



N° 11-522-XIF au catalogue

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

Symposium 2003 : Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

2003



Statistique
Canada

Statistics
Canada

Canada

Recueil du Symposium 2003 de Statistique Canada
Défis reliés à la réalisation d'enquêtes pour la prochaine décennie

LE RECENSEMENT CANADIEN DE L'AN 2006 : DIRECTION STRATÉGIQUE ET PROBLÉMATIQUE

Anil Arora¹

RÉSUMÉ

Après une période de 30 années assez stable, la façon dont le recensement est réalisé au Canada va être modifiée fondamentalement, pour répondre aux pressions externes et profiter des possibilités qu'offre l'amélioration des techniques de traitement et des technologies de collecte des données. Un changement important des méthodes de collecte et de traitement des données réduira significativement les craintes exprimées par les Canadiens et les Canadiennes en ce qui concerne la protection de la vie privée, donnera un moyen de remédier à la diminution de la capacité de saisie des données, offrira une option de réponse en ligne, réduira les coûts à long terme et raccourcira la durée du traitement des données. Le présent article décrit les changements qu'il est prévu d'apporter aux méthodes de collecte et de traitement des données pour le Recensement de 2006 au Canada et met l'accent sur trois aspects importants, à savoir les changements de contenu, la collecte de données sur Internet et l'impartition.

MOTS CLÉS : Impartition, internet, recensement, vie privée.

1. CONTEXTE

Le dernier remaniement important du processus de collecte des données du Recensement du Canada a été entrepris pour le Recensement de 1971. Depuis, lors de chaque recensement, des améliorations ont été apportées à divers modules. Suit une description des méthodes de collecte et de saisie des données employées en 2001 pour le Recensement de la population et le Recensement de l'agriculture, qui sont réalisés simultanément.

La tâche attribuée à chaque agent recenseur correspondait à un secteur pour lequel il ou elle était responsable de toutes les activités de dénombrement. Deux semaines avant le jour du recensement, chaque agent recenseur a commencé à faire la tournée des logements compris dans le secteur qui lui était assigné afin de livrer une trousse contenant le questionnaire et d'énumérer l'adresse du logement sur un document de contrôle appelé Registre des visites (RV). La liste d'adresses figurant dans le RV a été utilisée pour contrôler la livraison et le retour des questionnaires pour tous les logements compris dans le secteur de dénombrement assigné.

L'agent recenseur a demandé aux répondants de remplir le questionnaire et de le renvoyer par la poste. Les questionnaires retournés par la poste ont été transmis à l'agent recenseur approprié (c'est-à-dire celui qui avait livré le questionnaire) aux fins de leur vérification. L'agent recenseur a vérifié chaque questionnaire pour s'assurer qu'il était rempli complètement. Pour chaque questionnaire incomplet, il a procédé à un suivi par téléphone et/ou sur place pour obtenir l'information nécessaire. L'agent recenseur a en outre effectué un suivi des cas de non-réponse pour toute adresse énumérée dans le RV pour laquelle aucun questionnaire n'a été retourné par la poste dans le délai indiqué et a procédé à une interview sur place pour recueillir les réponses à ce questionnaire.

Après qu'il ait achevé le recensement de tous les logements énumérés dans le RV et que toutes les vérifications de qualité concernant la tâche assignée aient été satisfaisantes, l'agent recenseur a transmis tous les questionnaires pour le secteur de dénombrement (SD) à l'un des centres de saisie des données. Pour le Recensement de 2001, la saisie des données a été réalisée sous contrat par l'Agence des douanes et du revenu du Canada (ADRC). En fait, l'ADRC est l'organisme qui s'occupe de la saisie des données du recensement depuis 1981. Cette solution s'est avérée fort rentable, puisque l'ADRC dispose d'équipement inactif et d'employés ayant la formation requise durant la période

¹ Anil Arora, directeur, chef du Recensement de 2006, Statistique Canada, 1-B2, immeuble Jean Talon, Pré Tunney, Ottawa, ON, K1A 0T6, arorani@statcan.ca

de « ralentissement » qui suit l'échéance pour la production des déclarations de revenus, période qui coïncide précisément avec les besoins du cycle de recensement. Pour le Recensement de 2001, le codage manuel des variables d'industrie et de profession a également été réalisé sous contrat par l'ADRC, de nouveau pour les raisons susmentionnées.

Statistique Canada a introduit l'imagerie des questionnaires en 2001 et toutes les étapes de vérification et de correction subséquentes ont été réalisées en utilisant des images plutôt qu'en devant manipuler des questionnaires imprimés. Depuis plus de 30 ans maintenant, l'approche suivie par Statistique Canada pour réaliser le recensement a donné des résultats de haute qualité. La robustesse et l'expérimentation prudente sont des qualités essentielles, étant donné l'énormité de la tâche (qui met en jeu chaque ménage du Canada, nécessite 45 000 employés et occasionne des dépenses importantes) et le fait que nous n'ayons qu'une seule chance « de bien nous y prendre ».

Notre méthode consistant à morceler le pays en secteurs de dénombrement représentant des tâches faisables, chacune adaptée à un agent recenseur, suscite chez celui-ci un sentiment de prise en charge, d'engagement et de fierté à l'idée de produire un travail de bonne qualité (couverture et complétude) à un niveau d'agrégation géographique assez faible. Inversement, les problèmes éventuels peuvent être limités à ce même petit secteur.

Jusqu'à 2001, Statistique Canada a consciemment favorisé un système manuel robuste de collecte des données plutôt qu'une option « dépendante de la technologie », dont certains aspects n'ont pas encore été éprouvés dans le contexte d'un recensement. « Nous atteignons maintenant le stade où une confluence d'événements nous pousse dans une nouvelle direction, alors que la technologie arrive à maturité et que nous estimons que les risques ont diminué au point d'être gérables². »

2. LES MOTEURS D'UN CHANGEMENT DE MÉTHODE DE COLLECTE DES DONNÉES

2.1 Protection de la vie privée

En 2001, comme lors des recensements précédents, les questionnaires dûment remplis ont été retournés directement aux agents recenseurs locaux qui étaient chargés d'en vérifier la complétude et de faire un suivi, au besoin. Toutefois, les répondants sont de moins en moins à l'aise à l'idée que les agents recenseurs locaux aient accès aux données personnelles qu'ils fournissent lors du recensement, particulièrement dans les régions rurales. La Division du recensement constate une multiplication des craintes exprimées par les membres de la population par la voie de l'Assistance téléphonique du Recensement, d'appels au Commissariat à la protection de la vie privée et de nos employés sur le terrain. La sensibilisation à la protection de la vie privée n'est plus limitée à quelques cas isolés et nos assurances de protection n'apaisent pas chaque ménage canadien. Répondre à ces préoccupations par l'inaction aurait un effet négatif sur la qualité des données et les taux de réponse, et pourrait coûter des millions de dollars en activités de suivi.

2.1 Option Internet

Compte tenu de l'usage croissant d'Internet pour des transactions bancaires, commerciales et avec tous les paliers de gouvernement, il est prévisible que d'ici à 2006, les répondants au recensement demanderont une option Internet et, peut-être, d'autres options électroniques pour remplacer le renvoi par la poste du questionnaire imprimé. En effet, le gouvernement du Canada a annoncé son intention d'offrir une option Internet aux Canadiens pour tous les services gouvernementaux d'ici à 2005 (Gouvernement en direct).

2.2 Capacité de saisie des données

Les données du recensement sont enregistrées manuellement par saisie clavier dans les installations de l'ADRC depuis 1981. Étant donné l'expansion de la production électronique des déclarations de revenus grâce à la TED (transmission électronique des déclarations), IMPÔTEL et IMPÔTNET, nous ne pouvons pas nous attendre à ce que

² Ivan Fellegi, statisticien en chef du Canada

l'Agence des douanes et du revenu du Canada maintienne sa capacité de saisie manuelle au niveau et au coût requis pour répondre aux exigences des futurs recensements.

La reconnaissance intelligente de caractères (RIC) est une technologie qui offre une solution éprouvée pour remplacer la saisie manuelle. Les renseignements qui figurent sur les questionnaires imprimés du Recensement de 2006 seront saisis au moyen d'un logiciel de balayage optique et de reconnaissance d'images (ROM et RIC). La saisie des réponses aussitôt que les questionnaires individuels sont retournés par la poste, au lieu d'attendre qu'un agent recenseur ait achevé sa tâche, favorisera l'exécution rapide de la vérification automatisée de la complétude et du suivi téléphonique assisté par ordinateur pour les questionnaires rejetés à la vérification. Bien qu'elle offre la possibilité de réduire le temps de traitement, cette approche place celui-ci carrément dans la trajectoire critique des activités de collecte et entremêle deux processus qui, habituellement, sont séquentiels et indépendants.

3. POSSIBILITÉS

3.1 Importante main-d'œuvre décentralisée

Il devient de plus en plus difficile de recruter l'importante main-d'œuvre temporaire, décentralisée, nécessaire pour réaliser le recensement et de retenir les employés durant la très courte période que dure le recensement sur un marché du travail concurrentiel offrant une meilleure rémunération et des perspectives d'emploi à long terme. L'adoption de la nouvelle méthode nous permettra de dépendre moins d'employés temporaires. Pour le Recensement de 2001, nous avons eu besoin de plus de 45 000 employés sur le terrain et, pour le Recensement de 2006, nous estimons que nous en utiliserons 27 000.

3.2 Contrôle des coûts à long terme

Une part importante du coût du recensement est imputable aux opérations sur le terrain. Les agents recenseurs sont rémunérés d'après le nombre de logements et leur salaire reflète les composantes salariales et non salariales du dénombrement réussi de chaque logement compris dans le secteur qui leur est assigné. Aussi le coût global du recensement est-il fort sensible à des facteurs comme le nombre de logements, le taux salarial, le taux de réponse et l'augmentation du coût de facteurs tels que le carburant, le transport et l'hébergement.

La méthode qui sera appliquée pour le Recensement de 2006 permettra de réduire l'effet de ce genre de facteurs sur le coût global du recensement. Celui-ci devrait diminuer à mesure que la participation par Internet s'accroîtra. En outre, à mesure que les techniques de saisie de données s'amélioreront, les interventions manuelles nécessaires pour corriger les données deviendront moins fréquentes. Enfin, à mesure que la couverture du Registre des adresses s'étendra, l'envoi du questionnaire du recensement par la poste plutôt que sa livraison par un agent recenseur pourra prendre de l'importance.

4. DESCRIPTION DE LA MÉTHODE PROPOSÉE DE COLLECTE POUR LE RECENSEMENT DE 2006

Les changements proposés comprennent le remaniement des activités de collecte et de saisie de données.

Les principaux éléments de la proposition sont les suivants :

- créer un fichier d'adresses complet et étendre sa couverture à des secteurs comprenant 80 % des unités de logements du Canada;

- envoyer le questionnaire par la poste dans les secteurs où cette opération est jugée faisable³ (environ les deux tiers des unités de logements du Canada);
- dresser la liste d'adresses/livrer le questionnaire dans les secteurs où la livraison ne peut se faire par la poste (environ le tiers des unités de logement du Canada);
- offrir aux répondants l'option de renvoyer leurs données (par la poste, Internet, téléphone) directement à un centre de traitement des données (CTD);
- introduire le traitement continu et un registre automatisé des contrôles (traitement et contrôle selon le logement plutôt que selon la tâche affectée à l'agent recenseur);
- automatiser l'enregistrement, la lecture et la saisie des données figurant sur les questionnaires renvoyés;
- automatiser la vérification des questionnaires dont les données sont saisies et, pour les questionnaires rejetés, faire un suivi téléphonique assisté par ordinateur (ITAO);
- faire le suivi des cas de non-réponse (SCNR) sur le terrain.

Les **activités antérieures au jour du recensement** seront axées sur la couverture. Elles comprendront la mise à jour du Registre des adresses (RA) et la livraison d'une trousse pré-identifiée contenant le questionnaire à chaque logement. Environ huit mois avant le jour du recensement, le RA sera fourni aux agents recenseurs afin qu'ils procèdent à la prospection des îlots. (Les agents recenseurs parcourent chaque îlot porte-à-porte pour repérer chaque unité de logement et comparer chaque adresse avec celles énumérées dans le RA.) Les agents recenseurs feront les ajouts, corrections et suppressions pertinents. Cette activité durera environ huit semaines. Un fichier d'adresses sera fourni aux imprimeurs afin de préparer les trousse d'envoi par la poste. Dans les secteurs d'envoi par la poste pré-identifié, les trousse comprenant le questionnaire seront envoyées aux chefs de ménages environ deux semaines avant le jour du recensement. Statistique Canada se fondera sur les mises à jour du Fichier du réseau routier et des dossiers administratifs (permis de bâtir, fichiers de remboursement de la taxe sur les produits et services (TPS) pour habitations neuves, fichiers téléphoniques) pour repérer toute croissance de ces secteurs durant la période assez courte de six mois entre la prospection des îlots et l'envoi par la poste.

Dans les secteurs non couverts par le RA et dans certains secteurs couverts par le RA où l'envoi par la poste n'est pas possible, Statistique Canada continuera d'utiliser des agents recenseurs pour (lister/livrer) les trousse contenant le questionnaire pour obtenir une couverture complète. Dans ces secteurs, les agents recenseurs commenceront à livrer les trousse contenant le questionnaire deux semaines avant le jour du recensement. Les répondants auront le choix d'envoyer leurs données par Internet, par la poste ou par téléphone. Les données seront transmises directement au CTD plutôt qu'à l'agent recenseur travaillant sur le terrain, ce qui éliminera pratiquement toutes les craintes concernant la divulgation de renseignements personnels à l'agent recenseur local.

Les **activités postérieures au jour du recensement** viseront principalement à assurer qu'un questionnaire dûment rempli soit recueilli et traité pour chaque logement. Elles incluront la saisie des données sur le ménage et le suivi des cas de questionnaires incomplets et des cas de non-réponse. Tous les questionnaires retournés par la poste seront transmis directement aux CTD où auront lieu l'enregistrement, le balayage et l'imagerie, la vérification et le codage automatisés. Toutes les données figurant sur les questionnaires retournés par Internet seront enregistrées et saisies automatiquement.

Les données provenant des questionnaires balayés optiquement et électroniques seront soumises à un ensemble d'algorithmes automatisés (vérification de la complétude) conçus pour confirmer que les questionnaires satisfont aux règles de vérification requises. Les questionnaires rejetés à la vérification seront transmis à un centre régional d'interview téléphonique assistée par ordinateur aux fins d'un suivi.

Les tâches de suivi des cas de non-réponse (logements pour lesquels aucun questionnaire n'a été retourné par la poste, par téléphone ou par Internet) seront générées à partir du registre de contrôle deux semaines après le jour du

³ Les secteurs pour lesquels l'envoi par la poste ne serait pas considéré faisable sont ceux pour lesquels :

- les adresses ne sont pas reconnues par Postes Canada (principalement les régions rurales et les petites régions urbaines);
- il existe une grande instabilité et le Registre des adresses est vraisemblablement de mauvaise qualité;
- l'envoi par la poste serait opérationnellement peu commode, par exemple parce que le secteur ne compte qu'un nombre assez faible d'adresses pour lesquelles l'envoi par la poste est possible entourées par des zones où la livraison doit être faite par un recenseur (enclaves isolées).

recensement et les activités seront gérées par environ 33 bureaux locaux du recensement (BLR) répartis dans tout le pays. Les questionnaires remplis avec l'aide du personnel sur le terrain seront enregistrés par les BLR, puis transmis régulièrement au CTD.

4.1 Répétition générale du recensement en mai 2004

Une répétition générale du recensement est prévue pour mai 2004 dans le cadre de la stratégie de réduction des risques. Cette répétition générale a pour objectif d'essayer les divers systèmes et processus interdépendants qui sont essentiels au succès du Recensement de 2006. Contrairement au test du recensement réalisé lors des cycles de recensement précédents, la répétition générale n'est pas une occasion de « tester » et d'évaluer une série d'options et de faire un choix parmi celles-ci. Les systèmes et processus sont planifiés, conçus, développés, testés et mis en œuvre en ayant pour objectif qu'ils fonctionnent comme ils le feront en 2006, la seule différence étant l'échelle et la portée de l'échantillon visé par la répétition générale. En fait, il est prévu d'augmenter proportionnellement la portée des systèmes pour 2006 et de ne modifier la conception que dans les domaines où des défauts importants justifient un changement. La répétition générale servira aussi d'évaluation quantitative du contenu du questionnaire du Recensement de 2006.

Trois régions concentrées du pays⁴, comprenant 300 000 logements, ont été sélectionnées pour la répétition générale, par envoi par la poste et par listage-livraison en proportions à peu près égales. En outre, 20 000 exploitations agricoles ont été sélectionnées pour assurer l'intégration appropriée des processus du Recensement de l'agriculture et du Recensement de la population.

Les opérations sur le terrain, avec les activités de prospection des îlots (vérification et mise à jour du Registre des adresses sur le terrain) ont débuté cet automne (septembre 2003) et la répétition générale du recensement est prévue pour le 11 mai 2004.

5. TROIS ASPECTS CLÉS

Statistique Canada a adopté une approche novatrice pour trois aspects clés du Recensement de 2006, à savoir le contenu, l'utilisation d'Internet comme outil de collecte des données et le recours à l'impartition pour le développement des systèmes et les activités de traitement. La suite de l'article traite de ces trois aspects, y compris de l'avancement des travaux et de certains défis qu'il faudra relever et de certaines questions qu'il faudra résoudre.

5.1 Contenu du Recensement de 2006

Au Canada, le contenu final du questionnaire du recensement est approuvé par le Cabinet, mais l'élaboration des questions soumises à ce dernier est le résultat de consultations approfondies avec les intervenants importants, chaque modification faisant l'objet d'une évaluation cognitive, qualitative et quantitative exhaustive. Voici certains défis que pose l'élaboration du contenu du Recensement de 2006 :

- mesure dans laquelle la population est disposée à continuer de fournir les renseignements délicats qu'il est essentiel de recueillir pour satisfaire aux exigences de nombres de lois et règlements et aux besoins des utilisateurs de données;

⁴ Région Atlantique – Halifax

Secteurs d'envoi par la poste : parties d'Halifax et de Kentville

Secteurs de listage-livraison : parties de l'Annapolis Valley et du Cumberland County

Région du Québec – Montréal ou Longueuil

Secteurs d'envoi par la poste : parties de Longueuil, Marieville et St-Mathias-Sur-Richelieu

Secteurs de listage-livraison : partie de la Montérégie

Région des Prairies – Winnipeg

Secteurs d'envoi par la poste : parties de Winnipeg, Selkirk et Yorkton (recensement par interview de secteurs urbains – partie de Winnipeg)

Secteurs de listage-livraison : parties du centre du Manitoba et du sud-est de la Saskatchewan

- prise en compte des attentes des lobbys et des groupes d'intérêts qui souhaitent voir figurer certaines questions dans le recensement, souvent pour des raisons symboliques plutôt que spécifiquement en réponse à des exigences juridiques ou politiques;
- rétention du contenu des recensements antérieurs (séries chronologiques), tout en répondant aux besoins de nouvelles données sans accroître le fardeau de réponse;
- maintien de la comparabilité des données chronologiques tout en se conformant aux nouvelles classifications types (p. ex. passage de la CTI au SCIAN, du principal domaine d'études (PDE) à la CPE, etc.);
- incertitude quant aux fonds nécessaires pour recueillir, traiter et diffuser l'information du recensement;
- maintien de la qualité des données, tout en introduisant des changements importants dans la méthode de collecte et de traitement;
- réduction au minimum des effets potentiels que pourrait avoir la collecte de données selon divers modes (questionnaires imprimés, Internet, téléphone).

Conscient des risques que posent les changements importants dans les méthodes de collecte et de traitement des données pour 2006, Statistique Canada a élaboré une stratégie de consultation sur le contenu articulée sur une approche de changement minimal. Elle s'appuie sur les éléments fondamentaux suivants :

- fournir d'autres sources de données (enquêtes) susceptibles de répondre aux besoins des utilisateurs, parfois en ne modifiant que peu, voire aucunement, le contenu courant ou les tailles d'échantillon;
- intégrer les consultations sur les besoins concernant la diffusion des données de 2001, la géographie et le contenu du Recensement de 2006;
- adopter un horizon plus lointain pour les besoins anticipés de données (ultérieur au Recensement de 2006);
- adopter une approche d'« analyse de rentabilisation » pour examiner l'ajout de contenu interne (demandé par SC) et externe ou les demandes de modification.

Cette approche s'est avérée fructueuse et a permis de maintenir la pertinence du processus de consultation, tout en réduisant au minimum les demandes d'ajout et de modification du contenu. Les intervenants consultés ont apprécié l'idée d'intégrer les discussions sur le contenu, la géographie et les produits en un processus unique efficace. Ils se sont dits d'accord avec le contexte et prêts à examiner d'autres sources de données pour combler les lacunes statistiques, s'il y a lieu, au lieu d'introduire des changements de contenu importants.

Une courte liste de modifications et d'améliorations qu'il est recommandé d'apporter a été dressée aux fins d'examen. Les éléments le plus généralement recommandés sont :

- une nouvelle question sur les taux salariaux;
- la mise à jour de l'ensemble de questions sur l'éducation (y compris une question sur le lieu où a été obtenu le grade le plus élevé);
- une nouvelle question sur la durée du déplacement domicile-travail;
- une nouvelle question sur l'état de santé général.

En outre, deux questions ont été soulevées à l'interne par les employés de Statistique Canada, à savoir :

- offrir aux répondants l'option de consentir à ce que Statistique Canada utilise l'information figurant dans leur déclaration de revenus au lieu de répondre à une question détaillée sur le revenu;
- obtenir la permission d'autoriser l'accès aux données de recensement après 92 années⁵.

Ces six éléments, qui forment le noyau du programme de mise à l'essai du contenu du Recensement de 2006, sont complétés par plusieurs modifications mineures apportées aux questions existantes. Le programme de mise à l'essai comprend une composante d'évaluation cognitive (groupes de discussion et interviews individuelles qualitatives) et une composante quantitative. Cette dernière comprendra une analyse détaillée des données recueillies durant la répétition générale de 2004.

Parmi les éléments susmentionnés faisant l'objet d'une évaluation cognitive pour le Recensement de 2006, les questions concernant le taux salarial, l'état de santé général et le temps de déplacement pour se rendre au travail ont

⁵ En réponse à un projet de loi que l'on s'attend à être déposé devant le Parlement du Canada.

été abandonnées au printemps de cette année et ne feront pas l'objet d'une évaluation plus approfondie. Les premiers cycles d'évaluation cognitive ont révélé des problèmes de données éventuels pour toutes ces questions. Il semble qu'elles nécessiteraient des instructions détaillées, inacceptablement longues et qu'il faudrait vraisemblablement poser plus d'une question pour obtenir des résultats significatifs.

Les autres éléments (consentement à l'utilisation des données fiscales pour le revenu, module de l'éducation et permission de diffuser les données) continuent d'être évalués. Si, d'après les évaluations cognitives, on peut s'attendre à obtenir des données de bonne qualité, ces éléments seront inclus dans le questionnaire de la répétition générale. Une décision finale quant à l'inclusion des questions dans le document soumis au Cabinet sera prise après l'analyse des données de la répétition générale de 2004.

En plus du processus de détermination du contenu, l'introduction de la reconnaissance intelligente de caractères nécessite l'apport de changements importants de format au questionnaire du recensement. Le questionnaire détaillé (2B) du Recensement de 2001 ne comprenait pas la segmentation des réponses en toutes lettres et des demandes d'information sur six membres du ménage en une matrice plutôt qu'un format séquentiel sur un seul questionnaire. L'inclusion de boîtes segmentées pour 2006, qui augmente considérablement l'exactitude du balayage pour la RIC, ne laisse suffisamment d'espace que pour les réponses concernant cinq membres du ménage⁶. Cette réduction du nombre de membres du ménage traités au moyen d'un seul questionnaire fait passer le nombre de ménages nécessitant un questionnaire supplémentaire (plus de cinq membres) à environ 250 000 pour 2006.

Le questionnaire abrégé (2A) du Recensement de 2001 était une brochure à impression tête-pied anglais/français (anglais d'un côté et français de l'autre). Plusieurs problèmes associés au « perçage des réponses à travers le papier », la préparation des questionnaires pour l'insertion dans les balayeurs optiques, le balayage de millions de pages blanches et ses répercussions sur les exigences de stockage sur disque et les coûts connexes, nous obligent à étudier l'utilisation d'un questionnaire abrégé sous forme d'un feuillet plié unique. Plusieurs essais qualitatifs sont en cours pour examiner la possibilité de passer d'une brochure bilingue à un format unilingue à feuillet unique pour le questionnaire 2A.

5.2 Recensement de 2006 sur Internet

L'initiative du Gouvernement en direct (GED) du Canada précise que d'ici à 2005, toutes les transactions que les Canadiens font avec le gouvernement fédéral devront être offertes en ligne. En vue de se conformer à cette initiative et, compte tenu de l'usage croissant d'Internet par le public, de répondre aux attentes d'une option en ligne, Statistique Canada (SC) offrira à tous les Canadiens et Canadiennes la possibilité de remplir le questionnaire du recensement en ligne en 2006. En principe, il semble s'agir d'une initiative assez simple, mais en pratique, elle pose des défis importants dans les domaines de l'intégration, des communications, de la sécurité, de l'extensibilité, du coût et de la conception. Le fait que l'expérience en la matière dans le contexte d'un recensement à grande échelle, outre de petits essais et des offres limitées (au Canada et ailleurs), soit limitée accentue encore le défi.

Une petite étude pilote a été réalisée dans le cadre du Recensement du Canada de 2001 dans deux petites régions géographiques pour évaluer les réactions des gens à l'offre d'une option Internet et d'étudier les questions connexes. Bien qu'environ 5 % seulement des ménages admissibles dans les deux régions aient choisi cette option, Statistique Canada a obtenu concernant le processus des commentaires précieux qui ont orienté l'élaboration de l'option Internet pour le Recensement de 2006. Les principaux enseignements tirés et les objectifs pour l'application du Recensement de 2006 sont les suivants :

- nécessité d'intégrer le système et les processus sur le terrain pour tenir compte des modes de réponses multiples et de la notification rapide des réponses;
- nécessité d'offrir une application ne demandant aucun téléchargement et ne laissant aucune « empreinte » permanente sur la machine du répondant;
- nécessité d'offrir une application simple et efficace;
- utilisation d'éléments de conception Web standard dans le questionnaire;

⁶ Un essai préliminaire des formats (auprès de 21 000 ménages en avril 2002) a montré que le format matriciel restait préférable au format séquentiel du point de vue de la qualité, de la rapidité de réponse et de la convivialité, et qu'il satisfaisait aux exigences de la RIC.

- fourniture d'une extensibilité rapide;
- nécessité d'assurer la sécurité des transmissions de données aux répondants et de celles faites par ces derniers;
- nécessité d'une campagne de communication publique efficace en vue de promouvoir la « valeur ajoutée » qu'offre aux répondants l'utilisation de l'option Internet.

D'ici à 2006, il est prévu que de 60 % à 70 % de ménages soient raccordés à Internet que de 30 % à 40 % de ces ménages choisissent de répondre au questionnaire du recensement en ligne. Si ces hypothèses sont vérifiées, nous pourrions nous attendre à ce qu'environ 18 % de ménages répondent par Internet en 2006. Nos hypothèses budgétaires sont fondées sur ce taux projeté d'adoption. Il est important de noter qu'il faudrait que plus de 32 % des ménages répondent par Internet pour compenser les coûts de développement et d'intégration de l'application. Étant donné que les investissements nécessaires pour traiter la part prévue des réponses sur questionnaire imprimé doivent être faits d'avance, un taux d'adoption d'Internet nettement plus élevé ne donnera pas nécessairement lieu à une économie énorme.

Nombres de Canadiens et Canadiennes utilisent déjà Internet pour produire leur déclaration de revenus, pour faire leurs opérations bancaires et pour acheter des marchandises en ligne. En 2001, plus de 5,8 millions de ménages, soit 49 % d'entre eux, comptaient au moins un membre qui utilisait régulièrement Internet à la maison. Ce chiffre représente une hausse de 1,1 million (+23 %) par rapport à 2000. Ce progrès s'inscrit dans la foulée d'une croissance encore plus importante de 1,4 million (+42 %) de 1999 à 2000 (source : Enquête sur l'utilisation d'Internet par les ménages, SC). De plus en plus de Canadiens et de Canadiennes s'attendent à ce qu'on leur offre une option électronique pour la plupart des services gouvernementaux. L'Agence des douanes et du revenu du Canada (ADRC) a reçu plus de 10 millions (plus de 40 %) de déclarations de revenus sous forme électronique en 2003 (pour l'exercice 2002)⁷.

Les Canadiens et Canadiennes utilisent Internet parce que celui-ci leur offre certains avantages matériels, à savoir un remboursement plus rapide de la part de l'ADRC, plus de commodité et le choix des horaires pour les transactions bancaires et l'achat de marchandises. Le défi, dans le cas du recensement, est de trouver des moyens d'offrir des incitatifs comparables aux répondants.

Les progrès en matière de solutions de sécurité permettront à SC d'offrir une application robuste, conviviale, assurant un niveau de protection des données fournies par les répondants qu'il était difficile d'atteindre auparavant.

5.3 Développement et intégration de l'application

Conformément à la stratégie du recensement consistant à impartir les activités que le secteur privé a prouvé qu'il peut accomplir convenablement, le développement et la mise en œuvre de l'option Internet ont été inclus dans un contrat accordé à Lockheed Martin. Cette firme, qui est l'entrepreneur principal, a choisi IBM comme sous-traitant chargé d'établir une conception d'application en association avec SC. Ensuite, IBM développera et mettra en œuvre cette conception. En engageant un intégrateur d'application tel que Lockheed Martin, qui concevra et développera la majorité des applications de traitement, nous minimiserons les risques et les problèmes associés à l'intégration de données recueillies à partir de plusieurs circuits.

L'un des changements importants dans la méthode de collecte qui faciliterait l'intégration d'Internet est l'utilisation d'un registre de contrôle automatisé. L'existence d'une base de données comportant un paramètre fictif pour chaque logement avant le jour du recensement nous permettra d'imprimer un code d'accès à Internet unique sur chaque questionnaire⁸. Grâce à ce code d'accès, les répondants pourront être authentifiés, utiliser le logiciel de cryptage des données du recensement (qui offrira une connexion sécuritaire entre le navigateur Internet du répondant et les

⁷ Il est toutefois important de noter que l'ADRC offre un remboursement d'impôt plus rapide, s'il y a lieu, aux personnes qui produisent leur déclaration électroniquement. Dans le cas du recensement, par contre, il est impossible d'offrir ce genre d'encouragement matériel pour la déclaration électronique et il serait donc trompeur de faire des comparaisons avec les déclarations reçues en appliquant cette méthode.

⁸ Les questionnaires individuels et ceux destinés aux ménages vivant outre-mer n'auront pas de code d'accès à Internet pré-imprimé. Les destinataires de ces questionnaires pourront obtenir un code d'accès à Internet par l'entremise de l'Assistance téléphonique du Recensement.

serveurs du recensement) et obtenir le type de questionnaire approprié, et la Division du recensement pourra enregistrer automatiquement les réponses du ménage. Ceci nous permettra d'éliminer certains suivis inutiles, donc de réduire les coûts et le fardeau de réponse. Les réponses seront vérifiées durant leur saisie, ce qui limitera encore les suivis dus au rejet de questionnaires à la vérification. Les réponses électroniques seront introduites automatiquement dans le flot de données saisies et intégrées aux réponses saisies à partir des questionnaires imprimés.

5.4 Évaluation cognitive

Les travaux de SC en vue de concevoir une interface utilisateur réduisant au minimum les effets de mode, tout en offrant à l'utilisateur l'expérience positive nécessaire pour atteindre les niveaux d'adoption souhaités ont débuté par un petit essai visant 200 000 logements dans deux régions durant le Recensement de 2001. À ce moment-là, afin d'obtenir le niveau de sécurité souhaité pour les données des répondants, Statistique Canada a implémenté une application téléchargeable que les utilisateurs ont installée sur leur ordinateur. En plus des enseignements tirés de cet essai, SC a commandé une série d'essais qui ont été réalisés par des spécialistes du Center for Human-Computer Interaction de l'Université Carleton. Le « centre d'innovation » de Lockheed Martin/IBM, qui a conçu avec succès certains des sites Web les plus importants et les plus applaudis par les critiques du monde collabore avec SC à la conception de l'interface utilisateur.

Un autre aspect important de l'approche que nous avons adoptée pour concevoir l'interface de l'application est que nous avons intégré les bonnes idées de certains de nos collègues étrangers en nous efforçant d'éviter les embûches qu'ils ont connues. SC a entretenu un dialogue avec des représentants des bureaux de la statistique de la Suisse, de la Nouvelle-Zélande, de l'Australie, des États-Unis et de Singapour, et a partagé avec eux les meilleures pratiques et les enseignements tirés en ce qui concerne la collecte de données du recensement sur Internet. Afin de favoriser cette communication, Statistique Canada a organisé une conférence sur la collecte de données du recensement par Internet qui s'est tenue dans la région d'Ottawa en octobre 2003.

5.5 Sécurité

Parmi les raisons originales d'envisager de modifier la méthode de collecte de données du recensement figuraient les questions de la confidentialité (en particulier, les réserves quand à l'utilisation d'agents recenseurs locaux) et la sécurité des renseignements personnels. Ces questions préoccupent de plus en plus les Canadiens, qu'il s'agisse de leurs rapports avec des établissements publics ou privés. Les données que les répondants transmettront par Internet seront protégées par des méthodes approuvées, y compris l'authentification et le cryptage bidirectionnel « client à serveur et serveur à client » des transactions. Ces mesures de protection demandent l'exploitation de la « voie de communication protégée » de GED – établie d'après le concept de l'infrastructure à clé publique (ICP) – et la mise en œuvre de fonctions particulières, qui incluent l'utilisation de certificats anonymes d'usage limité, l'isolement du réseau interne de Statistique Canada grâce à l'utilisation du concept de la « lame d'air » et l'assurance que l'application et les services qui l'appuient auront la capacité requise pour traiter les volumes de pointe du recensement.

5.6 Fonctions de l'application

Afin d'atteindre les objectifs souhaités d'adoption, Statistique Canada devra encourager les répondants à utiliser l'application Internet. Les mesures prises incluront la promotion de l'option Internet sur la page de couverture du questionnaire, ainsi que des campagnes de communication publique et de marketing.

Le défi suivant consistera à s'assurer qu'un aussi grand nombre que possible de répondants qui essayeront l'application comprennent comment l'utiliser et jugent l'expérience suffisamment plaisante pour finir de répondre au questionnaire. À cette fin, nous avons inclus un certain nombre de fonctions destinées à accroître la convivialité de l'application. Par exemple, cette dernière comprendra un processus d'ouverture de séance simple, en une seule étape, durant lequel le répondant entrera un code figurant à la première page du questionnaire imprimé pour faire apparaître le type de formulaire approprié.

Les répondants qui doivent répondre au questionnaire « détaillé » pourront suspendre la séance et la reprendre plus tard là où ils l'ont interrompue. Ils pourront aussi reprendre la séance à partir d'une machine différente. Par exemple, ils pourraient commencer à répondre au questionnaire à la maison, arrêter la séance, puis la reprendre à partir d'un café Internet. L'application comportera des instructions « passez à » automatisées et des vérifications en direct, et utilisera les données entrées par le répondant pour personnaliser le questionnaire. Par exemple, si le répondant indique que trois personnes occupent le logement, le questionnaire créera trois espaces pour les réponses.

5.7 Impartition

Compte tenu de la nécessité de trouver une solution de remplacement rentable pour la saisie manuelle des données des questionnaires du recensement et de l'arrivée à maturité de la technologie de la reconnaissance intelligente de caractères, il a été décidé de faire appel à des spécialistes externes tout au début de la planification du Recensement de 2006. Le succès remporté par d'autres pays ayant utilisé des spécialistes externes pour la saisie des données et le traitement des questionnaires du recensement retourné (États-Unis, Royaume-Uni, Australie) a montré que le risque pouvait effectivement être transféré et géré par impartition. Nous avons établi un ensemble de critères d'impartition, dont le coût, la rapidité d'exécution, les risques, les effets à long terme sur la qualité et le transfert des connaissances, que nous avons appliqué à la gamme d'activités de développement des systèmes et de traitement requises pour 2006. Une démarcation nette a été établie afin que les répondants n'aient des contacts qu'avec les employés de Statistique Canada, malgré le fait que tous les employés contractuels doivent se soumettre à tous les contrôles de sécurité nécessaires et prêter serment aux termes de la *Loi sur la statistique*, donc sont sujets aux pénalités prévues par la loi pour la divulgation de toute information confidentielle. L'application des critères aux activités individuelles et au groupe d'activités interdépendantes nous a permis de définir un ensemble d'activité à impartir qui a été approuvé par la haute direction de Statistique Canada. Concomitamment, nous avons mené des consultations approfondies auprès du secteur privé afin d'évaluer sa capacité, ses connaissances et son expérience relatives aux diverses activités qu'il était envisagé d'impartir. La liste finale inclut :

- l'impression des questionnaires du recensement afin d'assurer que le type de papier, d'encre, de pliage et d'autres caractéristiques satisfassent aux exigences de la reconnaissance intelligente de caractères (RIC);
- la saisie automatisée des données pour tous les questionnaires imprimés;
- le codage de toutes les réponses en toutes lettres;
- les opérations du centre de traitement des données, y compris les installations et la gestion des ressources humaines*;
- le développement, le déploiement et la tenue à jour de l'application Internet, ainsi que de l'application d'interview téléphonique assistée par ordinateur (ITAO).

Ce groupe d'activités est intéressant sous plusieurs angles. L'industrie apprécie le fait qu'il s'agisse d'un ensemble varié d'activités interdépendantes qui lui permet de gérer les interfaces, donc de contrôler les risques d'intégration. L'ensemble est également intellectuellement stimulant, ce qui incite les chefs de file de l'industrie à former des consortia dont l'expertise est éprouvée.

La gestion des activités de pair avec le fournisseur externe a lieu dans un environnement de projet avec équipes de projets intégrées (représentation du fournisseur et de Statistique Canada). Le projet d'impartition compte trois phases distinctes. La phase I comprend une étape de conception détaillée qui permettra de comprendre clairement les exigences, et une étape de présentation et de description d'une solution détaillée décrivant la qualité, les risques, les coûts et les échéances. La phase II comprend le développement, la mise en œuvre et l'évaluation des systèmes et processus pour la répétition générale. La phase III inclut la modification, la mise à l'échelle et le déploiement des systèmes et processus nécessaires pour le Recensement de 2006. Le passage d'une phase du projet à la suivante ne peut se faire que selon un processus d'approbation officiel (auquel participent les organismes centraux du gouvernement du Canada).

Statistique Canada a entamé les consultations avec les représentants de l'industrie en 2001, a diffusé officiellement une demande de proposition (DP) en octobre 2002 et, après une période d'évaluation rigoureuse, a octroyé le contrat à Lockheed Martin à titre d'entrepreneur principal en février 2003. En juin 2003, nous avons reçu l'autorisation de procéder de la phase I à la phase II. Chaque phase du projet comporte un ensemble bien défini d'exigences à

* Cette partie du contrat est à l'étude à l'heure actuelle pour le Recensement de 2006.

satisfaire en respectant un budget fixe, avec des encouragements et des facteurs de dissuasion fondés sur des jalons particuliers de coût, de qualité et de respect des délais.

6. QUESTIONS EN SUSPENS

Pour chacun des trois aspects examinés dans le présent document, à savoir le contenu, l'utilisation d'Internet et l'impartition, plusieurs problèmes se posent encore.

Un certain nombre de questions du recensement, quoi qu'importantes et pertinentes pour certains groupes, semblent figurer davantage pour des raisons symboliques que pour répondre directement à des exigences stratégiques, juridiques ou de programmes. Vu que, dans certains cas, ces questions jouent un rôle de catalyseur permettant d'obtenir un taux relativement élevé d'intérêt, donc un taux de réponse élevé, comment devrait-on gérer les attentes des utilisateurs des données relativement aux exigences stratégiques/de programmes et au fardeau de réponse? En outre, au Canada et dans d'autres pays, comme l'Australie et la Nouvelle-Zélande, un recensement est réalisé tous les cinq ans. Alors que les systèmes statistiques et les utilisateurs de données ont adapté leurs besoins à cette fréquence historique, les organismes centralisés de financement exercent des pressions et créent des incertitudes qui augmentent le risque d'un programme d'une telle envergure. Quelles mesures pourraient, ou devraient, être prises pour faire concorder les exigences des utilisateurs de données avec les contraintes et les exigences financières imposées par les organismes centraux lors de chaque cycle de recensement, tout en introduisant des méthodes plus rentables de collecte et de traitement des données?

Étant donné l'initiative du Gouvernement en direct, l'établissement d'une option Internet soulève plusieurs questions. La fourniture d'une option en ligne sécurisée à chaque ménage du pays exige un investissement considérable dans le développement de systèmes, les campagnes de communication publique, le soutien en ligne et une infrastructure sécurisée. On peut soutenir que, même si le taux de réponse sur Internet est de 18 % en 2006, Statistique Canada ne récupérera pas l'investissement fait en vue de répondre à l'attente d'un mode de réponse en direct de la part du public. Des investissements lourds doivent être faits dans les systèmes de manutention et de traitement des questionnaires imprimés pour les 82 % prévus d'autres répondants. Quelles mesures peuvent être prises afin de maximiser l'investissement dans les systèmes électroniques ainsi que dans le traitement des questionnaires papier? En outre, une composante importante de l'investissement dans l'application Internet concerne le matériel et les logiciels requis pour traiter d'énormes volumes de réponses en une période très courte (volume de pointe prévu le jour du recensement). Aucun modèle de prédiction fiable n'existe pour établir une correspondance entre les profils de réponse et les investissements dans l'infrastructure. Quel devrait être le profil d'investissement dans la construction de l'infrastructure idéale par opposition à la mise en place d'une stratégie de communication publique ayant pour but d'uniformiser la courbe de réponse?

L'offre de services en ligne est un objectif clé du gouvernement du Canada, y compris le développement et la mise en place d'une infrastructure sécurisée. Alors que d'autres applications gouvernementales peuvent inviter les membres du public à réutiliser une clé publique/un certificat pour de multiples transactions, la nature infrequente du recensement, ainsi que les préoccupations concernant la confidentialité, empêchent la pleine utilisation de ce genre de disposition. À long terme, dans quelles mesures les bureaux de la statistique devraient-ils continuer de demander une différenciation claire au lieu de tirer profit des grandes tendances et des investissements du gouvernement?

Les efforts de l'Agence des douanes et du revenu du Canada en vue de convaincre les Canadiens et Canadiennes d'utiliser les applications et les outils en direct pour produire leur déclaration de revenus ont été couronnés de succès. L'un des principaux encouragements est la promesse d'un remboursement d'impôt plus rapide. Pour la saison fiscale 2002, plus de 10 millions de Canadiens et Canadiennes (40 %) ont produit leur déclaration de revenus électroniquement. Dans le cas du recensement, le but est d'atteindre un taux de réponse sur Internet en 2006 de 20 % à 25 %, mais il semble qu'un incitatif clairement articulé fasse défaut. Quels devraient être les messages clés communiqués pour convaincre les Canadiens et les Canadiennes de remplir le questionnaire du Recensement de 2006 en ligne?

L'offre de systèmes et de solutions produits à l'interne comporte une informalité implicite. Si le processus n'est pas entièrement discipliné et dépourvu d'échéances fixes pour les produits livrables, les délais courts sont plus facilement tolérés et les modifications, acceptées plus volontiers. Nous estimons que l'impartition des activités

principales de développement des systèmes et de traitement érodera considérablement toute informalité ou capacité de faire des changements à un stade avancé du processus sans encourir des coûts énormes. Nombres d'organismes statistiques ont remporté un succès relatif en engageant des partenaires externes pour réaliser un recensement. Alors que la capacité du secteur privé à répondre aux demandes spécifiques d'un recensement semble croître, celle des bureaux de la Statistique diminue ou reste non développée. Dans le long terme, quel devrait être le juste équilibre, en ce qui concerne tant la portée que le type, entre l'expertise interne et celle demandée du secteur privé? Les organismes statistiques devraient-ils travailler en collaboration afin d'assurer la durabilité à l'interne et à l'externe?

Les trois aspects du recensement discutés en profondeur plus haut ne représentent que quelques-uns des défis à relever afin d'apporter des changements fondamentaux à un programme d'une aussi grande envergure et complexité que le recensement. Les procédures sur le terrain doivent être remaniées, et l'expérience accumulée au fil des ans en ce qui a trait à l'organisation régionale du travail sur le terrain n'est pas forcément d'une pertinence directe dans le nouvel environnement. Les changements dans la technique de saisie des données, le suivi centralisé des questionnaires rejetés à la vérification et l'autocodage des réponses en toutes lettres sur l'industrie et la profession auront tous, sur le profil des erreurs dans les données, des répercussions qu'il faudra identifier et comprendre. Toutes ces considérations confirment la nécessité de procéder à une mise à l'essai et à une évaluation minutieuse, ainsi que le rôle important que jouera le test du Recensement de 2004 à cet égard.