

# Sicore embarqué pour améliorer les libellés et limiter les questions supplémentaires : utilisation dans la future enquête Emploi

*Sophie DESTANDAU<sup>1</sup>, Romain WARNAN<sup>2</sup>*

Sicore, système informatisé de codification des réponses aux enquêtes, est un outil qui permet de codifier dans une nomenclature donnée un libellé en s'appuyant sur des bases de connaissance. Cette codification s'effectue en deux étapes : la première étape est une reconnaissance du libellé en le décomposant en bigrammes et en définissant des priorités dans les bigrammes. Sicore normalise également le libellé (synonymes et mots vides).

En cas de succès de cette 1<sup>re</sup> phase, un précode est obtenu. La 2<sup>e</sup> étape consiste alors à affecter un code final à partir d'une table de décision indiquée par le précode obtenu lors de la 1<sup>re</sup> étape. Cette table de décision donne les règles à suivre pour réaliser la codification à l'aide d'éventuelles variables annexes. Les variables annexes améliorent la qualité du codage mais n'améliorent que marginalement le taux de codage automatique. Sicore fournit un écho (nature du succès ou de l'échec du codage), un code en cas de réussite, et une liste de variables annexes qui ont manqué pour le codage en cas de succès partiel.

Jusqu'en 2008, Sicore n'était utilisé que pour la codification automatique, en aval de la collecte. Or, sur la codification de la PCS par exemple, le taux de codage automatique s'élève à 80 % environ. Les libellés qui ne sont pas codés sont pour la plupart des libellés non reconnus pour différentes raisons : libellé imprécis, plusieurs métiers combinés, secteur d'activité plutôt que métier, faute d'orthographe, etc. Or certains de ces problèmes pourraient être résolus, en signalant à l'enquêteur que le libellé n'est pas codé, et en suscitant le cas échéant un échange avec l'enquêté pour améliorer l'intitulé du métier.

Depuis 2006, Sicore a été installé sur le portable (poste de collecte Capi) des enquêteurs pour permettre à Blaise, logiciel de programmation des enquêtes ménages de l'Insee, d'accéder à ses fonctionnalités. Sicore embarqué a été utilisé pour la 1<sup>re</sup> fois pour coder le libellé de la profession en CS dans le tronc commun des enquêtes ménages (TCM). L'utilisation de Sicore embarqué permet de profiter du face-à-face enquêteur-enquêté en demandant un 2<sup>nd</sup> libellé lorsque le 1<sup>er</sup> libellé n'est pas reconnu.

En outre, le codage de la profession utilise le codage de l'activité de l'employeur comme variable annexe. Jusqu'en 2009, le code de l'activité était choisi par l'enquêteur parmi une liste déroulante de modalités. Depuis 2009, le TCM utilise également Sicore embarqué pour coder l'activité selon la NAF au niveau division (deux chiffres). L'originalité de ce Sicore embarqué activité (dit Sicape) est de proposer plusieurs codes lorsqu'il y a ambiguïté ; l'enquêteur peut alors proposer ces différents secteurs à l'enquêté pour qu'il choisisse le plus pertinent.

Depuis 2009, en s'appuyant sur l'expérience du TCM, dans le cadre du projet de Refonte de l'enquête Emploi (Reflee), la Division Emploi a cherché à optimiser le fonctionnement de Sicore embarqué sur la profession sous un autre angle : se servir de cet outil pour cibler les variables annexes nécessaires, et ainsi diminuer le nombre de questions posées dans le questionnaire Emploi (à partir de 2013). Pour atteindre cet objectif, le rôle de l'écho Sicore

---

<sup>1</sup> Insee, DSDS, Division Emploi

<sup>2</sup> Insee, DSDS, Division Emploi

est essentiel car il renseigne justement sur la qualité du codage. En fonction des échos Sicore, on posera telle ou telle question permettant de remplir la variable annexe utile à Sicore.

La refonte a permis d'étendre l'utilisation de Sicore embarqué à tous les libellés de l'enquête Emploi, en espérant un gain de qualité sur les codages ainsi qu'une baisse de la reprise manuelle. Les libellés de professions sont traités de manière plus complexe à celle du TCM : au maximum trois passages de Sicore sont programmés avec un nombre croissant de variables annexes (mais on ne demande qu'au plus deux fois le libellé à l'enquêteur). Les activités, hors celle de la profession principale qui nécessite un appariement avec Sirene (pour obtenir le Siret, l'activité en NAF sur cinq positions, la taille et la catégorie juridique), sont entièrement codées en interactif. Enfin, les diplômes et niveaux de formation bénéficieront également de Sicore embarqué.