

Changer la collecte - Rénovation des PCS et auto-complétion dans l'enquête Emploi

Olivier Chardon, Division Emploi, DSDS, Insee

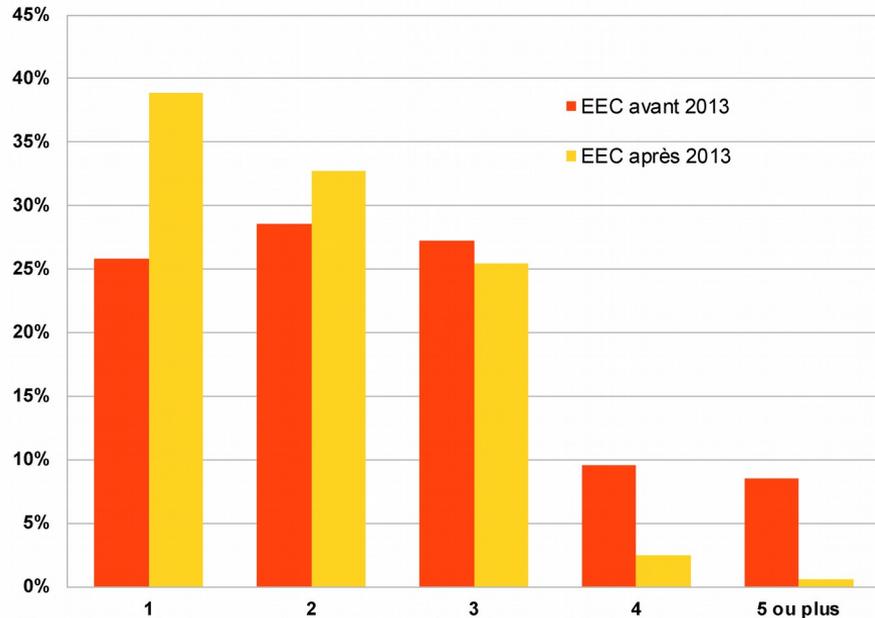
Séminaire Big Data
14 janvier 2020

[Cliquez pour ajouter un titre](#)

- Une nomenclature de 1982 qui repose sur le libellé de profession, une rénovation en 2003. Une codification automatique et une reprise manuelle.
- Au niveau détaillé = **les situations professionnelles** (comme les FAP de la Dares ou Isco au niveau international)
- Au niveau agrégé (groupes et catégories socioprofessionnelles) = **les positions sociales**, (comme les groupes socio-économiques européens EseG)

- Un groupe de travail du Cnis d'une cinquantaine de personnes pendant 2 ans.
- Principaux résultats
 - Une actualisation de la nomenclature au niveau fin et une simplification de son codage
 - Des agrégats complémentaires pour disposer d'un bouquet de grilles d'analyse de la société
 - Des outils à la disposition de tous sur un site internet dédié.
- Une PCS 2020 collectée depuis le 7 janvier 2020 dans le pilote de l'EEC, puis 1^{er} janvier 2021 pour la nouvelle EEC
- Rapport complet sur le site du Cnis

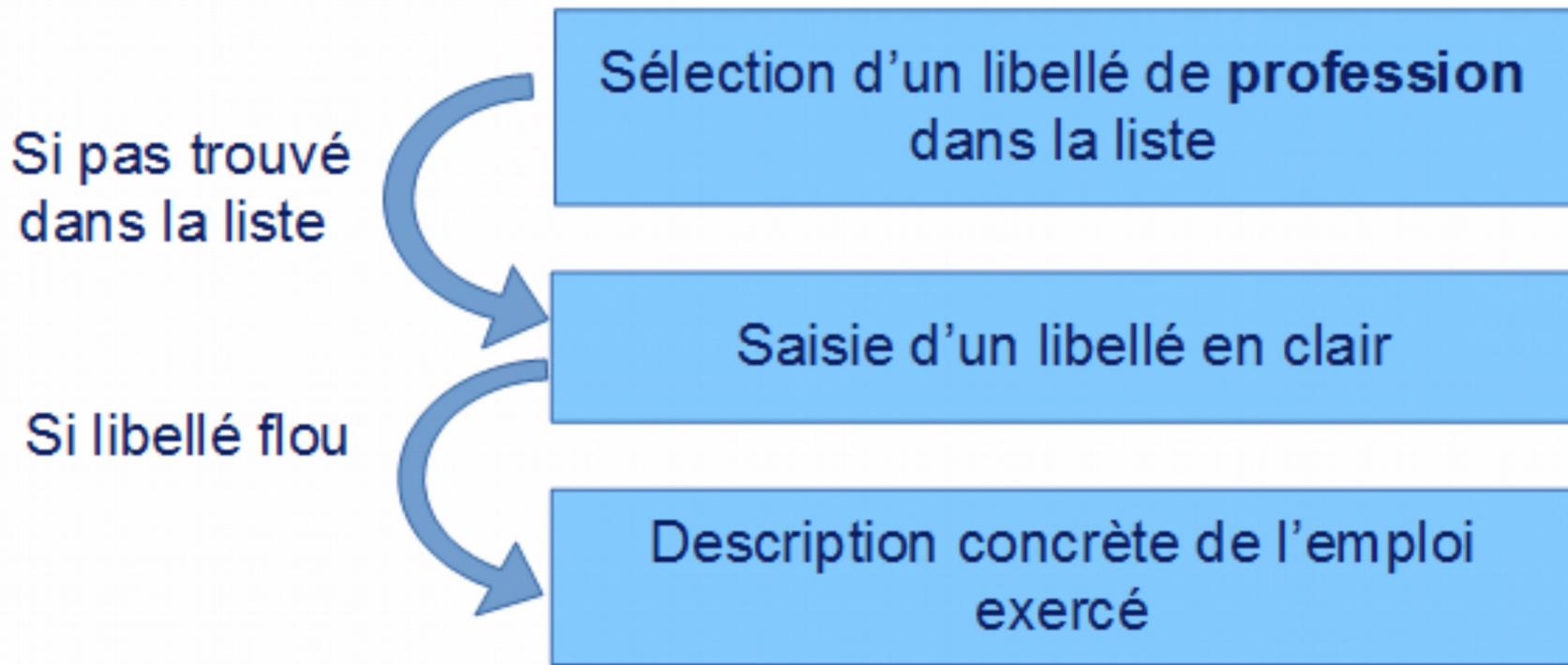
Distribution des libellés de professions selon le nombre de mots signifiants



- Des libellés trop « pauvres » ou trop « riches »
- Beaucoup de variables « annexes » pour la codification automatique ou de la reprise manuelle
- Objectif : avoir des libellés riches et un taux de codification automatique élevé

- Une collecte de « libellés enrichis » dans une liste fermée pour les questionnaires informatisés (face à face, téléphone et internet).
- Une réduction des variables annexes d'une dizaine à trois.
- Une nomenclature simplifiée de 486 à 316 rubriques élémentaires.
- Un libellé choisi dans la liste = un codage automatique au niveau fin

Une collecte du libellé en trois étapes



**!/ ** important pour la reprise manuelle

- Les libellés en clair ont vocation en intégrer la liste ...s'ils sont codables avec les trois variables annexes

Une **ergonomie simple** : dès que l'enquêté a saisi trois caractères, les libellés de la liste qui contiennent la chaîne de caractères saisie sont affichés (par ordre alphabétique, avec les caractères entrés en gras)

- Un **moteur de recherche intelligent**. Mise en place d'une API (**Melauto**)
qfmelautoIht01.ad.insee.intra/melauto/profession?saisie=inge

Nous allons désormais décrire votre emploi.

B1a. Quelle est votre profession ?

Saisissez le libellé de votre profession et sélectionnez-le dans la liste.

 Je n'ai pas trouvé dans la liste

pol

47 professions correspondent a votre saisie

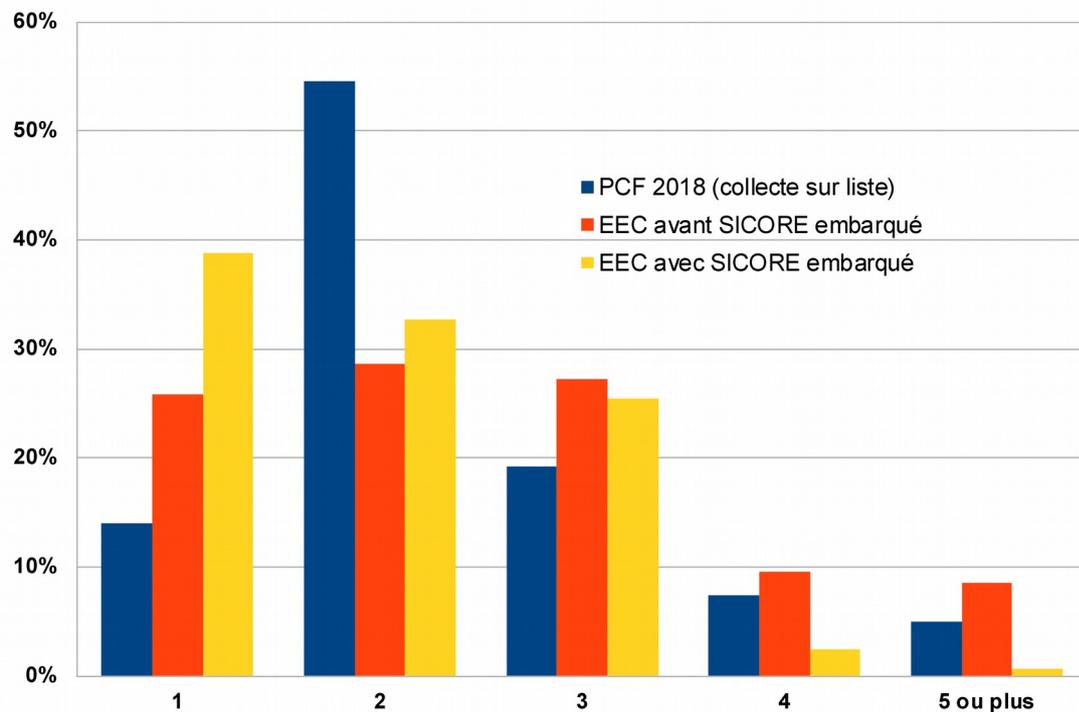
- Agent **poly**valent
- Agent de traitements dé**poll**uants
- Agent de vidéosurveillance de **pol**ice municipale
- Agricultrice en **poly**culture-élevage
- Analyste de la **poll**ution de l'air
- Anthrop**olo**gue
- Brigadier (chef) de **pol**ice municipale
- Brigadier de la **pol**ice nationale
- Cadet de la **pol**ice nationale
- Capitaine de **pol**ice

Pour avoir moins de résultats, précisez votre saisie.

◀ Précédent Suivant ▶

- **Un recours à la liste qui oscille fortement lors des tests** : entre de 83% à 97% pour une collecte en face à face ; de 67% à 87% pour le téléphone et de 62% à 100% pour internet.
- Dans la plupart des cas les libellés « hors liste » sont des libellés « trouvables » dans la liste.
- La présence de la liste entraîne un enrichissement des libellés « hors liste »

Distribution des libellés de professions selon le nombre de mots signifiants



- Un taux d'utilisation des « mots-clefs » souvent utilisés pour décrire sa profession proche de l'ancien protocole

| | Nombre de libellés commençant par le mot dans la liste | Pourcentage de libellés (choisis sur liste) commençant par le mot dans l'enquête PCDF 2018 => 4000 premiers répondants | Pourcentage de libellés (choisis sur liste) commençant par le mot dans l'enquête PCDF 2018 => 7400 répondants | Pourcentage de libellés (déclarés spontanément) commençant par le mot dans l'EAR 2017 |
|--------------|--|---|--|---|
| TECHNICIEN | 246 | 4,9% | 4,8% | 4,8% |
| AGENT | 194 | 9,1% | 8,1% | 7,3% |
| CHEF | 171 | 4,4% | 4,7% | 3,9% |
| INGENIEUR | 167 | 3,4% | 3,6% | 3,3% |
| RESPONSABLE | 147 | 2,4% | 2,6% | 5,5% |
| DIRECTEUR | 123 | 2,3% | 2,3% | 2,8% |
| OPERATEUR | 122 | 0,4% | 0,7% | 1,0% |
| VENDEUR | 101 | 0,8% | 0,6% | 1,2% |
| OUVRIER | 98 | 2,3% | 2,3% | 2,0% |
| AIDE | 85 | 0,8% | 0,5% | 0,6% |
| MONTEUR | 76 | 0,4% | 0,3% | 0,5% |
| ASSISTANT | 63 | 0,7% | 0,7% | 0,8% |
| CONSEILLER | 56 | 1,1% | 1,1% | 1,0% |
| EMPLOYE | 55 | 1,5% | 1,4% | 1,2% |
| TOTAL | 1704 | 34,5% | 33,6% | 35,9% |

- **Un effet «liste »** qui augmente avec le nombre d'échos renvoyés par le mot saisi

| | Réponses collectées avec la première moitié des résultats affichés lorsque le premier mot-clé saisi renvoie... | | | |
|--------------------------|--|------------------|------------------|-------------------|
| | plus de 10 échos | plus de 20 échos | plus de 50 échos | plus de 100 échos |
| 4000 premiers répondants | 61% | 63% | 66% | 71% |
| 7400 répondants | 62% | 63% | 66% | 70% |

- Les améliorations apportées au protocole :
 - l'interface de collecte pour internet
 - le moteur de recherche dans la liste
 - la formation des enquêteurs

- **Exploitation du T1 2020 du pilote l'EEC** (2000 ménages en première interrogation en face à face, 16000 ré-interrogés par internet ou par téléphone)
 - Estimation du recours à la liste par internet
 - Confrontation du recueil du libellé dans l'ancien et le nouveau protocole pour les ré-interrogés

Mise au point d'une méthode de codification pour les questionnaires « papier » du Recensement

- Le défi = le volume de libellé et la forte diversité

Concentration des libellés bruts de la profession actuelle dans l'EAR 2016

| 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% | 100% |
|-----|-----|-----|-----|-----|-------|-------|--------|---------|---------|
| 13 | 44 | 106 | 238 | 553 | 1 465 | 5 099 | 28 924 | 155 580 | 327 090 |

- La contrainte : ne pas augmenter la reprise manuelle (13 %)
- L'objectif : optimiser le recours aux gestionnaires (envoyer en reprise uniquement les libellés qui n'ont « jamais » été repris par le passé... 300000 libellés sont repris annuellement depuis 2004 dans le RP).

Retrouvez-nous sur :

insee.fr



Cliquez pour ajouter un titre