

# Codification automatique en ligne

*HUMBERT Stéphane, INSEE, France*

## 1. Introduction

L'ensemble des enquêtes auprès des ménages de l'INSEE sont réalisées grâce à une application CAPI qui permet d'intégrer les différents questionnaires écrits en Blaise. Certaines variables telles que la profession sont collectées dans des zones de textes qui permettent à un enquêteur de relever la réponse exacte donnée par l'enquêté. Elles doivent donc être codifiées systématiquement pour pouvoir être exploitées.

L'opération était jusqu'à présent réalisée à l'INSEE après la collecte en deux étapes :

- une codification automatique à l'aide d'un outil (SICORE) assurant d'une part la reconnaissance des libellés et d'autre part l'application de règles de codage s'appuyant sur d'autres variables appelées variables annexes,
- puis une reprise dans un service de l'INSEE pour les libellés qui ne peuvent être codifiés automatiquement.

Il y avait alors un nombre important de libellés envoyés en reprise et l'ensemble des variables annexes pouvant éventuellement servir à la codification devaient être collectés. Pour réaliser des gains de productivité et de délai, la codification automatique de ces variables a été insérée sur le poste de collecte.

Après une rapide présentation du fonctionnement de l'outil SICORE, nous verrons quelles fonctionnalités ont été développées sur le portable des enquêteurs. Nous expliquerons ensuite comment ces fonctionnalités ont été implantées d'un point de vue technique. Enfin, nous exposerons quelques applications de cette codification.

## 2. Présentation de SICORE :

SICORE est un outil qui permet de codifier dans une nomenclature des libellés saisis en toutes lettres en s'appuyant sur des bases de connaissance. Cette codification s'effectue en deux étapes. La première étape est une reconnaissance du libellé qui s'appuie sur une recherche en bigramme. En cas de succès de cette première phase, un précode est obtenu. La deuxième étape consiste alors à affecter un code final à partir d'une table de décision indiquée par le précode obtenu lors de la première étape. Cette table de décision donne les règles à suivre pour réaliser la codification à l'aide d'éventuelles variables annexes. Si toutes les variables annexes utiles ont bien été collectées, on obtient en final une codification.

Un "serveur SICORE" a été installé sur le portable de l'enquêteur pour permettre à Blaise d'accéder à ces fonctionnalités. Ce serveur fonctionne de la façon suivante : au démarrage, les bases de connaissance sont chargées en mémoire vive. Le serveur reste alors en attente des demandes de connexion de postes clients. Lorsqu'une telle connexion est établie, il reçoit les requêtes en provenance du client, effectue ces requêtes et renvoie le résultat au client. Du fait du chargement de l'environnement en mémoire vive la réponse est quasi-instantanée sur un portable enquêteur.

La requête et la réponse à la requête sont formées de chaînes de caractères aux formats prédéfinis. La chaîne de la requête contient notamment le nom de l'environnement SICORE à utiliser pour effectuer la codification, le libellé à codifier et la liste des variables annexes renseignées avec leur valeur. La chaîne de la réponse contient quant à elle le résultat de la codification, et éventuellement le précode, le code final et la liste des variables manquantes (variables dont SICORE a eu besoin et qui n'étaient pas dans la liste précédente).

### **3. Fonctionnalités de l'outil de codification en ligne**

Avec la codification en ligne des variables, nous avons cherché à exploiter au mieux les possibilités offertes par SICORE. La codification en ligne permet ainsi de s'assurer de la reconnaissance du libellé, de vérifier la qualité du libellé, de ne poser que les variables annexes utiles à SICORE et enfin d'obtenir le code final.

#### **3.1 Reconnaissance du libellé par SICORE**

Lorsqu'un libellé est saisi ou modifié par l'enquêteur, une requête est envoyée au serveur SICORE afin de contrôler si le libellé est reconnu et pourra donner lieu à une codification. En retour, on exploite le code résultat SICORE pour savoir si oui ou non le libellé a été reconnu : si le libellé a été reconnu par SICORE, la codification pourra avoir lieu. Sinon, on indique à l'enquêteur que le code n'est pas reconnu et on lui propose de vérifier sa saisie.

#### **3.2 Vérification de la qualité des libellés**

Certains libellés peuvent être reconnus par SICORE mais peuvent ne pas être satisfaisants car ils ne donnent pas une information suffisamment précise pour aboutir à une codification satisfaisante.

Ainsi si on prend l'exemple des libellés de profession, « OUVRIER » est reconnu par SICORE mais il n'est pas très précis.

Pour éviter ce problème qui risque de faire perdre de l'information au moment de la collecte, une phase de vérification du précode a donc été introduite. Le principe retenu pour effectuer cette opération est de vérifier que le précode fourni par SICORE n'est pas dans une liste de précode à proscrire. Si le libellé conduit à utiliser une table de connaissance réputée peu précise, on demande à l'enquêteur de préciser le libellé qu'il vient de saisir.

#### **3.3 Questionnement sur les variables annexes**

Après la phase de reconnaissance du libellé, SICORE s'appuie sur une table de décision pour obtenir la codification finale à l'aide de variables annexes.

Ainsi, pour codifier la profession, SICORE a besoin de 14 variables annexes. En pratique, SICORE n'a pas toujours besoin de l'ensemble de ces variables pour effectuer la codification. Une seule de ces variables peut parfois suffire à obtenir le code final.

Une exploitation du résultat de la requête envoyée à SICORE qui contient notamment la liste des variables annexes manquantes, permet de ne poser que les questions nécessaires. Pour cela, on récupère la première variable manquante non posée et on pose à l'enquêté la ou les questions qui correspondent à cette variable SICORE puis, on effectue à nouveau un appel à SICORE. Cette opération est répétée autant de fois que nécessaire, c'est à dire tant qu'on a des variables annexes manquantes qui correspondent à une question présente dans le questionnaire de l'enquête.

Parmi les variables nécessaires à la codification, le traitement n'est pas toujours identique. En effet, certaines variables sont de toutes façons nécessaires pour l'enquête alors que certaines ne sont présentes que pour codifier la variable. Pour chaque variable, on peut donc décider au moment de la conception d'une enquête si elle est toujours posée ou si elle n'est posée que dans le cas où elle serait nécessaire à la codification.

#### **3.4 Obtention du code Final**

Lorsque le libellé est reconnu par SICORE et que toutes les variables annexes nécessaires ont été posées, on récupère le code final correspondant à ce libellé.

## 4. Réalisation technique

Outre l'utilisation du serveur SICORE, le développement d'un tel outil s'est appuyé d'une part sur des développements en DELPHI et d'autre part sur des développements Blaise. La partie développée DELPHI a en charge les interactions avec le serveur et la partie développée en Blaise permet d'assurer le questionnement, la mise en forme de la requête pour SICORE et l'exploitation de cette requête.

### 4.1 Développements DELPHI pour l'intégration de SICORE dans CAPI

L'interaction avec le serveur SICORE comprend 3 opérations : démarrer le serveur, l'éteindre et créer un poste client qui envoie une requête au serveur.

Le lancement de ces opérations est géré dans les programmes Blaise par l'intermédiaire de deux bibliothèques nommées capi\_sic.dll et Sicore.dll. Pour assurer la communication entre les programmes Blaise et la couche DELPHI, des fonctions et méthodes de la bibliothèque BLDIIObj.dll ont été utilisées.

Capi\_Sic.dll contient les deux procédures de lancement et de fermeture du serveur SICORE. Ces deux procédures sont lancées par un manipula qui permet à l'enquêteur de réaliser la collecte. Ainsi au moment où l'enquêteur accède à la liste de ses questionnaires, la procédure de lancement du serveur est activée. Cette procédure renvoie un code retour pour indiquer si l'opération s'est passée correctement. De même à la fermeture du questionnaire, la procédure de fermeture est lancée.

Sicore.dll contient quant à elle la procédure qui crée un poste client, envoie le message de la requête et réceptionne le message indiquant le résultat de la requête. Cette procédure reçoit en entrée le libellé de profession, la liste des variables annexes déjà renseignées pour SICORE et le nom de l'environnement à utiliser. Elle écrit alors le message de la requête en respectant le format attendu par SICORE et envoie cette requête. Lorsqu'elle réceptionne le résultat de cette requête, elle renvoie au programme Blaise différentes chaînes de caractères contenant le résultat de la requête, le précode, la liste des variables manquantes et le code final s'il existe.

### 4.2 Développements Blaise pour l'intégration de SICORE dans CAPI

Du côté Blaise, différentes procédures ont été écrites pour assurer la communication avec la procédure DELPHI, la mise en forme des informations pour SICORE et l'exploitation du résultat renvoyé par SICORE. En plus de l'écriture de ces procédures, il a fallu adapter le questionnaire.

#### 4.2.1 Communication avec DELPHI

Pour assurer l'échange entre DELPHI et Blaise, on utilise une méthode « **ALIEN** » dans une procédure spécifique. Cette procédure reçoit en entrée toute l'information qui doit être transmise à la procédure DELPHI et réceptionne en retour toute l'information que la procédure DELPHI doit retransmettre à Blaise. Elle assure donc la passerelle entre le datamodèle Blaise et le programme DELPHI. Elle s'écrit de la façon suivante :

```

PROCEDURE CallPrecodage
  PARAMETERS
    IMPORT
      LibProfession           : STRING
      ...
    EXPORT
      CodeSicore              : STRING
    EXPORT
      CodeFinal               : STRING

    ALIEN('C:\Proggen\dll\SICORE.DLL','Precodage_CS')
ENDPROCEDURE

```

## 4.2.2 La mise en forme des données pour SICORE

Pour fonctionner, SICORE attend d'une part, une chaîne de caractères correspondant au libellé de profession et d'autre part, une chaîne de caractères lui indiquant quelles sont les valeurs prises par chaque variable annexe dont il a besoin.

La deuxième chaîne de caractères se présente sous la forme suivante : « variable1=valeur1, variable2 = valeur2... ». Dans le questionnaire de l'enquête, chaque variable SICORE correspond à une ou plusieurs questions pouvant prendre différentes modalités. Une procédure de transcodification a donc été écrite pour chaque variable annexe utilisée par SICORE. Ces procédures reçoivent en entrée une liste de variables et donnent en sortie la chaîne de caractères correspondante.

Ainsi, si on prend l'exemple de la profession, SICORE a besoin de savoir si la personne travaille dans une entreprise publique ou privée pour obtenir certains codes. Dans le questionnaire, cette information peut être obtenue par la variable Statut. La procédure s'écrit alors de la façon suivante:

```

PROCEDURE TransCodePUB

  PARAMETERS
    IMPORT
      Statut : StatuType
    EXPORT
      Libelle : STRING

  RULES
    Libelle := ''
    IF STATUT = etat THEN
      Libelle := 'PUB=1,'
    ELSEIF STATUT = coll THEN
      Libelle := 'PUB=2,'
    ELSEIF STATUT IN [entr,part,fam,chef,indep] THEN
      Libelle := 'PUB=5,'
    ELSEIF STATUT = DK OR STATUT=RE THEN
      Libelle := 'PUB=*, '
    ENDIF
ENDPROCEDURE

```

Pour le libellé de profession, les bases de connaissance de SICORE ne prennent pas en compte d'éventuels accents. Une procédure a donc été écrite pour convertir la chaîne de caractère en une chaîne en majuscule sans accent. Cette procédure supprime également certains caractères spéciaux non pris en compte par SICORE.

### 4.2.3 L'exploitation des données fournies par SICORE

Une procédure générale gère les différentes opérations liées à l'intégration de SICORE. Elle permet ainsi la mise en forme de la liste des variables annexes et du libellé, l'envoi des informations à la procédure assurant l'interface entre Blaise et DELPHI, la vérification de la validité du libellé, la mise à jour de la liste des variables annexes à poser et l'affectation du code PCS obtenu.

Elle reçoit en paramètres le libellé de profession, l'ensemble des variables annexes ainsi que des variables « Pose\_var » indiquant si une variable donnée est posée ou non. Après avoir effectué les différentes étapes, elle renvoie le résultat de la codification, les variables Pose\_Var mises à jour et éventuellement le code final.

La première phase consiste en l'appel des différentes procédures de mise en forme décrites précédemment et la constitution des chaînes de ces chaînes de caractères pour l'envoi à SICORE.

La deuxième phase consiste à appeler la procédure assurant l'interface entre Blaise et DELPHI. Pour cela, on fournit le libellé de profession reçu en paramètre, la liste des variables annexes mise en forme, l'environnement à utiliser ainsi que les tailles des différents éléments renvoyés par SICORE. Pour une variable à codifier donnée, les valeurs de ces différents paramètres sont des constantes stockées au niveau de la procédure.

Les informations fournies en retour par la procédure assurant l'interface entre Blaise et DELPHI sont récupérées dans des variables locales afin d'être exploitées dans les étapes suivantes. Pour cela, la procédure vérifie d'abord la validité du libellé, puis exploite la liste des variables annexes manquantes et enfin renvoie le code final.

La vérification de la validité du libellé se découpe en deux parties : on interprète le « code retour SICORE » et on vérifie si le précode appartient à une liste de précodes « ambigus ». Le « code retour SICORE » permet de savoir si la reconnaissance du libellé a pu aboutir et s'il manquait ou non des variables annexes pour effectuer la codification finale. Lorsque le code retour SICORE indique un échec, on stocke dans la variable Res\_Sicore la nature de l'échec. Si le résultat indique une reconnaissance du libellé par SICORE, on vérifie si le précode est satisfaisant ou non. Cette phase consiste à vérifier si le précode correspondant au libellé appartient ou non à une liste fournie dans un fichier externe. S'il est trouvé dans cette liste alors on affecte à Res\_Sicore une valeur indiquant que ce libellé est ambigu, sinon le libellé est considéré comme bon. Cette variable sera retournée au bloc de questionnement et sera affichée après la saisie du libellé. Les différentes modalités de Res\_Sicore sont ainsi :

- **LibOk** "Le libellé est reconnu et pourra donner lieu à une codification.",
- **LibAmb** "Le libellé est reconnu mais il est ambigu. Il ne pourra donc pas donner une codification précise.",
- **LibNon** "Le libellé n'a pas été reconnu.",
- **Pbxxx** "Un incident s'est produit durant l'opération. Par exemple, non connexion au serveur",

Lorsque le libellé est reconnu, on exploite la liste éventuelle de variables annexes pour mettre à jour la liste des questions à poser dans le questionnaire. Cette opération est la plus complexe à gérer. En effet, lorsqu'une variable est manquante, SICORE prend une valeur par défaut pour continuer sa

codification et trouver si d'autres variables sont manquantes. Ceci implique que, suivant la valeur qu'on donne à cette variable, le cheminement peut être différent et SICORE peut avoir besoin d'une variable qui n'était pas indiquée dans la liste précédente. Il faut donc consulter la liste fournie par SICORE et mettre à jour la liste des questions à poser en ajoutant une seule question supplémentaire à chaque appel. De même, en cas de modification de valeur, si une variable est déjà déclarée comme étant à poser mais n'a pas encore été renseignée, il ne faut pas rajouter une nouvelle variable.

Pour gérer la liste des variables à poser, une variable Pose\_Var a été créée pour chaque question. Elle peut prendre 3 valeurs : « Oui », « Non » ou « déjà posé ». « Oui » indique que la question est à poser mais n'est pas encore renseignée. « Non » indique qu'elle n'est pas posée et « déjà posé » indique qu'elle est renseignée.

On commence d'abord par regarder si on a une valeur de Pose\_Var à « Oui ». Si c'est le cas, on ne modifie pas les valeurs des autres variables Pose\_Var. Sinon, on consulte la liste des variables annexes manquantes fournie par SICORE. On récupère en fait une chaîne de caractères contenant différents noms de variables séparés par des espaces. Il faut tout d'abord isoler chaque mot contenu dans cette chaîne. Ensuite, on analyse les variables manquantes dans l'ordre d'apparition : Pour la première variable rencontrée, si la question permettant d'obtenir cette information n'est pas posée et existe dans le questionnaire, on bascule la variable Pose\_Var correspondante à « Oui » et on arrête l'exploitation de la chaîne de caractère. Sinon, on passe à la deuxième variable et on continue jusqu'à trouver une nouvelle variable à poser ou la fin de la chaîne de caractère.

Lorsque le code retour SICORE nous indique que la phase de codification a eu lieu sans ambiguïté (cas où la liste des variables annexes est vide), la procédure renvoie dans une variable le code obtenu. Ce code pourra alors être affiché dans le bloc ou simplement affecté dans une variable récupérée en sortie.

#### **4.2.4 L'adaptation du questionnement à la codification**

Il s'agit d'une adaptation d'un bloc de questionnaire Blaise classique permettant de prendre en compte les appels à SICORE.

Concernant la structure, plusieurs variables ont été ajoutées pour gérer le questionnement et les modifications.

Ainsi, les variables Pose\_Var, indiquant si une question est à poser ou non, ont été ajoutées pour chaque question.

De plus, le lancement de SICORE s'effectue uniquement lorsqu'une variable utile à la codification a été modifiée pour ne pas alourdir les temps de réponses. Pour cela, chaque variable utile à la codification apparaît en double, la première variable étant la question posée et la deuxième une variable prenant la même valeur. Pour vérifier si on a eu une modification, on teste si ces deux variables sont égales : si elles ne sont pas égales, cela signifie que la question a été modifiée ; on lance alors un appel à SICORE avant d'affecter la valeur de la question dans l'autre variable.

Une variable SIC\_INFO a également été ajoutée après la saisie du libellé pour indiquer le résultat de l'appel à SICORE. Elle permet de dire à l'enquêteur s'il doit ou non vérifier le libellé.

Enfin, une variable Code\_var a été ajoutée pour stocker le résultat de la codification.

Concernant le déroulement du questionnaire, les règles de routage doivent également être adaptées.

En début de bloc, on commence par affecter toutes les valeurs de pose\_Var à la valeur choisie par défaut : « non » dans le cas courant, « déjà posée » si cette variable doit toujours être posée pour les besoins de l'enquête.

Une fois le libellé et le premier appel à SICORE réalisés, la question SIC\_INFO indique à l'enquêteur si le libellé a été reconnu ou non. Selon la valeur de cette variable, on se retrouve ensuite soit dans un déroulement classique (libellé non reconnu ou ambigu), soit dans le nouveau déroulement (libellé reconnu).

Dans le nouveau déroulement, chaque question est posée si la valeur de Pose\_Var correspondante vaut « oui » ou « déjà posée ». Ces valeurs sont mises à jour par la procédure décrite précédemment.

De plus, un test est présent au niveau de chaque question pour vérifier si le libellé est modifié. S'il est modifié, la procédure d'appel à SICORE est lancée.

Le filtrage des questions par les variables Pose\_Var implique que les questions sont posées dans l'ordre où SICORE en a besoin. Cet ordre n'est pas toujours le même que l'ordre « normal » du questionnaire. Il est donc possible d'avoir des « retours en arrière » dans le questionnement : on pose la troisième question du bloc puis la deuxième. Dans la pratique, les questions sont souvent posées dans le même ordre. Cependant, une variable Pause a été programmée pour indiquer à l'enquêteur un éventuel retour en arrière dans le questionnement.

Grâce à l'utilisation de procédures standards gérant tous les aspects de communication avec SICORE, l'adaptation du questionnement consiste simplement à faire appel à la procédure de codification et à filtrer le questionnement en fonction des besoins de SICORE. L'adaptation du questionnement est donc relativement simple.

## **5. Applications de la codification en ligne.**

Le développement de la codification en ligne étant récent, les utilisations de ce procédé sont pour l'instant peu nombreuses mais elles permettent déjà de voir les possibilités qu'offre un tel outil. Le procédé a ainsi été appliqué à deux domaines : le domaine de la profession et le domaine des produits de consommation. Ces deux exemples sont intéressants car la problématique n'est pas forcément similaire : le premier domaine s'appuie sur un nombre important de variables annexes alors que dans le deuxième, aucune variable annexe n'est pour l'instant utilisée. Les efforts d'adaptation se concentrent donc uniquement sur le libellé et l'obtention d'un code le plus précis possible.

### **5.1 Les libellés de profession**

Cette variable est présente dans toutes les enquêtes réalisées par l'INSEE. Pour les exploitations, un code CS ou PCS est utilisé. Ce code est obtenu à partir du libellé de profession et d'une liste de 14 variables annexes. Jusqu'à présent, la codification de masse en fin de collecte permettait d'obtenir un taux de codification automatique de l'ordre de 75 à 80%. Les libellés non codés étaient ensuite envoyés "en reprise" dans un service de l'INSEE qui pouvait traiter ces cas selon 3 grandes méthodes. Si l'échec de codification était dû à une erreur de saisie de libellé, il était modifié et on relançait un appel à SICORE. Si l'échec était dû à une incohérence dans la liste des variables annexes, celles-ci étaient modifiées et SICORE était également relancé. Enfin, si l'échec était dû à une "non-connaissance" de SICORE (le libellé est bon mais n'est pas dans les bases de connaissance), la codification était effectuée manuellement avec l'aide d'un expert qui relevait le cas pour une éventuelle mise à jour des bases de connaissance.

L'outil de codification en ligne a été utilisé sur l'enquête SILC pour la reconnaissance du libellé de profession. L'utilisation de cet outil a permis d'obtenir un taux de codification automatique de 90% et donc de diviser par deux les cas de traitement en reprise et les délais dus à cette opération.

Un bloc de questionnement standard a été développé pour l'ensemble des enquêtes et il est intégré dans la plupart des enquêtes à venir de l'INSEE. Suivant l'importance de la profession dans le thème de l'enquête, les concepteurs choisissent soit de poser toutes les questions sur la profession, soit de ne poser que les questions utiles à SICORE. Dans le premier cas, ils n'utilisent que le module de reconnaissance du libellé pour limiter la reprise alors que dans le deuxième cas, ils limitent également le temps du questionnement nécessaire pour obtenir le code CS désiré. Cette deuxième version est appréciée par les enquêteurs car pour une variable comme la NAF2 (NACE). En effet, cette variable est posée de façon standard dans les enquêtes sous forme d'une classification hiérarchique mais elle reste cependant assez compliquée à obtenir.

Les gains attendus de l'utilisation de SICORE dans le cas du libellé de profession sont donc d'une part une diminution du temps de reprise et d'autre part une diminution du temps de questionnement pour l'enquêteur.

## **5.2 Les produits de consommation et magasins**

Pour l'enquête "Budget de Famille 2005", des carnets de consommation sont remplis par les personnes enquêtées puis saisis par l'enquêteur qui peut notamment leur demander des compléments d'information si nécessaire. Dans ces carnets de consommation, on note un libellé de dépense. Ce dernier est ensuite codifié par SICORE selon la nomenclature COICOP. Le problème rencontré lors de l'édition de l'enquête était que le niveau de précision donné par le ménage n'était pas forcément celui attendu par SICORE. Le libellé pouvait être riche tout en n'ayant pas forcément tous les éléments nécessaires. De plus, dans certains cas, le manque de précision pouvait obliger à remonter assez haut dans la nomenclature pour finalement obtenir un code très partiel.

Avec la codification en ligne, le code obtenu par la codification est exploité dès la collecte. Il permet de demander à l'enquêteur de préciser le libellé ou d'améliorer la codification.

Pour assurer une meilleure aide aux enquêteurs, des codes supplémentaires ont été rajoutés dans la nomenclature pour traiter des cas particuliers. Lorsque la codification en ligne aboutit à un code, on vérifie si ce code est présent dans un fichier externe. On peut récupérer dans ce fichier soit une aide spécifique, soit une liste de cas clairement identifiée permettant à l'enquêteur d'améliorer la qualité du code final qu'on obtiendra. Par exemple, si l'enquêteur saisit "boisson", l'aide va lui indiquer de préciser le type de boisson et notamment s'il s'agit d'une boisson alcoolisée ou non alcoolisée. L'enquêteur peut alors modifier le libellé en conséquence.

En cas de codification partielle, une autre solution est proposée à l'enquêteur : il s'agit de terminer la codification de façon manuelle sans modifier le libellé. Pour cela un programme DELPHI a été développé. Il permet de parcourir la nomenclature en rentrant au niveau du code partiel obtenu à l'aide du libellé et donc de compléter la codification. Par exemple, si l'enquêteur saisit "pantalon", le code partiel obtenu correspondra à "vêtement". Si l'enquêteur qui est en contact avec le ménage sait si ce pantalon est destiné à un homme, une femme ou un enfant ; il peut soit compléter le libellé soit terminer directement la codification.

L'utilisation de la codification en ligne dans ce cas nous permet donc d'envisager une amélioration de la qualité de l'information saisie et donc de limiter d'une part le travail de reprise et d'autre part le travail d'imputation. Pour l'instant, un seul test a été réalisé. Il ne permet pas de conclure sur les gains chiffrés d'un tel outil. On peut néanmoins déjà dire que les principes mis en œuvre sont appréciés par les enquêteurs.



## 6. Conclusion

L'utilisation de cet outil de codification en ligne nous permet donc d'envisager un gain de temps en limitant les questions posées d'une part et en diminuant les échecs de codage d'autre part. Il permet également d'améliorer la qualité du libellé saisi en donnant à l'enquêteur des consignes adaptées pour traiter des cas particuliers.

Cependant, cet outil modifie la façon de travailler des enquêteurs et il faut être vigilant sur les consignes données pour l'utilisation. En effet, le but n'est pas d'obtenir 100% de codage automatique. Un libellé n'est pas forcément faux parce qu'il n'est pas reconnu par SICORE : il peut s'agir d'une nouvelle dénomination non prise en compte par les bases de connaissance. Dans ce cas, il faut bien sûr laisser le libellé pour permettre d'enrichir ces bases. Il ne faut corriger le libellé que dans les cas où ce dernier est trop vague ou saisi avec une erreur. Le premier libellé saisi est pour l'instant conservé afin d'analyser la nature des modifications effectuées par l'enquêteur.

Enfin, l'utilisation d'un tel outil implique un investissement important pour la création des bases de connaissance. Cette opération est d'autant plus compliquée dans le cadre d'un nouvel environnement que les bases de connaissance doivent être réalisées avant la collecte.