

N° 74
août 1995

Courrier des statistiques

informations sur le système statistique public



Courrier des statistiques

Réalisé à l'INSEE par la Direction de la coordination statistique et des relations internationales, le Courrier des statistiques est une publication **interministérielle** donnant des informations sur l'ensemble des activités du système statistique public - INSEE, services statistiques de ministères, autres organismes de production ou d'étude.

"**Grâce à cette revue**, précisait E. Malinvaud dans l'éditorial du premier numéro, **chaque statisticien devrait, même s'il a une fonction très spécialisée, être averti des travaux de ses collègues et des problèmes qu'ils rencontrent**". Information générale donc sur les différentes activités du système statistique, mais aussi information plus pointue sur l'évolution des méthodes et des moyens de l'activité statistique.

Le Courrier des statistiques souhaite rendre compte chaque trimestre des **travaux** réalisés par les statisticiens, et plus précisément de la manière dont ces travaux ont été menés, depuis les concertations qui en ont fixé les objectifs jusqu'à la publication des résultats, en passant par la mise au point des méthodes de recueil, puis de traitement de l'information.

La revue présente également les **services** statistiques eux-mêmes, grandes unités de l'INSEE, services statistiques des ministères, autres organismes publics : structure, situation dans le ministère, missions, dispositif de publications ...

Opérations et services sont décrits par les statisticiens qui en ont la charge, mais des **points de vue extérieurs** sont également souhaités, avec l'expérience des statisticiens étrangers, les débats du Conseil national de l'information statistique (CNIS), les comptes rendus des colloques et séminaires, l'avis des utilisateurs de l'information statistique.

Des **tribunes** permettent plus explicitement encore d'ouvrir un dialogue sur les diverses formes prises par l'investigation statistique dans le domaine économique et social, et sur l'utilisation qui peut en être faite : études, recherches, critiques ...

ARGUMENTS

- Yves FRANCHET 3 Statistique publique et société de l'information :
risques et opportunités

DOSSIER : DÉMOGRAPHIE (II)

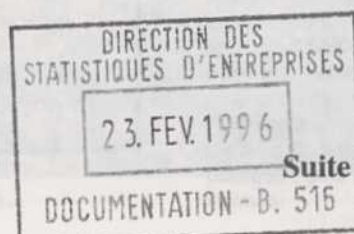
- Michel JACOD 11 La statistique démographique :
tour d'ivoire ou tour de guet ?
- DINH Quang Chi 17 Des pyramides d'aujourd'hui à celles de demain
et Christine COUET (*Projections et estimations nationales de population*)
- Suzanne THAVE 23 Venus d'ailleurs : une statistique encore lacunaire
(*Immigrés et/ou étrangers*)
- Jean-Claude LABAT 29 Les estimations localisées de population
- Jean-Christophe FANOUILLET 33 Télédétection et démographie : l'expérience
de délimitation urbaine à Bordeaux
- Philippe HOUSSAY 40 Géographie et cartographie à l'INSEE
- Alain MALMARTEL 45 Le Limousin place la démographie au coeur de
son devenir

INTERNATIONAL

- Jan-Robert SUESSER 51 Transition statistique ou statistique de la transition ?
(*Les systèmes d'information dans les pays de l'Est*)

OUTILS ET MÉTHODES

- Michèle MANSUY 59 L'Observatoire national des entrées
dans la vie active
- Pascal RIVIÈRE 65 SICORE, un outil et une méthode
pour le chiffrage automatique à l'INSEE



Suite au verso ...

SÉMINAIRES ANNUELS

INSEE 70 Séminaire Recherche
Séminaire Solidarité et protection sociale

COLLOQUES ET CONGRÈS

Pascale GENIER 71 XLVII^e Colloque de l'Association d'économétrie
appliquée : Vieillesse et vieillesse
Athènes, 18-19 mai 1995

NOMINATIONS

74

à ne pas manquer

Colloque L'information économique et sociale aujourd'hui besoins, représentations, usages

14 février 1996

Organisé à l'initiative
de la CFDT et de la CGT
de l'INSEE, ce colloque se
tiendra le 14 février 1996,
dans les locaux de l'ASIEM,
6 rue Albert de Lapparent,
75007 Paris.

Pour tout renseignement,
contacter :

Sophie BOURREL
tél. : 41 17 58 91

Christine BONNAYS
tél. : 41 17 58 90

8h45 Accueil

9h Ouverture

*Quelle information pour quelle
société ? : un débat d'actualité*

9h30 - 12h30

*Atelier 1 : Représentations de
la société et de l'économie*

*Atelier 2 : Outils et méthodes :
normer ou comprendre ?*

14h - 17h

*Atelier 3 : La demande sociale
et le service public de l'informa-
tion économique et sociale*

*Table ronde : Les usagers de
l'information économique et
sociale*

*Atelier 4 : International et
local, quelles évolutions, quels
besoins d'information ?*

*Table ronde : Statistiques et
études locales*

17h15 - 18h30 Table ronde

*Information et démocratie : instances de concertation et
action des personnels*

18h30 Conclusions

SICORE, un outil et une méthode pour le chiffrement automatique à l'INSEE

SICORE (Système Informatique de Codage des Réponses aux Enquêtes) est un système de chiffrement automatique développé par le Département des projets de l'INSEE, qui a fait l'objet de tests sur de nombreuses variables et a déjà été appliqué en production avec succès. Son usage est appelé à se généraliser à l'INSEE, et éventuellement à d'autres services statistiques, en France ou dans d'autres pays.

Pascal RIVIÈRE définit le codage et évoque la méthode QUID, qui a servi de référence, avant de présenter SICORE, sa genèse, sa structure et son environnement. Il aborde ensuite les conditions de son utilisation en production et évoque en conclusion quelques-unes de ses applications.

Le **codage**, ou chiffrement, est une opération qui consiste à interpréter le sens d'un libellé, en attribuant à ce dernier un code. Le libellé doit décrire un concept, i.e. une variable, et le code doit appartenir à une nomenclature, que l'on suppose connue. Il se peut aussi que le texte seul ne suffise pas pour le chiffrement, et que l'on ait besoin de *variables annexes* pour coder : ainsi, pour le chiffrement de la profession, le *statut* ou la *qualification* font partie

des variables supplémentaires généralement utiles.

Le codage est indispensable, dans une enquête, dès lors qu'il existe des questions ouvertes dans le questionnaire et que l'on veut les traiter statistiquement : en effet, il est nécessaire de diviser les réponses en plusieurs catégories, et placer une réponse dans une catégorie n'est rien d'autre que coder.

Malheureusement, le chiffrement "à la main" a l'inconvénient d'être très coûteux ; c'est pourquoi les services statistiques cherchent de plus en plus à le réaliser soit de façon assistée, avec des logiciels interactifs d'aide au codage, soit avec des programmes de codage automatique.

La méthode QUID

L'INSEE s'est préoccupé de codage automatique depuis longtemps : l'algorithme QUID, mis au point en 1979 par Jacques Lorigny à l'unité de recherche de l'INSEE, a porté ses fruits. Le logiciel qui met en pratique cette méthode a été utilisé sur de nombreuses applications, en particulier le codage de la *profession* dans l'enquête Emploi, et le codage de la *catégorie socio-professionnelle* et de la *commune* dans les déclarations annuelles de données sociales.

La méthode QUID est générale, au sens où elle est indépendante de la variable à coder. Mais pour la mettre en œuvre, il est nécessaire de disposer d'un **fichier de référence reliant les libellés aux codes**, appelé **fichier d'apprentissage** : il s'agit en quelque sorte d'une nomenclature

Extrait du fichier d'apprentissage des communes

59	FERRIERE	GRANDE		59230
59	FERRIERE	GRANDE		59230
59	FERRIERE	PETITE		59231
59	FERRIERE	PETITE		59231
59	FLAMENGRIE			59232
59	FLAUMONT	WAUDRECHIES		59233
59	FLERS	ESCREBIEUX		59234
59	FLESQUIERES			59236
59	FLETRE			59237
59	FLINES	MORTAGNE		59238
59	FLINES	RACHES		59239
59	FLOURSIES			59240
59	FLOYON			59241
59	FONTAINE	AU	BOIS	59242
59	FONTAINE	AU	PIRE	59243
59	FONTAINE	NOTRE	DAME	59244

Les "connaissances" dans SICORE

Le "fichier d'apprentissage", que QUID utilisait déjà, est l'une de ces "connaissances". Mais SICORE en emploie d'autres, notamment les "règles de normalisation" (synonymes, mots vides, troncatrices, caractères à éliminer), les "règles logiques" (prise en compte de variables annexes, comme le statut, la qualification, l'activité principale de l'établissement, dans le cas de la profession.), et les "paramètres d'apprentissage" (permettant à l'utilisateur d'orienter l'algorithme d'apprentissage comme il le désire).

Toutes ces "connaissances" dépendent de la variable à coder, mais il s'agit de **connaissances générales** : ce sont toujours les mêmes quelle que soit l'enquête. Pour atteindre cette généralité, les "règles logiques" (ou règles de décision), qui mettent en jeu les variables annexes, font intervenir la modalité "manquant", ce qui permet de coder des enquêtes où la variable annexe est absente, et revient à faire une imputation de celle-ci. Il existe également des **connaissances spécifiques à l'enquête** : "dessin du fichier à coder", "table de passage" entre les variables annexes de l'enquête et celles des règles logiques.

fondée sur les réponses effectives des enquêtés.

Dans un fichier d'apprentissage, les libellés sont divisés en groupes de deux caractères (lettres ou chiffres), appelés **bigrammes** : le premier bigramme de "FONTAINE AU BOIS" est "FO", le second "NT",...

Pour préparer le codage, il faut d'abord "apprendre" ce fichier, en tirer la substantifique moelle ; on obtient une structure résumant le fichier ainsi digéré, appelée *arbre de questionnement* : à chaque noeud de l'arbre, l'algorithme choisit le bigramme apportant le plus d'information, afin que l'arbre soit le moins profond possible (par exemple, dans le cas des communes, le premier bigramme choisi est le 2^e bigramme du premier mot ; la suite du parcours en dépend : voir figure page précédente).

Afin de pallier les erreurs possibles sur les bigrammes eux-mêmes, QUID vérifie, à l'issue du parcours de l'arbre de questionnement, un certain nombre de bigrammes appelés "bigrammes de redondance".

Une fois l'arbre créé, le programme devient capable de coder, en parcourant simplement cette structure : l'arbre généré est remarquablement adapté au chiffrement, très rapide avec cette méthode.

SICORE, du projet au système

Le codage automatique **ne permet jamais de traiter tous les libellés** : il reste une proportion de non-codés, qui n'est pas toujours négligeable. Une analyse de ces échecs devrait permettre d'améliorer petit à petit le fichier d'apprentissage, en y incorporant les intitulés les plus fréquents parmi ceux que l'on n'a pu chiffrer. En pratique, cette amélioration régulière et raisonnée du fichier d'apprentissage n'a pas réellement eu lieu avec QUID ; par exemple, l'enquête Emploi n'a pas changé ce fichier de référence depuis la création de la chaîne de codage automatique, en 1990.

On peut aisément expliquer cette lacune. En fait, l'enrichissement d'un fichier d'apprentissage nécessite une organisation adaptée : des experts, spécialisés dans le chiffrement d'une variable, des logiciels facilitant le travail de mise à jour, une centralisation des "connaissances" (voir encadré ci-dessus) sur chaque domaine, et une prise en compte réelle des chiffrements automatiques déjà effectués. Or ce n'était pas le cas : le chiffrement automatique proprement dit fonctionnait bien, mais le travail *autour* du chiffrement restait à faire.

SICORE a donc été lancé, courant 1993, avec **deux objectifs** essentiels : d'abord, **bâtir une organisation et mettre au point des outils autour du codage** ; mais aussi **généraliser l'utilisation du codage automatique** à l'INSEE, le rendre plus facilement accessible aux non-informaticiens.

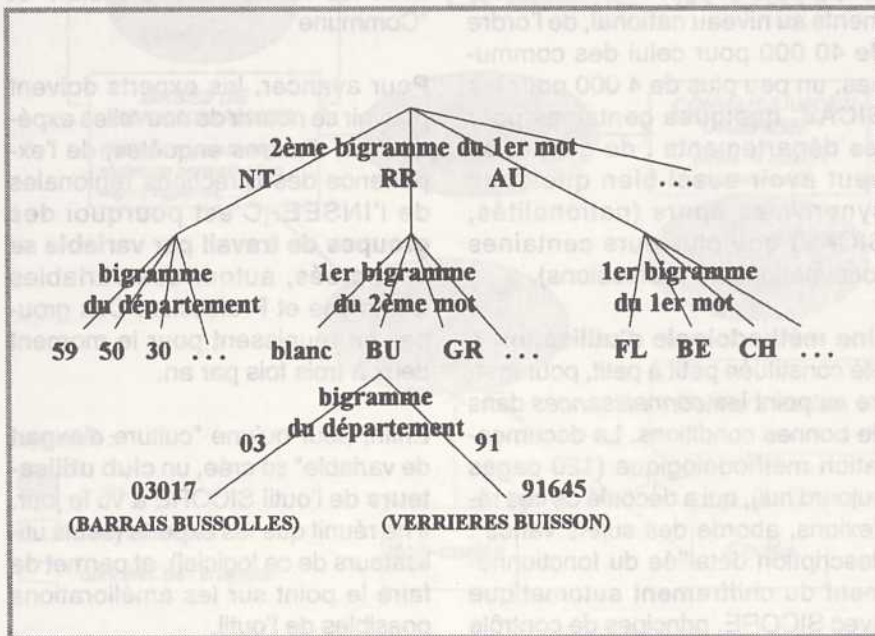
La volonté a donc été affichée dès le départ de créer un système simple, unificateur et efficace pour le chiffrement automatique, dont l'algorithme de codage s'appuierait, mathématiquement, sur QUID, mais qui en assouplirait et en généraliserait l'approche. Comme l'efficacité n'est jamais de 100 %, il fallait préparer le cycle vertueux d'enrichissement des connaissances.

Cela passait d'abord par l'écriture d'outils, liés par exemple à la visualisation des libellés non-codés, ou bien au contrôle de cohérence du fichier d'apprentissage. Cela nécessitait aussi que les connaissances expertes sur le chiffrement existent, et qu'elles évoluent régulièrement sur la base de ces analyses. Enfin, cet objectif ambitieux exigeait également que soient mis en place des repères méthodologiques et organisationnels pour la mise à jour de ces connaissances.

Cet ensemble "programmes + connaissances statistiques + méthodologie + organisation" constitue ce que nous appellerons un système de codage automatique, prenant en compte l'ensemble des tenants et aboutissants.

Pour y arriver, le principe de base du projet a toujours été de "faire d'abord", c'est-à-dire de fabriquer le plus tôt possible des prototypes informatiques et de les appliquer. Ceci permet en effet d'anticiper sur les outils, de se donner la possibilité de les ajuster, de créer méthodologie et organisation de façon naturelle. Il ne fallait pas imposer *a priori* des outils et une méthodologie : ceux-ci devaient se dégager de l'utilisation courante. De même pour l'organisation : elle ne se décrète pas, elle naît de l'action.

Exemple d'arbre de questionnement : la commune



Origine : simple extrait de l'arbre (gigantesque !) obtenu à partir du fichier d'apprentissage des communes.

Lecture : il n'y a que deux communes ayant pour 2^e bigramme du premier mot "RR" et pour 1^{er} bigramme du deuxième mot "BU" : "BARRAIS BUSSOLLES" et "VERRIERES BUISSON". On les distingue par le bigramme de département.

En cette fin d'année 1995, SICORE existe effectivement, mais il ne peut et ne doit être définitif : il faut que ce soit un système vivant, en perpétuel renouvellement.

En effet, on dispose bien des 4 composantes préconisées, mais elles ont toutes vocation à évoluer : les **outils** sont amenés à changer en fonction des desiderata des utilisateurs (ce qui relève de la maintenance de SICORE) ; les **connaissances** s'enrichissent pour améliorer l'efficacité et la qualité des futurs chiffrements ; de nouvelles **règles méthodologiques** émergent d'une utilisation régulière du codage automatique ; enfin, les **organisations** sont malléables et éphémères, leur existence dépen-

dant de leur utilité réelle et de la volonté des acteurs.

Distinguer et conjuguer programmes et connaissances

Les programmes de codage automatique de SICORE sont généraux : ils ne concernent aucune variable particulière ; à eux seuls, ils sont donc incapables de réaliser le moindre chiffrage. Pour fonctionner, il leur faut des **connaissances** sur la variable à coder, complètement indépendantes des programmes généraux.

C'est toujours un couple [SICORE, bases de connaissances] qui opère le codage. La distinction nette qui est faite entre programmes généraux et **connaissances** offre au statisticien l'avantage de bien maîtriser ce qu'il fait : en quelque sorte, les **connaissances** sont des spécifications écrites à l'extérieur des programmes.

Pour coder, SICORE utilise toutes les **bases de connaissances** dont il dispose, et opère en trois temps.

En premier lieu, il simplifie le libellé, grâce à des règles de normalisation (mots vides, caractères à éliminer, synonymes de mots ou de groupes de mots,...) : on passe d'un texte à un autre texte, "apuré".

Dans un deuxième temps, il faut passer de l'univers des textes à celui des codes, transformer ce libellé simplifié en un code (éventuellement intermédiaire) après "apprentissage" du fichier de référence. Sur le plan théorique, cette composante de SICORE s'inspire largement de QUID¹. Cependant, plusieurs modifications ont été apportées à l'algorithme originel, ce qui a permis d'améliorer l'efficacité de codage, d'accélérer considérablement l'apprentissage (SICORE est **100 fois plus rapide** sur certains essais !), et d'obtenir une bien meilleure stabilité du codage lorsque le fichier de référence évolue.

Enfin, si le code obtenu est un code intermédiaire, on lève l'ambiguïté sur le code définitif grâce aux règles logiques de codage intégrant les variables annexes.

Avec SICORE, le codage d'un libellé est très rapide, même si les **bases de connaissances** sont de taille importante : plusieurs milliers de libellés par seconde (temps CPU) sur site central, quelques centaines par seconde sur un PC de type 486.

A l'issue d'un chiffrage avec SICORE, les résultats sont analysés, afin de perfectionner les **bases de connaissances**, ce qui améliorera le prochain codage. Le processus "codage – étude des non-codés et des mal-codés – mise à jour des **bases de connaissances** – nouveau codage" est ce que l'on appelle la **boucle SICORE** (cf. figure).

Un logiciel, des bases, ... et une méthodologie

SICORE est un **ensemble de programmes qui sont des briques**

1. Mais pas sur un plan informatique : aucun programme du logiciel QUID n'a été réutilisé. SICORE et QUID ont en fait pour seul point commun la trame mathématique de l'algorithme d'apprentissage (et donc de codage). C'est à la fois beaucoup et peu.

élémentaires pour le codage : lecture de *connaissances*, apprentissage, normalisation, reconnaissance de libellé, traitement de variables annexes,... Ces briques sont écrites en langage C, et fonctionnent aussi bien sur site central que sur PC.

Sur site central, il n'y a rien de plus que les briques élémentaires, qui sont appelables par des programmes.

Sur PC, toutes ces briques sont réunies au sein d'un seul et même logiciel, fonctionnant sous Windows. Outre les programmes de chiffrement automatique, il comporte **une interface permettant de créer, d'analyser et de mettre à jour des connaissances.** L'objectif majeur de ce logiciel, appelé **outil SICORE**, est de **mettre au point des connaissances** ; le codage automatique fait partie des outils qui contribuent à cette mise au point, mais il n'est pas ici une fin en soi. L'outil SICORE a très peu d'utilisateurs (il est utilisé sur 5 postes aujourd'hui) : ce sont les experts de variables.

SICORE a été appliqué à de nombreuses variables, ce qui a nécessité, dans plusieurs cas, **la création de bases de connaissances.** A l'heure actuelle, de telles bases existent pour les variables suivantes : SICAV, communes, départements, nationalités, professions et catégories sociales, occupations, lieux de séjour, raisons sociales et adresses d'établissements.

Toutes ont conduit à des tests sur des données réelles : par exemple, le codage de la profession-catégorie socioprofessionnelle (PCS) par SICORE a été appliqué à des fichiers de quelques dizaines de milliers de libellés, provenant du Recensement de la population, des Déclarations annuelles de données sociales (DADS), de l'état civil, de l'enquête Emploi, ou de l'enquête Permanente conditions de vie (PCV).

Ces *bases de connaissances* sont de tailles très variables : plus de 4

millions de libellés pour le fichier d'apprentissage des établissements au niveau national, de l'ordre de 40 000 pour celui des communes, un peu plus de 4 000 pour les SICAV, quelques centaines pour les départements ; de même, on peut avoir aussi bien quelques synonymes épars (nationalités, SICAV) que plusieurs centaines (occupations ou professions).

Une méthodologie d'utilisation a été constituée petit à petit, pour mettre au point les *connaissances* dans de bonnes conditions. La documentation méthodologique (120 pages aujourd'hui), qui a découlé de ces réflexions, aborde des sujets variés : description détaillée du fonctionnement du chiffrement automatique avec SICORE, principes de contrôle de cohérence, détection d'erreurs dans une *base de connaissances*, choix des *connaissances* à modifier, mesures de qualité, mise à jour des *connaissances* au fur et à mesure du traitement de l'enquête, codage de libellés hétérogènes,...

La méthodologie ne doit surtout pas être considérée comme une chose définitive et figée : SICORE a dégagé un certain nombre de principes généraux, mais il est probable que l'on en découvrira de nouveaux, ou que l'on en viendra à amender certaines règles existantes.

Une organisation pour deux objectifs

SICORE nécessite l'existence d'une organisation pour deux raisons bien distinctes : d'une part, pour faire en sorte que **les bases de connaissances évoluent et s'enrichissent** régulièrement ; d'autre part, pour être prêt à assurer la **mise en place du chiffrement automatique** dans une enquête ou une source quelconque.

- Pour le premier objectif, SICORE inaugure la notion d'**expert de variable**, dont le rôle est de mettre au point les *bases de connaissances* relatives à une variable. Pour cela, chaque expert dispose de **l'outil SICORE**, sur PC. Le travail d'exper-

tise, destiné à évoluer, a commencé avec les variables "Profession" et "Commune".

Pour avancer, les experts doivent pouvoir se nourrir de nouvelles expériences, d'autres enquêtes, de l'expérience des Directions régionales de l'INSEE. C'est pourquoi des **groupes de travail par variable** se sont créés, autour des variables Commune et Profession. Ces groupes se réunissent pour le moment deux à trois fois par an.

Enfin, pour qu'une "culture d'expert de variable" se crée, un **club utilisateurs** de l'outil SICORE a vu le jour. Il ne réunit que les experts (seuls utilisateurs de ce logiciel), et permet de faire le point sur les améliorations possibles de l'outil.

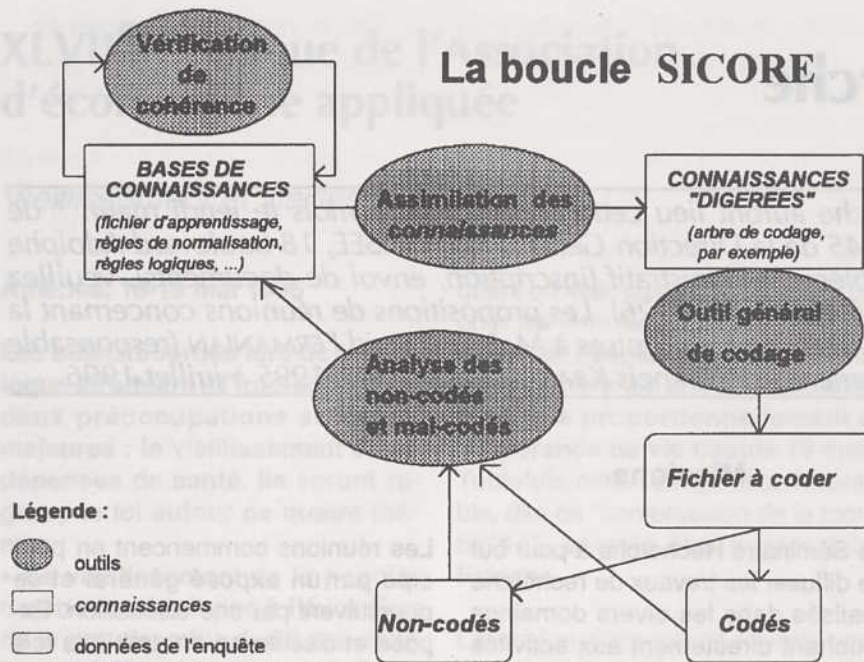
Tous ces réseaux (groupes de travail par variable, club utilisateurs) sont animés par **l'expert SICORE**, membre de l'unité de méthodologie de l'INSEE, plaque tournante de l'ensemble de l'organisation.

- Le deuxième objectif, la mise en place du chiffrement automatique dans une application donnée, nécessite une organisation complètement différente, à nouveau centrée autour de l'expert SICORE. Il s'agit ici de mettre en correspondance trois entités : l'enquête (représentée par un responsable statistique et un responsable informatique), SICORE (représenté par l'expert SICORE et le responsable informatique de SICORE), et la variable à coder (représentée par l'expert de variable).

La difficulté de cette mise en place varie énormément selon la variable et selon l'enquête considérée : quelques jours, ou beaucoup plus longtemps, notamment quand les *bases de connaissances* n'existent pas déjà.

SICORE en pratique

SICORE, en tant qu'outil général, n'est pas (et ne peut être) un outil clé en main : ce sont des *modules*



La codification des CS a porté sur 9 992 libellés, celle des lieux de séjour sur 12 239. **SICORE a codé automatiquement 76 % des CS et 92 % des lieux de séjour²**. Il restait à savoir si les codes SICORE étaient "bons"³ : rechercher dans combien de cas le code automatique différait du code manuel, puis analyser une par une ces divergences.

Dans le cas des lieux de séjour, on s'aperçoit qu'il y a très peu de différences : seulement 3 % des libellés codés différemment du codage manuel. En étudiant un par un les libellés en question, on s'aperçoit que **le code SICORE est juste dans 82 % des cas, et le code manuel dans les cas restants**. Parmi les erreurs de SICORE, citons l'exemple de "CANARIES" (qui n'était pas dans le fichier d'apprentissage), codé comme la ville corse "CANARI".

Le codage d'une CS est en revanche une opération plus délicate : parmi les intitulés de profession codés par SICORE, plus de 1 sur 6 (18 %) diffèrent du code manuel. **L'étude des divergences montre de nouveau que c'est plutôt le codage automatique qui donne les meilleurs résultats** : dans 62,5 % des cas, le code SICORE est le bon code, et le codage manuel a raison dans 17,5 % des cas de divergence. Enfin, dans les 20 % restants, les deux codes peuvent convenir.

Bien entendu, les principes de base de SICORE ont été immédiatement mis en pratique : toutes les erreurs de codage ont été réanalysées une par une, de même que les libellés non-codés, et les bases de connaissances améliorées en conséquence.

Pascal RIVIÈRE

Division "techniques et organisation du travail"
INSEE

Pascal Rivière, qui était Responsable du projet SICORE, est nommé Chef de la Division "harmonisation d'enquêtes auprès des entreprises" de l'INSEE-DG, à compter du 1.11.1995

élémentaires de codage, qu'il faut intégrer dans une application, et auxquels il faut adjoindre les fameuses bases de connaissances. Ainsi le module de codage automatique (proprement dit) reçoit-il, en entrée, un libellé plus d'éventuelles variables annexes ; en sortie, il renvoie un résultat de codage.

Mais ce n'est pas tout : outre l'intégration des modules élémentaires, le codage automatique engendre plusieurs travaux, en amont et en aval, dont on ne peut sous-estimer l'ampleur. Ainsi, l'on doit prévoir la saisie des libellés, dont le coût n'est pas négligeable. Il faut également écrire des logiciels conviviaux pour le traitement des "rejets" du codage automatique, c'est-à-dire des outils de codage assisté. Enfin, la mise en place du codage nécessite un minimum de préparation de la part de spécialistes de la variable ; dans le cas de "nouvelles" variables, cela conduit à la création de bases de connaissances ex nihilo, ce qui peut s'avérer long (cela dépend en fait de

l'existence ou non d'une première nomenclature reconnue).

SICORE a été appliqué à l'enquête sur les transports de la Direction régionale Rhône-Alpes de l'INSEE, appelée SYTRAL, où il fallait coder des communes, en l'occurrence des points de départ et d'arrivée pour chaque utilisation d'un transport en commun. Sur quelques milliers de libellés, seuls trois n'ont pas été codés par SICORE : le chiffrage assisté n'a donc pas été véritablement nécessaire.

Il est difficile d'en tirer des conséquences sur un plan statistique, car le chiffrage de la commune n'est certes pas le plus difficile. En revanche, informatiquement, on a pu voir que dans ce cas (où il n'y avait pas véritablement d'intégration dans une chaîne), le temps de travail nécessaire était très minime (un à deux jours).

SICORE a également servi à chiffrer la **catégorie socioprofessionnelle (CS)** et les **lieux de séjour** dans l'enquête Permanente conditions de vie (PCV) de l'INSEE. La situation était idéale pour une évaluation de SICORE, puisqu'un codage manuel avait été fait au préalable.

2. C'est-à-dire lieu de vacances : ville, département, pays étranger, circuit, ...

3. Pour que tous ses chiffrements soient exacts, le seul moyen serait de ne coder que des libellés strictement identiques à ceux du fichier d'apprentissage.