

Article

Dossiers administratifs et recensement

par Fritz Scheuren

Décembre 1999



Dossiers administratifs et recensement

Fritz Scheuren¹

Résumé

Dans de nombreux pays européens, l'utilisation de dossiers administratifs dans les applications de recensement qui était, au départ, une mesure accessoire a évolué pour devenir aujourd'hui une mesure essentielle. Comme on le précise dans le présent article, cette transition pose de plus grands défis au Canada et aux États-Unis. On commence par examiner de façon assez approfondie les progrès réalisés aux États-Unis en vue de définir un nouveau paradigme de recensement administratif et on les compare à ce qui a déjà été fait ailleurs, particulièrement au Canada. Puis, on définit un plan de recherche et on établit certains liens entre toute une gamme de programmes statistiques associés au recensement des États-Unis, y compris les enquêtes courantes, les estimations intercensitaires et la mesure du sous-dénombrement. On accorde une place importante à l'examen des questions que soulève la protection des renseignements personnels. On souligne aussi le rôle du traitement informatique à faible coût et des logiciels perfectionnés de couplage des enregistrements. Enfin, on examine brièvement l'évolution du rôle du bureau central de la statistique à mesure que l'on s'enfonce davantage dans l'ère de l'informatisation.

Mots clés : Couplage des enregistrements; protection des renseignements personnels; enquêtes et estimations intercensitaires.

1. Introduction

L'utilisation des dossiers fiscaux et d'autres dossiers administratifs gouvernementaux pour exécuter le recensement remonte assez loin dans le temps, même si la plupart des progrès réels ont eu lieu les 25 dernières années – autrement dit, à peu près la même période que celle durant laquelle a été publiée la revue *Techniques d'enquête*.

Il semble que l'introduction des méthodes modernes d'échantillonnage – l'échantillonnage représentatif, comme l'a appelé Kiaer (Bellhouse 1988) – ait été liée à l'appariement d'échantillons de déclarations de revenus aux données du recensement norvégien de 1890 (Johnson et Kotz 1997). Naturellement, les travaux ultérieurs ont porté surtout sur le fondement mathématique de la méthode de Kiaer, plutôt que sur ses applications particulières (perspective qui continue d'être généralement adoptée dans *Techniques d'enquête*). Cette approche était justifiée puisqu'il était souvent difficile d'avoir accès aux dossiers administratifs et de les utiliser.

Toutefois, au fil du temps, de nombreux pays ont adopté de nouvelles lois (comme la Loi sur la statistique du Canada) qui ont donné au bureau central de la statistique l'accès systématique aux dossiers administratifs. S'ils ont suscité de nouveaux problèmes, les progrès réalisés par les administrations publiques et d'autres entités dans le domaine du traitement informatique et de la tenue de dossiers ont aussi facilité la consultation des dossiers administratifs à des fins statistiques et cette tendance semble se poursuivre, voir même s'accélérer (par exemple, Kenessey 1994).

Dans certains pays nordiques, les données de recensement traditionnelles ont été remplacées complètement ou partiellement par des données administratives (par exemple, Myrskla 1991; Thomsen et Holmoy 1998). Par contre, cela

n'est le cas ni au Canada ni aux États-Unis – d'une part à cause de la nature des dossiers administratifs existants et d'autre part à cause de l'énormité et de la complexité de la tâche. En fait, même les auteurs de l'une des propositions nord-américaines de recensement fondé sur les dossiers administratifs (RDA) les plus ambitieuses (par exemple, Alvey et Scheuren 1982) recommandent de ne pas éliminer complètement la collecte classique de données de recensement, mais plutôt d'adopter une stratégie mixte.

Le présent article vise à faire brièvement le « point » sur les propositions de recensement fondé sur des dossiers administratifs (pour des renseignements supplémentaires, voir Steffey et Bradburn 1994). Il porte principalement sur la situation aux États-Unis, mais établit un parallèle avec la situation au Canada (Leyes et Elsl-Culkin 1994). Il sert, en quelque sorte, de suite ou de mise à jour à un article publié dans *Techniques d'enquête* il y a dix ans (Scheuren 1990).

Avant d'entrer dans les détails, il serait bon de situer dans leur contexte les changements requis pour réaliser une certaine forme de recensement fondé sur les dossiers administratifs. Il serait sans doute sensé de parler pour commencer de la nature générale des révolutions scientifiques. À cet égard, Bellhouse (1988) cite Kuhn (1970) pour ce qui est de l'échantillonnage proprement dit, tandis que Scheuren (1990) s'inspire indépendamment de cet auteur en ce qui concerne les RDA.

Par exemple, ce n'est vraiment qu'à partir de la publication de l'article de Neyman (1934) – appuyé par Sukhatme (1935), ainsi que par la défense de Deming et l'article très important de Hansen et Hurwitz (1943) – que l'on a pu soutenir que la méthode représentative fondée sur la randomisation proposée par Kiaer était reconnue.

1. Fritz Scheuren, The Urban Institute. Mail: 1402 Ruffner Road, Alexandria, VA 22302, États-Unis. Courriel : scheuren@aol.com.

Du moins en Amérique du Nord, les idées concernant le RDA n'ont pas encore trouvé leur Neyman. Pourtant, de nombreux succès, remportés au prix de lourds efforts, méritent d'être applaudis et, grâce à la mise au point d'un plus grand nombre de technologies porteuses (comme le couplage des enregistrements et le traitement informatique à faible coût), on peut affirmer que les perspectives d'avenir sont encourageantes.

Le présent article est divisé en huit sections, la première étant l'introduction (section 1). La section 2 fournit certains renseignements contextuels sur le RDA et la section 3, quelques hypothèses concernant les problèmes qui dépassent le cadre de la faisabilité opérationnelle d'un RDA. Le reste de l'article comprend des propositions dans quatre domaines, à savoir le Recensement de l'an 2000 aux États-Unis (section 4), le programme d'estimations démographiques intercensitaires pour la décennie à venir (section 5), le programme d'enquêtes courantes (section 6), et la planification du Recensement de 2010 aux États-Unis (section 7). La section 8, réservée aux conclusions, comprend un examen des priorités. Elle est suivie par une série de références bibliographiques.

2. Contexte de la proposition originale de RDA

Dans plusieurs pays européens, l'élaboration de recensements fondés sur des dossiers administratifs a débuté durant les années 70 et est déjà bien amorcée (par exemple, Jensen 1983; Jabine et Scheuren 1987; Redfern 1989; Blum 1999). Toutefois, comme ces pays sont beaucoup plus petits que les États-Unis ou que le Canada et s'en distinguent de bien d'autres façons, dont le contrat social qui sous-tend la réalisation du recensement, il est difficile de transposer directement les résultats de leurs travaux. Néanmoins, on peut tirer de nombreuses leçons de leurs efforts de pionniers.

L'idée d'un RDA partiel aux États-Unis a été rendue publique pour la première fois à l'assemblée de 1982 de l'American Statistical Association (Alvey et Scheuren 1982). La proposition s'inspirait notamment des travaux de John Leyes et de Doug Norris, de Statistique Canada.

Fondamentalement, les auteurs de la communication recommandaient d'étudier les moyens de coupler les données fiscales du U.S. Internal Revenue Service (IRS) aux enregistrements sur les salaires, les revenus de pension et les prestations de chômage, ainsi qu'au fichier administratif des U.S. Health and Human Services (HHS) pour obtenir un recensement « embryonnaire » de la population. (Traduction)

L'élément clé de cette proposition consistait à rechercher un moyen de remplacer partiellement plutôt que complètement un recensement classique. Qui plus est, l'utilisation des dossiers administratifs était limitée à ceux auxquels le

U.S. Census Bureau avait déjà légalement accès, entièrement ou partiellement. Les auteurs se demandaient aussi s'il serait possible d'apporter certains changements aux systèmes administratifs pour permettre d'exécuter un RDA; toutefois, il n'a jamais été envisagé dans la proposition que le RDA représente, « en matière de contenu », beaucoup plus qu'un dénombrement embryonnaire de population.

Cependant, les auteurs de la proposition s'attendaient à ce que la couverture du RDA soit bonne, quoique imparfaite. En fait, ils ont toujours supposé qu'une forme ou l'autre d'échantillonnage serait nécessaire en vue de corriger les chiffres de population pour s'assurer de la complétude des données. Ils prévoyaient que le RDA proposé couvrirait nettement plus de 95 % de la population observée grâce à un recensement classique. Les articles publiés par Sailer et ses collaborateurs en 1993 et en 1998 l'ont confirmé (Sailer, Weber, et Yau 1993; Czajka, Moreno, et Schirm 1997; Sailer et Weber 1998).

Le fait qu'on n'ait pris aucune disposition concernant le recensement du logement qui est réalisé parallèlement au recensement de population courant des États-Unis est sans doute le meilleur exemple de l'aspect embryonnaire du projet. Aux termes de la proposition, la question du logement devrait être traitée d'une autre façon. L'une des faiblesses de la proposition, admise à l'époque, tient à la qualité des données sur la race qui figurent dans les dossiers administratifs et au fait de disposer d'adresses postales plutôt que d'adresses résidentielles réelles.

Étant donné ces limitations, pourquoi poursuivre le projet ? Parce que le RDA proposé au départ permettrait non seulement de réduire le coût et le fardeau du recensement décennal, mais aussi de produire un chiffre global de population plus fréquemment que tous les dix ans. Il permettrait aussi d'améliorer le champ d'observation de certaines populations sous-dénombrées à l'heure actuelle lors du recensement décennal. De surcroît, les travaux réalisés par Bye en 1997 (Bye 1997), ainsi que son examen détaillé des données sur la race et le groupe ethnique tirées des fichiers de la Social Security Administration permettent de placer les faiblesses mentionnées dans leur contexte et de proposer des moyens de les surmonter ou, du moins, de s'en accommoder (voir aussi Bye 1999.)

L'aspect le plus important de la proposition tient au fait que ses auteurs recommandent d'étudier la possibilité de réaliser un RDA dans dix, voir vingt ans. Ils ne proposent pas de mettre en œuvre un RDA, même si certaines réactions ont poussé à soulever la question. La proposition accorde aussi beaucoup d'attention à la protection des renseignements personnels et à la confidentialité, les auteurs insistant sur le fait que des études devaient être poursuivies dans ce domaine.

Il est inutile de poursuivre ici l'exposé détaillé de l'évolution de la situation depuis la publication du premier article d'Alvey et Scheuren jusqu'à présent, puisqu'il a été fait ailleurs (Scheuren 1995a). Il est toutefois important de

mentionner le changement de ton de la recherche au fil des ans, cette dernière visant au départ à « prouver » qu'un RDA ne pourrait pas donner de bons résultats pour en arriver aujourd'hui à déterminer par quels moyens on pourrait réaliser un tel recensement. Par exemple, dans l'excellent rapport qu'il a présenté au Census Bureau, Bye expose complètement un moyen de mettre en œuvre un RDA (Bye 1997). Même si nombre de ses aspects méritent d'être étudiés plus en détail, la méthode de Bye montre que le concept est concrétisable d'un point de vue opérationnel.

3. Hypothèses

Il est certain que la technique du couplage des enregistrements et l'accès généralisé à des systèmes d'enregistrement de grande envergure, entièrement informatisés, permettent d'envisager d'autres solutions que les recensements classiques à l'extérieur du secteur fédéral américain. Les gouvernements des divers États ont de bonnes raisons de s'assurer que chaque résident soit recensé (Biskupick 1998) et pourraient sans aucun doute mettre sur pied des RDA partiels en se servant de leurs propres systèmes d'enregistrements. Le transfert de la responsabilité de certaines activités fédérales et les incitatifs financiers intégrés aux programmes fédéraux de subventions motivent sans aucun doute les divers États à contester le monopole du Census Bureau. En effet, presque 200 milliards de dollars sont répartis annuellement au titre des programmes d'aide fédéraux en fonction du chiffre de population.

3.1 Ensembles énormes de données

Les spécialistes du marketing de masse et les organismes spécialisés dans les enquêtes téléphoniques possèdent également d'importants ensembles de données susceptibles d'être exploités. Les sources privées de données, dont il n'était nullement question il y a quelques années (par exemple, même en provenance de chaînes d'épicerie !) prennent rapidement de l'expansion et l'exploitation de ces sources privées à des fins statistiques est déjà un fait concret (National Academy of Sciences 1996). Compte tenu de la révolution mondiale des méthodes électroniques de tenue de dossiers, les jeunes industries de l'information compteront de nombreux nouveaux joueurs, le prix « minimal » à payer pour entrer sur le marché ayant baissé alors que la valeur de l'information ne cesse d'augmenter.

Certaines études réalisées récemment pour l'État du Connecticut illustrent les points généraux qui viennent d'être énoncés. White, Mulrow et Scheuren (1999) décrivent des travaux commissionnés par cet État en vue d'utiliser les dossiers administratifs publics pour améliorer le système de sélection des jurés. Il est important de souligner d'emblée que l'objectif de ces travaux n'était pas de mettre sur pied un RDA. Néanmoins, l'exercice montre à quel point il serait facile d'élaborer un RDA partiel pour un État.

Auparavant, le Connecticut se servait des fichiers d'enregistrement des électeurs et des véhicules à moteur et,

par un processus exigeant beaucoup de main-d'œuvre, éliminait des doubles comptes entre les deux systèmes en vue de produire une liste d'après laquelle étaient sélectionnés les jurés possibles. Le nouvel effort, décrit par White et coll. (1999), se fonde sur l'application de techniques probabilistes de couplage des enregistrements (Jaro 1989) à quatre fichiers de données recueillies au niveau de l'État, à savoir les deux fichiers susmentionnés, ainsi que les fichiers des déclarations de revenus et des prestations de chômage du Connecticut. Ces fichiers ont tous été créés au début de 1998.

Aux fins de leur évaluation, on a comparé les données couplées du Connecticut aux projections démographiques par canton établies par le Census Bureau en 1996 et reportées à 1997. La couverture de population obtenue grâce à la combinaison des fichiers de dossiers administratifs a été étonnamment satisfaisante, puisque l'objectif n'était pas de réaliser un RDA. En fait, les dénombrements par canton obtenus par couplage des dossiers administratifs étaient fortement corrélés aux projections de population. Le coefficient de corrélation simple était égal à $\rho = 0,946$. Après élimination de quatre des 169 cantons parce que les valeurs correspondantes paraissaient aberrantes, la valeur du coefficient est passée à $\rho = 0,977$.

3.2 Considérations concernant le respect des renseignements personnels

Bien sûr, l'utilisation directe de fichiers administratifs et les couplages entre ces fichiers sont assujettis à des contraintes visant à assurer la protection des renseignements personnels. Manifestement, dans le cas d'un RDA, les considérations relatives à la protection des renseignements personnels auront une incidence sur les couplages de données provenant de diverses sources en vue de produire un recensement. Les premiers résultats de l'étude sur la protection des renseignements personnels dans le contexte du couplage des enregistrements ont été présentés en 1985 (Scheuren 1985) et suivis par de nombreux autres travaux, notamment ceux du Census Bureau, publiés, par exemple, par Gates et Bolton (1998), Gates (1999) et par Singer (1999).

Malgré certaines réserves, dont celles exprimées par Scheuren (1997), il semble raisonnable de penser que le recours graduel à un nombre de plus en plus élevé de couplages aboutira à la sanction généralisée de la méthode en tant que stratégie. En fait, Statistique Canada a déjà entrepris l'expérience dans le cadre de son Enquête sur la sécurité financière, pour laquelle les répondants peuvent choisir d'autoriser la consultation de leur dossier fiscal ou de leur dossier de retraite au lieu de répondre à certaines questions de l'enquête (Statistique Canada 1999). Jusqu'à présent, pour cette enquête, les résultats indiquent que le concept est fort bien accepté.

En revanche, aux États-Unis, il se pourrait qu'un sixième des membres de la population s'opposent à l'idée, mais qu'on ne tienne pas compte de leur opinion. Il est néanmoins vraisemblable qu'il n'y ait aucune protestation et que

cette « violation » du droit à la protection des renseignements personnels aura lieu sans grand incident. Dans le discours d'ouverture qu'il a prononcé à la conférence internationale sur le couplage des enregistrements de 1997, Fellegi (1997) a bien analysé cette éventualité.

3.3 Considérations concernant l'accès aux données

Pour que le Census Bureau puisse procéder à un RDA, il faudrait adopter de nouveaux textes législatifs exigeant que divers organismes gouvernementaux collaborent avec le Bureau. À l'heure actuelle, l'IRS mise à part, le Bureau ne peut recevoir de données administratives que si d'autres organismes sont disposés à les lui fournir; toutefois, contrairement à la situation au Canada, où la Loi sur la statistique est en vigueur, il n'existe aucun texte législatif exigeant que les organismes collaborent avec le Bureau. Or, le maintien d'une telle situation serait insoutenable si le Census Bureau devait essayer de réaliser un RDA.

L'élaboration et la mise en application de ce genre de loi donnerait l'occasion d'organiser un débat public portant sur les concepts d'un recensement basé sur des données administratives, exercice qui doit avoir lieu avant d'exécuter tout recensement de ce genre. En tout cas, il faut adopter une loi exigeant la collaboration des chercheurs qui se penchent sur la question du RDA; sinon, le Bureau n'arrivera peut-être jamais à exécuter les travaux préparatifs. Par conséquent, il faudrait que la législation soit adoptée « maintenant », pour que le Census Bureau puisse préparer un RDA pour 2010.

3.4 Progrès techniques

À cet égard, on suppose que les progrès dans le domaine du couplage des enregistrements se poursuivront, sous la direction de Bill Winkler au U.S. Census Bureau et de Martha Fair, entre autres, à Statistique Canada. L'« engouement » pour l'exploration en profondeur des données aboutira vraisemblablement à la dissémination à très grande échelle de ces techniques. On multipliera les combinaisons de grands ensembles de données appartenant à des entreprises privées, et ce, d'une façon de plus en plus satisfaisante du point de vue statistique. Parallèlement à cette croissance, on se rendra compte, comme l'ont fait Scheuren et Winkler (par exemple, 1993 et 1997), que l'objectif fondamental du couplage des enregistrements n'est pas d'apparier des données, mais de trouver un moyen de combiner des sources disparates pour produire des renseignements que l'on ne pourrait obtenir autrement à cause de leur coût.

Il sera de plus en plus facile d'avoir accès aux logiciels améliorés des systèmes d'information géographique (SIG) et de les appliquer, particulièrement dans le cas de l'estimation par petite région, aussi bien au sein des bureaux centraux de la statistique qu'à l'extérieur. Nous entrons dans un nouveau monde « regorgeant de données », où on assistera à une véritable explosion de renseignements régionaux. La plupart de ces renseignements seront des estimations, mais la qualité globale sera superbe. De plus en plus souvent, les estimations isolées (comme dans Schaible

1996) seront remplacées par des ensembles de covariables interdépendantes formant un tout cohérent. Selon toute vraisemblance, le marché sera la force motrice de cette évolution. L'accès à des méthodes moins coûteuses et plus puissantes de calcul signifiera que le traitement de très gros fichiers et l'exécution de calcul fastidieux ne seront plus perçus comme des obstacles, même par les administrations publiques – quoique le secteur public accusera un retard dans ce domaine.

Si ces scénarios se concrétisent, la collecte fort onéreuse de données (comme dans le cas d'un recensement classique) fera de plus en plus place à la réutilisation frugale de données – souvent obtenues de façon automatisée (par exemple, comme le prédisait l'Institut International de Statistique (1994). Les applications généralisées de réutilisation des données donneront naissance à d'encore meilleures techniques qui, conjuguées à la concurrence et à la possibilité d'exécuter des calculs à bon marché, réduiront énormément le pouvoir des producteurs de données, y compris les bureaux centraux de la statistique.

4. Propositions concernant le recensement de l'an 2000

De nombreux travaux de recherche sont de toute évidence nécessaires à la mise au point d'un recensement fondé sur des dossiers administratifs. La présente section décrit certaines propositions de travaux de recherche à effectuer sur les fichiers de données administratives dans le cadre de la préparation du Recensement de l'an 2000 des États-Unis. Les propositions sont classées dans les catégories du processus (section 4.1) et du contenu (section 4.2).

4.1 Observations concernant le processus

Le Recensement de l'an 2000 sera le « coup d'envoi » d'une décennie d'activités éventuelles en prévision du Recensement de 2010. Les observations faites ici quant à ces activités sont groupées en quatre catégories, à savoir l'acquisition de grandes quantités de données administratives, le renforcement des mesures de protection relatives à leur utilisation, l'élaboration d'ententes de coopération visant les échanges d'employés et la création d'un précédent de modification des systèmes administratifs existants en vue de faciliter leur utilisation comme source d'information.

4.1.1 Acquisition de données

Les expériences passées montrent que la collaboration des autres organismes gouvernementaux s'intensifie au moment du recensement. Même s'il y a eu de nombreuses complications, le fait que le premier extrait du fichier maître des déclarations des particuliers (Individual Master File) que l'IRS a transmis au Census Bureau ait été obtenu d'après les déclarations de revenus remplies pour l'année 1969 durant

l'année du recensement décennal de 1970 n'est pas une coïncidence. On devrait par conséquent profiter de l'occasion du Recensement de l'an 2000 (et on l'a fait) pour faire progresser les préparatifs du Recensement de 2010 en acquérant des données et en déterminant comment les utiliser pour mettre sur pied un recensement fondé sur les dossiers administratifs (RDA).

Le précédent établi récemment par le Census Bureau en obtenant l'application complète du Social Security Number (SSN), ou fichier Numident, de la Social Security Administration (SSA) est un exemple particulièrement important du genre d'acquisition nécessaire, puisque le fichier contient l'âge et d'autres éléments de données démographiques sur toutes les personnes qui possèdent un SSN. Comme l'exposent Prévost et Leggieri (1999), de nombreux efforts déployés à l'heure actuelle ont permis au Census Bureau d'avoir accès à un nombre encore plus grand de systèmes d'enregistrement fédéraux.

L'obtention de l'accès pilote aux dossiers des programmes de soins médicaux offerts aux personnes indigentes (Medicaid) devrait être un objectif prioritaire, malgré les problèmes de qualité que pose ce genre de système. Cependant, il ne faut pas commettre d'erreur; une politique d'acquisition en gros des données pourrait être dangereuse d'un point de vue perceptuel (par exemple, violation de l'hypothèse de la protection des renseignements personnels mentionnée à la section 3). Seuls les systèmes pour lesquels il existe des objectifs de recherche précis et défendables (et un appui financier) devraient être visés.

Il est important de souligner qu'un recensement non seulement exige un dénombrement complet de la population, mais doit aussi inclure l'emplacement géographique correct à un moment particulier dans le temps. Pour la répartition des ressources, il est nécessaire d'obtenir des données au niveau de l'État; pour procéder à un nouveau découpage par district, les données sur l'emplacement géographique doivent être obtenues à un niveau d'agrégation beaucoup plus fin que celui de l'État. En ce qui concerne l'acquisition des données, les conséquences sont de deux types. En premier lieu, on doit s'efforcer de créer des fichiers de données administratives qui, collectivement, couvrent «l'ensemble» de la population. En deuxième lieu, les fichiers doivent fournir de bons renseignements à un faible niveau d'agrégation géographique à des moments précis dans le temps. Parfois, certains fichiers devraient même être obtenus simplement parce qu'ils fournissent de meilleurs renseignements géographiques pour des segments particuliers de la population (pour plus de renseignements, voir Bye 1997).

Les acquisitions de données produites par l'IRS pourraient revêtir deux formes, soit des ajouts par petit incrément, soit l'acquisition de fichiers complets de données fiscales déjà obtenues par le Census Bureau. Les petits ajouts, de nature technique et procédurale, nécessitent uniquement l'intervention d'employés exécutants; les acquisitions plus importantes, qui comportent des éléments de politique,

exigent une approche différente, avec la participation d'employés de plus haut niveau.

Depuis 1969, le Census Bureau obtient régulièrement un extrait du fichier maître des déclarations des particuliers de l'IRS. Les déclarations tardives, produites trop tard pour être incluses dans l'extrait, acquièrent de plus en plus d'importance et leur contenu pourrait être ajouté aux données transmises par l'IRS. On devrait aussi obtenir les déclarations produites l'année précédente et les intégrer aux échantillons longitudinaux recommandés plus loin aux sections 5 et 6. Il est également proposé d'augmenter marginalement le contenu des données élémentaires afin d'y inclure un plus grand nombre de catégories de revenu. Il est recommandé d'obtenir tous les fichiers maîtres de données, ou du moins un grand échantillon de ceux-ci, sous forme électronique. L'obtention de tous les enregistrements de données sur la rémunération et la sécurité sociale, comme cela a été fait, est un excellent point de départ qui représente assurément un compromis plausible pour l'an 2000, mais les enregistrements sur les intérêts et les dividendes sont importants également.

En tout cas, on devrait s'efforcer par tous les moyens de budgéter des fonds durant les années intercensitaires en vue de soutenir ce système, obstacle auquel se heurte traditionnellement le programme des données administratives du Census Bureau. En fait, on pourrait soutenir que, jusqu'à récemment, le Bureau obtenait déjà plus de données administratives qu'il ne peut pleinement traiter avec les ressources humaines et financières dont il dispose.

4.1.2 Sécurité physique et perceptuelle

L'amélioration de la sécurité physique des données administratives va de pair avec l'acquisition d'un plus grand nombre de données. Récemment, le Census Bureau a créé un environnement d'accès restreint et sécurisé pour ses données démographiques d'origine administrative (Clark et Gates 1999), mais il ne doit pas en rester là. Il devrait recourir aux services d'un cabinet de vérificateurs externes pour mettre à l'épreuve les nouveaux dispositifs de sécurité. En fait, ce genre d'effort devrait faire partie intégrante de ses nouvelles fonctions de gardien de données administratives. Assurer la protection des données est un élément indispensable au succès d'un RDA. Il est important de reconnaître que les bases de données qui résultent du couplage d'enregistrements administratifs sont intrinsèquement plus précieuses que les fichiers individuels des organismes; les employés sont donc soumis à de plus nombreuses tentations ou, du moins, soupçonnés d'être plus vulnérables. En fait, les infractions commises par des employés de l'IRS découvertes il y a plusieurs années (voir Scheuren 1995b) ont mené à l'adoption d'une loi interdisant le survol des données qui visent tout spécialement les données fiscales. Le Census Bureau doit prendre toutes les précautions nécessaires pour que ces règlements soient appliqués à tous ses fichiers de données administratives. Il faut aussi poursuivre les sondages d'opinion et multiplier les

groupes de discussion en profondeur avec divers intervenants et les membres du grand public, ainsi qu'avec les employés du Bureau. Le coût de la tenue à jour de systèmes énormes de données administratives englobe le maintien physique et perceptuel de la sécurité des données. Et ni l'un ni l'autre n'est bon marché.

4.1.3 Élaboration d'ententes de coopération visant les échanges d'employés

L'amélioration du capital humain est également un élément clé de tout projet d'utilisation de données administratives. Trop souvent, les statisticiens professionnels qui travaillent à l'extérieur du bureau administratif ne pensent qu'aux produits de données qu'ils obtiennent plutôt qu'au système dans son ensemble. D'aucuns soutiendront que l'expression malheureuse « exploitation des données administratives » est la conséquence de cette façon de voir étroite (et dénigrante). Quelle soit malheureuse ou non, cette phrase est le reflet de la phase de chasse et de cueillette de l'utilisation des dossiers administratifs à des fins statistiques. Mais, cette époque est révolue.

L'objectif réel (ou nouveau) devrait consister à transformer les « systèmes administratifs en systèmes d'information » (Scheuren et Petska 1993). Autrement dit, nous devons passer, pour poursuivre l'analogie, à la phase suivante, ou agricole, de l'utilisation des données administratives.

Un moyen d'accélérer cette nouvelle phase consisterait à recourir au programme de boursiers de recherche (fellows program) de l'American Statistical Association. Le Census Bureau pourrait offrir une bourse de congé sabbatique à certains employés des services des opérations d'organismes administratifs, venant de partout dans le monde. Le projet pourrait consister à accueillir en stage au Census Bureau pour de courtes périodes des personnes responsables des fichiers administratifs de l'IRS, de la SSA et d'autres organismes. L'un des objectifs serait de les aider à saisir l'importance des services d'information que leurs systèmes administratifs rendent possibles. En retour, les employés des organismes administratifs offriront des éclaircissements précieux sur les hypothèses applicables à leurs données et à leur utilisation à des fins statistiques.

L'échange inverse, c'est-à-dire l'envoi d'employés du Census Bureau en stage prolongé dans les services des opérations des organismes concernés serait encore plus important. Contrairement à la situation au Canada, où les mouvements de professionnels entre les diverses divisions de Statistique Canada, ainsi qu'entre le Bureau et les organismes administratifs sont fréquents, les échanges de ce genre sont fort peu nombreux aux États-Unis. En tout cas, il est nécessaire de les multiplier. Il suffit de penser à la stimulation statistique que ce genre d'échange pourrait représenter pour les personnes envoyées en stage. Deming a dit qu'il fallait penser en termes de systèmes (Deming 1986). Or, y aurait-il un meilleur moyen que ces programmes d'échange, répétés périodiquement tous les deux ou trois

ans, pour établir cette façon de penser dans le cas des systèmes administratifs.

4.1.4 Créer un précédent de modification des systèmes administratifs existants en vue d'améliorer leur utilisation à des fins d'information

On peut améliorer de nombreuses façons les produits de données statistiques tirés des systèmes administratifs. L'une d'entre elles consiste à ajouter un élément de données (et le fardeau connexe) à un système administratif existant. Naturellement, il s'agit d'une épée à double tranchant. L'obtention des adresses résidentielles qui figurent sur les déclarations de revenu, par exemple, comme cela a été fait en 1981, est un exemple évident; cependant, Bye (1997) propose une autre – et peut-être meilleure – approche qui comporterait un suivi direct dans le cas des adresses qui ne sont de toute évidence pas résidentielles.

On pourrait peut-être aussi ajouter aux données des déclarations de revenus un numéro de téléphone résidentiel conventionnel (ligne terrestre). L'usage croissant des téléphones cellulaires pourrait rendre la valeur de ces numéros temporaires, mais cela vaudrait quand même la peine de les obtenir. Du moins aux États-Unis, l'adoption du téléphone cellulaire n'est pas encore conjuguée à l'abandon des technologies moins récentes. En tout cas, on peut prédire que le rendement de l'utilisation administrative de ces numéros pourrait plus qu'excéder leur valeur en tant qu'instrument statistique du couplage des enregistrements durant le recensement et, ultérieurement, dans un programme permanent d'enquête. De surcroît, les numéros publiés offriraient un moyen précieux de vérifier les adresses.

Bien que cela soit probablement très difficile à réaliser, la modification des déclarations de rémunération remplies par une tierce partie (formule W-2 de l'IRS ou T-4 de Revenu Canada) afin d'y faire figurer la date de la dernière période de paie serait très fructueuse du point de vue d'un RDA. L'ajout de cette date éliminerait en grande partie l'ambiguïté due aux multiples adresses qui figurent sur ces documents. Évidemment, l'accès aux enregistrements salariaux trimestriels du système de chômage, par l'entremise du U.S. Bureau of Labor Statistics, serait une encore meilleure solution qui n'augmenterait pas le fardeau existant.

Quoique réalisable, la mise en œuvre de ces propositions exigera beaucoup d'efforts, puisqu'il faudra tenir compte de nombreux autres intervenants et des coûts. Dans son rapport de 1997 sur l'exécution éventuelle d'un RDA, Bye indique que l'on pourrait obtenir auprès de la U.S. Social Security Administration les fichiers d'adresses postales utilisés par cette administration pour envoyer les SSN aux parents des nouveau-nés. Ici, le fardeau est léger et la valeur, grande, puisqu'on obtiendrait ainsi l'accès aux adresses courantes des familles comptant un nouveau-né.

Certains compromis sont évidemment plus faciles à réaliser que d'autres. Il est essentiel, dans la mesure du possible, de trouver des moyens de relier un objectif

d'information à un objectif administratif existant, donc de produire un élément de valeur pour chacun.

4.2 Propositions d'études

Les études valables que l'on pourrait recommander sont nombreuses. Deux idées importantes consistent à 1) obtenir les SSN pour les échantillons sélectionnés pour la vérification postcensitaire de la qualité, de sorte qu'on puisse obtenir une estimation du sous-dénombrement fondée sur trois systèmes et 2) produire une estimation restreinte fondée sur un RDA durant le Recensement de l'an 2000 aux fins de la contre-vérification au moyen des « dénombrements » officiels. Pour que ces mesures soient pleinement efficaces, les résultats obtenus dans l'un et l'autre cas doivent figurer sur les mêmes formulaires que ceux utilisés pour les dénombrements officiels du Census Bureau et pour les estimations corrigées en vue de tenir compte du sous-dénombrement, lesquels doivent être produits en décembre 2000 et en mars 2001, respectivement.

4.2.1 Estimation fondée sur trois systèmes

Il est recommandé depuis longtemps (par exemple, Scheuren 1995a) d'essayer de produire des estimations en se fondant sur trois systèmes (Zaslavsky et Wolfgang 1993). Les trois systèmes correspondraient à l'échantillon de vérification de la qualité, au recensement proprement dit et à la liste amalgamée des enregistrements administratifs dont on a éliminé les doubles comptes. Pour que l'estimation à système triple donne de bons résultats, tous les appariements réalisés doivent être de haute qualité. Sans les SSN obtenus grâce à l'enquête postcensitaire de vérification de la qualité, l'appariement des enregistrements administratifs sera beaucoup plus difficile et, dans les cas douteux, produira peut-être des résultats ambigus. Sans les SSN, les cas d'adresses multiples avec un nom de famille très courant seront particulièrement problématiques.

4.2.2 RDA partiel concurrent au niveau de l'État

Même sans vouloir appliquer une méthode à trois systèmes, l'exécution d'un RDA limité concurrent est prometteuse lors de tout examen postcensitaire. La vérification immédiate des dénombrements nécessaires au niveau de l'État pourrait être faite selon la méthode déjà employée deux fois par Sailer, Webber et Yau 1993 et Sailer et Webber 1998 et ce, rapidement, si on lui accorde une priorité suffisante. En principe, les enregistrements administratifs de l'IRS nécessaires seront en place au début de l'automne 2000 et pourraient être traités par le Census Bureau selon le principe de flux.

Plus précisément, il est recommandé que le Census Bureau reçoive mensuellement les extraits habituels du fichier maître sur les particuliers de l'IRS, afin que l'appariement puisse commencer tôt. Les renseignements sur les travailleurs rémunérés et sur les bénéficiaires de la sécurité sociale pourraient aussi être transmis par flux par la Social

Security Administration (même avant qu'ils ne soient dépouillés par l'IRS). Nombre de personnes manquées lors du recensement décennal, mais qui figurent dans les bases de données de l'IRS, pourrait être des bénéficiaires du crédit d'impôt sur le revenu gagné qui ont déménagé. Les processus d'appariement continu et de vérification d'échantillon seront des éléments essentiels au dépistage de ces personnes. Les déclarants qui bénéficient du crédit d'impôt sur le revenu gagné et qui demandent un prêt (par anticipation du remboursement) mériteront une attention particulière, si l'on veut que le suivi donne de bons résultats.

Naturellement, de nombreuses autres idées de recherche applicables au Recensement de l'an 2000 pourraient préparer le terrain pour un éventuel recensement fondé sur des données administratives aux États-Unis (voir Provost et Leggieri 1999). Toutefois, les deux idées susmentionnées semblent être de loin les plus importantes. Pour une étude récente sur l'utilisation de données administratives dans le cadre du recensement au Canada, consulter Carter et McClean 1996.

5. Conséquences au niveau intercensitaire

Aussi paradoxal que cela puisse paraître, la réalisation de progrès révolutionnaires en ce qui concerne l'utilisation de données administratives exige une stratégie évolutionniste, particulièrement pour le programme des estimations intercensitaires.

5.1 Partie des enregistrements administratifs annuels du RDA

Avant tout, le Census Bureau devrait poursuivre l'estimation annuelle, fondée sur un RDA, au moins au stade d'échantillon, des totaux pour les États mentionnée à la section 4.2. En bout de ligne, selon les données acquises et les fonds disponibles, on pourrait étendre l'exercice et obtenir des renseignements à un niveau géographique plus fin.

5.2 Grand échantillon longitudinal de données administratives

En deuxième lieu, on devrait assembler de grands échantillons longitudinaux de données administratives. En suivant l'exemple du Canada, le Census Bureau pourrait commencer par créer un simple échantillon longitudinal d'enregistrements de données fiscales, que l'on appairerait au fichier Numident de la Social Security Administration, fichier qui contient des renseignements démographiques sur toutes les personnes qui possèdent un SSN. Depuis longtemps, Statistique Canada a créé un échantillon longitudinal à 10 % de déclarations de revenus T1 (Leyes et Elsl-Culkin 1994), qui, aux États-Unis, se traduirait par un échantillon à 1 %, étant donné les tailles relatives des deux pays. En fait, le couplage de cet échantillon longitudinal à

l'échantillon continu d'antécédents professionnels à 1 % (1 % Continuous Work History Sample) de la Social Security Administration produirait effectivement une très longue empreinte (temporelle).

En dernière analyse, cet échantillon longitudinal pourrait être étendu aux autres systèmes administratifs fédéraux (à l'IRS et à la SSA, mais peut-être aussi ailleurs). Toutefois, une mise en garde est nécessaire. La tâche consistant à appairer des unités administratives qui évoluent au fil du temps pourrait nécessiter plus de ressources et de patience qu'on ne pourrait l'imaginer et devrait donc être exécutée graduellement, par petites étapes, disons, par exemple, l'échantillon continu d'antécédent professionnel à 0,1 %, comme cela a déjà été réalisé en partie (Czajka et Walker 1989).

5.3 Échantillon statistique intégré de dossiers administratifs

Il y a 20 ans, Scheuren (1979) a formulé une proposition beaucoup plus ambitieuse concernant un ensemble d'échantillons de données administratives interreliés qu'il serait peut-être bon d'examiner de nouveau et de mettre à jour aujourd'hui. Cette proposition comporte des travaux indépendants ainsi que des travaux susceptibles d'appuyer directement les programmes classiques d'enquêtes intercensitaires et d'enquêtes courantes. Contrairement au concept fondamental du RDA, son objectif principal serait d'augmenter le contenu des éléments de données plutôt que de réaliser la couverture complète ou quasi-complète de la population. Elle pourrait aussi servir de point de départ pour divers projets d'enquête courante, comme dans le cas de l'enquête à base de sondage double sur les finances des consommateurs (Survey of Consumer Finances (SCF)) mise au point par le Federal Reserve Board (Kennickell et Woodburn 1997).

5.4 Système fondé sur les transactions

En quatrième lieu, à long terme, le programme intercensitaire courant fondé sur les dossiers administratifs devrait passer, dans la mesure du possible, d'un système de données annuelles avec appariement année après année à un système de rajustements directs, fondés sur les transactions, des dénombrements d'après les données administratives. Considérons, par exemple, le suivi d'un échantillon de SSN au cours de la décennie. Il s'agirait manifestement d'un pas vers la création d'un registre partiel de population. Malgré les réserves que pourrait soulever le public à l'égard des bases de données gigantesques, la création d'un registre statistique de la population pourrait être un bon moyen de simplifier les activités intercensitaires et d'en établir la priorité. Les travaux entrepris par le Census Bureau en vue d'améliorer le système du fichier maître d'adresses semblent déjà avoir pour objectif la création d'un registre d'adresses et de ménages mis à jour à mesure des transactions.

Ces idées concernant l'utilisation de dossiers administratifs dans le cadre du recensement décennal peuvent être

unies à des propositions visant le programme d'enquêtes courantes du Census Bureau, comme on en discutera plus bas. De toute façon, la coordination (et la synergie positive) entre ces deux programmes distincts doit devenir beaucoup plus importante qu'elle ne l'a été jusqu'à présent. Consulter Alexeter et Chet (1999) pour le genre d'effort qui pourrait réellement donner de bons résultats.

6. Conséquences concernant le programme d'enquêtes courantes

L'introduction de dossiers administratifs au stade de la conception et de l'estimation des enquêtes permanentes du Census Bureau semble une étape naturelle vers la réalisation d'un RDA. Voici quelques moyens d'y arriver.

6.1 Utilisation des bases de sondage

L'American Community Survey (ACS) pourrait être le point de départ naturel d'un effort en vue d'utiliser des dossiers administratifs dans le contexte d'une base de sondage multiple. L'utilisation du fichier maître d'adresses dans le cadre de l'ACS pourrait être complétée, par exemple, par les adresses figurant sur les déclarations de revenus et, éventuellement, celles des bénéficiaires de la sécurité sociale – et ce, pour d'autres raisons que la simple mise à jour des adresses.

6.2 Appariement d'échantillons stratifiés a posteriori

Il y a quelques années, Scheuren (1980) ont proposé que l'on apparie systématiquement les enregistrements de la CPS aux enregistrements administratifs et que l'on se serve de variables administratives comme instruments de stratification a posteriori. L'essai pilote a été réalisé au moyen de l'étude d'appariement exact (Exact Match Study) de 1973. La méthode serait beaucoup plus facile à utiliser aujourd'hui. Des travaux, comme ceux de Thomsen et Zhang (1999), pourraient constituer un prototype à jour. Une méthode apparentée est décrite par Kennickell et Woodburn (1997).

6.3 Couplage du programme d'enquêtes courantes aux objectifs intercensitaires

Que l'on prenne pour point de départ une base de sondage administrative ou que l'on procède par après à un appariement avec un fichier d'enregistrements administratifs, l'effort accompli fournira des renseignements sur les lacunes de couverture des dossiers administratifs qui permettront de mieux utiliser ces dossiers durant la période intercensitaire. En outre, ce genre d'opérations conjointes permettra de déterminer où il convient de concentrer les études de couverture pour le Recensement de 2010.

Il est essentiel d'intégrer dans le programme d'enquêtes courantes un programme permanent destiné à améliorer les méthodes démographiques déjà excellentes (programme qui

semble être envisagé par Prévost et Leggieri 1999). Des expériences répétées et ouvertes au cours de la décennie, conjuguées à l'amélioration permanente de la couverture et du contenu permettra d'éroder la résistance aux ajustements. On devrait viser à mettre au point un RDA annuel, complètement projeté, à compter de 2005 au plus tard. Il faudra s'efforcer d'obtenir les fonds nécessaires pour financer la production d'un échantillon suffisamment grand pour compléter les dossiers administratifs, peut-être par l'entremise du programme de l'American Community Survey. Toutefois, à vrai dire, cette stratégie budgétaire pourrait exiger que le Census Bureau promette de réaliser des économies importantes en 2010 – proposition risquée, mais nécessaire d'un point de vue psychologique et financier.

7. Conséquences concernant les études supplémentaires relatives au recensement de 2010

Il est difficile de faire des recommandations précises quant aux études supplémentaires sur le Recensement de 2010, puisque nombre d'entre elles dépendront du succès avec lequel le Census Bureau arrivera à utiliser les autres dossiers administratifs en vue de réaliser des objectifs multiples. Néanmoins, deux observations valent la peine d'être mentionnées, puisqu'il n'en a pas été question plus haut.

7.1 Échantillon de dépistage tiré du Recensement de l'an 2000

Un échantillon de dépistage important devrait être suivi au cours de la décennie. Le point de départ pourrait être l'échantillon postcensitaire de vérification de la qualité, après appariement avec les dossiers administratifs et ajout, dans la mesure du possible, des cas repérés uniquement dans les fichiers du recensement ou uniquement dans les fichiers administratifs. Le suivi engloberait le genre d'étapes administratives décrites plus haut, ainsi que l'application de méthodes réelles de dépistage, au moins à un sous-échantillon. Des travaux sur le terrain seraient nécessaires pour résoudre certains problèmes en vue d'assurer un « chevauchement » satisfaisant d'un recensement à l'autre. La plupart des travaux consisteraient à apparier des dossiers administratifs par vagues successives (de façon similaire à celle décrite brièvement à la section 5.2). De nouveau, la protection des renseignements personnels représenterait un problème important (déjà décrit à la sous-section 3.2).

7.2 Recensements spéciaux

En prévision du Recensement de l'an 2010, il faudra réaliser des recensements spéciaux fondés au départ sur des dossiers administratifs et s'efforcer de les compléter par échantillonnage. Du point de vue de la structure, ces recensements différeraient fort peu des essais préliminaires

réalisés avant chaque recensement. Cependant, comme le recensement fondé sur des dossiers administratifs constitue un nouveau paradigme, il faudra procéder à un plus grand nombre d'essais et, particulièrement, disposer de plus de temps pour les essais. Les premiers essais relatifs à l'an 2010 devraient être intégrés au Recensement de l'an 2000 et se poursuivre sans interruption pendant toute la décennie. Les problèmes généraux de faisabilité devraient être abordés très tôt. En voici quelques exemples.

7.2.1 Mettre au point une méthode pour utiliser efficacement comme base de sondage un amalgame d'adresses et de noms individuels extraits de dossiers administratifs, de sorte que les adresses qui ne figurent pas dans le fichier administratif soient suréchantillonnées.

7.2.2 Mettre au point une méthode pour traiter efficacement les logements à unités multiples, puisque les adresses des dossiers administratifs ne comportent habituellement pas de numéro d'appartement. (Il se pourrait qu'une partie de l'échantillonnage doive être réalisée indépendamment des listes, comme dans le cas de l'échantillonnage de vérification de la qualité du recensement, puis que les enregistrements soient appariés par la suite.)

7.2.3 Mettre au point une méthode pour traiter les populations problématiques (par exemple, enfants appartenant à un groupe minoritaire à faible revenu). Cet effort nécessitera une attention particulière (comme mentionné plus haut, les données de Medicaid pourraient être utiles ici, mais on n'en est pas certain).

7.2.4 Mettre au point une méthode pour résoudre les problèmes de localisation lors du Recensement de l'an 2000 (par exemple, quartiers déshérités). La question pourrait devoir être examinée individuellement.

L'idée de concevoir ces recensements spéciaux sous forme d'échantillon avec renouvellement (par exemple, Kish 1990) permet d'obtenir – disons, d'ici à 2005 – un « étalon-or » pour évaluer la stratégie évolutive du RDA. À noter que rien n'empêche d'essayer simultanément deux méthodes ou plus pour accélérer le processus d'essai. En effet, il se pourrait qu'on doive appliquer plusieurs méthodes simultanément pour le Recensement de 2010, y compris l'échantillonnage. Étant donné la diversité des circonstances, une stratégie multiple pourrait s'avérer intrinsèquement supérieure à toute stratégie unique.

8. Sommaire et priorités éventuelles

Les efforts importants déployés par le U.S. Census Bureau (par exemple, Prévost et Leggieri 1999) en vue d'étudier la mise au point d'un recensement fondé sur des dossiers administratifs méritent des applaudissements. Bien que les travaux concernant le RDA soient maintenant bien en train, il demeure utile de réitérer les points importants et les priorités.

8.1 Constance de l'objectif

Sur la fameuse liste en 14 points de Deming figure « constance de l'objectif », autrement dit « ne pas quitter le prix des yeux ». Étant donné tous les autres défis que pose l'exécution du Recensement de l'an 2000, continuer à se concentrer sur l'avenir est sans doute la tâche la plus ardue dont doit s'acquitter l'excellente équipe du projet. On doit renoncer à la tentation de comprimer les budgets ou de réaffecter des personnes importantes après le recensement décennal. Des fonds distincts devraient être obtenus et le montant devrait être important. Il est incorrect de penser que les grands efforts sont liés au Recensement de 2000 et qu'un certain relâchement sera permis par la suite. Les efforts de recherche devront s'intensifier sans cesse.

8.2 Surveillance de l'environnement

Bien que la responsabilité du recensement officiel n'est pas près d'être transférée (si tant est qu'elle le soit jamais), l'exécution du recensement ne sera désormais plus le monopole qu'elle a été. Les moyens utilisés pour intégrer les sources indépendantes de renseignements seront des étalons fondamentaux utilisés pour évaluer le succès du Census Bureau. On peut s'attendre à ce que l'obligation de rendre des comptes s'intensifie. L'étude de cas du Connecticut, exposée à la section 3, en est un exemple.

Ironiquement, il est essentiel de se rendre compte que les bureaux centraux de la statistique, y compris le Census Bureau, pourraient très bien être à la traîne à l'ère de l'information. La part du marché du Census Bureau dans le secteur de l'information n'a cessé de diminuer depuis plusieurs décennies et, toutes autres choses étant égales, il est vraisemblable que la diminution s'accroît au cours des dix prochaines années, compte tenu de la lenteur caractéristique de l'évolution des organismes gouvernementaux. Les travaux du Bureau portant sur les dossiers administratifs et ses efforts permanents en vue d'être un chef de file dans le domaine des technologies-clés de l'information pourraient tempérer cette tendance, mais sûrement pas la renverser.

8.3 Hypothèses

Comme le soutiennent depuis longtemps les cadres du Bureau, comme Gates, les hypothèses concernant la protection des renseignements personnels sont celles dont il faut se soucier le plus. À cet égard, il est nécessaire de bien observer et écouter. On pourrait envisager de consulter un conseil d'examen institutionnel (Institutional Review Board), non mentionné antérieurement, en vue d'obtenir un avis indépendant tenant compte de l'intérêt du public, particulièrement en ce qui concerne le couplage des enregistrements. Sinon, maintenant qu'il existe un conseil de surveillance du recensement (Census Monitoring Board), ce dernier pourrait être l'organisme naturel sur lequel se concentrer pour obtenir des conseils concernant la résolution des questions de protection des renseignements personnels et l'établissement des priorités.

La préoccupation réelle tient non pas au fait que le Census Bureau pourrait procéder trop agressivement, mais plutôt au fait qu'il pourrait se montrer trop timide. Comme l'a dit Emerson, soyons audacieux, mais pas trop audacieux. Le Census Bureau devrait de toute évidence s'efforcer de faire approuver un texte législatif comme la Loi sur la statistique du Canada, afin de s'assurer de la collaboration des organismes administratifs en vue d'obtenir les données nécessaires pour un RDA. Comme cela a été mentionné précédemment, l'élaboration et la mise en application d'une loi de ce genre donneraient l'occasion d'organiser un débat public sur les concepts du recensement fondé sur des dossiers administratifs, débat qui a lieu avant que l'on ne puisse exécuter un tel recensement.

8.4 Propositions concernant le Recensement de l'an 2000

L'inclusion d'une question sur le SSN dans l'enquête de vérification de la qualité du Recensement de l'an 2000 est indispensable si l'on veut faire le lien entre le passé et l'avenir. La proposition consistant à produire simultanément une estimation partielle basée sur un RDA permettra non seulement d'éprouver une méthode qui sera nécessaire en 2010, mais aussi de valider les données du RDA et du Recensement proprement dit. En ce qui concerne d'autres études fondées sur les dossiers administratifs, carpe diem – saisissons l'instant présent – particulièrement l'occasion d'acquérir des fichiers essentiels (effort déjà bien en train selon Prevost et Leggieri 1999) et de renforcer les partenariats de longue durée qui constituent le capital humain.

8.5 Étapes intercensitaires

Il est également essentiel d'arriver à une synergie positive plus grande entre les programmes courants d'enquêtes mensuelles et annuelles, d'une part, et le programme des estimations intercensitaires, d'autre part. Une autre priorité consisterait à transformer le programme d'estimations intercensitaires en un programme axé sur les transactions, plutôt qu'un programme essentiellement transversal. L'intégration de résultats particuliers de recensement serait également cruciale.

8.6 Étapes concernant les enquêtes courantes

Les occasions d'utiliser des dossiers administratifs dans le cadre de l'American Community Survey sont excellentes, à condition de continuer à les saisir suffisamment rapidement. Cependant, la période entre le début et la production des résultats des nouvelles enquêtes par recensement est souvent trop longue. On pourrait soutenir que les recensements sont caractérisés par des périodes cycliques qui s'étendent sur plus de dix ans. L'adoption d'une nouvelle base de sondage pour les programmes d'enquêtes courantes a progressé. Toutefois, quelles que soient les mesures qui seront prises après le prochain recensement, elles devront l'être beaucoup plus rapidement qu'après le dernier.

8.7 Étapes supplémentaires éventuelles

Des deux propositions faites à la dernière section, celle de l'échantillon de dépistage est celle qui paraît la plus prometteuse du point de vue de la recherche fondamentale. On peut s'inspirer ici de l'expérience canadienne, mais le résultat final est incertain. Le dépistage permettrait de résoudre plus facilement les problèmes que posent les mouvements d'immigration, tant légaux qu'illégaux. Manifestement, il sera indispensable de réaliser des recensements spéciaux en vue de prouver le concept et de rendre le RDA opérationnellement réalisable. Le budget requis sera beaucoup plus lourd durant les premières années de la décennie que cela n'a été le cas antérieurement pour les recensements pilotes et les répétitions finales effectuées durant les années 80 et 90. On devrait aussi donner la priorité à la planification des travaux de recherche à poursuivre après 2010 et peut-être commencer cette planification maintenant et la réexaminer au moins une fois par an.

Le numéro du 26 mars 1999 de la revue *Science* (Cohen 1999) comprend une nouvelle rubrique intitulée « The March of Paradigms » qui a pour objet de repérer le nombre croissant d'articles scientifiques dont le titre ou le résumé comprend l'expression « new paradigm ». On y ajoute que nombre de ces allégations pourraient toutefois ne pas correspondre réellement au genre de développements scientifiques auxquels pensait le philosophe Thomas Kuhn quand il a rendu l'expression new paradigm célèbre dans l'ouvrage traitant du changement de paradigme publié en 1962 sous le titre *The Structure of Scientific Revolutions*.

Malgré cette mise en garde, le passage d'un recensement classique à un RDA partiel mérite d'être considéré comme un changement de paradigme et d'être étudié sous cet angle. À cet égard, il convient de féliciter tous ceux et celles qui, partout dans le monde, ont déjà essayé de l'accomplir ou y sont parvenus. Je souhaite au U.S. Census Bureau de mener à bien les travaux qu'il exécute à ce sujet à l'heure actuelle.

Remerciements

Plusieurs personnes méritent des remerciements pour les idées continus dans cet article. Parmi les principales, il faut mentionner Wendy Alvey, Bonny Bye, John Leyes et Peter Sailer. Mes remerciements s'adressent aussi à l'éditeur adjoint ainsi qu'aux deux examinateurs très constructifs.

Bibliographie

- Alvey, W., et Scheuren, F. (1982). Background for an administrative records census. (Avec discussion par John Leyes, Statistique Canada). *Statistics of Income and Related Administrative Record Research*. Washington, DC : U.S. Department of the Treasury, Internal Revenue Service.
- Alexander, C., et Chand, N. (1999). Indirect estimation with administrative records and the American Community Survey. *1999 Federal Committee on Statistical Methodology Proceedings*. Washington DC : U.S. Bureau of the Census.

- Bellhouse, D. (1988). *Handbook of Statistics: Sampling, A brief history of random sampling methods*. New York : North-Holland.
- Biskupick, J. (1998). Division of representation, funds at stake in census feud. *The Washington Post*. 27 Novembre 1998.
- Blum, O. (1999). Combining register-based and traditional census processes as a pre-defined strategy in census planning. *1999 Federal Committee on Statistical Methodology Proceedings*. Washington DC : U.S. Bureau of the Census.
- Bye, B. (1997). Administrative Record Census for 2010: Design Proposal. Préparé pour le U.S. Bureau of the Census. Rockville MD : Westat Inc.
- Bye, B. (1998a). Race and Ethnicity Modeling with SSA Numident Data: File Development and Tabulations. Préparé pour le U.S. Bureau of the Census. Rockville MD : Westat Inc.
- Bye, B. (1998b). Race and Ethnicity Modeling with SSA Numident Data: Individual-level Regression Model - Version 2. Préparé pour le U.S. Bureau of the Census. Rockville MD : Westat Inc.
- Bye, B. (1999). Race and Ethnicity Modeling with SSA Numident Data: Two-level Regression Model. Préparé pour le U.S. Bureau of the Census. Rockville MD : Westat Inc.
- CARTER, R., et McClean, K. (1996). Using administrative data in the Canadian census: Experiences and plans. *Statistical Journal of the United Nations*, 13, 4, 375-383.
- Clark, C., et Gates, G. (1999). *Memorandum on Restricted Access Policy for Administrative Records*. U.S. Bureau of the Census, 25 juin 1999.
- Cohen, J. (1999). The march of paradigms. *Science*, 283, 1998-99.
- Czajka, J.L., et Walker, B. (1989). Combining panel and cross-sectional selection in an annual sample of tax returns. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 463-468.
- Czajka, J.L., Moreno, L. et Schirm, A. (1997). On the Feasibility of Using Internal Revenue Service Records to Count the U.S. Population. Washington, DC : Mathematica Policy Research.
- Deming, W. (1986). *Out of the Crisis*. Cambridge MA : MIT Press.
- Edmonston, B., et Schultze, C. (Éds.) (1995). *Modernizing the U.S. Census, Panel on Census Requirements in the Year 2000 and Beyond*. Committee on National Statistics, National Research Council, Washington, DC : National Academy Press.
- Fellegi, I. (1997). Record linkage and public policy - a dynamic evolution. *Record Linkage Techniques*, 3-12. Arlington VA : Ernst and Young, LLP.
- Gates, G. (1999). Data Mining, Panel on Privacy and Statistics in the New Millennium. Présentation du panel au Joint Statistical Meetings, Baltimore, MD.
- Gates, G., et Bolton, D. (1998). Privacy research involving expanded statistical uses of administrative records. *Proceedings of the Government and Social Statistics Section, American Statistical Association*.
- Hansen, M., et Hurwitz, W. (1943). On the theory of sampling from finite populations. *Annals of Mathematical Statistics*, 14, 333-62.
- Institut International de Statistique (1994). *The Future of Statistics*, (Z. Kenessey, éd.). ISBN : 90-73592-11-9.

- Jabine, T., et Scheuren, F. (1987). Record linkages for statistical purpose: Methodological issues. *Journal of Official Statistics*, 2, 255-277.
- Jaro, M. (1989). Advances in record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida. *Journal of the American Statistical Association*.
- Jensen, P. (1983). Towards a register-based statistical system - some Danish experiences. *Statistical Journal of the United Nations*, 341-365.
- Johnson, N., et Kotz, S. (Éds.) (1997). *Leading Personalities in Statistical Science: From the Seventeenth Century to the Present*. New York : John Wiley & Sons, Inc.
- Kenessey, Z. (Éd.) (1994). *The Future of Statistics: An International Perspective*. Voorburg : Institut International de Statistique.
- Kennickell, A., et Woodburn, L. (1997). Consistent Weight Design for the 1989, 1992, and 1995 SCF, and the Distribution of Wealth, disponible au site <http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html>.
- Kish, L. (1990). Recensements par étapes et échantillons avec renouvellement complet. *Techniques d'enquête*, 16, 1, 67-77.
- Kuhn, T. (1970). *The Structure of Scientific Revolutions*. Chicago IL : University of Chicago Press.
- Leyes, J., et Elsl-Culkin, J. (1994). Administrative social data in Canada: Some results and some implications. *Statistics of Income: Turning Administrative Systems into Information Systems*, U.S. Internal Revenue Service : Washington, DC.
- Myrskla, P. (1991). Census by questionnaire – census by registers and administrative records: The experience of Finland. *Journal of Official Statistics*, 7, 457-74.
- National Academy of Sciences (1996). Massive Data Sets: Proceedings of a Workshop Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences, National Research Council : Washington, DC.
- Neyman, J. (1934). On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97, 558-625.
- Prevost, R., et Leggieri, C. (1999). Expansion of administrative record uses at the census bureau; a long-range research plan. *1999 Federal Committee on Statistical Methodology Proceedings*, Washington DC : U.S. Bureau of the Census.
- Redfern, P. (1989). Population registers: Some administrative and statistical pros and cons. *Journal of the Royal Statistical Society, Series A*, 153, 1-41.
- Sailer, P., et Weber, M. (1998). The IRS population undercount: An update. *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Sailer, P., Weber, M. et Yau, E. (1993). How well can the IRS count the population? *Proceedings of the Section on Survey Research Methods*, American Statistical Association.
- Schaible, W. (Éd.) (1996). *Indirect Estimators in U.S. General Programs*. New York : Springer-Verlag.
- Scheuren, F. (1979). Integrated linked administrative statistical sample. LASS Working Notes, U.S. Social Security Administration.
- Scheuren, F., Oh, H.L., Vogel, L. et Yuskavage, R. (1981). Methods of Estimation for the 1973 Exact Match Study. Studies from Interagency Data Linkages. U.S. Department of Health and Human Services, Social Security Administration, Publication 13-11750.
- Scheuren, F. (1985). Methodological issues in linkage of multiple databases. Dans *Record Linkage Techniques - 1985: Proceedings of the U. S. Internal Revenue Service*, 155-178.
- Scheuren, F. (1990). Commentaires sur l'article de Kish (1990). *Techniques d'enquête*, 16, 1, 78-86.
- Scheuren, F. (1995a). A U.S. administrative records census. *Chance*, 8, 2, 43-45.
- Scheuren, F. (1995b). Private lives and public policies: Confidentiality and accessibility of government services. Dans *Journal of the American Statistical Association*, 90, 386-387. Washington, DC : National Academy Press (1993).
- Scheuren, F. (1997). Linking health records: Human rights concerns. *Record Linkage Techniques*. Washington DC : Ernst and Young, LLP.
- Scheuren, F., et Petska, T. (1993). Turning administrative systems into information systems. *Journal of Official Statistics*, 9, 109-119.
- Scheuren, F., et Winkler, W. (1993). Analyse de régression des fichiers de données couplés par ordinateur. *Techniques d'enquête*, 19, 45-65.
- Scheuren, F., et Winkler, W. (1997). Analyse de régression des fichiers de données appariés par ordinateur - Partie II. *Techniques d'enquête*, 23, 171-180.
- Singer, E. (1999). Data Mining, Panel on Privacy and Statistics in the New Millennium. Présentation du panel au Joint Statistical Meetings, Baltimore, MD.
- Statistics Canada (1993-1996). *Enquête sur la dynamique du travail et du revenu: documents de recherche*. Catalogue No. 75F0002F, 93-01, 94-03, 94-11, 95-19 et 96-12.
- Statistics Canada (1999). *L'enquête sur la sécurité financière de Statistique Canada: mise à jour - juillet 1999*. Catalogue No. 13F0026MPF - 99006.
- Steffey, D., et Bradburn, N. (Éds.) (1994). *Counting People in the Information Age, Panel to Evaluate Alternative Census Methods*. Committee on National Statistics, National Research Council, Washington, DC : National Academy Press.
- Sukhatme, P. (1935). Contributions to the theory of the representative method. *Journal of the Royal Statistical Society Supplement*, 2, 263-68.
- Thomsen, I., et Holmoy, A.M.K. (1998). Combining data from surveys and administrative record systems, the Norwegian experience. *Revue Internationale de Statistique*, 66, 2, 201-221.
- Thomsen, I., et Zhang, L. (1999). The effects of using administrative registers in economic short term statistics: The Norwegian Labour Force Survey as a case study. 1999 Federal Committee on Statistical Methodology Proceedings. Washington DC : U.S. Bureau of the Census.
- White, G., Mulrow, E. et Scheuren, F. (1999). *Connecticut Jury Record Linkage Research*. Washington DC : Ernst and Young, LLP.
- Zaslavsky, A.M., et Wolfgang, G.S. (1993). Triple-system modeling of census, post-enumeration survey, and administrative list data. *Journal of Business and Economic Statistics*, 11, 279-288.