

No 11-522-XIF au catalogue

**La série des symposiums internationaux
de Statistique Canada - Recueil**

**Symposium 2005 : Défis
méthodologiques reliés aux
besoins futurs d'information**



2005



Statistique
Canada

Statistics
Canada

Canada

LA REFONTE DU REGISTRE DES ENTREPRISES DE STATISTIQUE CANADA

Paul Hunsberger, Yanick Beaucage et Stuart Pursey¹

RÉSUMÉ

La structure du Registre des entreprises (RE) de Statistique Canada a été élaborée au milieu des années 80. Le Registre a pour rôle principal de fournir une base de sondage qui couvre plus de 2,3 millions d'entreprises actives au Canada, pour plus de 90 enquêtes récurrentes auprès des entreprises. Au cours des dernières décennies, des changements énormes se sont produits à l'intérieur et à l'extérieur de Statistique Canada. La mondialisation et l'évolution des politiques fiscales dans l'économie mondiale ont imposé à Statistique Canada de réévaluer le cadre conceptuel du RE. L'évolution de la technologie informatique a fourni l'occasion de revoir l'infrastructure informatique du Registre. Le projet de refonte du RE vise à simplifier, optimiser et harmoniser les processus et les méthodes. La présente communication fournit un aperçu de la refonte du RE et met l'accent sur les enjeux qui affectent la méthodologie des enquêtes auprès des entreprises. Parmi les principaux enjeux méthodologiques figurent l'utilisation de la structure d'exploitation des entreprises pour l'élaboration d'un plan de sondage pour les enquêtes auprès des entreprises, des procédures en vue de mettre à jour et de maintenir les données sur les entreprises, des méthodes pour déterminer la création et la disparition d'entreprises, et l'élaboration d'une stratégie d'assurance de la qualité.

MOTS-CLÉS : Registre des entreprises, refonte, base de sondage

1. INTRODUCTION

Le Registre des entreprises (RE) de Statistique Canada a été mis au point au milieu des années 80. Un examen détaillé de son infrastructure et de ses approches se trouve dans Colledge (1987) et Cuthill (1990). Le RE vise principalement à fournir une base de sondage de plus de 2,3 millions d'entreprises actives au Canada, pour plus de 90 enquêtes récurrentes auprès des entreprises.

La présente communication vise à décrire la refonte du Registre des entreprises, un projet conçu pour réviser et mettre à jour complètement l'infrastructure et les approches du RE actuel. D'autres détails se trouvent dans Bérard et coll. (2005), Gagné (2004), Pursey et coll. (2005) et Rancourt et coll. (2005). La section 2 du présent document fournit un aperçu du RE actuel et décrit les raisons de la refonte. La section 3 examine la façon dont les éléments d'échantillonnage peuvent être tirés du RE. La section 4 décrit l'utilisation des données administratives pour la mise à jour du RE. La section 5 porte sur l'élaboration d'une stratégie d'assurance de la qualité.

2. APERÇU DU REGISTRE DES ENTREPRISES DU CANADA

Les bases et les concepts du Registre des entreprises, c'est-à-dire l'ensemble du programme de statistiques économiques de Statistique Canada, sont axés sur les besoins du Système de comptabilité nationale du Canada. Le Registre des entreprises a commencé à prendre forme au début des années 80, au moment où des efforts ont été déployés pour établir une base de sondage centrale pouvant être utilisée pour la plupart, sinon la totalité, des enquêtes auprès des entreprises. Graduellement, au fil des ans, de plus en plus d'enquêtes ont utilisé le RE — il s'agit maintenant de la pierre angulaire des programmes de statistiques sur les entreprises à Statistique Canada.

Depuis les années 80, des changements se sont produits à l'intérieur et à l'extérieur de Statistique Canada. Les politiques fiscales, la mondialisation et la structure des entreprises ont évolué. La base conceptuelle qui sous-tend le

¹ Yanick Beaucage, Paul Hunsberger et Stuart Pursey, Statistique Canada, Division des méthodes d'enquêtes auprès des entreprises, Immeuble R.-H. Coats, 11^e étage, Pré Tunney, Ottawa (Ontario), Canada K1A 0T6.

RE est devenue périmée. La révolution de la technologie informatique a fait en sorte que l'infrastructure informatique du RE est difficile et coûteuse à maintenir. De nouvelles sources de données administratives utiles ont vu le jour. Les fonctions du RE se sont élargies au fil du temps. Les tâches et les processus servant à maintenir et à mettre à jour le RE sont de plus en plus nombreuses. Ce qui importe encore davantage, c'est qu'avec tous ces changements, les besoins des utilisateurs ont évolué. Le RE nécessite d'être repensé, remanié.

2.1 Le Registre des entreprises

Le Registre des entreprises est une liste des entreprises actives au Canada. Il sert de base de sondage centrale pour la presque totalité des enquêtes auprès des entreprises de Statistique Canada, et plus particulièrement les enquêtes qui fournissent des données au Système de comptabilité nationale. Les variables importantes du RE pour la conception de l'enquête et de l'échantillon sont les champs qui comprennent des données sur la géographie, l'industrie, l'effectif, le revenu et les personnes-ressources.

On compte environ 2,29 millions d'entreprises actives dans le RE, et 2,27 millions d'entre elles sont des entreprises « simples ». Pour celles-ci, une seule entité représente bien la structure juridique, la structure d'exploitation, l'unité de production, l'unité d'observation, l'unité d'échantillonnage et l'unité de collecte des données de ces entreprises. Même si ces entreprises correspondent à 99,1 % des entreprises au Canada, elles ne sont à l'origine que de 40 % de l'activité économique. Les autres entreprises sont « complexes ». On en compte environ 21 000 seulement, mais elles sont à l'origine de 60 % de l'activité économique. Il existe deux représentations utiles d'une entreprise : la structure d'exploitation et la structure statistique (voir la section 3). La structure d'exploitation correspond à la perception qu'a l'entreprise d'elle-même, c'est-à-dire la façon dont elle fonctionne et dont elle est structurée. Cette structure est établie par suite d'une interview exhaustive avec des représentants de l'entreprise. Ce « profil » de l'entreprise est important pour Statistique Canada, parce qu'il permet de déterminer les types d'unités de production des entreprises complexes, la disponibilité des types de données et de nombreuses mesures requises pour la conception de l'enquête (taille, géographie et industrie).

Les structures légales et d'exploitation d'une entreprise complexe sont représentées par un ensemble d'entités (unités de production) organisées selon un modèle hiérarchique à plusieurs niveaux, qui illustre les rapports hiérarchiques et la disponibilité des données des unités de production. On compte cinq types d'unités de production dans la structure d'exploitation. L'entité commerciale (EC) représente l'ensemble de l'entreprise et des données. Le centre d'investissement (CI) comprend des données sur le revenu, les dépenses et les investissements. Le centre de profit (CP) comprend des données sur le revenu et les dépenses et permet le calcul de la « valeur ajoutée ». Le centre de profit correspond à ce que l'on appelle habituellement un établissement. Le centre de coût (CC) comprend des données sur les dépenses seulement. Le centre de ventes/revenu (CR) comprend des données sur le revenu seulement. Le diagramme figurant ci-après illustre la structure d'exploitation d'une entreprise complexe donnée. Trois niveaux sont séparés en deux branches qui contiennent plusieurs types d'unités de production.

Figure 1a : Entreprise complexe

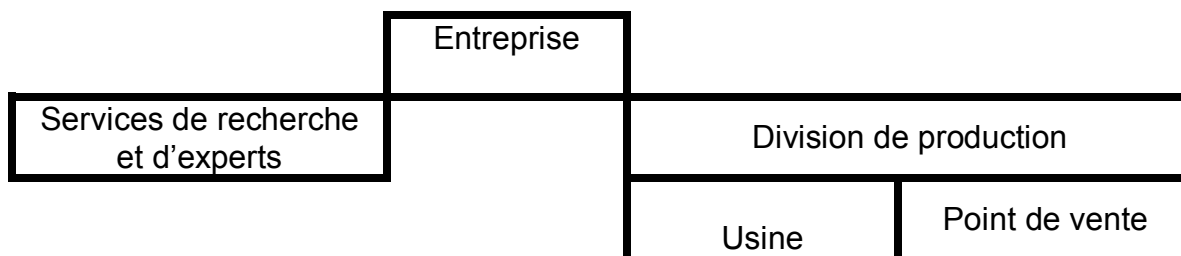
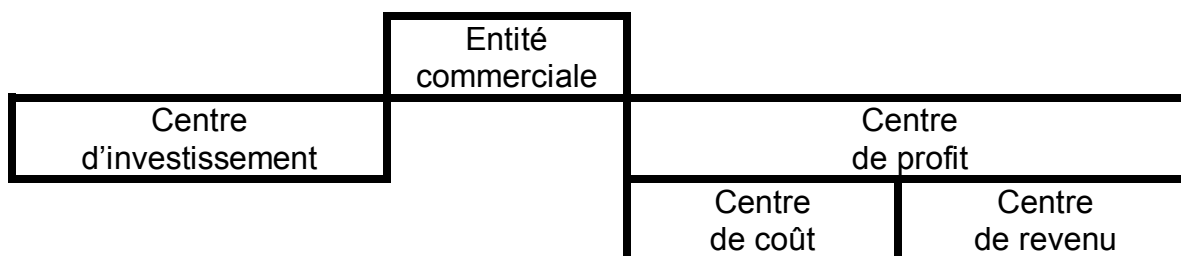


Figure 1b : Structure d'exploitation d'une entreprise complexe



3. ÉLÉMENTS D'ÉCHANTILLONNAGE

3.1 Approche actuelle

La structure d'exploitation d'une entreprise sert de base pour établir la deuxième représentation de l'entreprise : la structure statistique. Celle-ci est hiérarchique et comporte quatre niveaux : entreprise, compagnie, établissement et emplacement. Chaque niveau correspond à un type de disponibilité de données : le tableau 1 montre les rapports entre la disponibilité des données et la structure d'exploitation. L'objectif de la structure statistique est de fournir un cadre qui facilite la détermination et la sélection des éléments d'échantillonnage. Par exemple, la base de sondage d'une enquête auprès des entreprises fondée sur l'établissement est établie à partir du troisième niveau de la structure statistique (les établissements).

Tableau 1 : Structure d'exploitation et structure statistique – Types d'unités

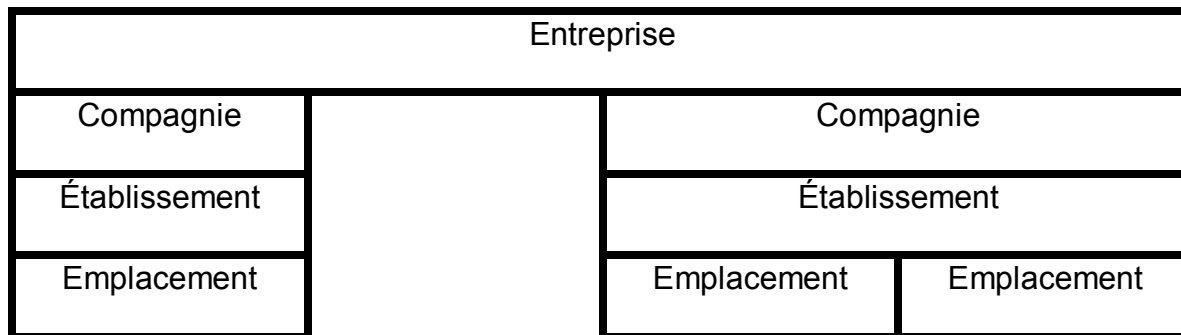
Structure d'exploitation	Structure statistique	Disponibilité des données
Entité commerciale (EC)	Niveau 1 : Entreprise	Toutes les données
Centre d'investissement (CI)	Niveau 2 : Compagnie	Revenu, dépenses et investissement
Centre de profit (CP)	Niveau 3 : Établissement	Revenu et dépenses
Centre de coût (CC)	Niveau 4 : Emplacement	Dépenses
Centre de ventes/revenu (CR)		Revenu

STATGEN est l'algorithme utilisé pour produire la structure statistique à partir de la structure d'exploitation. Pour les 2,27 millions d'entreprises « simples », STATGEN met en correspondance l'unité de production unique de l'entreprise avec chacun des quatre niveaux. Pour les 21 000 entreprises complexes (deux niveaux ou plus dans la structure d'exploitation), l'algorithme élargit ou simplifie la structure d'exploitation, en vue d'obtenir quatre niveaux. De façon générale, même s'il existe des complexités lorsque la structure d'exploitation ne comporte pas

quatre niveaux (ou plus exactement lorsque chaque branche de la structure d'exploitation ne comporte pas quatre niveaux), les centres d'investissement deviennent des compagnies, les centres de profit deviennent des établissements, et les centres de revenu et centres de coût deviennent des emplacements.

La figure 2 montre l'élaboration de la structure statistique pour l'exemple figurant dans les figures 1a et 1b. Dans cet exemple, la branche de gauche doit passer de deux à quatre niveaux, et celle de droite, de trois à quatre niveaux. Dans les entreprises plus complexes, qui comportent de nombreux niveaux dans la structure d'exploitation, ceux-ci doivent être ramenés à quatre. Toutefois, parmi les 21 000 entreprises complexes, seulement 2 000 comportent plus de deux niveaux dans leur structure d'exploitation.

Figure 2 : Structure statistique d'une entreprise complexe (d'après les figures 1a et 1b)



3.2 Difficultés liées à la structure statistique

Le principal avantage de la structure d'exploitation est qu'elle fournit un aperçu réaliste, précis et à peu près intact de la structure d'une entreprise, c'est-à-dire la structure de l'entreprise selon la perception de l'entreprise même. Si nous fondons nos plans d'échantillonnage et nos modalités de collecte des données directement sur la structure d'exploitation, plutôt que sur la structure statistique hautement structurée, nous nous attendons à réaliser des gains au chapitre de la qualité des données recueillies et à réduire le fardeau de réponse.

Le principal avantage de la structure statistique est qu'elle fournit une façon simple de sélectionner des unités pour une enquête fondée sur l'entreprise, la compagnie, l'établissement ou l'emplacement. STATGEN organise les diverses unités de production de la structure d'exploitation et les intègre à l'endroit approprié dans la structure statistique, selon les lignes directrices sur la disponibilité des données figurant dans le tableau 1. La structure statistique comporte trois principaux inconvénients, qui se sont manifestés au fil des ans, au fur et à mesure que nous avons acquis de l'expérience dans l'utilisation de la structure statistique, et que nos priorités en ce qui a trait à la réduction du fardeau de réponse et à l'amélioration de la qualité des données ont augmenté.

Le premier inconvénient est que la structure statistique repose sur notre infrastructure informatique. Cela représente des coûts importants, particulièrement lorsque l'on tient compte du fait que STATGEN est utilisé uniquement pour 21 000 entreprises complexes, dont 19 000 ne comportent que deux niveaux.

Le deuxième inconvénient vient de ce que les estimations doivent être fournies selon la province pour le Système de comptabilité nationale. Toutefois, il arrive souvent qu'une entreprise ne fonctionne pas sur la base de la province et ne tienne pas ses livres sur cette base. Prenons la figure 1b. Même si cela n'est pas indiqué implicitement dans le diagramme, le centre de revenu et le centre de coût sont dans la même province, ce qui nous amène à la figure 2, dans laquelle un établissement unique se situe au-dessus des deux emplacements. Supposons que le centre de revenu et le centre de coût soient situés dans des provinces différentes. C'est donc dire que le centre de profit a des activités dans deux provinces différentes. La solution, dans la perspective d'une enquête fondée sur l'établissement, consiste à créer deux pseudo-établissements, dont aucun n'est réel, ce que montre la figure 3.

Figure 3 : Structure statistique d'une entreprise complexe (à partir de la figure 1) : le cas des pseudo-établissements

Entreprise			
Compagnie		Compagnie	
Établissement		Pseudo- établissement (pour l'Ontario)	Pseudo- établissement (pour le Québec)
Emplacement		Emplacement (en Ontario)	Emplacement (au Québec)

Les pseudo-établissements peuvent poser des problèmes au chapitre de la collecte des données, parce que certains répondants tentent de fournir des données sur le revenu et les dépenses pour une unité de production qui n'existe pas. Il reste que, d'une façon ou d'une autre, des données sont recueillies ou imputées, mais que la qualité des données peut être faible en raison de la non-réponse ou de la difficulté de savoir exactement à quoi les données recueillies se rapportent. C'est ce que l'on appelle un pseudo-établissement « fractionné entre des provinces », mais il existe deux autres types. Lorsqu'un emplacement dessert un établissement et que les deux se situent au même niveau de la structure d'exploitation, l'emplacement est intégré (combiné) à l'établissement et appelé pseudo-établissement « intégré ». Lorsqu'un emplacement dessert plus d'un établissement et qu'il se situe au même niveau dans la structure d'exploitation, l'emplacement devient un établissement à part entière (et est appelé pseudo-établissement auxiliaire).

Le troisième élément constitue en fait une généralisation du deuxième. La structure statistique modifie chaque entreprise, afin qu'elle s'intègre dans la structure à quatre niveaux, peu importe le nombre réel de niveaux. Dans le cas des entreprises simples (un niveau), cela ne pose pas réellement de problème (outre le coût du traitement et de l'entreposage des données), parce que le résultat est simplement une répétition de l'unité de production simple : une pour chacun des quatre niveaux. Pour la presque totalité des entreprises complexes, il existe une distorsion réelle dans la structure d'exploitation. Les 19 000 entreprises à deux niveaux sont transformées en entreprises à quatre niveaux. Celles qui comportent de nombreux niveaux sont ramenées à quatre. Il se peut que deux structures d'exploitation très différentes aient des structures statistiques identiques.

Il arrive souvent qu'une entreprise, lorsqu'elle examine sa structure statistique, ne s'y reconnaisse pas. La difficulté en ce qui a trait à la structure statistique vient de ce qu'elle rend les modalités de collecte des données plus complexes et confuses, ce qui peut mener à un fardeau de réponse inutile et à une faible qualité des données. Compte tenu de l'expérience actuelle dans le contexte économique d'aujourd'hui, nous nous rendons compte que la structure statistique est, dans les faits, une distorsion de la réalité d'une entreprise qui, au mieux, n'a pas de répercussions sur les modalités de collecte des données, le fardeau de réponse et la qualité des données et, au pire, entraîne des difficultés à ce chapitre.

3.3 Utilisation de la structure d'exploitation directement pour créer des éléments d'échantillonnage

La figure 1b montre la structure d'exploitation d'une entreprise, une fois son profil créé; cette structure est identique à celle de la figure 1a. Dans le cadre d'une enquête axée sur l'établissement, l'objectif est de faire enquête auprès des centres de profit (afin de mesurer la valeur ajoutée = revenu – dépenses) et de couvrir entièrement, sans chevauchement, l'entité commerciale. Ainsi, l'élément d'échantillonnage est le centre de profit. Deux difficultés se posent mais peuvent être facilement surmontées.

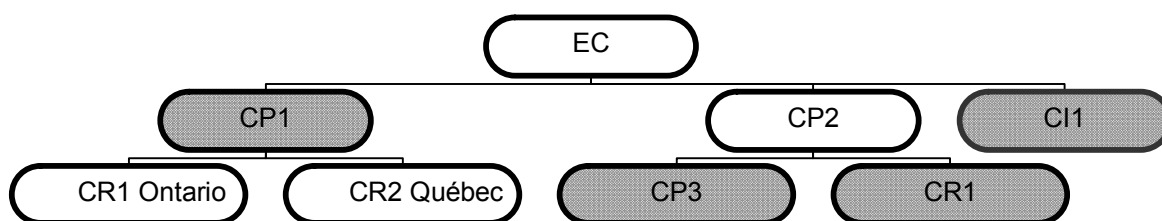
La première a trait aux pseudo-établissements. Peut-on atteindre l'objectif de représenter la dimension provinciale des données sans créer d'éléments d'échantillonnage artificiels? Comme le montre la section 3.2, il existe trois types de pseudo-établissements. La solution dans le cas des pseudo-établissements fractionnés entre des provinces consiste à étiqueter le centre de profit comme ayant des activités dans plusieurs provinces et à y joindre un vecteur de

données correspondant à l'importance des activités dans chaque province. Il s'agit là des données nécessaires pour créer un plan d'échantillonnage selon la province. (Une autre option consiste à coupler le centre de profit avec les CR et les CC qui en relèvent.) Tant pour les pseudo-établissements intégrés qu'auxiliaires, les CR et les CC deviennent des éléments d'échantillonnage à part entière. Cela est parfois dommage parce que les CR et les CC ne peuvent fournir de données sur la valeur ajoutée, mais cela correspond à la réalité de l'entreprise. Grâce à un aperçu réaliste, nous pouvons plus facilement élaborer des modalités de collecte des données avec les répondants, en tenant compte des unités de production qui ne sont pas des centres de profit.

La deuxième préoccupation a trait au fait que, dans une entreprise comportant de nombreux niveaux, au fur et à mesure que nous progressons dans chaque branche, nous pouvons retrouver plus d'un centre de profit. La règle consiste à sélectionner le « CP le plus bas » de la branche comme élément d'échantillonnage. C'est ce qui contribue le mieux à produire la somme la plus importante de données sur l'industrie et de données géographiques.

La figure 4 montre les éléments d'échantillonnage (ombrés) pour une enquête fondée sur l'établissement à partir de la règle du CP le plus bas. L'élément d'échantillonnage de la branche de gauche est CP1 (un vecteur de données est nécessaire pour illustrer les activités provinciales). Les éléments d'échantillonnage de la branche du milieu sont CP3 et CR1 (qui doit être inscrit comme un centre de revenu). L'élément d'échantillonnage de la branche de droite est CI1. L'entité commerciale est complètement couverte sans chevauchement. Des modalités de collecte des données peuvent être conclues avec les répondants et on peut calculer la valeur ajoutée. L'organisation de la collecte des données pour la branche du milieu posera des défis, mais cette difficulté n'est pas cachée par la création d'un établissement artificiel.

Figure 4 : Éléments d'échantillonnage (ombrés) d'une enquête fondée sur l'établissement auprès d'une entreprise.



L'échéancier relatif à la refonte de l'univers d'échantillonnage comprend la création et l'analyse de bases d'échantillonnage d'essai pour les principales enquêtes auprès des entreprises de Statistique Canada, en 2006 et 2007, avant que l'on mette la dernière main au nouveau processus de création d'éléments d'échantillonnage et qu'on l'intègre, en 2008.

4. UTILISATION DES DONNÉES ADMINISTRATIVES

4.1 Mesure de la taille de l'entreprise

À l'heure actuelle, on utilise un modèle pour estimer le revenu d'une entreprise simple. Dans le cas des entreprises avec salariés, le modèle est fondé sur un fichier mensuel de données administratives appelé PAYDAC, qui comprend les remises de l'entreprise à l'Agence du revenu du Canada (ARC). Ces chiffres servent à estimer le revenu annuel et le nombre annuel d'employés d'une entreprise. Dans le cas des entreprises sans salariés, le revenu est fondé sur un fichier de données administratives de l'ARC, le fichier de la taxe sur les produits et services (TPS). Celui-ci comprend des données mensuelles, trimestrielles ou semestrielles sur l'entreprise. Il arrive parfois que PAYDAC et le fichier de la TPS ne sont pas disponibles; si l'entreprise est constituée en société, son revenu annuel (T2) obtenu de l'ARC est utilisé dans le modèle.

Dans le cadre de la refonte du RE, nous voulons éviter l'utilisation de modèles et produire plutôt des mesures de la taille directement, sans modification, outre les habituels contrôles, imputations et détection des valeurs aberrantes.

Le fichier de la TPS servira à élaborer une mesure annuelle du revenu, tant pour les entreprises avec salariés que pour celles sans salariés. De même, à partir des données de l'ARC, nous produirons, à titre de variables distinctes, le revenu annuel T1 (pour les entreprises non constituées en société) et le revenu annuel T2 (pour les entreprises constituées en société). Un fichier relativement nouveau est disponible auprès de l'ARC : le fichier mensuel des retenues sur la paie (PD-7). Il permet d'obtenir directement le nombre d'employés, et il servira à calculer une mesure annuelle du nombre d'employés d'une entreprise.

Dans le cas des entreprises complexes, l'établissement de profils continuera de servir de source de données pour les mesures de taille dans le RE.

4.2 Création d'entreprises

Le RE enregistre en moyenne 18 000 créations d'entreprises par mois. À l'heure actuelle, les entreprises créées sont chargées dans le RE tous les trois mois, dans les cas des entreprises sans salariés, et tous les mois, dans les cas des entreprises avec salariés. Le processus utilisé cause parfois des « avalanches d'entreprises nouvellement créées ». Par exemple, une entreprise doit être codée selon son code complet à six chiffres du SCIAN (Système de classification des industries de l'Amérique du Nord) avant d'être créée, ce qui entraîne souvent un arriéré d'entreprises qui ont besoin d'être codées. Lorsque ces arriérés deviennent trop importants, on affecte un plus grand nombre de personnes à la tâche de codage, ce qui donne lieu à des périodes de surcharge de travail. Une ou deux fois par année, on assiste à un afflux important d'entreprises nouvellement créées à partir de sources administratives externes. La difficulté en ce qui a trait aux processus de création d'entreprises irréguliers est la possibilité de périodes de surcharge de travail subites pour les estimations découlant des enquêtes auprès des entreprises. Même s'il est difficile de gérer les diverses sources de données administratives, du point de vue de l'activité de la création des entreprises, nous devons explorer des façons de surmonter cette difficulté et, autrement, des façons de réduire les répercussions des créations tardives. Par exemple, une possibilité consisterait à accepter le codage des industries aux niveaux à trois chiffres ou à quatre chiffres du SCIAN, ce qui nécessite moins de temps que le codage au niveau à six chiffres. Cela contribuerait probablement à accélérer les mises à jour mensuelles des créations. Des négociations minutieuses avec les intervenants, avec comme objectif de modifier les processus à l'intérieur et à l'extérieur de Statistique Canada, pourraient faire en sorte que les « avalanches » se produisent à des intervalles moins fréquents.

4.3 Disparition d'entreprises

Les règles actuelles servant à déterminer la disparition d'une entreprise sont conservatrices. Il est plus facile de traiter les entreprises qui disparaissent au moment de l'estimation (à partir de l'estimation par domaine), que de traiter une entreprise considérée comme disparue, mais qui fonctionne toujours (ce qui mène au sous-dénombrement). Toutefois, il est coûteux de conserver les entreprises disparues dans le RE, les plans d'échantillonnage n'étant pas aussi efficaces qu'ils pourraient l'être, et les fonds consacrés à la collecte des données n'étant pas dépensés efficacement. Nous devons trouver un équilibre entre la protection contre le biais dû au sous-dénombrement et le maintien de l'efficacité des plans d'échantillonnage et de la collecte des données. Il existe plusieurs sources de données administratives précises pour déterminer, avec beaucoup de certitude, la disparition d'une entreprise. Ces fichiers seront intégrés dans la refonte du RE et un processus sera créé pour les « supprimer » deux fois par année. La rétroaction d'enquête sert aussi de source pour déterminer les unités disparues, mais elle se limite aux entreprises sélectionnées dans l'échantillon. Afin d'éviter le biais dans les estimations, la source indiquant que l'entreprise est disparue (c.-à-d., l'enquête ou le fichier de données administratives) devrait être conservée et utilisée pour mettre à jour les bases de sondage, uniquement lorsqu'elle est indépendante de l'enquête.

Il existe un autre type de disparition d'entreprises, c'est-à-dire une entreprise qui est en voie de terminer ses activités et qui sera considérée comme supprimée au moment de la collecte des données. Par ailleurs, il se peut qu'une entreprise soit inactive (même si les fichiers de données administratives indiquent qu'elle est toujours en activité) et qu'elle reprenne ses activités. Dans le cadre de la refonte du RE, nous explorerons l'élaboration de méthodes, en vue de gérer les entreprises affichant des signes de disparition peu concluants. Voici trois exemples d'approches possibles. Dans le cadre d'une enquête, on peut décider d'utiliser une règle de décision; sur la base de signes peu concluants, exclure une entreprise d'une base de sondage, mais la conserver dans le RE. Cette règle tiendrait compte à la fois du risque de surdénombrement et du risque de sous-dénombrement. Dans le cadre d'une enquête, on peut utiliser un processus relativement peu coûteux de contacts préalables avant la collecte des données, afin de

confirmer si l'entreprise est disparue ou si elle est toujours en activité. Parmi les autres utilisations possibles des signes peu concluants figurent la stratégie d'imputation et l'estimation de la qualité des données sur les entreprises actives/disparues dans le RE.

5. STRATÉGIE D'ASSURANCE DE LA QUALITÉ

Le refonte du Registre des entreprises fournit l'occasion de mettre en place une stratégie d'assurance de la qualité complètement nouvelle et à jour. Cette stratégie devrait être axée sur les six aspects de la qualité : exactitude, pertinence, actualité, accessibilité, intelligibilité et cohérence. À l'heure actuelle, les rapports de qualité concernant les activités du RE ne sont ni centralisés ni uniformisés. Il s'agit principalement de rapports ponctuels et spéciaux. Les mises à jour du RE sont effectuées chaque mois, à partir de sources administratives. La rétroaction découlant de ces mises à jour doit faire l'objet de rapports clairs, afin que nous puissions évaluer le processus au fil du temps. Un certain nombre d'éléments de données devraient faire partie de ce rapport, y compris le nombre de mises à jour des variables clés et l'ampleur de ces mises à jour, le nombre de nouvelles entreprises (création) et les nouveaux enregistrements considérés comme disparus. Afin d'évaluer la qualité du codage selon le SCIAN pour les nouvelles entreprises, un processus informel de contrôle de la qualité est actuellement en place. Dans le cas des nouveaux codeurs, un contrôle de la qualité à 100 % est effectué jusqu'à ce qu'ils atteignent un taux de 5 % de codage erroné. Par la suite, on procède uniquement à des vérifications ponctuelles.

Une autre mesure est en place pour évaluer la qualité du codage et de la couverture du RE, à savoir l'Enquête sur l'assurance de la qualité (EAQ), qui est effectuée de temps à autre à l'égard du RE. L'enquête a été mise en œuvre pour l'année de référence 2002 et devrait se tenir à nouveau en 2006. L'EAQ produit un bulletin évaluant la proportion d'entreprises codées correctement selon le SCIAN dans le RE. Lorsque les unités sont mal codées, cela peut être attribué à deux raisons possibles : l'unité a été codée de façon incorrecte, ou l'unité a changé récemment de code. Des proportions peuvent être calculées pour chaque type de codage erroné, ce qui mène à des mesures qui pourraient être utilisées pour évaluer la qualité et l'actualité du SCIAN. L'EAQ mesure en outre le taux d'unités disparues qui étaient considérées comme actives dans le RE. Cela se révèle utile lorsqu'on tente de prédire la taille d'échantillon optimale.

5.1 Plans pour le nouveau Registre des entreprises

Nous aimerions uniformiser les procédures et nous assurer que toutes les étapes font l'objet de rapports de ce qui a été changé ou mis à jour, du point de vue des comptes et de l'ampleur. Il devrait s'agir de rapports systématiques, qui seraient conservés dans une base de données pour l'analyse longitudinale. Toute mise à jour des mesures de taille (revenu, nombre d'employés, etc.) devrait être analysée, et les cas extrêmes devraient être examinés soigneusement, particulièrement lorsque les mises à jour proviennent d'une source administrative qui pourrait comprendre des valeurs non corrigées. Encore une fois, des rapports réguliers devraient être produits, en vue d'évaluer les changements qui touchent la mesure de la taille, laquelle pourrait être analysée au fil du temps.

Par ailleurs, nous souhaitons mettre en place des contrôles de la qualité plus officiels, dans le cas des activités cruciales du RE, comme le codage selon le SCIAN et l'établissement de profils. Le codage selon le SCIAN doit être précis, et l'une des façons d'assurer son exactitude continue est le contrôle de la qualité. Cela est valable à la fois pour le codage manuel (évaluation du travail des codeurs) et le codage automatisé (évaluation de la qualité du codage automatisé). Nous souhaitons mettre en place un programme que les gestionnaires des opérations du RE pourront modifier, selon l'expérience du codeur ou toute autre contrainte externe. Dans le cas du codage automatisé, il sera extrêmement important de vérifier les nouveaux codes automatiques mis en place, ainsi que les secteurs dans lesquels le codage pourrait susciter de la confusion. Par exemple, le codage automatique des « banques alimentaires » dans le secteur bancaire ne devrait pas se produire, et une vérification devrait être effectuée afin de prévenir ce type d'erreur.

Aux fins du codage manuel, quelques outils sont disponibles à Statistique Canada, mais cela est encore perçu comme un exercice subjectif. Il existe toutefois un nouvel outil élaboré de concert avec l'Agence du revenu du Canada. Il repose sur une description de l'activité et sert à susciter un ensemble de questions qui contribueront à sélectionner le code approprié du SCIAN. Cet outil est actuellement disponible pour près de 80 % des codes du SCIAN, et on prévoit qu'il sera précisé et élargi, afin d'englober davantage d'industries du SCIAN. Il devrait être

utilisé par tous les codeurs, afin d'uniformiser les procédures et produire des résultats cohérents.

À Statistique Canada, on mène un exercice appelé établissement de profils (décrit dans la section 2.1), dans le cas des grandes entreprises. L'exercice d'établissement du profil se fait à intervalles réguliers et sert à mettre à jour la structure d'exploitation de l'entreprise. Selon l'importance de l'entreprise, l'établissement du profil peut se faire plus régulièrement. Parmi les aspects importants de l'établissement des profils figure la détermination de la responsabilité de chaque partie de l'entreprise. Il existe autant de façons de le faire qu'il y a de responsables de l'établissement des profils. Dans le cadre du nouveau RE, nous voulons assurer la formation et trouver des façons d'uniformiser le processus. Nous voudrions peut-être aussi envisager l'inclusion d'un processus de contrôle de la qualité dans l'exercice d'établissement des profils, afin d'en évaluer la qualité. Cela sera probablement plus difficile à mettre en œuvre qu'un processus régulier de contrôle de la qualité, l'établissement des profils étant très long et complexe.

L'amélioration de la qualité du Registre des entreprises, grâce à l'uniformisation, à la formation, à l'ajout de mécanismes de contrôle de la qualité, à des mises à jour opportunes ou à l'établissement de profils, aura des répercussions sur toutes les enquêtes reposant sur le RE. Des rapports de qualité sur une base régulière et un suivi des changements apportés au Registre des entreprises au fil du temps contribueront à informer les gestionnaires et les utilisateurs du RE. Ces données seront utiles pour prendre des décisions concernant la mise à jour ou l'amélioration du RE; elles peuvent aussi être utiles pour les nouvelles enquêtes pour lesquelles on envisage l'utilisation du RE comme base de sondage. La refonte du RE fournit une excellente occasion d'augmenter la qualité des données, de l'officialiser et d'en rendre compte pour les années qui viennent.

6. DÉFIS POUR L'AVENIR

La modification de l'élément d'échantillonnage pour les enquêtes axées sur l'établissement pourrait avoir des répercussions à différents niveaux du processus d'enquête, à partir de l'échantillonnage jusqu'à l'estimation. Nous devons évaluer les changements qui se produiront selon notre nouvelle définition des éléments d'échantillonnage, particulièrement en ce qui a trait aux estimations selon l'industrie et la province. Ce changement majeur devrait aussi toucher les opérations et les programmes d'enquêtes. Les responsables de chaque enquête devront évaluer individuellement les répercussions de cette nouvelle approche sur leur système.

Au moment de la refonte du RE, l'un des objectifs consiste à effectuer des mises à jour plus fréquentes et à optimiser notre utilisation des données administratives. Mais, utilisons-nous au maximum ces sources de données non coûteuses? Maintenant que ces sources sont en place depuis un certain temps, nous devrions les utiliser pour d'autres activités, comme les changements de structures ou la vérification du SCIAN. Il s'agit de l'une des façons possibles d'améliorer la qualité du nouveau RE. Nous avons aussi fait état d'un certain nombre d'autres possibilités pour améliorer la qualité des données ou en rendre compte, mais sont-elles suffisantes? La seule certitude que nous avons c'est que, plus nous connaissons notre base de sondage, plus nous pouvons avoir confiance en nos estimations.

RÉFÉRENCES

- Bérard, H., Gagné, P., Rancourt, E. and Pursey, S. (2005), "La Refonte Du Registre des Entreprises de Statistique Canada", Recueil du Colloque francophone sur les sondages, Québec, Canada
- Colledge, M. J. (1987), "The Business Survey Redesign Project: Implementation of a New Strategy at Statistics Canada", Recueil de la Third Annual Research Conference, Bureau of the Census, 550-576.
- Cuthill, I. (1990, révisé 1997), "The Statistics Canada Business Register", document interne, Secteur de l'informatique, Statistique Canada
- Gagné, P. (2004), "Projet de refonte du registre des entreprises", document interne, Statistique Canada.

Pursey, S., Beaucage, Y. and Hunsberger, P. (2005), "The Redesign of the Statistics Canada's Business Register"
Recueil 2005 de la American Statistical Association (à paraître)

Rancourt, E., Bérard, H. and Pursey, S. (2005), "Re-Thinking Statistics Canada's Business Register"; Federal
Committee on Statistical Methodology Research Conference, Arlington, VA. Sur CD-ROM.