

Contrôle automatisé de la qualité des données provenant de dossiers administratifs

JAMES R. JONAS et PAUL S. HANCZARYK¹

RÉSUMÉ

Aux fins de son programme de la statistique économique, le Census Bureau fait grande utilisation des dossiers administratifs. Le nombre de dossiers traités chaque année est considérable, mais il l'est encore plus durant les années de recensement. Pour assurer la qualité de toutes les données qu'il reçoit, le Census Bureau a mis au point des programmes de contrôle sur gros ordinateur. Toutefois, étant donné le très grand nombre de tableaux de vérification requis, et compte tenu du fait que les ressources nécessaires à la programmation sont limitées et coûteuses, il serait difficile d'adopter un système exhaustif de contrôle exploité entièrement sur gros ordinateur. En outre, vu le caractère confidentiel des données et les conséquences possibles des erreurs, le Census Bureau a conclu qu'il fallait mettre au point un système de contrôle plus «sophistiqué», qui s'appuie sur la technologie des micro-ordinateurs. La division des enquêtes économiques est en train de mettre au point un tel système qui sera utilisé pour traiter les fichiers de données administratives de 1987. Le système intègre la technologie des gros ordinateurs à celle des micro-ordinateurs. Les fichiers de données administratives qui arrivent chaque semaine au Bureau sont d'abord traités sur gros ordinateur. Les données ainsi traitées sont transférées sur micro-ordinateur et sont formatées pour être exploitées par un tableur. La vérification systématique de la qualité des données (examen de données, détection des erreurs, production de rapports) est faite sur micro-ordinateur, par le tableur. L'utilisation combinée d'un gros ordinateur et de micro-ordinateurs permet d'alléger la charge des programmeurs, de faciliter le travail des analystes (plus de souplesse) et de réduire les coûts du traitement sur gros ordinateur. Enfin, la composante «contrôle-vérification» du système assure une vérification rigoureuse et exhaustive des données.

MOTS CLÉS: Intégration de la technologie des micro-ordinateurs et des gros ordinateurs; vérification systématique des données; actualité des données.

1. INTRODUCTION

Aux fins de son programme de la statistique économique, le *Bureau of the Census* fait grande utilisation des données administratives, en particulier des données provenant des dossiers des déclarations d'impôt sur le revenu des entreprises du *Internal Revenue Service* (IRS) et, dans une moindre mesure, des dossiers de la *Social Security Administration* (SSA). Dans les années où ont lieu le recensement agricole et le recensement économique, le Census Bureau reçoit une quantité beaucoup plus considérable de données administratives. Ces données nous permettent de mener nos recensements économique et agricole de manière efficace et dans les meilleurs délais de même que de réduire au maximum le fardeau de réponse des entreprises et des exploitants agricoles. La réussite de nos programmes de la statistique économique et de la statistique agricole dépend, dans une large mesure, de la qualité et de l'actualité de ces données administratives.

Il est indispensable que le Bureau vérifie la qualité de toutes les données qu'il reçoit. À cette fin, nous avons mis au point et utilisé au cours des derniers recensements économiques des programmes de contrôle sur gros ordinateur. Toutefois, étant donné le très grand nombre de tableaux de vérification requis et compte tenu du fait que les ressources nécessaires à la

¹ James R. Jonas et Paul S. Hanczaryk, Economic Surveys Division, U.S. Bureau of the Census, Washington, D.C., U.S.A., 20233.

programmation sont limitées et coûteuses, il serait difficile d'adopter un système exhaustif de contrôle exploité entièrement sur gros ordinateur. En outre, vu le caractère confidentiel des données et les conséquences possibles des erreurs, nous avons conclu qu'il fallait mettre au point un système de contrôle plus «sophistiqué». Le Census Bureau a réussi à élaborer un système complet pour gérer les différentes étapes du processus de vérification des données administratives. Ce système informatisé nous permettra d'effectuer une vérification plus approfondie des données malgré les contraintes budgétaires qui nous sont imposées et les ressources de programmation limitées dont nous disposons.

Le système intègre la technologie des gros ordinateurs à celle des micro-ordinateurs. Nous avons défini des critères ou exigences de base quant au genre de données administratives que nous désirons obtenir. Nous stockons ces critères dans la mémoire des micro-ordinateurs. Après exécution du programme de contrôle de la qualité sur gros ordinateur, nous faisons transférer les résultats sur les micros-ordinateurs. Les modes de déclaration des données sont ensuite comparés à nos exigences. La vérification proprement dite (examen des données, détection des erreurs, production de rapports) est faite sur micro-ordinateur. L'utilisation des micros-ordinateurs par opposition au recours exclusif à un gros ordinateur, permet d'alléger la charge des programmeurs, de faciliter le travail des analystes (plus de souplesse) et de réduire les coûts de traitement sur gros ordinateur. En outre, la composante «contrôle-vérification» du système assure une vérification rigoureuse et exhaustive des données. Enfin, bien que ce système ait été conçu spécialement pour la vérification des données provenant des déclarations d'impôt sur le revenu des entreprises aux fins du recensement économique, il peut être modifié (et il le sera après 1988) pour effectuer la vérification de toutes les données administratives que nous recevons.

2. APERÇU DU SYSTÈME DANS UNE PERSPECTIVE DE GESTION

Le Census Bureau a beaucoup recours aux dossiers administratifs pour produire des statistiques. En fait, l'importance de cette source de données n'a cessé de croître avec les années. Cet état de choses tient à la nécessité de produire des statistiques plus nombreuses et de meilleure qualité, de réduire au maximum le fardeau de réponse des entreprises privées et d'optimiser l'utilisation de nos ressources humaines et financières.

Ces dernières années, les données provenant des dossiers administratifs ont été, dans l'ensemble, jugées d'excellente qualité. Toutefois, les données tirées des déclarations d'impôt des entreprises qui nous ont été fournies en 1982 par l'IRS nous ont causé certains problèmes. Nous pensons, en particulier, à la qualité des principaux codes d'activité économique pour les entreprises à propriétaire unique. C'est en raison de ce problème, qu'au recensement de 1982, le Census Bureau n'a pu produire qu'une quantité limitée de statistiques sur les non-employeurs. Si nos programmes de contrôle de la qualité avaient été plus perfectionnés, nous aurions pu cerner plus rapidement les erreurs et en réduire les effets.

Lorsqu'est venu le moment de préparer le recensement de 1987, nous nous sommes rendu compte qu'il fallait prendre des mesures supplémentaires pour assurer la qualité des données administratives fournies par l'IRS. Ce dont nous avons besoin, c'était d'un système global de contrôle nous permettant de composer avec les trois principaux facteurs qui, dans le passé, ont compromis la qualité de ensembles de données administratives. Ces trois sources de problèmes sont les suivantes:

1. Volume élevé des données administratives

L'IRS nous fournira des enregistrements provenant des déclarations d'impôt produites par les entreprises pour 1987 (reçues en 1988). Ces enregistrements sont tirés des déclarations d'entreprises ayant différentes formes juridiques, notamment, des sociétés de capitaux, des sociétés dites «S Corporations» (petites sociétés), des sociétés étrangères, des sociétés de personnes, des entreprises à but non lucratif et des entreprises à propriétaire unique. En 1988, le Census Bureau prévoit recevoir au-delà de 75 millions d'enregistrements

Tableau 1

Nombre approximatif d'enregistrements tirés de dossiers administratifs utilisés aux recensements économique et agricole de 1987, selon le genre de formule et l'année d'imposition

Genre de fichier	Nombre d'enregistrements		
	1985	1986	1987
Déclaration d'impôt sur le revenu des entreprises:	2,617,000	20,051,000	30,881,000
Formule 1040, annexe C	—	11,750,000	12,500,000
Formule 1040, annexe F	2,450,000	2,450,000	—
Formule 1040, annexe SE	—	—	10,000,000
Formule 1120	42,000	2,550,000	2,650,000
Formule 1120-A	—	200,000	210,000
Formule 1120F	—	11,000	11,000
Formule 1120S	17,000	900,000	950,000
Formule 1065	108,000	1,750,000	1,800,000
Formule 990	—	380,000	400,000
Formule 990-PF	—	35,000	35,000
Formule 990-T	—	25,000	25,000
Formule 1120S, annexe K-1	—	—	700,000
Formule 1065, annexe K-1	—	—	1,600,000
Fichiers des données fiscales (données annuelles):	41,950,000	43,500,000	45,050,000
IRS Fichier principal des entreprises	24,000,000	25,000,000	26,000,000
IRS Fichier de la paye et du personnel	17,000,000	17,500,000	18,000,000
SSA Fichier des nouvelles entreprises	950,000	1,000,000	1,050,000
Total	44,567,000	63,551,000	75,931,000

au total. On trouvera dans le tableau 1 le nombre approximatif d'enregistrements qui seront tirés, aux fins des recensements économique et agricole de 1987, des diverses formules d'impôt. Il est clair que le nombre d'enregistrements que nous recevons durant une année de recensement est très élevé. Toutefois, la raison pour laquelle nous avons envisagé de mettre au point un système de contrôle de la qualité aussi perfectionné ne tient pas uniquement à la quantité des données à traiter mais aussi à la complexité du traitement. Un enregistrement contient souvent plusieurs données élémentaires et chacune d'elles vient accroître la complexité (niveau de détail) de l'enregistrement lui-même et de l'ensemble du fichier. En outre, les différentes formules d'impôt ne contiennent pas toutes les mêmes éléments d'information et les revenus n'y sont pas tous déclarés de la même façon. Par conséquent, non seulement faut-il vérifier la qualité de plus de 75 millions d'enregistrements mais aussi la qualité des données élémentaires qu'ils contiennent. Soulignons enfin que les entreprises transmettent leurs déclarations d'impôt à l'un des dix centres de traitement de l'IRS. Chaque centre assure le traitement des déclarations qu'il reçoit et la qualité des données qui nous sont transmises peut varier. Le Bureau vérifie donc individuellement la qualité des données transmises par chacun d'eux.

2. Contraintes budgétaires

Les contraintes budgétaires sont un autre facteur important qui vient accroître la difficulté du contrôle de la qualité des données administratives. Compte tenu de la politique générale de restriction des dépenses du gouvernement, le Bureau essaie d'offrir des services plus nombreux à un moindre coût. La charge de travail en programmation est beaucoup plus lourde pendant les années de recensement, mais le nombre des programmeurs n'est pas augmenté proportionnellement. La vérification de la qualité, qui dépend dans une large mesure des diverses ressources informatiques existantes, peut en souffrir. Il faut également souligner que jusqu'à maintenant, la plupart des tâches de vérification de la qualité

ont été faites sur gros ordinateur. L'utilisation de l'ordinateur central coûte cher, d'autant plus que les fichiers sont de plus en plus importants.

3. Manque de communication entre les organismes

Par le passé, le manque de communication entre les organismes intéressés a nui à la qualité des données. Pour corriger ce problème, il était indispensable de définir clairement les axes de communication devant exister entre le Bureau et l'organisme qui lui fournit les données, et cela, pour chacune des étapes du travail. Le Bureau doit d'abord définir les fichiers de données et les enregistrements dont il a besoin puis s'entendre avec l'organisme sur les données qui peuvent lui être fournies. Certaines des données demandées ne sont peut-être pas disponibles ou encore peuvent être trop coûteuses à produire. Les difficultés doivent être aplanies au fur et à mesure car des retards dans la transmission des données pourraient compromettre l'utilité. En outre, les organismes doivent s'entendre sur la qualité et la quantité des données désirées. Les exigences du Bureau touchant les données requises doivent être clairement quantifiées.

C'est pour apporter une solution globale au problème posé dans le passé par la qualité des données administratives que le Bureau a décidé d'élaborer et de mettre en oeuvre son système informatisé de contrôle de la qualité: le système permet de vérifier des fichiers importants et complexes de l'IRS, favorise les échanges et permet la détection immédiate des erreurs. La composante «vérification informatisée» est la plus importante du système. En gros, le Bureau définit les critères ou exigences de base auxquels doivent répondre les données fournies par l'IRS. Il compare ensuite les modes de déclaration aux critères établis puis procède, sur micro-ordinateur, à la vérification systématique des données. Le Bureau produit ensuite un rapport de la situation dans lequel il indique si les données répondent ou non aux exigences.

Le personnel du Bureau définit ses exigences bien avant que les données ne lui soient transmises. L'IRS dispose donc de tout le temps voulu pour déterminer s'il peut raisonnablement fournir les données demandées et, au besoin, pour modifier les demandes. Les exigences définies par le Bureau portent sur les deux points suivants: fréquence de transmission et qualité des données. Pour ce qui est de la fréquence de transmission, le Bureau établit une estimation du nombre total de déclarations qu'il souhaite obtenir pour chaque type d'entreprise avec la date à laquelle il désire les recevoir. Pour ce qui est des exigences qualitatives, le Bureau définit les modes de déclaration désirés pour chaque type d'élément d'information.

Le système de vérification automatisé facilite l'analyse des données. Une série de tableaux de résultats sont produits afin de comparer les données fournies aux critères établis. Des indicateurs permettent de relever les données qui ne satisfont pas aux exigences. Cette façon de procéder réduit les risques d'omission pendant le processus d'analyse.

Chaque mois, le Bureau envoie à l'IRS des rapports de la situation dans lesquels les modes de déclaration des données sont comparés aux critères établis. Ces rapports constituent des sous-ensembles des tableaux détaillés des résultats et ne contiennent que les exigences de base pour les ensembles de données fournis par l'IRS. Ces rapports favorisent la communication, entre le Bureau et l'IRS. Les données posant des problèmes y sont indiquées et le Bureau et l'IRS doivent décider immédiatement des mesures correctives ou des mesures de récupération à prendre afin d'éviter de compromettre les résultats du recensement. À cet égard, il est indispensable que les mesures nécessaires soient prises dans les plus brefs délais parce que l'IRS ne conserve pas ses bandes de données indéfiniment. Si les erreurs ne sont pas décelées rapidement et si les mesures correctives ne sont pas prises à temps, la récupération risque d'être impossible ou encore d'être extrêmement coûteuse.

Le système de contrôle de la qualité ne nous permet pas de nous assurer que les données administratives ne poseront plus jamais de difficultés. Toutefois, il nous permet de définir très clairement nos exigences, de sorte que les caractéristiques des ensembles de données ne sont pas laissées au hasard. Il nous permet par ailleurs de déceler les erreurs le plus rapidement possible.

Tableau 2
Répartition pondérée des formules 1040,
annexe C selon le centre de traitement et la tranche de revenu net¹

Centre de traitement	Total	Tranche de revenu net (000)				
		< 0	en blanc ou 0	1— 2,499	2,500— 4,999	5,000— 9,999
Ensemble des centres	1,327,100	200	52,200	149,300	73,900	98,100
Atlanta	133,200	0	5,100	16,500	6,300	11,000
Philadelphie	132,100	100	4,200	11,300	5,300	9,600
Austin	147,600	0	6,300	20,900	9,900	12,900
Cincinnati	153,100	0	5,300	14,900	8,700	9,800
Kansas City	119,500	0	5,500	16,700	7,500	8,500
Andover	111,100	0	3,800	9,800	6,700	8,200
Ogden	162,300	0	7,500	20,200	7,900	11,600
Brookhaven	119,700	0	4,400	12,600	7,100	10,000
Memphis	111,900	100	4,700	14,700	6,700	8,600
Fresno	136,500	0	5,400	11,700	7,800	7,900
Autres	100	0	0	0	0	0

¹ Revenus bruts moins les rendus et rabais sur les ventes.

Centre de traitement	Tranche de revenu net (000)					
	10,000— 24,999	25,000— 49,999	50,000— 99,999	100,000— 249,999	250,000— 499,999	500,000 +
Ensemble des centres	168,600	185,500	225,100	243,400	87,400	43,400
Atlanta	17,000	19,800	22,200	22,200	8,400	4,700
Philadelphie	17,800	19,800	22,700	27,000	10,100	4,200
Austin	18,700	18,500	22,000	24,900	9,100	4,400
Cincinnati	20,500	20,700	27,300	30,500	9,600	5,800
Kansas City	16,200	15,900	20,700	18,300	6,400	3,800
Andover	13,600	16,700	19,500	20,000	8,800	4,000
Ogden	17,800	19,500	28,800	33,600	11,200	4,200
Brookhaven	16,400	19,700	20,400	19,400	6,400	3,300
Memphis	15,100	14,700	18,600	19,000	6,800	2,900
Fresno	15,500	20,200	22,900	28,400	10,600	6,100
Autres	0	0	0	100	0	0

3. PARTICULARITÉS DU SYSTÈME

Les fichiers de données que le Bureau reçoit chaque semaine sont d'abord soumis à un contrôle qualitatif sur gros ordinateur. Les programmes utilisés sont mis au point bien avant l'arrivée des fichiers et ils servent à produire les premiers tableaux de contrôle de la qualité, qui constituent la pierre angulaire de tout le système. Auparavant, les programmeurs travaillant sur gros ordinateur étaient chargés de créer les tableaux dans leur ensemble, ce qui comprend les cellules de données et le texte qui s'y rapporte (en-têtes et colonnes principales). Toutefois, dans les programmes de production de tableaux pour le recensement de 1987, ces deux composantes seront traitées séparément. La production des tableaux continuera de se faire sur gros ordinateur tandis que le texte accompagnant les tableaux sera produit à l'aide de micro-ordinateurs par des employés qui ne sont pas des spécialistes de la programmation. Une procédure a été mise au point pour permettre la production de tableaux pour tous les fichiers de

données administratives. Cette procédure a permis au Census Bureau de concevoir un programme pour micro-ordinateur permettant de construire des images de tableaux pour n'importe quel genre de fichier de données administratives. Une fois achevées, les images de tableaux sont transférées au gros ordinateur et sont utilisées par les programmeurs pour aligner les données contenues dans les fichiers. Le travail de programmation qu'exigent les tableaux de contrôle de la qualité est de beaucoup simplifié du fait que la construction des images de tableaux est effectuée par des non-spécialistes. Les programmeurs travaillant sur gros ordinateur peuvent donc concentrer leurs efforts uniquement sur l'établissement des totalisations. Le tableau 2 est un exemple des divers tableaux produits sur gros ordinateur pour chaque type d'entreprise. Ce tableau fait voir la répartition pondérée des formules 1040, annexe C, selon le centre de traitement et la tranche de revenu net.

Le gros ordinateur n'effectue que les totalisations de base (mise à jour des totalisations). Ces dernières sont ensuite transférées sur des micro-ordinateurs où toutes les autres opérations de vérification sont effectuées. Parmi celles-ci, mentionnons le calcul des pourcentages dans les nouvelles totalisations, la production de totalisations cumulatives, la vérification de données clés et la production de rapports sur l'état d'avancement des travaux de vérification. Cette méthode de traitement systématique, qui s'appuie essentiellement sur l'utilisation de micro-ordinateurs, offre plus de souplesse pour l'examen des données et permet de réduire la charge de travail des programmeurs.

Les totalisations produites sur gros ordinateur sont transférées dans un tableur préformaté exploité sur micro-ordinateur. Ce tableur contient également les critères définis par le Census Bureau relativement aux ensembles de données administratives. Le micro-ordinateur effectue automatiquement la vérification des données du tableau. Les écarts entre les données et les exigences établies sont indiqués dans les tableaux de résultats. Les deux principaux avantages du système de vérification sont les suivants:

1. Il nous permet de déceler aisément les problèmes. Les données qui ne satisfont pas aux normes sont signalées par un indicateur afin qu'un analyste les examine. On risque donc moins d'oublier des erreurs.
2. Il attire notre attention sur des données qu'il faudrait examiner de plus près. Souvent, bien que toutes nos exigences aient été remplies, les tableaux de résultats nous mettent sur la piste de certains problèmes (par exemple, le système nous permet d'examiner plus en détail certaines tendances inattendues). En fait, les tableaux de résultats nous permettent de concentrer notre attention sur les données qui peuvent poser des difficultés. Il peut s'agir de demander à un centre de traitement de l'IRS d'examiner à nouveau les données ou encore, de transférer du gros ordinateur sur micro les enregistrements présentant certaines caractéristiques. Ensuite, nous examinons ces enregistrements à la main afin de cerner les problèmes.

Comme nous l'avons souligné, nous avons défini les exigences qualitatives de base auxquelles doivent satisfaire les données pour les recensements économique et agricole de 1987. Notre système automatisé (comparaison des données fournies aux critères établis), nous permet de déterminer immédiatement si les données sont de qualité acceptable.

Après examen et vérification des données du cycle en cours, des tableaux cumulatifs sont produits sur micro-ordinateur. Le fait d'avoir recours à des micros plutôt qu'à un gros ordinateur pour la production de ces tableaux permet une utilisation plus efficace de nos ressources. Premièrement, le système nous évite d'avoir à garder en mémoire du gros ordinateur des fichiers cumulatifs, ce qui réduit les coûts d'ordinateur. Auparavant, ces fichiers cumulatifs étaient conservés dans le gros ordinateur et étaient combinés aux fichiers du cycle en cours pour produire de nouveaux tableaux cumulatifs. En utilisant les micro-ordinateurs, nous avons pu intégrer au tableur des formules simples qui nous permettent de créer des tableaux cumulatifs à très peu de frais. Deuxièmement, la production de ces tableaux cumulatifs n'engage pas la création de programmes pour gros ordinateur. Des imprimés des tableaux cumulatifs sont produits et conservés pour analyse ou référence.

Outre cet ensemble complet de tableaux cumulatifs, nous produisons également un ensemble de tableaux de résultats. Ces tableaux contiennent des comparaisons détaillées de certaines données clés pour le cycle en cours. Le tableau 3 est un exemple des nombreux tableaux des résultats produits par le système. On y trouve le nombre et le pourcentage pondérés de formules 1040 — annexe F par centre de traitement de même que le pourcentage attendu de formules. Comme on peut le voir, les chiffres cumulatifs sont raisonnables et répondent aux critères établis. Si des incohérences avaient été relevées, un indicateur aurait été affiché en regard du nom du centre de traitement concerné. Le dernier type de rapport produit par le système automatisé est le rapport de la situation qui est un bilan détaillé du nombre cumulatif de fichiers envoyés par l'IRS. Dans ce rapport, on compare la qualité globale des ensembles de données aux critères établis (délais pour la transmission des données et exigences qualitatives). Ces rapports sont transmis à l'IRS une fois par mois environ. Comme nous l'avons souligné, ces rapports, qui constituent un état récapitulatif de la qualité des données, favorisent les échanges entre le personnel du Bureau et celui de l'IRS.

4. RÉSULTATS DU CONTRÔLE DE LA QUALITÉ

Les rapports de la situation permettent au Census Bureau et à l'IRS de déceler les problèmes au fur et à mesure et ils facilitent la collaboration. Il est à noter cependant que lorsque nos critères ne sont pas respectés (délais ou qualité), la plupart du temps c'est en raison d'un changement dans les modes de déclaration des entreprises. Des cas semblables ne posent aucune difficulté pour l'IRS (il n'a pas de mesure corrective à prendre) mais, pour le Bureau, ils peuvent

Tableau 3
Répartition pondérée des formules 1040, annexe F (1986),
en pourcentage, selon le centre de traitement

Année d'imposition	Annexes F reçues	Centre de traitement				
		Atlanta	Philadelphie	Austin	Cincinnati	Kansas City
1986						
En nombre	2,087,200	176,700	71,600	374,900	262,100	358,600
En Pourcentage	100.0	8.5	3.4	18.0	12.6	17.2
Objectif visé						
En pourcentage	100.0	8.5	3.0	18.5	11.5	17.5
Objectif non atteint ¹						
Année d'imposition		Centre de traitement				
		Andover	Ogden	Brookhaven	Memphis	Fresno
1986						
En nombre	118,800	343,200	40,300	288,100	52,500	400
En pourcentage	5.7	16.4	1.9	13.8	2.5	0.0
Objectif visé						
En pourcentage	5.5	16.5	2.0	14.0	2.5	0.0
Objectif non atteint ¹						

¹ Objectif atteint: écart maximum de plus ou moins 2%.

Tableau 4a

Répartition pondérée des formules 1120 (1986), en nombre, selon la date

Date	Formules 1120		Objectif non atteint
	Nombre reçu	Nombre exigé	
Fin mars 1987	326,500	303,000	
Fin avril 1987	697,600	760,000	Objectif non atteint
Fin mai 1987		988,000	
Fin juin 1987		1,190,000	
Fin juillet 1987		1,418,000	
Fin août 1987		1,621,000	
Fin janvier 1988		2,077,000	
Fin octobre 1988		2,533,000	

Tableau 4b

Répartition pondérée des formules 1120S (1986), en nombre, selon la date

Date	Formules 1120S		Objectif non atteint
	Nombre reçu	Nombre exigé	
Fin mars 1987	103,350	90,000	
Fin avril 1987	328,850	225,000	
Fin mai 1987		292,000	
Fin juin 1987		352,000	
Fin juillet 1987		420,000	
Fin août 1987		480,000	
Fin janvier 1988		615,000	
Fin octobre 1988		750,000	

entraîner des frais de traitement informatique supplémentaires. Aux recensements économique et agricole, le Bureau a relevé un cas semblable (voir tableaux 4a et 4b): à la fin de mai 1987, le Census Bureau avait reçu environ 697,600 formules 1120 (sociétés); le nombre requis étant de 760,000. L'objectif n'a donc pas été atteint. Par contre, le Bureau a reçu un nombre beaucoup plus élevé de formules 1120S («S Corporations») que le nombre exigé (environ 328,850 au lieu de 225,000). La cause de ce changement est la suivante: avec l'adoption de la nouvelle loi fiscale, il devenait plus avantageux pour les entreprises de remplir une formule 1120S qu'une formule 1120. Ainsi, bien qu'il n'y ait eu aucune erreur dans les données, il était important que le Bureau relève ce changement pour pouvoir modifier ses méthodes de traitement. Nous sommes en train de mettre au point des mesures pour tenir compte de cet accroissement du nombre des «S Corporations». Le tableau 5 est un exemple des divers tableaux que nous utilisons pour étudier la qualité des données. Comme on peut le voir, la qualité des données répond aux exigences. Si, pour un élément d'information donné, nos critères n'étaient pas respectés, un indicateur serait affiché pour qu'un analyste fasse les recherches nécessaires.

Le système informatisé de contrôle de la qualité sera tout à fait prêt pour effectuer le traitement des fichiers de l'IRS de 1987. Des systèmes pilotes ont été utilisés (et continuent de l'être) pour évaluer les fichiers de 1985 et de 1986. C'est grâce à ces systèmes que le Bureau a pu, en vue des recensements agricole et économique, mener à bien la vérification des données administratives de 1985 et de 1986 transmises par l'IRS.

Tableau 5

Modes de déclaration des éléments d'information des formules 1120S, 1986 (chiffres pondérés)

Éléments d'information	Pourcentage de formules 1120S		Objectif non atteint
	reçu	exigé	
EIN			
Espaces laissés en blanc, zéros, données non numériques	0.0	Moins de 1.0	
Code IRD invalide	0.0	Moins de 1.0	
CODE PBA			
Espaces laissés en blanc ou données non numériques	0.0	Moins de 6.0	
Espaces laissés en blanc, données non numériques, codes non classés ou codes PBA invalides	11.5	Moins de 18.0	
REVENUS OU CHIFFRE DES VENTES BRUTS MOINS RENDUS ET RABAIS			
Espaces laissés en blanc, zéros ou données non numériques	20.9	Moins de 40.0	
Enregistrements pour lesquels un chiffre est fourni, pourcentage dans chacune des tranches ci-après:			
- moins de \$100,000	45.7	30.0 — 60.0	
- \$100,000 ou plus mais moins de \$500,000	36.9	20.0 — 50.0	
- \$500,000 ou plus	17.4	10.0 — 30.0	
EXERCICE FINANCIER			
Espaces laissés en blanc, zéros ou données non numériques	0.0	Moins de 1.0	

L'utilisation combinée de micro-ordinateurs et d'un gros ordinateur a permis au Census Bureau de vérifier, d'une manière efficace et exhaustive, les très gros fichiers de données de l'IRS et d'en assurer la qualité. En outre, le système a allégé la charge des programmeurs et il continuera de le faire parce qu'une bonne partie du travail peut être faite sur micro-ordinateurs, par des non-spécialistes. Par ailleurs, puisque le système a permis de réduire l'utilisation du gros ordinateur et la charge de travail des programmeurs, ces derniers peuvent concentrer leurs efforts sur les totalisations de base. Enfin, en ce qui concerne l'analyse des données, le système permet que la révision puisse être effectuée par du personnel de différents niveaux. Les états récapitulatifs permettent aux gestionnaires de déterminer rapidement et efficacement si les délais de transmission sont respectés et si la qualité des données est satisfaisante. Les analystes, pour leur part, consultent les rapports hebdomadaires plus détaillés qui leur indiquent, le cas échéant, les ensembles de données nécessitant un examen plus approfondi.

5. RÉSUMÉ

Le Census Bureau a mis au point un système global de contrôle de la qualité qui permet de déceler les problèmes risquant de nuire à la qualité des données. Avec le système, il est possible de traiter de très gros fichiers de données de l'IRS. En outre, comme le système favorise la collaboration, il permet au Census Bureau de régler rapidement les problèmes de manière à assurer la transmission adéquate des données. Les critères de qualité et les délais de transmission sont définis conjointement par le Bureau et l'IRS. Le respect de ces critères est vérifié par l'ordinateur. Par ailleurs, grâce à l'informatisation, le Bureau est en mesure d'utiliser au mieux

les programmeurs et les ressources budgétaires limitées dont il dispose pour effectuer la vérification. De plus, le recours aux micro-ordinateurs, parce qu'il offre beaucoup de souplesse d'utilisation, a permis d'accroître le rôle des analystes et de diminuer la charge de travail des programmeurs. Les échanges entre le Census Bureau et l'IRS étant nombreux, les personnes concernées sont plus sensibles à l'importance de la qualité des données. Ces échanges permettent également de mettre au point des procédures efficaces de transmission des données. Enfin, grâce à cet effort collectif pour régler les problèmes, le Census Bureau est maintenant assuré de recevoir les données requises pour mener à bien les recensements économique et agricole de 1987.

REMERCIEMENTS

Nous voulons remercier les arbitres pour leurs commentaires et suggestions.