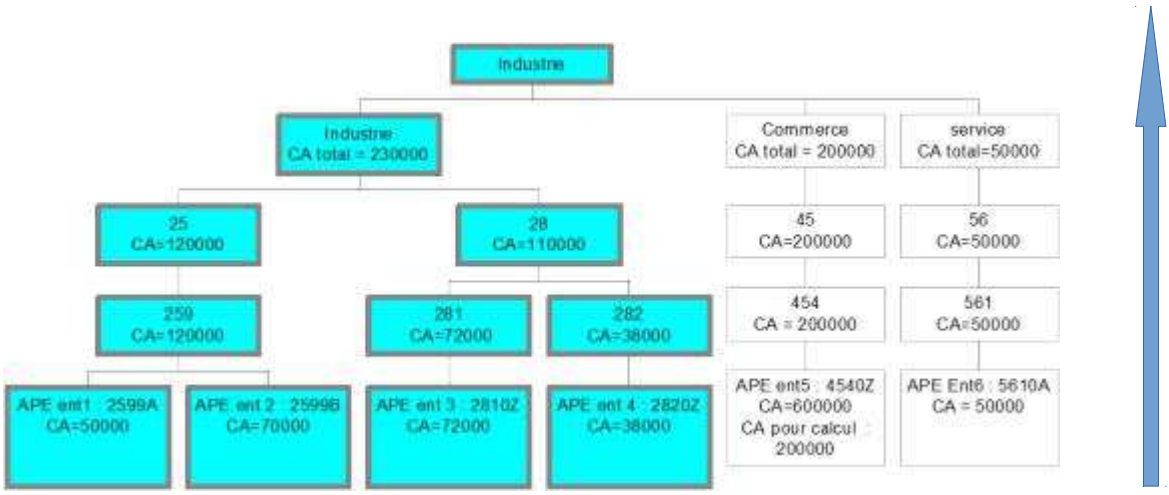
Example of initialization of the Nace code in the SBR for an ENT to be treated automatically :

The first step consists of a bottom up approach establishing the main activity sector of an ENT.

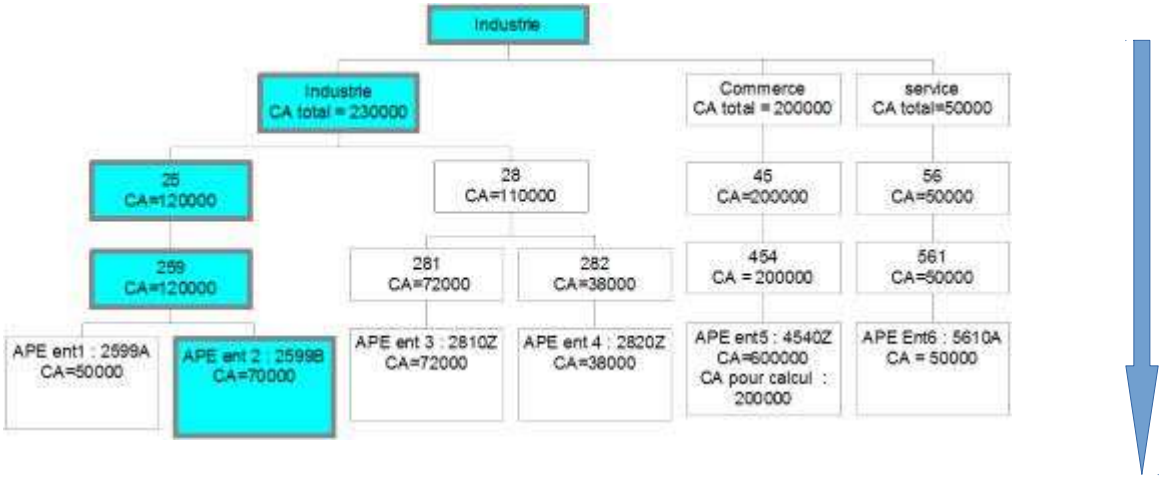
The following ENT is made of 6 LEUs. The characteristics of the LEUs are detailed below.

- LEU1 : NAF/Nace code 2599A (belong to the industrial sector) and records a 50 000 € turnover
- LEU2 : NAF/Nace code 2599B (belong to the industrial sector) and records a 70 000€ turnover
- LEU3 : NAF/Nace code 2810Z (belong to the industrial sector) and records a 72 000€ turnover
- LEU4 : NAF/Nace code 2820Z (belong to the industrial sector) and records a 38 000€ turnover
- LEU5 : NAF/Nace code 4540Z (belong to the trade sector) and records a 600 000€ turnover. As this LEU belongs to the trade sector, the proxy of the value added is estimated by dividing the turnover by 3. Thtat's why for the calculation of the NAF/Nace code, the turnover used in the algorithm for the LEU5 is 200 000€
- LEU6 : NAF/Nace code 5610A (belongs to the sector of services) and records a 50 000€ turnover

First step: a bottom up approach establishing the main activity sector of the ENT.



The **second step** is a top down approach to obtain the Nace code within ENT's main acitivity sector. In this step the chosen branch of the tree where is the branch with the biggest turnover.



As it shown in the graphic above, the ENT NAF/Nace code is 2599B. However, it should be noticed that according to this algorithm the ENT's NAF/Nace code is not necessarily the NAF/Nace code of the LEUs belonging to the perimeter having the biggest turnover !

4 – The calculation of the consolidated turnover

4.1 – Description of the algorithm

The LEUs' characteristics will be used to define the typology used for the consolidation. A consolidated account is created in consecutive steps :

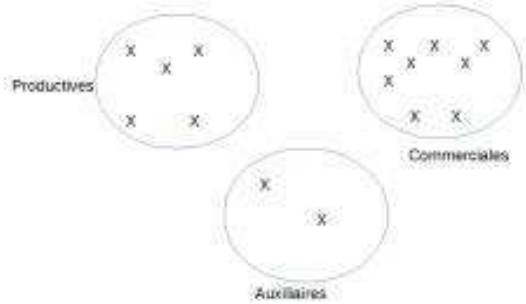
Step 1 : the ancillary, commercial or productive nature of the LEUs that make up the group is determined on the basis of the LEU's Nace code, according the following method :

- a LEU is commercial if the Nace code is between 45 and 47
- a LEU is ancillary if the Nace code belongs to a list of activities, and if the employment of the LEU among the enterprise is lower than a threshold :

  10 % : 4778C, 4941B, 6430, 8559A, 62, 66, 68, 77

  20 % : 7810, 691, 692, 711

- a LEU is productive if it is not commercial, nor ancillary



Step 2 : The aim of this step is to calculate the consolidated turnover of the 2 productive and commercial sub-groups independently and then to consolidate these two sub-groups.

For this purpose, the French classification of activities division has provided 4 lists of integrated activities.[2]

- The first one is a list of upstream and downstream productive activities (integrated production principle which is already employed for the algorithm used to consolidate groups of 2 activities [3]). In this case, the turnover of the upstream LEU is consolidated to the maximum extent with the commodity purchases of the downstream LEU.
- The second one is a list of wholesale and retail commercial activities. In this case, the wholesale trading enterprise's turnover is consolidated to the maximum extent with the retail trading LEU's purchases of merchandises.
- The third one is a list of productive activities that sell to the commercial ones. In this case, the productive activity's turnover is consolidated to the maximum extent with the commercial activity's purchases of merchandise.
- A fourth list of activities is being produced for commercial LEUs that sell to productive ones. This list has not been integrated into the calculation of consolidation for the moment. The memo will be updated with the results relating to this new list of activities.

The list of integrated activities for productive activities could lead to the same LEU being situated downstream of a second but upstream of a third. For example, 0220 (logging) sells to 1610 (sawmilling and

2 For more information : see Annexes.

planing of wood), which sells to 1621 (manufacture of veneer sheets and wood-based panels). In this case, we shall keep a single flow. This will be the one for which the LEU has the highest purchases of raw materials. This will allow for the maintenance of the maximum consolidation effect, without adding complexity for managers who will be responsible for verifying the consolidation links. In 2014, only 17 ENTs were in this situation.

The major change in the new algorithm is the way to ID the links between two LEUs. The calculation of the flow between two LEUs is based on a typology that is explained below.

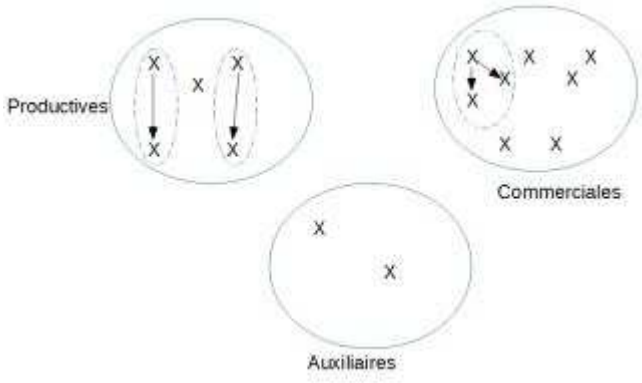
We recall the typology used to define the 13 types of links between LEUs :

Type of link	**By LU**	Accounting relations between LEUs in the ENT
A1	A and (P or C)	Turnover of the auxiliary is recorded under “other operating costs ”of the productive (or the commercial) LU
A2	A and (P or C)	APE of the auxiliary LU(except 64.2, 68, 77)
A3	A and (P or C)	APE of the auxiliary LU=64.2 (activities of holding companies)
A4	A and (P or C)	APE of the auxiliary LU=68 (real estate activities) or 77 (renting or renting and leasing activities)
B1	P and C	APE of the commercial LU =46.1 (intermediaries in wholesale trade)
B2	P and C	Turnover of productive LU is less than purchase of merchandise by commercial LU
B3	P and C	Turnover of productive LU is greater than purchase of merchandise by commercial LU
C1	P and P	Upstream sales of goods by the productive LU are less than item “purchase of commodities and othersupplies”by the productive LU downstream
C2	P and P	Upstream sales of goods by the productive LU are greater thanitem “purchase of commodities and other supplies” by the productive LU downstream
D1	C and C	Wholesale sales of merchandise are less than retail purchases of merchandise
D2	C and C	Wholesale sales of merchandise are greater than retail purchases of merchandise
E1		2 LEUs have the same APE
E2		Simple consolidation: none of the preceding criteria is applied

Effect of consolidation on turnover according to type of link :

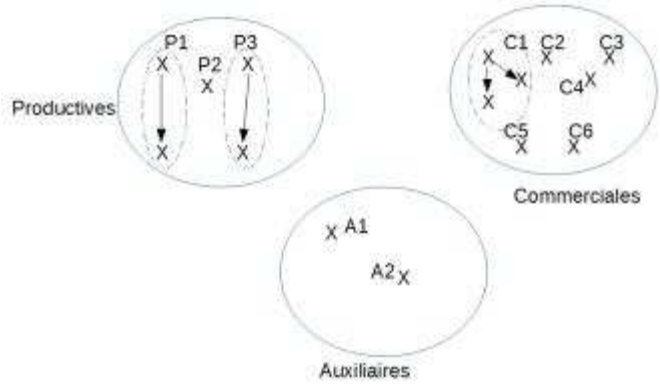
Type of link	**By LU**	Components to be withdrawnfrom the turnover of the profiled enterprise
A1	A and (P or C)	turnover of the auxiliary LU
A2	A and (P or C)	turnover of the auxiliary LU
A3	A and (P or C)	turnover of the auxiliary LU
A4	A and (P or C)	turnover of the auxiliary LU
B1	P and C	“Sales of services” component by the commercial LU
B2	P and C	“Sales of goods” and “Sales of merchandise” components by the productive LU
B3	P and C	“Purchase of merchandise” component by the commercial LU
C1	P and P	Upstream “Sales of goods” component by the productive LU
C2	P and P	Downstream “Purchases of commodities and other supplies” component by the productive LU
D1	C and C	Wholesale “Purchase of merchandise” component
D2	C and C	Retail “Purchase of merchandise” component

The same method is used for the commercial LEUs, with wholesale trading LEUs that sell to retail trading LEUs. In this way, P upstream towards P downstream and C wholesale towards C retail pairings are obtained, which will be consolidated.
Graph 1 :



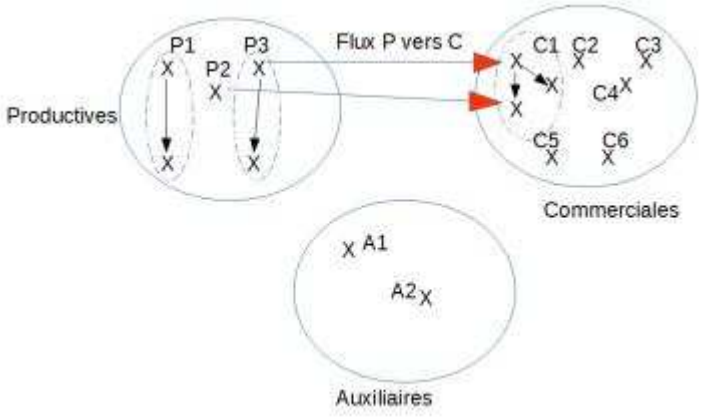
In this way, paired and non-paired LEUs are obtained in each set of LEUs (P, C or A). This action can be represented in the following manner : for P1, P3 and C1, these "sub-groups" are consolidated.

Graph 2 :



Step 3 : in the same way, and based on the Nace code list of productive that sell to commercial LEUs, we will be seeking to identify the Productive – Commercial pairings within the ENT. If at least one LEU of a productive sub-group is potentially associated with a LEU of the commercial sub-group according to list 3 mentioned above, we will then consolidate the set of 2 sub-groups already consolidated (red arrows in the Graph 3). Here, we will therefore consolidate the set (P2+P3) with set C1.

Graph 3 :



Step 4 : finally, we will consolidate the turnover of the sub-group of ancillary LEUs with the sub-groups consolidated in the previous stages containing all of the commercial and productive LEUs.

Example:

The following table presents the tax return variables for the 15 LEUs that make up a profiled enterprise :

	type_LU	Sub-groupe	Purchase of merchandises	Purchase of raw materials	Sell of goods	Sell of merchandises	Sell of services	Turnover
1	P	P1	0	110	136	0	0	136
2	P	P1	0	130	157	0	0	157
3	P	P2	0	15	17	0	0	17
4	P	P3	0	180	215	0	0	215
5	P	P3	0	190	238	0	0	238
6	C	C1	285	0	0	295	0	295
7	C	C1	170	0	0	198	0	198
8	C	C1	50	0	0	85	0	85
9	C	C2	180	0	0	207	0	207
10	C	C3	120	0	0	155	0	155
11	C	C4	170	0	0	201	0	201
12	C	C5	50	0	0	81	0	81
13	C	C6	10	0	0	16	0	16
14	A	A1	0	0	0	0	100	100
15	A	A2	0	0	0	0	200	200

The variable “type_LU” is a variable that characterizes each LEU as productive P, commercial C or ancillary A. The sub-groups, identified as P1, P2, etc. are based on the same sub-groups presented above. Each sub-group that we will be consolidating is identified by a colour.

The turnover for the sum of the LEUs is 2,301.

The consolidation starts with the "sub-groups" P1, P3 and C1 : in P1 and P3, the commodity purchases of LEUs downstream are withdrawn. The turnover (as for the sales of goods) is therefore subtracted : -130 for P1, and -190 for P3. In C1, purchases of merchandises by the retail trading LEUs are withdrawn. -170-50=-220 is therefore subtracted from the turnover.

The consolidation continues with the flow of P toward C : this is the link shown by the red arrow between (P2+P3) and C1 in the graph 3. The sales of merchandise and goods of the productive LEUs are withdrawn, amounting to : 17+215+238 – 190=280. The term -190 corresponds to the consolidated proportion of the sales of goods in "sub-group" P3. The turnover of the remaining, non-paired LEUs is simply added together.

Finally, the consolidation with the ancillary LEUs is carried out. Their turnover is subtracted : -300.

In all, to obtain the turnover of the profiled enterprise, the turnover calculated by simple addition is reduced by -130-190-220-280-300=-1,120. The turnover of the profiled enterprise is therefore 2,301-1,120 = 1,181.

4.2 - Some Results

The following table summarises the intra-group turnover flows per type of flow: auxiliaries, productive to commercial, productive to productive and commercial to commercial.

Table 1 : consolidation effect on all groups in target 2 per type of flow, in K€

	Intra turnover
A	-58,198,624
P+C	-86,149,509
P+P	-8,948,884
C+C	-10,493,127
Total	**-163,790,144**

Half of the total intra turnover flow of €-164 billion is explained by the intra-flows between productive and commercial LEUs.

The following table 2 presents the consolidation effects per major sector. The turnover of the profiled enterprises is compared with the turnover obtained from the sum of the LEUs that make up the profiled enterprise.

Table 2: consolidation effect in A10 per Nace for the profiled enterprise

	Intra flows
Industry	-8,8%
Construction	-5,0%
Trade, transportation accomodation food service activities	-6,6%
Information and communication	-3,9%
Real estate activities	-3,4%
Specialised, scientific & technical activities, and administrative & support service activities	-6,4%
General governement, education, human health and social work services	-2,9%
Other service activities	-1,5%
Total	**-7,0%**

The table 3 below shows, for 2014, the effect of taking account of the Nace on the sector of the LEU or the profiled enterprise and the consolidation effect.

- The first column is the turnover calculated per Nace of the LEU,
- The second column shows the turnover calculated at the Nace level of the profiled enterprise,
- The third column shows the consolidated turnover calculated at the Nace level of the profiled enterprise.

Table 3: distribution of turnover

	Turnover by APE of the LU	**%**	**Turnover by APE of the ENT**	**%**	**Consolidated turnover by APE of the ENT**	**%**
AZ	250,736	0	0	0	0	0
BE	697,306,497	30	808,623,964	35	737,262,256	34
FZ	110,169,044	5	113,514,974	5	107,866,650	5
GI	1,122,107,257	48	1,049,373,019	45	979,719,363	45
JZ	98,845,895	4	94,419,472	4	90,703,858	4
KZ	20,169,711	1	0	0	0	0
LZ	35,846,163	2	36,213,276	2	34,994,444	2
MN	190,986,137	8	169,480,843	7	158,613,436	7
OQ	27,371,075	1	28,775,444	1	27,948,775	1
RU	31,349,041	1	34,000,564	1	33,502,630	2
Total	**2,334,401,556**	**100**	**2,334,401,556**	**100**	**2,170,611,411**	**100**

The difference between the 2 first columns show the reallocation effect.

The difference between the 2 last columns show the “pure” consolidation effect.

The upward sectoral distortion of industry is due to the reallocation of the commercial or service LEUs that belong to the group and which are changed to industry. The downward sector distortion is due to a greater degree of consolidation than for the other sectors, as we have already seen.

This result is in line with the findings observed for the largest groups manually profiled.

5 – The calculation of the employment

The employment is an additive variable. That means the employment of a profiled enterprise is the simple sum of the employment of the legal units which belong to the perimeter of the ENT.

6 – Registration in the SBR

The perimeter and the main characteristics of those profiled ENTs are registered in the French SBR. Their characteristics (main activity, turnover and employment) will be updated by the results of our Structural Business Statistics process called ESANE [4]. The perimeters of the enterprises are updated twice a year.

7 – Conclusion

This article shows that in France, we decide to implement a top down approach to delineate automatically the profiled enterprise and a bottom up approach to estimate the characteristics of these units. This decision allows to have data at the enterprise level without an important growth of the statistical burden. In fact we can use all the administrative data available at legal unit level. The only consequence is that we need to distinguish the study unit (the enterprise) and the reporting unit (the LEU). We must develop algorithm of consolidation and the SBR must allow to make the links between these two types of unit.

References :

[1] The French enterprise group register, H. Mariotte, 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] Report to Eurostat, Analysis of the impact of the changes in the definition of the statistical unit enterprise on SBS - Theoretical description of the breaks in series that could occur when the definition of enterprise will be changed Lot 2 Task 1 NoLI/LU/L1908/ J.M. Béguin, February 2014

[4] *Reengineering French structural business statistics : an overview,* Depoutot R., paper presented at the Q2010 conference, Helsinki

Annexes :

Example of upstream and downtream productive activities :

upstream															
0111	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0112	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0113	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0114	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0115	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0116	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0119	1031	1032	1039	1041	1042	1061	1062	1071	1072	1073	1081	1085	1091	1092	1200
0121	1041	1042	1082	1101	1102	1103	1104	1105	1106	1107					
0122	1041	1042	1082	1101	1102	1103	1104	1105	1106	1107					

Example of wholesale and retail trade commercial activities :

upstream					
4531	4532	4789			
4540	4759	4789			
4621	4721	4729	4773	4776	4781
4622	4776				
4623	4722	4776			
4624	4772				
4631	4721	4729	4776	4781	
4632	4722	4781			
4633	4722	4729			

Example of productive activities that sell to commercial activities :

upstream								
1011	4632	4633	4639	4722	4781			
1012	4621	4632	4639	4649	4722	4778	4781	4789
1013	4632	4639	4722	4781				
1020	4638	4639	4723	4781				
1031	4631	4639	4721	4781				
1032	4634	4725	4781					
1039	4621	4631	4639	4721	4729	4781		
1041	4633	4676	4729	4781				
1042	4633	4729	4781					