

# Variables de Profession et secteur d'activité, calculées par Sicore

---

Version de janvier 2020

## Contenu

1. Pré-traitement des intitulés de professions et de secteurs d'activité .....	1
2. Questionnaire Expositions professionnelles – Codage de la profession.....	2
3. Questionnaire Calendriers professionnels – Codage de la profession et du secteur d'activité .....	3
4. Mise en garde importante :.....	4
5. Quelques effectifs (avril 2016) .....	4
6. Références choisies .....	5

Sicore est un logiciel de codage automatique utilisé par l'Insee. Il permet le codage automatique de la profession en PCS, version 2003 et du secteur d'activité en NAF, version 2008.

## 1. Pré-traitement des intitulés de professions et de secteurs d'activité

Les informations de professions et de secteurs d'activité sont recueillies par des questions de type champ libre. Tous les intitulés ont été prétraités pour éviter des échecs de codage par Sicore. Ces pré-traitements sont les suivants :

- Substitution des tirets '-', apostrophes "'", double apostrophe '"', pourcentage '%', parenthèse '(', ')', étoiles '\*', virgules ',', points ':', slash '/', point d'interrogation '?' et numéro '0' en 'espace'
- Substitution des lettres accentuées ('À', 'Â', 'Ç', 'ç', 'É', 'É', 'È', 'Ê') en lettres non accentuées ('A', 'A', 'C', 'C', 'E', 'E', 'E')
- Remplacement du signe positif '+' en ' ET '
- Remplacement du signe '&' en ' ET '

## 2. Questionnaire Expositions professionnelles – Codage de la profession

Les informations recueillies dans la partie III du questionnaire « Expositions professionnelles » permettent de fournir à Sicore un certain nombre de variables annexes qu'il utilise parfois pour conforter le codage.

Il s'agit :

- du statut de l'emploi AQ\_EXPOACT\_Statut :
  - Indépendant
  - Salarié de sa propre entreprise, gérant mandataire, PDG
  - Salarié
- du statut de l'employeur AQ\_EXPOACT\_Employeur :
  - Etat
  - Une collectivité territoriale, un hôpital, un office HLM
  - La sécurité sociale ou un organisme de sécurité sociale
  - Une entreprise publique nationalisée
  - Une entreprise privée
- de la classification professionnelle AQ\_EXPOACT\_Classif :
  - Manœuvre, ouvrier spécialisé
  - Ouvrier qualifié, ouvrier hautement qualifié, technicien d'atelier
  - Agent de maîtrise
  - Directeur général, adjoint direct au directeur
  - Technicien, dessinateur, VRP
  - Instituteur, assistante sociale, infirmier, personnel de catégorie B de la fonction publique
  - Ingénieur, cadre
  - Professeur, personnel de catégorie A de la fonction publique
  - Employé de bureau ou de commerce, agent de service, aide-soignant, garde d'enfants, personnel de catégorie C ou D de la fonction publique

A l'issue du codage de la profession par Sicore, les variables retournées sont les suivantes :

- **CC\_RES** : Code Sicore : 'C-' = un code PCS est fourni, 'T-' = ce n'est pas un code PCS mais un précode non directement signifiant ou non signifiant
- **PCS\_RES** : Code PCS2003 ; à utiliser UNIQUEMENT si CC\_RES='C-', sinon, c'est un précode

### 3. Questionnaire Calendriers professionnels – Codage de la profession et du secteur d'activité

Le codage par Sicore se déroule en deux étapes. Dans une première étape, le secteur d'activité est codé en NAF 2008 ; et, dans une deuxième étape, le code ainsi obtenu est pris en compte pour le codage de la profession en PCS 2003.

Il existe une particularité pour le codage de la NAF 2008. Le résultat du codage de Sicore est un ensemble de codes NAF 2008 (jusqu'à 5 codes différents) au niveau le plus fin, 5 caractères, assortie d'un indice de fréquence (ou de fiabilité). Selon l'Insee, le résultat au niveau le plus fin n'est pas suffisamment fiable et l'Insee recommande d'extraire uniquement les deux premiers caractères et de ne retenir que cette information. Un traitement est alors réalisé pour sélectionner, parmi les 5 codes à 5 caractères, le code à 2 caractères le plus plausible en sommant les indices. C'est ce code à 2 caractères qui est fourni.

Néanmoins, lorsqu'un seul code NAF à 5 caractères est fourni par Sicore, ce code est également mis à disposition dans les données.

Le code NAF sur 2 caractères obtenu est alors proposé à Sicore comme variable annexe pour le codage de la profession, en complément du seul statut (Salarié ou saisonnier-intérimaires/A votre compte).

A l'issue du codage par Sicore, les variables retournées sont les suivantes :

- **CODAGE\_NAF2** : Sicore a codé la NAF2 O/N
- **NAF2** : NAF2 obtenue par Sicore
- **NAF5** : NAF5 obtenue par Sicore (uniquement si Sicore ne fournit qu'un code)
- **CODAGE\_PCS** : Sicore a codé la PCS O/N
- **CC\_RES** : Code Sicore : 'C-' = un code PCS est fourni, 'T-' = ce n'est pas un code PCS mais un précode non directement signifiant ou non signifiant
- **PCS\_RES** : Code PCS2003 ; à utiliser UNIQUEMENT si CC\_RES='C-', sinon, c'est un précode

## 4. Mise en garde importante

Le codage est une activité délicate, très sensible à la qualité de l'information disponible. Sicore ne permet pas de remédier au manque potentiel de qualité de l'information recueillie dans les autoquestionnaires « Expositions Professionnelles » et « Calendriers professionnel », même s'il fournit, dans plus de 7 cas sur 10, un code PCS ou NAF. Par ailleurs, la qualité du codage fournie par Sicore fait l'objet de débats (voir la documentation en annexe).

Pour le codage de la PCS, Sicore nécessite des variables annexes qui ne sont pas toutes recueillies dans les 2 autoquestionnaires. Lorsqu'il manque des variables annexes, Sicore peut quand même fournir un code. Le fait que Sicore fournisse un code ne signifie donc pas forcément codage de qualité.

Ces variables calculées sont donc fournies sous toutes réserves. De plus, Sicore ne fournit pas un code pour tous les libellés et aucune analyse qualitative des libellés non codés par Sicore n'a été faite pour le moment. Ces libellés non codés par Sicore peuvent bien sûr être codés manuellement. **Il est donc exclu de pouvoir considérer que les volontaires pour lesquels un code est fourni par Sicore sont semblables aux volontaires pour lesquels Sicore n'a pas réussi à affecter un code.**

## 5. Quelques effectifs (avril 2016)

Pour indication, les codes sont fournis par Sicore dans les proportions suivantes en avril 2016<sup>1</sup> :

- Expositions professionnelles : 71 243 questionnaires dont,
  - 50 020 avec un libellé de profession actuelle complété dont,
    - 36 751 avec un code PCS 2003 fourni (73,5%)
- Episodes professionnels : 232 570 dont,
  - 201 954 avec un libellé de secteur d'activité complété dont,
    - 61 411 avec un code NAF5 fourni (30,4%)
    - 160 010 avec un code NAF à 2 caractères fourni (79,2%)
  - 226 136 avec un libellé de profession complété dont,
    - 176 150 avec un code PCS 2003 fourni (77,9%)
  - 201 247 avec un libellé de secteur d'activité ET de profession complété dont
    - 125 294 avec un code NAF2 ET un code PCS (62,3%)

---

<sup>1</sup>  

```
select count(*) from Sicore_PCS_EXPOACT;  
select count(*) from Sicore_PCS_EXPOACT where LIB_RES is not null;  
select count(*) from Sicore_PCS_EXPOACT where LIB_RES is not null and substr(CC_RES,1,2)="C-";  
select count(*) from Sicore_PCS_NAF_CPROFF;  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_ACT is not null;  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_ACT is not null and NAF5 is not null;  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_ACT is not null and CODAGE_NAF2="Oui";  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_RES is not null;  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_RES is not null and CODAGE_PCS="Oui";  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_ACT is not null and LIB_RES is not null;  
select count(*) from Sicore_PCS_NAF_CPROFF where LIB_ACT is not null and LIB_RES is not null and CODAGE_PCS="Oui" and  
CODAGE_NAF2="Oui";
```

## 6. Références choisies

1. Pascal RIVIERE. « SICORE: système de chiffrage automatique » Actes des Journées de Méthodologie Statistique 1995
2. Contribution des variables annexes au codage des libellés de profession par le logiciel Sicore : [http://www.invs.sante.fr/publications/2007/note\\_sicore/note\\_sicore\\_complet.pdf](http://www.invs.sante.fr/publications/2007/note_sicore/note_sicore_complet.pdf)
3. Cedo F. et Lopez A. « Codifier la pcs à 4 chiffres : une ambition raisonnable ? analyse d'un flou statistique sur la relation formation-emploi », Insee-Méthodes : Actes des Journées de Méthodologie Statistique 2002
4. Chenu. A. et Guglielmetti F., «Coder la profession : nouvelles procédures, nouveaux enjeux », Paris, INSEE, VIIèmes Journées de Méthodologie Statistique, décembre 2000
5. Destandau Sophie et Warman Romain « Sicore embarqué pour améliorer les libellés et limiter les questions supplémentaires : utilisation dans la future enquête Emploi », Journées de Méthodologie Statistique, 2012.
6. Guglielmetti F. « Autoclassement versus classement objectivé ? Petit exercice sur la robustesse d'une classification socioprofessionnelle », Insee-Méthodes : Actes des Journées de Méthodologie Statistique 2002.
7. Chardon O. « L'amélioration du codage du groupe socioprofessionnel dans l'enquête de santé en milieu scolaire 2012-2013 en utilisant SICORE et uniquement un libellé de profession. » Insee-Méthodes : Actes des Journées de Méthodologie Statistique 2015.