

## **Application en vue : la révision 4 de la CITI et la version 2 de la CPC sont examinées dans la région de la CESAO**

La Division de statistique de l'ONU et la Commission économique et sociale pour l'Asie occidentale (CESAO) ont tenu au Caire, du 10 au 13 décembre 2007, un séminaire sur les classifications économiques internationales, organisé en coopération avec l'Agence centrale égyptienne pour la mobilisation de l'opinion publique et la statistique et l'Organisation des ingénieurs-conseils des industries du Golfe.

Les participants – au nombre de 57 – venaient de pays de la région de la CESAO et du Maghreb, dont ils représentaient pour l'essentiel les bureaux de statistique nationaux et les ministères de l'industrie.

L'objet était avant tout de renforcer encore la compréhension actuelle des classifications des activités et des produits récemment modifiées (révision 4 de la CITI et version 2 de la CPC) de façon à aider les pays de la région à adopter ces classifications et à promouvoir la coopération régionale et l'harmonisation dans le domaine du classement. À ce titre, le séminaire s'inscrivait dans le prolongement d'une précédente réunion sur les classifications organisée à Beyrouth en 2004.

Les débats ont porté sur des aspects très variés de l'application des nouvelles classifications, tels que le cadre réglementaire et juridique nécessaire et la description détaillée des méthodes les plus probantes de numérotation des catégories de classement. Beaucoup de temps a été consacré à la tenue de groupes de travail au sein desquels les représentants nationaux et régionaux ont pu débattre d'un certain nombre de questions sur le classement et échanger directement des données d'expérience. L'une des sessions du séminaire a été l'occasion de faire le point de l'état d'avancement de la révision des Recommandations internationales concernant les statistiques industrielles, qui a été approuvé par la suite à la session de février-mars 2008 de la Commission de statistique de l'ONU.

Une grande partie du séminaire a été mise à profit pour présenter la structure définitive des versions révisées de la CITI et de la CPC aux participants et appeler leur attention sur les nouveaux concepts qui ont été exploités et les directives qui ont été suivies lors de l'élaboration de ces versions. Une meilleure connaissance de ce contexte aidera en effet les pays à aligner les versions nationales des classifications qu'ils établiront sur les principes appliqués pour la révision 4 de la CITI et la version 2 de la CPC.

Le traitement de la sous-traitance dans la CITI a notamment été abordé lors du débat sur les questions d'application : il devient de plus en plus important de traiter ce phénomène de manière cohérente dans les classifications, étant donné la part croissante dans l'économie de la sous-traitance, en particulier la délocalisation à l'étranger. Si ce traitement est assez évident pour les activités du secteur tertiaire, des difficultés surgissent notamment dans le cas de la sous-traitance des activités manufacturières. Dans la CITI, ces activités reçoivent le même traitement que les biens de transformation dans le Système de comptabilité nationale et le Manuel de la balance des paiements. Ainsi, le classement est fonction de la propriété des matières premières. Dans un accord de sous-traitance de ce genre, l'activité du mandant est classée à la rubrique Activités de fabrication à la condition expresse que le mandant soit propriétaire des matières premières. Sinon, l'activité considérée est en substance celle d'un grossiste ou d'un détaillant et entre dans la catégorie G de la révision 4 de la CITI (à moins que ne prévalent d'autres activités exercées dans la même entreprise). La CPC associe les produits voulus aux services de fabrication assurés et aux biens produits. L'accent a été mis avant tout sur le fait qu'en pratique, il faut veiller à n'enregistrer la production des biens obtenus qu'une seule fois pour éviter les doublons, ce qui est particulièrement difficile lorsque la production est sous-traitée à l'étranger.

En ce qui concerne la CPC, les nouveaux concepts liés à l'économie de l'information ont été passés en revue, outre les modifications générales touchant à la portée, à la structure et au niveau de détail de la classification. Aux fins de son application, il a été demandé d'établir des correspondances entre la CPC et d'autres classifications, à savoir la CITI, la Classification élargie des services de la balance des paiements et la liste GNS/W/120 de l'Organisation mondiale du commerce. On a signalé que ces correspondances existaient déjà et qu'il était prévu de les tenir à jour, même si certaines n'étaient intrinsèquement pas aussi précises qu'on pouvait le souhaiter, du fait des différences de portée et d'objet entre les classifications rapprochées.

Pour ce qui est de l'enchaînement des mesures d'application, les participants au séminaire ont admis la difficulté d'appliquer simultanément les classifications et la version révisée du Système de comptabilité nationale. À cet égard, on a soulevé la question de l'influence conjointe sur les données de la modification des classifications et des changements de concepts et de méthodes touchant le Système de comptabilité nationale. En effet, comment les modifications des séries chronologiques pourraient-elles être appliquées ou transposées à ces deux systèmes de façon à rendre dûment compte de l'évolution de la situation économique des

pays? Cette question sera portée à l'attention des acteurs du processus de révision du Système de comptabilité nationale et sera également prise en considération dans les directives pour l'analyse rétrospective des données qui seront publiées parmi les documents d'aide à l'application de la CITI.

D'autres questions d'ordre pratique ont été débattues, comme le calendrier d'application au niveau national recommandé par le Groupe d'experts des classifications économiques et sociales internationales. Le premier conseil donné aux représentants des pays a été de s'employer à transposer la CITI à l'échelon national d'ici à 2009.

Les échanges qui ont eu lieu lors du séminaire ont finalement abouti à des recommandations invitant les États de la région et l'ONU à prendre certaines mesures à l'avenir. Pour les États, il s'agira d'instaurer les lois et règlements propres à faciliter l'application des classifications, de créer des services exclusivement chargés des classifications pour faciliter leur application par les organes nationaux compétents et la concertation entre ces organes, de promouvoir un mécanisme de coopération régionale et d'établir des calendriers pour l'application des classifications et la formation du personnel des bureaux de statistique et des ministères de tutelle. Quant à l'ONU, elle devra s'employer à organiser des stages de formation spécialisée pour faciliter l'application des versions révisées des classifications économiques, à numériser les classifications internationales afin qu'elles soient consultables sur le site Web de la Division de statistique de l'ONU et à fournir les documents d'application nécessaires en arabe.

Tous les documents examinés lors du séminaire peuvent être consultés sur les pages Web du site de la Division de statistique consacrées aux classifications, à l'adresse suivante : <http://unstats.un.org/unsd/class/> (rubrique « Workshops and Training »).

---

### **Adoption de la version de 2008 de la Classification internationale type des professions (CITP-08)**

Le Bureau international du Travail (BIT) a poursuivi en 2007 ses travaux de mise à jour et d'amélioration de la version de la Classification internationale type des professions datant de 1988 (CITP-88). Un projet de structure, distribué en février 2007, a été examiné aux réunions de la Commission de statistique de l'ONU, du Groupe d'experts des classifications économiques et sociales internationales de l'ONU et du Groupe d'experts du BIT chargé d'actualiser la CITP. En s'appuyant sur les avis de ces organes et les commentaires formulés par les

pays et autres parties prenantes, le BIT a établi une version révisée du projet pour examen par les participants à une réunion d'experts tripartite sur les statistiques du travail, organisée par l'Organisation internationale du Travail (OIT) en décembre 2007. Après avoir subi des modifications mineures, le projet de structure y a été adopté dans le cadre d'une résolution, en tant que version de 2008 de la Classification internationale type des professions (CITP-08). Ladite résolution a été entérinée en mars 2008 par le Conseil d'administration de l'OIT, qui a ainsi confirmé le statut de norme statistique internationale de la nouvelle classification. La CITP-08 a également été présentée à la trente-neuvième session de la Commission de statistique, qui s'est tenue en février et mars 2008. Cette session a également été l'occasion pour la Commission de prendre note des projets d'application de la classification.

La structure de la CITP-08 peut être consultée sur le site Web de l'Organisation internationale du Travail dans les trois langues officielles de l'Organisation, à savoir l'anglais, l'espagnol et le français (voir <http://www.ilo.org/public/french/bureau/stat/isco/index.htm>).

Les définitions des catégories ainsi qu'un index actualisé des professions et un tableau des correspondances entre la CITP-88 et la CITP-08 seront publiés sur le site Web de l'OIT en 2008. Un projet de définitions en anglais y est d'ailleurs déjà disponible afin que les parties intéressées donnent leur avis. La nouvelle classification et sa documentation complète seront publiées dans un ouvrage en anglais, en espagnol et en français à paraître dès que possible à compter de leur affichage sur le site Web de l'OIT. Pour tout renseignement, veuillez envoyer un message électronique à l'adresse suivante : [ISCO@ilo.org](mailto:ISCO@ilo.org).

---

### **Codage par secteur des entreprises**

*John Crysedale et Michael Pedersen  
Division des normes (Statistique Canada)*

À Statistique Canada, le codage par secteur des entreprises est une activité permanente et pluridimensionnelle qui s'articule autour du Registre des entreprises et qui bénéficie du concours de la Division des normes. La présente note renseigne sur les rôles respectifs du Registre et de la Division dans le codage des entreprises, les outils utilisés par les spécialistes du codage, le processus de codage proprement dit et, enfin, les méthodes de contrôle qualité appliquées au codage.

#### **Registre des entreprises**

Le Registre des entreprises est à l'origine des cadres exploités par Statistique Canada pour ses études sur les entreprises. Un « cadre d'étude » consiste à peu de chose

près en une liste des entreprises qu'il est possible d'utiliser pour constituer l'échantillon d'une étude. Le Registre des entreprises est une base de données qui regroupe les caractéristiques élémentaires des 2,3 millions d'entreprises en activité au Canada. Ces caractéristiques sont les suivantes : effectifs; recettes; dépenses; actifs; situation géographique; statut juridique; code à six chiffres SCIAN (Système de classification des industries d'Amérique du Nord).

L'Agence du revenu du Canada est la principale source de données du Registre aux fins de l'établissement de la population d'entreprises du pays. Elle assure l'administration des lois fiscales au nom du Gouvernement canadien et de la plupart des provinces et territoires. Elle est notamment chargée d'obtenir une description de l'activité de toute nouvelle entreprise se faisant immatriculer. Pour que les entreprises n'aient pas à communiquer plusieurs fois les mêmes informations, l'Agence du revenu du Canada partage les renseignements de base dont elle dispose sur chacune d'entre elles avec Statistique Canada, qui reçoit de l'Agence quelque 25 000 dossiers par mois aux fins du codage SCIAN des entreprises.

#### **Division des normes**

La Division des normes met au point et tient à jour un certain nombre de classifications sectorielles, de tables de concordance et d'outils de codage, ainsi qu'une base de données comprenant des descriptions détaillées des activités des entreprises. Elle prend également des décisions intéressant le codage par secteur. Elle collabore avec d'autres bureaux nationaux et internationaux de statistique, avec d'autres divisions de Statistique Canada et avec les usagers extérieurs. Les classifications sectorielles dont s'occupe la Division sont les différentes versions canadiennes du SCIAN, les différentes versions de la Standard Industrial Classification (SIC), les versions du SCIAN en vigueur aux États-Unis et au Mexique, la Classification internationale type, par industrie, de toutes les branches d'activité économique (CITI) et la Nomenclature statistique des activités économiques dans la Communauté européenne (NACE).

#### **Outils de codage**

Plusieurs outils distincts interviennent, directement ou non, dans les activités de codage par secteur. Par exemple, un outil de codage par lots permet de coder un grand nombre d'entreprises de manière informatisée. Deux outils de la sorte sont utilisés.

Le premier porte le nom de « concordance directe ». Dans ce cas, la phrase descriptive communiquée par l'Agence du revenu du Canada sur l'entreprise à coder est comparée aux données d'une base comportant de nombreuses descriptions des activités des entreprises et les codes SCIAN correspondants.

La « concordance indirecte » est le second de ces outils. À la différence de la concordance directe, la phrase descriptive est analysée du point de vue syntaxique, ce qui permet d'éliminer tout ce qui est extrinsèque pour ne conserver que les éléments essentiels de la phrase. L'analyse syntaxique est effectuée à l'aide d'un logiciel de codage automatisé par reconnaissance de texte. Le résultat de l'analyse est ensuite comparé aux données d'une base contenant les descriptions des activités des entreprises et les codes SCIAN correspondants.

Deux outils interactifs sont par ailleurs utilisés dans le cadre du codage par secteur, quand les entreprises ne se prêtent pas au codage automatisé ou lorsque les activités de codage portent sur un petit nombre d'entreprises, voire sur une seule.

Il existe tout d'abord le Système de codage pour les classifications des industries (SCCI). Cet outil permet à son utilisateur de saisir une brève description de l'activité qu'il souhaite coder puis de faire son choix parmi les possibilités de la liste qui lui est proposée en retour. À peu de chose près, cet outil est une version électronique du manuel de classement par secteur, néanmoins dotée d'une base élargie de descriptions des activités des entreprises, permettant le codage selon de nombreuses classifications sectorielles et offrant aux spécialistes du codage la possibilité de soumettre leurs requêtes à la Division des normes. Il est généralement utilisé lorsque le spécialiste du codage essaie d'attribuer un code à une entreprise mais n'est pas en mesure d'obtenir réponse à ses questions directement auprès du propriétaire ou d'un employé de l'entreprise. L'attribution de codes exige que l'utilisateur de l'outil ait une bonne connaissance du classement par secteur.

L'autre outil interactif est celui de codage par arbre de décision. Cet outil repose sur une série organisée de questions et de réponses conduisant à une description normalisée de l'activité de l'entreprise, description à laquelle correspond un code SCIAN à six chiffres unique. L'avantage que présente le codage par arbre de décision est qu'il suffit de connaître les activités de l'entreprise à coder, sans nécessairement être familiarisé avec la classification sectorielle, pour pouvoir aisément attribuer un code sectoriel précis à cette entreprise.

En règle générale, on a recours à cet outil lorsqu'on peut s'adresser en personne ou par téléphone à un représentant de l'entreprise. Un échange de questions et de réponses peut alors avoir lieu. En voici un exemple :

**Q** : Quelle est l'activité principale de l'entreprise?  
**R** : Il s'agit d'un restaurant.  
**Q** : Y assure-t-on le service à table? **R** : Oui.  
**Q** : Y sert-on de l'alcool? **R** : Oui.

Il existe deux autres sources propres à faciliter le codage par secteur. La première est le manuel de

#### 4 *Classifications Bulletin d'information*

classement lui-même, disponible en version imprimée ou sur support électronique. La version électronique est particulièrement intéressante, car associée à des outils de recherche performants. Les importantes connaissances spécialisées amassées au fil de nombreuses années de travail par les experts en classement de la Division des normes constituent la seconde ressource. Ces experts apportent leur concours à tous ceux qui, tant au sein qu'à l'extérieur de Statistique Canada, le demandent.

##### **Processus de codage par secteur**

On peut considérer que le processus de codage par secteur commence à l'Agence du revenu du Canada, où les efforts se concentrent sur l'enregistrement des numéros d'entreprise. En effet, un numéro est obligatoire pour toute entreprise qui souhaite exercer son activité au Canada. À l'inscription, les entreprises doivent donner une description de leur activité principale aux fins de l'attribution d'un code SCIAN.

Les entreprises dont un représentant demande l'inscription en se rendant dans un bureau de l'Agence du revenu du Canada, ou bien par téléphone, sont codées au moyen de l'arbre de décision. Ceux qui font la demande d'un numéro d'entreprise par d'autres moyens, comme Internet, ne sont pas soumis à l'arbre de décision et fournissent une description librement formulée de l'activité de leur entreprise. Si une suite est donnée par l'Agence, cette description est revue selon la méthode de l'arbre de décision. Les entreprises dont les comptes sont vérifiés font également l'objet d'un codage par arbre de décision. L'Agence du revenu du Canada communique ensuite les informations clefs, à savoir le numéro d'entreprise et la description de l'activité, à Statistique Canada.

À ce stade du processus, les fichiers concernant les entreprises sont d'abord analysés par concordance directe pour isoler toutes celles dont la description de l'activité coïncide avec l'une des phrases de la base, auquel cas un code SCIAN leur est attribué, le Registre est mis à jour et les dites entreprises sont écartées du processus de codage.

En revanche, les entreprises restantes dont la description n'a pas de concordance exacte sont analysées par concordance indirecte, ce qui suppose de procéder au découpage syntaxique du texte et à la notation du résultat en fonction du degré de concordance. En cas de concordance exacte ou d'obtention d'une note supérieure à un certain seuil, les entreprises se voient attribuer un code SCIAN, le Registre est mis à jour et les entreprises concernées sont exclues du processus de codage.

Pour les autres entreprises, on a recours au codage manuel, d'où l'exploitation du SCCI. Le fichier de référence associé au Système regroupe l'ensemble des descriptifs d'activité d'entreprise qui figurent dans le manuel de classement, ainsi qu'un grand nombre de

variantes. Les spécialistes du codage peuvent aussi recueillir des informations complémentaires sur les entreprises en consultant les annuaires spécialisés, en effectuant des recherches sur Internet et en joignant un représentant des intéressés par téléphone. Au final, on parvient toujours à attribuer un code SCIAN, sauf dans un petit nombre de cas. En pourcentage, les codages par concordance directe et par concordance indirecte représentent 50 % et 10 % des activités de codage, respectivement, les 40 % restants étant le fait d'un traitement manuel.

##### **Méthodes de contrôle qualité**

Statistique Canada mène un certain nombre d'activités en vue de valider ou mettre à jour les informations qui figurent dans le Registre des entreprises. Les spécialistes de Statistique Canada doivent notamment établir régulièrement le profil des grandes entreprises et de leurs unités; il s'agit aussi de recueillir en permanence des commentaires dans les domaines d'étude, ce qui peut se faire avant la diffusion d'une enquête (par exemple, en prévenant par téléphone les destinataires) ou après, à partir des réponses obtenues, y compris lorsque leurs auteurs font état d'un changement d'activité dans les questionnaires.

En outre, la Division des normes et le personnel chargé du Registre des entreprises contrôlent de concert la qualité du codage. Il existe deux méthodes de contrôle qualité : d'une part, les experts peuvent examiner un échantillon des entreprises dont le code SCIAN a été modifié au cours d'une période donnée; d'autre part, ils peuvent passer en revue un échantillon de toutes les entreprises, que des modifications aient été enregistrées ou non. Pour les besoins de cette évaluation, le personnel de Statistique Canada recontacte les sujets interrogés. L'arbre de décision joue un rôle important à cet égard : d'une part, les données peuvent être vérifiées d'une manière potentiellement différente de celle dont le code SCIAN a été attribué à l'origine; d'autre part, avantage peut-être plus remarquable encore, la vérification peut être effectuée par des non-spécialistes du classement SCIAN.

##### **Conclusion**

Le codage par secteur au sein de Statistique Canada est un processus pluridimensionnel en pleine évolution. Chaque mois, un grand nombre d'entreprises font l'objet d'un codage visant à faciliter la réalisation d'études par Statistique Canada. Beaucoup d'outils différents sont aujourd'hui utilisés, tandis que des nouveaux sont en train d'être mis au point, pour surmonter les nombreuses difficultés liées au besoin de recueillir un volume important de données de qualité.

## Utilisation d'une application de codage automatisé par secteur pour l'immatriculation des nouvelles entreprises en Australie

Michael Meagher

Bureau australien de statistique

### Introduction

Le Bureau australien de statistique exploite le Registre australien des entreprises comme principale source d'information pour recenser les nouvelles entreprises. Les renseignements concernant celles-ci sont en effet transmis au Bureau pour inscription dans son Registre des entreprises, qui sert de cadre pour les études que le Bureau mène sur les entreprises.

Lors de leur immatriculation auprès du Bureau australien de la fiscalité, les entreprises sont tenues de sélectionner dans une liste le secteur d'activité auquel elles appartiennent au premier chef et de définir par écrit, dans les termes de leur choix, leur activité principale. La liste des secteurs, qui sont l'équivalent des catégories de la CITI, s'inspire des distinctions sectorielles établies dans l'Australian and New Zealand Standard Industrial Classification (ANZSIC). Les informations recueillies sont ensuite utilisées par le Bureau de la fiscalité pour classer les entreprises dont elles émanent dans les catégories pertinentes respectives de l'ANZSIC.

L'application de codage automatisé mise au point par le Bureau de statistique en concertation avec le Bureau de la fiscalité doit contribuer à améliorer la qualité du codage ANZSIC et à réduire les activités manuelles ainsi que les délais de traitement de l'information reçue des clients. Cette application, baptisée Autocoder, a été exploitée pour le codage ANZSIC automatisé des entreprises employant des salariés, en 2004, et de celles n'en comptant aucun, en 2006.

Autocoder utilise un algorithme de concordance de type « expression exacte » qui s'appuie sur un fichier index comportant des codes régulièrement mis à jour. Ce principe garantit un meilleur contrôle des correspondances entre codes et descriptions. Il convient toutefois de noter que le Bureau australien de statistique continue actuellement d'étudier le moyen de relever les taux de concordance grâce aux systèmes de recherche d'informations, étant donné que des économies importantes pourraient ainsi être réalisées à l'échelle de l'organisme.

### Traitement des informations

Autocoder fait appel aux informations qui figurent dans le formulaire d'obtention d'un numéro australien d'entreprise quant à l'activité et au secteur d'activité

principaux. Ces informations sont réduites à une brève description accompagnée d'un code d'abréviation à trois caractères. Par exemple, les informations « cultivateur d'olives vertes » (description de l'activité principale) et « Agriculture » (secteur d'activité principal) sont condensées sous la forme suivante : « cultivateur d'olives vertes; agr ».

Autocoder compare ensuite la description complète ou partielle de l'activité aux descriptions contenues dans le fichier index et lui attribue un code ANZSIC à quatre chiffres. Lorsqu'aucune concordance ne peut être établie, un code ANZSIC fictif est associé à la description, l'entreprise correspondante étant codée manuellement d'après l'ANZSIC par le personnel du Bureau de statistique.

### Règles de syntaxe

La première étape du processus de codage automatisé consiste à appliquer certaines règles de syntaxe à la description de l'activité. Ces règles permettent de convertir la description en une séquence plus facile à exploiter par Autocoder, et donc d'accroître la probabilité qu'un code ANZSIC corresponde à la description de l'activité.

L'application des règles de syntaxe évoquées à une description d'activité se traduit par les modifications suivantes :

- Correction des fautes d'orthographe dans les termes courants (« chasse » et non « chase », par exemple);
- Suppression des termes dénués de pertinence (« le », « comme », « SARL », par exemple);
- Remplacement des termes spécifiques par le concept général dont ils relèvent (« nourrisseur »/« élevage » ou « cultivateur »/« production », par exemple);
- Conversion des symboles en mots (« + » devient « et », par exemple);
- Simplification des formules développées (« réparation automobiles » pour « réparation de véhicules automobiles », par exemple).

Ainsi, la description d'activité « cultivateur d'olives vertes; agr » deviendrait « production olives vertes; agr ». Grâce à ce procédé, il est inutile d'insérer plusieurs entrées pour une activité donnée dans le fichier index des descriptions.

### Fichier index des descriptions d'activité

Le fichier index regroupe des descriptions d'activité, c'est-à-dire des entrées se présentant sous la forme d'une chaîne de termes, qui reprennent les descriptions d'activité principale de l'ANZSIC. Toutefois, pour améliorer le taux de concordance, les entrées sont ajoutées compte tenu des connaissances acquises sur la façon dont les entreprises décrivent leur activité. Le fichier index compte aussi des entrées pour les descriptions d'activités courantes qui ne peuvent être

## 6 Classifications Bulletin d'information

codées de manière automatisée, un code ANZSIC fictif leur étant alors attribué. Cela contribue à réduire au minimum les erreurs de codage imputables à Autocoder.

Par ailleurs, les descriptions d'activité peuvent comporter plusieurs termes, aussi les entrées sont-elles décomposées en niveaux de codage de sorte que le bon code ANZSIC soit associé à chaque description. Les trois niveaux de codage des entrées de l'index sont les suivants : « termes clefs », « qualificatifs » et « discriminants ». Le terme clef est celui sur lequel le spécialiste du codage fait porter tout d'abord la recherche, en remplissant le champ Description de l'activité principale. Le qualificatif est le terme qui fait l'objet de la deuxième opération de recherche, associant les champs Description de l'activité principale et Secteur d'activité principal. Enfin, le discriminant est le dernier terme auquel s'intéresse le spécialiste du codage, qui effectue dans ce cas une recherche à l'aide du seul champ Description de l'activité principale. La décomposition d'une entrée pourra, par exemple, aboutir au résultat suivant : terme clef = tomate; qualificatif = agriculture; discriminant = serre.

Les entrées susmentionnées sont formatées et agencées selon un modèle unique pour s'assurer que le programme Autocoder puisse systématiquement comparer les descriptions d'activité aux descriptions du fichier index sur la base de règles de logique.

### Réduction des concordances multiples

Autocoder peut donner des concordances multiples lorsque, dans une description d'activité, un terme clef fait ressortir du fichier index plusieurs entrées associées à des codes ANZSIC distincts. Par exemple, le terme clef « papier imprimerie » renvoie à l'entrée « 1510 Papier d'imprimerie, fabrication » de l'index. Cela étant, Autocoder peut aussi établir une concordance entre le terme clef « imprimerie » et l'entrée « 1611 Imprimerie, fabrication » de l'index.

On a recours à des métadonnées pour éviter les concordances multiples courantes, ce qui revient à isoler certains termes clefs (ou certaines combinaisons de tels termes) et à leur donner un statut prioritaire de façon à permettre le codage d'associations courantes de descriptions d'activité. Ainsi, Autocoder sait que le terme « papier imprimerie » a une importance et une pertinence plus grandes que ses éléments considérés séparément. Il ne fera donc désormais correspondre la description d'activité « fabrication papier imprimerie » qu'avec l'entrée « 1510 Papier d'imprimerie, fabrication » de l'index et ne tiendra pas compte du terme « imprimerie » lors de l'établissement de la concordance.

### Qualité d'Autocoder

Il importe pour le Bureau australien de statistique de préserver la qualité et le degré de précision des entrées de l'index et des fichiers de codage connexes, d'où la nécessité de contrôles réguliers. En 2007, Autocoder a codé environ 50 % des entreprises nouvellement immatriculées, avec un taux de qualité avoisinant 97 %.

Le Bureau australien de statistique s'efforce de maintenir le haut niveau de qualité des opérations de codage ANZSIC effectuées par Autocoder en mettant à jour les données mensuellement. En outre, il faut régulièrement mettre à jour les fichiers de l'application pour tenir compte des méthodes en vigueur de codage ANZSIC, accroître les taux de concordance, assurer l'efficacité du codage et veiller à ce que sa qualité ne se dégrade pas.

---

## Note de la rédaction

Le *Bulletin Classifications* récapitule les faits nouveaux dans le domaine des classifications internationales, annonce les prochaines manifestations et signale la documentation disponible sur papier ou sur Internet. Il peut être consulté sur les pages Web du site de la Division de statistique de l'ONU consacrées aux classifications, à l'adresse suivante : <http://unstats.un.org/unsd/class> (rubrique « Newsletter »).

Pour recevoir le *Bulletin Classifications* par courrier électronique, veuillez vous abonner à la liste de diffusion à l'adresse suivante : <http://unstats.un.org/unsd/cr/registry/regmaillist.asp>, ou bien en cliquant sur « Mailing List » dans le menu de gauche, sur la page d'accueil de la section consacrée aux classifications. Pour toute information complémentaire, veuillez vous adresser au service de téléassistance chargé des classifications :

**United Nations Statistics Division**  
**Attn: Economic Statistics and Classifications Section**  
**United Nations**  
**New York, NY 10017 (États-Unis)**

**Adresse électronique : [chl@un.org](mailto:chl@un.org)**  
**Télécopie : +1 212 963 1374**

**Site Web (pages consacrées aux classifications) :**  
<http://unstats.un.org/unsd/class>