

4.4 - Le contrôle qualité de la codification

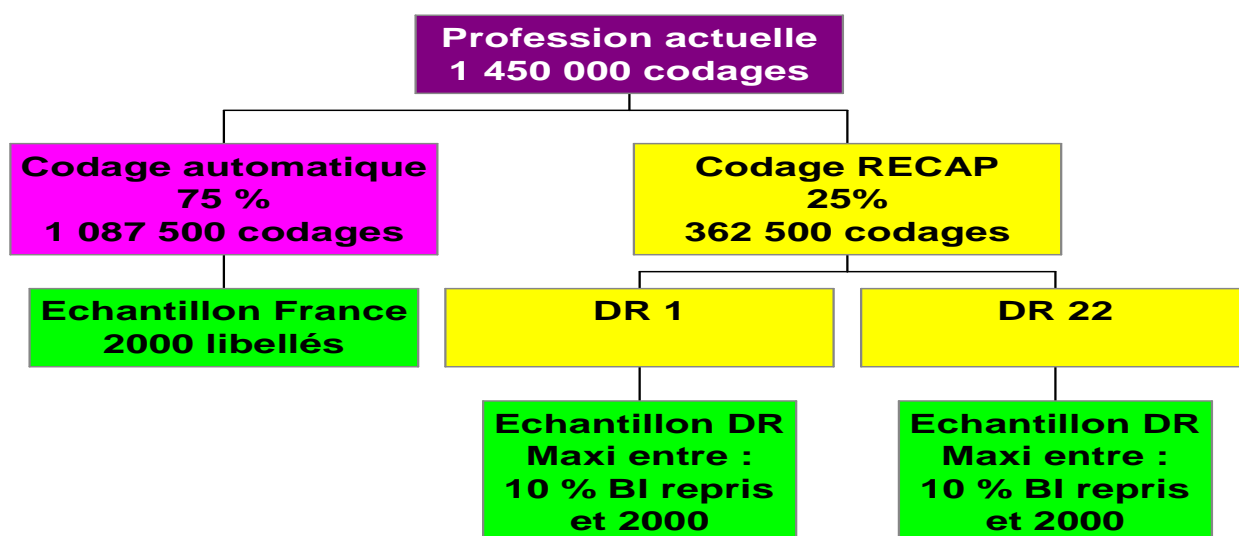
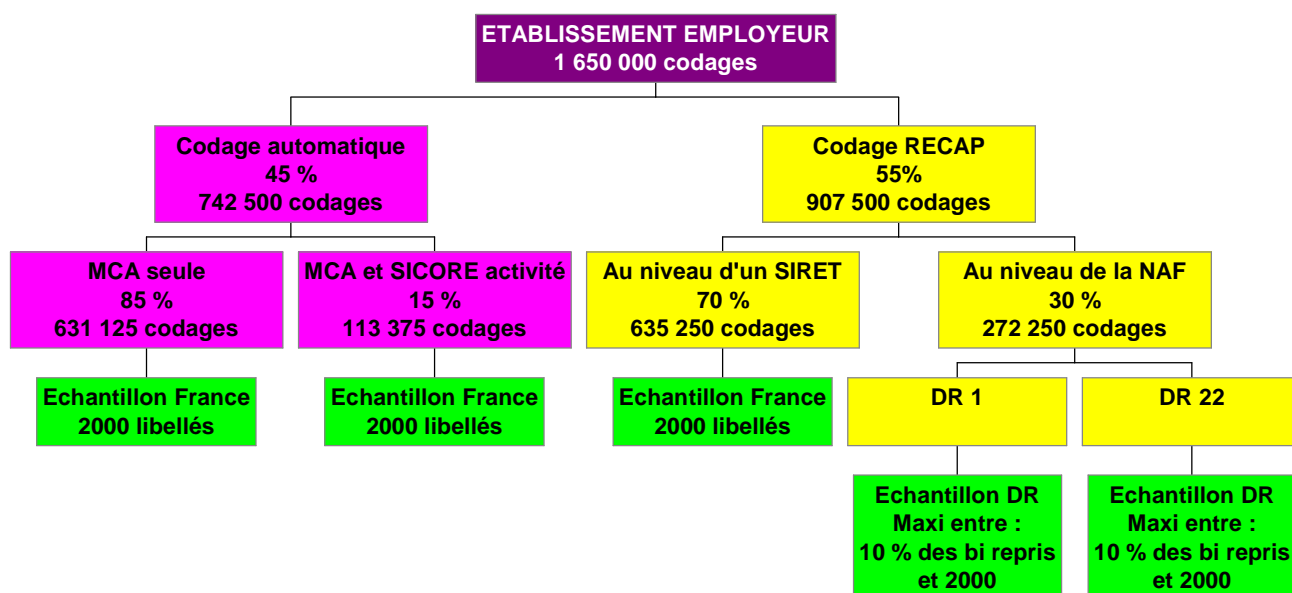
Cette phase permet de vérifier la qualité de la codification des deux principales variables que sont l'activité de l'établissement employeur et la profession actuelle. Contrairement à d'autres, ce contrôle qualité n'a pas pour objectif d'accepter ou de refuser des lots de codage, mais de mesurer la qualité globale du codage qu'il soit automatique ou non. En fonction des taux de codage erronés les actions mises en place sont les suivantes :

- faire un complément de formation dans les équipes RP pour les traitements de reprise ;
- proposer des corrections de la base d'apprentissage de SICORE (pour la profession) ;
- adapter les règles de décision de la MCA (pour l' établissement employeur).

L'opération mesure de la qualité des codages se déroule en trois étapes :

1) Tirage des échantillons élémentaires dans chaque sous-processus

L'ensemble des BI de chaque sous échantillon retenus au titre d'une variable constitue l'échantillon sur lequel la qualité du codage sera mesurée. Cet échantillon est tiré une fois que le traitement de reprise par RECAP des BI en échecs de codage automatique est terminé. Les différents sous-échantillons, sont présentés (en vert) dans les schémas suivants :



2) Réalisation d'un deuxième codage

Un deuxième codage des bulletins de cet échantillon est réalisé avec les mêmes outils que dans RECAP. Cette opération qui nécessite les mêmes compétences que RECAP est réalisée dans quelques DR (9 par exemple en 2006) en novembre et décembre. Les BI provenant de la codification automatique ou de la reprise manuelle sont mélangés, l'origine du codage initial (automatique ou reprise, DR de codage et codeur), de même que le codage initial ne sont pas connus lors de ce deuxième codage. Le mélange des BI des différents sous-processus DR lors de ce deuxième codage permet de préserver l'anonymat des codeurs et permet de déceler un éventuel effet DR lors du codage initial.

3) Arbitrage entre les deux codifications.

Il s'agit d'analyser tous les codages divergents entre le codage initial et le second codage et d'analyser un échantillon parmi les codages convergents. Ce travail demande une compétence d'expert sur les nomenclatures (établissement employeur - activité et PCS). Le poste de travail reprend l'ergonomie actuelle des écrans de l'application RECAP. Cette opération est réalisée dans un petit nombre de DR (3 par exemple en 2006) en janvier et février. Cette étape utilise un module qualité développé dans l'application RECAP.

4) Exemple de résultats sur le double codage

Pour le contrôle qualité réalisé en 2006, quand on compare les deux codages au niveau de chaque BI on trouve des divergences qui deviennent assez importantes au niveau le plus élémentaire des nomenclatures (PCS et NAF700) :

4.1 - Codage de l'établissement employeur - activité selon le processus initial de codage

Processus initial	NOMBRE BI codés		Echantillon	SIRET différent		NAF différent		DIVISION différent	
	Nombre	%		Nombre	%	Nombre	%	Nombre	%
MCA seule	631 000	38,2	2000	231	11,6	159	8,0	131	6,6
MCA + Sicore activité	113 500	6,9	2000	383	19,2	76	3,8	49	2,5
Choix SIRET dans RECAP	635 000	38,4	2000	663	33,2	312	15,6	178	8,9
Codage NAF dans RECAP	272 000	16,5	46181	11280	24,4	19460	42,1	10738	23,3

4.2 - Codage de la profession selon le processus initial de codage

Processus initial	Nombre de BI codés Processus global		Echantillon qualité	PCS différent		CS 42 différent		CS 24 différent		CS 8 différent	
	Nombre	%		Nombre	%	Nombre	%	Nombre	%	Nombre	%
SICORE	1 090 000	75,0	2000	515	25,8	322	16,1	285	14,3	175	8,8
RECAP	362 500	25,0	48520	23197	47,8	14899	30,7	12817	26,4	8826	18,2

5) - Exemple de résultats sur la phase d'arbitrage des divergences

5.1 - Arbitrage entre les codages activité selon le processus de codage initial

Processus initial	Codage 1	Codage 2	Codage autre	Codage flou
Automatique	16,8	74,0	8,4	0,8
Automatique + Sicore activité	57,1	34,7	4,1	4,1
Codage siret dans RECAP	55,6	29,8	10,7	3,9
Codage NAF dans RECAP	36,1	45,4	14,6	3,9

Lecture du tableau : pour les codages initiaux faits en automatique par la MCA, dans 74% des cas l'arbitrage retient le codage 2. Dans 8.4% un autre codage que le codage initial (codage 1) ou de mesure de la qualité (codage 2) est proposé. Le code du codage automatique n'est validé que dans 16.8% des cas.

Pour les deux derniers processus (lignes du tableau), les résultats sont très proches d'une DR à l'autre.

Globalement le taux d'erreur sur le codage de niveau division peut être estimé à 5.8%, avec une contribution au taux d'erreur venant du codage automatique et du codage en NAF sur RECAP de même ampleur (environ 2%). Pour calculer ce taux on utilise les pourcentages de codage par processus et les pourcentages de codages erronés par processus suite à l'arbitrage.

5.2 - Arbitrage entre les codages profession selon le processus de codage initial

Processus initial	Codage 1	Codage 2	Codage autre	Codage flou
Automatique	49,1	40,1	10,2	0,6
RECAP	40,8	42,0	15,4	1,8

Pour la profession, l'arbitrage confirme au mieux la moitié des codages 1, et ce pour les deux processus de codage initiaux (SICORE automatique ou reprise dans RECAP). Le taux de divergence entre codages au niveau de la CS étant de de 16.1 %, le taux d'erreur du codage par SICORE est d'environ 6.5%. Le taux d'erreur sur le codage au niveau CS peut être estimé à 9.8%, sur la CS 24 à 8.7% et enfin sur la CS 8 à 5.6%. Pour calculer ces taux on utilise les pourcentages de codage par processus et les pourcentages de codages erronés par processus suite à l'arbitrage.