

## Article

Symposium 2008 :  
Collecte des données : défis, réalisations et nouvelles orientations

### **Contrôle généralisé de la qualité de la saisie optique des données à Statistique Canada**

par Lampson Nguyen, Keith Davis, Cameron Oddy,  
Hélène St-Jean et Lillian Melki

2009



## **Contrôle généralisé de la qualité de la saisie optique des données à Statistique Canada**

Lampson Nguyen, Keith Davis, Cameron Oddy, Hélène St-Jean et Lillian Melki<sup>1</sup>

### **Résumé**

Statistique Canada a lancé un programme destiné à intensifier et à améliorer l'utilisation de la technologie d'imagerie pour le traitement des questionnaires d'enquête imprimés. Le but est d'en faire une méthode efficace, fiable et économique de saisie des données. L'objectif est de continuer de se servir de la reconnaissance optique de caractères (ROC) pour saisir les données des questionnaires, des documents et des télécopies reçus, tout en améliorant l'intégration du processus et l'assurance de la qualité/le contrôle de la qualité (CQ) du processus de saisie des données. Ces améliorations sont exposées dans le présent article.

Mots clés : Traitement des données, amélioration de la qualité du processus, contrôle de la qualité, reconnaissance optique des caractères.

### **1. Introduction**

Après le lancement concluant de la saisie des données généralisée par reconnaissance optique de caractères (ROC), le rôle important des procédures de contrôle de la qualité (CQ) dans le succès des opérations de saisie est devenu évident. Alors qu'elles fournissaient des renseignements précieux, les procédures de CQ (établies en dehors du système de saisie des données) ne faisaient pas partie intégrante du cycle de collecte, ce qui les rendait plus difficiles à tenir à jour et avait donc une incidence sur notre capacité de mettre en œuvre un programme de CQ complet.

Récemment, le secteur de programmes du bureau central de Statistique Canada responsable de la collecte des données a apporté des changements aux systèmes de CQ de ses opérations de saisie des données. Le défi consistait à personnaliser un logiciel du commerce, OCR for AnyDoc, y compris sa fonction de vérification Auditor existante, afin de développer un programme de CQ généralisé entièrement intégré qui répondait à nos besoins.

Le propos du présent article est de décrire les améliorations qui ont été apportées à la saisie des données et à l'approche de CQ générique, et de montrer les avantages et les incidences de l'utilisation de la ROC. Ces améliorations étaient axées sur les méthodes et procédures statistiques, tout en assurant la mise en place d'un processus de saisie des données offrant un niveau accru de fiabilité, de qualité, d'économie et de rapidité d'exécution.

### **2. Objectifs de qualité**

Les principaux objectifs de qualité du projet étaient les suivants : premièrement, continuer de mesurer, de contrôler et d'améliorer la qualité des opérations de saisie des données en continu; deuxièmement, accroître l'efficacité du processus de contrôle de la qualité afin de produire plus rapidement les résultats et de réduire au minimum la maintenance du système; troisièmement, améliorer l'environnement de développement en utilisant des modules logiciels préexistants et réduire la personnalisation des codes. Afin de réaliser ces objectifs, nous avons :

- a) accru l'efficacité de la stratégie d'échantillonnage pour le CQ et la représentativité des données obtenues grâce à l'utilisation des fonctions d'échantillonnage du système;

---

<sup>1</sup> Lampson Nguyen, Keith Davis, Cameron Oddy, Hélène St-Jean, Lillian Melki, Statistique Canada, Pré Tunney, Ottawa (Ontario) Canada, K1A 0T6

- b) remplacé les nombreuses procédures manuelles du processus de CQ par des procédés automatisés;
- c) modifié l'étape de vérification du CQ afin qu'elle soit intégrée dans le processus de saisie de production en utilisant une fonctionnalité logicielle intégrée.

La conception et les procédures du CQ ont été adaptées au logiciel AnyDoc en se servant des fonctions disponibles dans le module Auditor.

### **3. Système actuel**

Le processus de mise en œuvre de la saisie et du contrôle des données d'un nouveau questionnaire imprimé par la technologie d'imagerie ROC requiert les étapes préparatoires suivantes : la conception du questionnaire, l'élaboration d'un gabarit de questionnaire et la mise à l'essai du gabarit. Une fois ces étapes franchies, le processus de saisie des données et de CQ peut débuter. Les sections qui suivent décrivent ces étapes de manière plus détaillée dans le contexte du système actuel.

#### **3.1 Conception du questionnaire**

La clé du succès en ce qui concerne la qualité des données est de veiller à ce que le questionnaire imprimé soit conçu de manière optimale pour l'application de la technologie ROC. De nombreuses modifications simples peuvent être apportées à un questionnaire en vue d'améliorer significativement l'exactitude de la reconnaissance des données. Elles comprennent l'utilisation de codes à barres pour l'identification du document, la modification des champs d'entrée de données de manière à réduire au minimum la reconnaissance d'inscriptions manuelles (p. ex., utilisation de cases à cocher pour l'entrée de données), l'utilisation de couleurs pour séparer les données et l'impression de marques spéciales sur les questionnaires pour faciliter l'alignement de l'image. Statistique Canada a élaboré des lignes directrices pour la conception des questionnaires en vue d'optimiser l'exactitude de la ROC et cette documentation est fournie aux clients (en même temps que des commentaires concernant leurs ébauches de questionnaires) à cette étape.

#### **3.2 Élaboration d'un gabarit de questionnaire**

Dès que le questionnaire imprimé définitif, ainsi que les spécifications de saisie définitives ont été reçus, l'élaboration du gabarit de questionnaire débute. Ce gabarit est nécessaire, parce que le questionnaire à traiter est considéré comme un document structuré (le placement des données dans le questionnaire est statique) et, par conséquent, l'emplacement exact sur le questionnaire des champs où le système s'attend à extraire des données est connu. Ce gabarit, qui est basé sur chaque page en utilisant les images numérisées du questionnaire obtenues par balayage optique, requiert l'identification de toutes les données à extraire et de l'information clé dans chaque champ de données. Par exemple, l'emplacement du champ sur la page, le type de données (p. ex. numériques), l'ordre dans lequel la saisie et la sortie des données doivent avoir lieu, les contrôles s'appliquant au champ (p. ex. vérifier que la valeur est comprise dans une fourchette donnée) et les niveaux de confiance souhaités qui permettront au système de déterminer le degré de confiance avec lequel il arrive à interpréter les données. Durant l'élaboration du gabarit, toutes les consignes de CQ requises sont ajoutées, telles que les paramètres de CQ, la fonctionnalité d'échantillonnage et de vérification, ainsi que la sortie des statistiques de CQ et les résultats des décisions. Quand leur élaboration est terminée, tous les gabarits sont soumis à des essais approfondis.

#### **3.3 Saisie des données**

Une fois que l'enquête est placée dans le système de saisie des données pour la production, le processus débute par le balayage des questionnaires à mesure qu'ils sont reçus. Ce balayage produit des images numériques de toutes les pages du questionnaire. Le système se sert du gabarit qui a été élaboré pour repérer la zone de l'image où sont situés les champs de données et utilise son moteur de reconnaissance pour essayer d'interpréter les données. Durant ce processus, il évalue avec quel degré de confiance il est capable de déterminer correctement les valeurs des données (celles pour lesquelles le degré de confiance est faible sont souvent qualifiées de champs ou caractères « douteux »).

Les champs identifiés par le système de saisie des données rentrent dans deux catégories : les champs à saisie automatique des données (SAD) et les champs à saisie clavier à partir d'images (SCI).

Les champs SAD sont ceux dont les données sont saisies automatiquement par le système parce que le niveau de confiance qu'avait ce dernier quand il a essayé d'interpréter les données était élevé (supérieur à la valeur seuil préétablie). Dans la plupart des cas, ces champs sont des champs à reconnaissance optique de marques (ROM) (p. ex. cases à cocher).

Les champs SCI sont soumis à une opération de *saisie tête haute* (Heads up) des données exécutée par des opérateurs qui vérifient les champs sur un écran d'ordinateur. Ils sont de trois types : les champs présélectionnés pour la saisie clavier indépendamment du niveau de confiance du système (p. ex. champs clés/vérification à 100 %), les champs pour lesquels le niveau de confiance de la reconnaissance est inférieur au seuil préétabli pour le champ en question (c.-à-d. saisie SCI des données) et les champs pour lesquels le niveau de confiance qu'a le système dans son interprétation est élevé, mais qui sont rejetés à l'application d'une règle de contrôle particulière (p. ex. vérification de la fourchette de valeurs).

À chacune de ces étapes du processus de production, des erreurs peuvent être commises. Dans le cas de la saisie automatique des données des champs SAD, des *erreurs de substitution* peuvent survenir si le système interprète une autre valeur que la valeur réelle qui figure dans le champ. Ces erreurs sont graves, puisqu'elles peuvent être systématiques et avoir une incidence importante sur la qualité globale du processus de saisie des données. Il est important de pouvoir les dépister et de les éliminer pour essayer d'optimiser la qualité des données à la sortie. Dans le cas du processus de saisie tête haute des champs SCI, des *erreurs de saisie clavier* peuvent survenir si les opérateurs interprètent incorrectement les données et peuvent donner des niveaux élevés d'erreur (p. ex. erreur de saisie manuelle).

### **3.4 Contrôle de la qualité (CQ)**

Le processus de CQ a pour but de s'assurer d'obtenir un niveau donné de qualité durant le processus de collecte des données. Il comporte l'utilisation de mesures continues pour prendre des décisions concernant les processus, la résolution des problèmes observés et la comparaison à des normes de qualité. Les clients responsables des enquêtes demandent souvent d'obtenir l'assurance que certaines données recueillies seront entièrement exemptes d'erreurs. Ces champs de données particuliers sont appelés champs clés, puisqu'ils sont utilisés pour l'identification, l'agrégation ou des analyses essentielles. Dans le processus de vérification des données aux fins du CQ, tous ces champs clés sont soumis à une vérification totale (100 %) afin de s'assurer qu'ils sont exempts d'erreurs.

Par ailleurs, tous les champs de données qui ne sont pas considérés comme des champs clés sont échantillonnés, vérifiés et évalués pour confirmer que les erreurs de saisie clavier et les erreurs de substitution se situent dans les limites acceptables pour le questionnaire. Afin de respecter toutes ces exigences, une procédure de CQ en deux étapes a été mise en œuvre.

Durant l'exécution initiale de la saisie des données (étape 1), les opérateurs vérifient tous les champs clés et tous les champs douteux (p. ex. champs pour lesquels le degré de confiance avec lequel le système interprète les valeurs des données est faible). Durant la deuxième exécution (étape 2), nous procédons à la première étape de vérification de CQ des données. Un opérateur différent procède de nouveau à la saisie des champs clés sans voir les données saisies antérieurement, ce qui donne une vérification des entrées en double aveugle. Si les deux entrées consécutives sont identiques, la valeur est considérée comme étant correcte. S'il existe un écart entre les valeurs saisies à la première et à la deuxième exécution (étapes 1 et 2), les champs sont acheminés à la deuxième étape du CQ, où les valeurs sont examinées, confirmées ou corrigées.

Cette deuxième étape du CQ, également appelée phase d'audit, consiste à échantillonner et à vérifier tous les autres champs. Les procédures de CQ sont fondées sur l'approche de l'échantillonnage d'acceptation statistique en vertu de laquelle un échantillon distinct est requis pour surveiller indépendamment la qualité de la saisie des champs SAD et SCI.

Le taux d'échantillonnage des champs SAD est déterminé en se fondant sur l'enquête et ne varie pas. Le taux d'échantillonnage des champs SCI est déterminé en se fondant sur l'enquête, ainsi que sur les opérateurs individuels

et évolue au cours du cycle de l'enquête. Les taux d'erreur de substitution et de saisie clavier sont enregistrés pour les champs compris dans l'échantillon et sont utilisés aux fins suivantes :

1. établir, pour chaque opérateur, un taux d'échantillonnage qui peut être ajusté hebdomadairement en fonction de sa performance. Ces taux paramétrisés permettent une adaptation plus souple et plus rapide aux erreurs de saisie clavier;
2. procéder à une évaluation au moyen de cartes de contrôle en comparant les taux d'erreur à leurs limites supérieures de contrôle respectives (p. ex. niveaux de rejet). Ces limites sont calculées par le système en fonction de la cible de qualité espérée pour chaque enquête et de la taille de l'échantillon de champs sélectionné dans le lot. Une décision est alors prise quant à l'acceptation ou au rejet de l'échantillon comparativement à la norme de qualité prévue pour l'enquête.

#### **4. Inconvénients du système antérieur**

Le système de contrôle de la qualité antérieur présentait un certain nombre d'inconvénients importants qui ont rendu nécessaire un remaniement. Du point de vue de l'architecture et de l'implémentation, le système précédent était plus compliqué et requérait plus de ressources pour le développement, la mise à l'essai et la maintenance. Il nécessitait l'élaboration d'un deuxième ensemble de gabarits de questionnaire avec un codage personnalisé pour s'assurer que l'application de vérification de CQ saisisse de nouveau les champs requis pour le contrôle. L'élaboration de ces deux ensembles de gabarits pour chaque enquête nécessitait aussi que toute modification apportée aux gabarits ou aux scripts le soit dans les deux ensembles pour s'assurer de la cohérence des deux processus, puisqu'ils doivent être de conception identique pour que l'on puisse comparer exactement les données de CQ.

En outre, le processus de CQ tout entier constituait un système distinct. Les questionnaires devaient d'abord passer par les étapes de la saisie des données de production, puis être traités de nouveau (après une routine d'échantillonnage) par le système de CQ. La comparaison des valeurs des données n'était effectuée qu'après l'achèvement des processus de production et de CQ et la création des fichiers de sortie.

La stratégie d'échantillonnage utilisée pour le CQ était l'échantillonnage passe-lots avec vérification à 100 % (p. ex. chaque N° lot était sélectionné pour une vérification totale). Les décisions d'acceptation et de rejet étaient prises pour le lot échantillonné. Les lots rejetés entraînaient un examen manuel de tous les lots omis avant et après le lot rejeté, afin de déterminer s'il fallait les rejeter également.

Les clients recevaient les données du système de saisie des données de production, tandis que les données corrigées étaient sauvegardées dans un fichier de sortie distinct et n'étaient envoyées au client que sur demande. Les estimations de la qualité produites à partir du système ne permettaient de calculer que des estimations du taux d'erreur à l'entrée. Étant donné qu'aucune erreur découverte n'était corrigée, on estimait que ce taux d'erreur était le même que le taux d'erreur à la sortie.

Enfin, si à n'importe quel point, un lot devait être examiné, de nombreuses étapes manuelles étaient nécessaires pour dépister le problème, car les processus de saisie des données et de CQ n'étaient pas intégrés en un système complet.

#### **5. Mise en œuvre du contrôle de la qualité**

Après son rajustement à une version plus récente d'OCR for AnyDoc, notre logiciel comprenait une nouvelle fonction appelée module Auditor. Ce module offrait une couche supplémentaire d'examen de la qualité et la capacité de définir sélectivement la surveillance. Il apportait aussi une couche de base de données qui permettait le stockage des valeurs des données modifiées durant le processus de saisie des données.

Le module Auditor compte trois grandes composantes : les critères d'audit, la phase d'audit et l'enregistrement chronologique des données. Les critères d'audit sont les renseignements qui doivent être fournis pour l'étape d'audit. La phase d'audit est celle durant laquelle les champs de données marqués pour l'audit sont examinés et

corrigés au besoin. La composante d'enregistrement chronologique des données enregistre les données de tous les champs et traite l'information recueillie durant les étapes de saisie et d'audit des données.

## 5.1 Critères d'audit

Les critères d'audit consistent à préciser les taux d'échantillonnage des opérateurs au niveau de l'enquête et à spécifier quels champs de données doivent être vérifiés au niveau de la page. Pour la plupart des enquêtes, le plan d'échantillonnage d'acceptation comprend un échantillonnage par lot, selon lequel, pour chaque lot, on sélectionne aléatoirement un échantillon de champs pour l'inspection, par opposition à l'échantillonnage passe-lots, où un échantillon de lots est sélectionné sur la base d'une fréquence prédéterminée de saut de lots et où les lots restants ne subissent aucune inspection. L'échantillonnage par lot est préférable, parce que tous les lots sont uniques et que l'échantillonnage au niveau du champ produit un échantillon plus représentatif.

À l'heure actuelle, la fonctionnalité intégrée du module Auditor permet uniquement l'échantillonnage passe-lots ou l'échantillonnage systématique des documents (c.-à-d. chaque N° questionnaire). La personnalisation à l'aide de Visual Basic Scripting est par conséquent requise pour procéder à l'échantillonnage de lot. À cette fin, Statistique Canada a mis au point trois méthodes de contrôle de l'échantillonnage des champs : les drapeaux de vérification et de contrôle; les tableaux propres à chaque enquête et l'échantillonnage de CQ.

### 5.1.1 Drapeaux de vérification et de contrôle

Les drapeaux de vérification et de contrôle sont des champs supplémentaires ajoutés au gabarit de document pour dépister divers indicateurs utilisés dans le processus de CQ.

Les drapeaux de vérification dépistent trois attributs clés par champ : la phase, la sélection de l'échantillon et la valeur du champ. Un drapeau de vérification est attribué à chaque combinaison de ces attributs comme l'illustre le tableau 5.1.1-1.

**Tableau 5.1.1-1**  
**Drapeaux de vérification**

Drapeau de vérification	Phase	Sélection de l'échantillon	Valeur du champ à l'étape
0	Extraction	SAD	Non en blanc
1	Étape 1	SCI	Non en blanc
2	Extraction	SAD	En blanc
3	Étape 1	SCI	En blanc
4	Étape 2	SCI	Non en blanc
5	Étape 2	SCI	En blanc
6	Auditor	SAD	Non en blanc
7	Auditor	SCI	Non en blanc
8	Auditor	SAD	En blanc
9	Auditor	SCI	En blanc

Les drapeaux de contrôle attribués à chaque champ déterminent l'action que doit exécuter l'application, comme il est illustré au tableau 5.1.1-2.

**Tableau 5.1.1-2**  
**Drapeaux de contrôle**

<b>Drapeau de contrôle</b>	<b>Type d'action</b>	<b>Description</b>
0	Dans l'échantillon	Le champ peut être échantillonné s'il est sélectionné.
2	Champ clé	Le champ sera vérifié à 100 % (champ clé).
3	Champ non admissible	Le champ ne peut pas être échantillonné (p. ex. champs de commentaire)
4	Ignorer le champ	Le champ ne sera vérifié au cours d'aucune étape (champs internes)

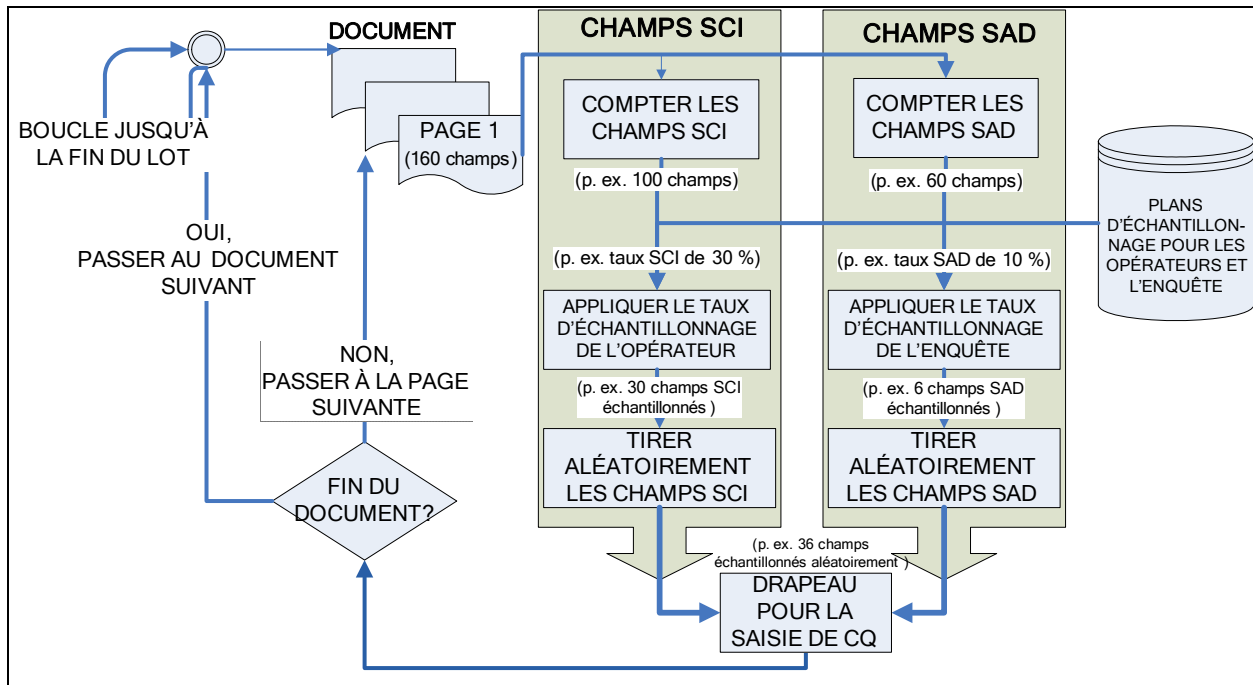
### **5.1.2 Tableaux propres aux enquêtes**

Les tableaux propres aux enquêtes sont sauvegardés dans une base de données et contiennent les données de CQ relatives aux enquêtes individuelles. Les taux d'erreur cibles pour le CQ spécifiés dans les tableaux pour la SAD et la SCI sont utilisés comme ligne médiane dans l'intervalle de plus ou moins trois écarts-types pour déterminer si un taux d'erreur de lot est accepté ou rejeté. En outre, les tableaux contiennent les taux d'échantillonnage applicables aux opérateurs pour l'enquête en question qui sont consultés durant le processus d'échantillonnage.

### **5.1.3 Échantillonnage de CQ**

Après l'étape de vérification, le processus de CQ consiste à échantillonner aléatoirement un certain nombre de champs SAD et SCI par page en fonction des taux d'échantillonnage SAD et SCI respectifs mémorisés dans les tableaux propres à l'enquête. Seuls les champs dont le drapeau de contrôle est égal à 0 (Dans l'échantillon) sont pris en considération pour le tirage de l'échantillon. Les champs marqués pour l'audit seront montrés à un opérateur durant l'étape d'audit. Ce processus d'échantillonnage de CQ, qui s'appuie sur la méthode d'échantillonnage de Bernoulli, est illustré au diagramme 5.1.3-1 qui contient des exemples par souci de clarté.

**Diagramme 5.1.3-1**  
**Échantillonnage pour le contrôle de la qualité**



## 5.2 Phase d'audit

La phase d'audit est celle durant laquelle les champs marqués pour l'audit sont effectivement examinés et corrigés, au besoin. Une fois que l'échantillonnage de CQ a été effectué et que la deuxième étape pour la vérification à 100 % des champs clés est achevée, l'étape d'audit commence et les champs SAD et SCI échantillonnés sont affichés sur l'écran d'un opérateur afin qu'il les vérifie. En outre, tout champ clé pour lequel existe un écart entre les valeurs de la première et deuxième étape est également vérifié une troisième fois durant cette phase d'audit.

## 5.3 Enregistrement chronologique des données et système de production de statistiques de CQ

Une fois que la phase d'audit d'un lot est terminée, celui-ci est sauvegardé et les données d'audit sont stockées dans une base de données. Une analyse plus approfondie est exécutée par le système de production de statistiques de CQ, une application développée par Statistique Canada, qui analyse les champs vérifiés, et extrait et totalise les statistiques de CQ afin de déterminer si le lot est accepté ou rejeté. En cas de rejet, les données concernant le lot sont supprimées et la saisie des données du lot est recommencée.

Les fonctions et capacités du système sont les suivantes :

- capacité de distinguer les champs SAD des champs SCI (et de les compter), ainsi que leurs tailles respectives d'échantillon pour chaque lot;
- capacité d'identifier et de compter tout champ comme un champ clé pour une vérification à 100 %;
- consignation complète des statistiques de CQ par lot pour tous les groupes de champs (p. ex. champs SAD, SCI et clés);
- possibilité de spécifier les cibles d'erreur pour l'évaluation de la qualité pour chaque enquête en fonction de la combinaison attendue de types de champs sur lesquels l'application effectuera la saisie;



- module de tirage d'échantillons qui permet de spécifier les taux d'échantillonnage souhaités selon l'opérateur et l'enquête pour l'échantillonnage des champs SCI, et selon l'enquête pour l'échantillonnage des champs SAD;
- capacité de choisir divers modes d'échantillonnage selon les champs SAD et SCI qui sont en blanc ou non en blanc;
- comparaison des données issues de la phase d'audit du processus de CQ aux données de saisie issues des étapes 1 et 2, et calcul des taux d'erreurs de substitution et de saisie clavier par lot;
- calcul des limites supérieures de contrôle par lot et analyse en ligne des cartes de contrôle pour repérer les échantillons qui ont été rejetés et requièrent un examen, et production quotidienne de la liste des lots rejetés;
- création de rapports contenant des renseignements administratifs et des renseignements statistiques de CQ par lot et par groupe de champs aux fins d'analyses supplémentaires. Ces sorties sont utilisées comme entrées dans un système distinct développé par Statistique Canada, le Système d'analyse des données de contrôle de la qualité (SADCQ ou QCDAS en anglais), qui les analyse et produit des estimations et des rapports de rétroaction (remontant à la source d'erreur) à l'intention du secteur des opérations.

## 6. Avantages du nouveau module de vérification Auditor

L'implémentation d'une nouvelle version d'OCR for AnyDoc en utilisant la fonction Auditor Module, ainsi que le développement des propres outils de Statistique Canada et la personnalisation ont permis d'améliorer le processus global de CQ. Les progrès les plus appréciables sont l'amélioration de la stratégie d'échantillonnage, la réduction des ressources de développement nécessaires, l'amélioration de la vérification du CQ et l'amélioration de la qualité des données et des estimations de la qualité.

Aujourd'hui, le système en place a la souplesse voulue pour traiter diverses unités d'échantillonnage, comme l'échantillonnage au niveau du document (questionnaire), du gabarit et (ou) du champ. En pratique, nous échantillonnons au niveau du champ dans chaque document et gabarit pour toutes nos enquêtes. Nous pouvons ainsi utiliser de meilleurs plans d'échantillonnage de CQ qui nous permettent de répondre aux divers besoins des clients et qui, en bout de ligne, se concrétisent par un meilleur contrôle moyennant des ressources supplémentaires minimales.

Du point de vue de l'élaboration des gabarits de questionnaire, les améliorations sont considérables. À l'heure actuelle, un seul jeu de gabarits est requis pour la production ainsi que le CQ, ce qui réduit considérablement l'effort et les coûts de développement et de maintenance, sans parler du risque d'erreur. Les modifications d'un gabarit ne doivent être faites qu'à un seul endroit, ce qui réduit sensiblement le temps global de développement.

Le nouveau système est moins complexe et plus facile à mettre en œuvre et à maintenir. Il peut être appliqué à tout type d'enquête, indépendamment des contraintes de taille ou de temps. En outre, l'utilisation de mesures statistiques paramétriques permet la prise de décisions objectives quant aux améliorations de la qualité du processus qu'il convient d'apporter dynamiquement tout au long du cycle de production.

Le processus de vérification du CQ est maintenant intégré au système de production sous forme d'un processus complet. Le fait que le processus de CQ soit en ligne permet au système de comparer les valeurs de CQ aux valeurs de production et de déterminer immédiatement si des écarts existent. Cette comparaison en ligne permet à l'opérateur de CQ d'examiner les valeurs entrées durant l'étape de saisie, de corriger ses erreurs ou de garder les valeurs saisies au départ. Cette fonction d'autocorrection élimine les erreurs faites par l'opérateur de CQ, si bien que seules les erreurs réelles sont enregistrées. Cette amélioration a réduit considérablement le nombre de divergences (p. ex. coquilles et erreurs d'interprétation) entre les valeurs de production et de CQ, le nombre de faux rejets de lot, ce qui à son tour a réduit la quantité de ressources requises pour examiner les lots rejetés.

De surcroît, les avantages s'étendent à la qualité des données des clients et aux estimations de la qualité produites. Le système actuel remplace les données de production par les données corrigées d'après le CQ. Les clients reçoivent maintenant toutes les corrections pour les données soumises au processus de CQ. Il est donc possible d'estimer les taux d'erreurs à l'entrée et à la sortie pour tout client. La souplesse et l'amélioration de la stratégie d'échantillonnage de CQ, du processus de décision d'acceptation/rejet et de la vérification de CQ en ligne ont eu un effet favorable sur

les estimations de la qualité, car le processus de CQ permet maintenant de produire le compte des erreurs réelles par opposition aux écarts produits par le système précédent.

## **7. Conclusion**

Nous avons réussi à mettre en œuvre le processus de CQ décrit en tant que méthode principale de CQ du processus de saisie des données des enquêtes à questionnaire imprimé. Cela nous a permis d'accroître la rapidité d'exécution du processus et d'améliorer les données transmises à nos clients. Dans le système précédent, il fallait deux à trois jours pour qu'un lot de questionnaires soit traité et que les données soient transmises aux clients. Grâce au processus amélioré, ce délai est réduit à environ 24 heures (1 jour). Bien que le système courant représente un progrès considérable, nous attendons que le fabricant du logiciel AnyDoc réponde à un certain nombre de demandes d'amélioration que nous lui avons soumises. Nous continuons d'entretenir une relation de travail avec ce fabricant afin de développer un programme de CQ idéal, comprenant par exemple l'intégration de la fonctionnalité que nous avons développée dans le système de production de statistique de CQ à titre de fonction possible dans de futures versions de AnyDoc. Nous estimons que la conception de notre processus de CQ est solide et nous espérons arriver à convaincre la firme AnyDoc d'adopter notre méthode de CQ dans son logiciel.

## **Bibliographie**

Mudryk, W. et Xie, H. (2004). Generalized Quality Control Approach for ICR Data Capture in Statistics Canada's Centralized Operations, *2004 European Conference on Quality and Methodology in Official Statistics*. Wiesbaden, Allemagne.