Nº 68-515-X au catalogue
ISBN 978-0-660-02487-5

Aperçu du Programme intégré de la statistique des enterprises

Date de diffusion : le 17 juin 2015

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à infostats@statcan.gc.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « À propos de nous » > « Notre organisme » > « Offrir des services aux Canadiens »

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

.	indisponible pour toute période de référence
..	indisponible pour une période de référence précise
...	n'ayant pas lieu de figurer
0	zéro absolu ou valeur arrondie à zéro
0^s	valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
^{p}	provisoire
^{r}	révisé
x	confidentiel en vertu des dispositions de la *Loi sur la statistique*
^{E}	à utiliser avec prudence
F	trop peu fiable pour être publié
*	valeur significativement différente de l'estimation pour la catégorie de référence ($p<0,05$)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2015

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'entente de licence ouverte de Statistique Canada.

Une version HTML est aussi disponible.

This publication is also available in English.

Table des matières

1. Introduction et contexte 4
2. Objectifs et principes directeurs du PISE 5
3. Les pierres angulaires d'un système d'infrastructure intégré 6
a) Modèle fondé sur des métadonnées 6
b) Création d'une infrastructure intégrée 7
c) Fonctionnalité de gestion de l'information 8
4. Une base de sondage commune : le Registre des entreprises 9
a) Source des données aux fins de la répartition 9
b) Source de l'information sur les biens, services et activités 11
5. Élaboration d'un modèle de contenu harmonisé 12
6. Gestion du fardeau de réponse 13
a) Stratégie de remplacement des données fiscales 13
b) Stratégie de remplacement intelligente 14
c) Gestion active de la collecte 14
7. Intégration d'un cadre d'analyse de cohérence 16
8. Les grandes entreprises complexes : un cas spécial 16
a) Les programmes GPE et SGEA 16
b) Collecte personnalisée 17
9. Approche méthodologique 18
a) Échantillonnage à deux phases 18
b) Calage 18
10. Stratégie d'analyse et de diffusion 19
a) Stratégie d'analyse 19
b) Intégration avec le Système de comptabilité nationale (SCN) 20
11. Modèle de gouvernance 21
a) Structure des comités 21
b) Système d'information de gestion 22
12. Conclusion 22
13. Références 23

1. Introduction et contexte

En 2010, Statistique Canada a lancé l'initiative de l'Architecture opérationnelle du Bureau (AOB). À ce moment, l'organisme subissait des pressions financières croissantes qui ont mené à un examen approfondi des méthodes opérationnelles, des processus statistiques et de l'infrastructure des systèmes. Les principaux objectifs consistaient à cerner des possibilités de réaliser des économies, déterminer des méthodes qui permettent d'améliorer l'assurance de la qualité et trouver des façons de renforcer la capacité à offrir des programmes statistiques.

Ces travaux se sont traduits par de nombreuses recommandations, notamment l'élaboration et l'utilisation obligatoire des services intégrés partagés et génériques pour la collecte, le traitement, le stockage et la diffusion des renseignements statistiques. Afin d'atteindre les objectifs visés, Statistique Canada a entrepris plusieurs projets, dont un grand projet de transformation de ses enquêtes sur la statistique économique, le Programme intégré de la statistique des entreprises (PISE).

Le PISE sert de cadre normalisé aux enquêtes économiques menées à Statistique Canada. La base de sondage commune de ces enquêtes repose sur le Registre des entreprises de Statistique Canada. Les questionnaires sont fondés sur du contenu et des concepts harmonisés. De plus, les enquêtes ont en commun des méthodologies d'échantillonnage, de collecte et de traitement fondées sur des métadonnées. En outre, des outils communs sont en place afin de modifier, de corriger et d'analyser les données.

Bien que le PISE ait été une entreprise ambitieuse, le nouveau programme n'a pas été créé à partir de zéro. En fait, il s'inscrit dans les efforts qui sont constamment déployés pour concevoir une méthodologie harmonisée d'enquêtes entreprises et qui a débuté à la fin des années 1990 avec le Programme de l'Enquête unifiée auprès des entreprises (EUE)[1]. Ce dernier portait au départ sur sept enquêtes pilotes et a été graduellement élargi de manière à inclure 60 enquêtes-entreprises annuelles dans les secteurs de l'agriculture, de la fabrication, du commerce et des services. Or, avec le temps, l'infrastructure des systèmes de l'EUE est devenue désuète. La maintenance des systèmes nécessitait beaucoup de ressources et il n'était pas facile d'adapter le modèle à l'évolution des exigences. Par conséquent, le moment était tout indiqué pour remanier le modèle et instaurer les principes de l'AOB.

Dans le cadre du PISE, l'infrastructure des enquêtes entreprises a été entièrement remaniée et des méthodologies et processus novateurs ont été instaurés afin de continuer d'améliorer le modèle de l'EUE. Bon nombre de ces innovations règlent des faiblesses de longue date de l'EUE, notamment la mise en œuvre d'un système suffisamment souple pour s'adapter aux nouvelles exigences. D'ici 2016, la majorité des enquêtes économiques de Statistique Canada seront intégrées dans le modèle du PISE.

Le présent aperçu décrit divers aspects du PISE et se veut une introduction au modèle à l'intention du personnel actuel et futur. C'est l'une des composantes d'un vaste cadre d'apprentissage comprenant de la documentation détaillée sur les méthodes et processus, les manuels des utilisateurs qui accompagnent les outils, des documents expliquant les stratégies de vérification et d'analyse, ainsi que des séances de formation en classe qui donnent au personnel l'occasion de travailler avec les outils et systèmes du PISE dans un environnement d'apprentissage avant qu'on leur demande de le faire dans l'environement de production.

1. Voir le document de Brodeur et al. dans les références.

2. Objectifs et principes directeurs du PISE

Pendant la conception du modèle du PISE, l'équipe s'est concentrée sur six objectifs principaux, que voici :

- améliorer la qualité des données en appliquant des méthodologies et des processus normalisés, en instaurant du contenu harmonisé et en favorisant l'analyse de cohérence;
- alléger le fardeau de réponse;
- moderniser l'infrastructure de traitement des données;
- intégrer la majorité des enquêtes économiques dans le nouveau modèle;
- simplifier et normaliser les processus de manière à réduire les courbes d'apprentissage et à raccourcir les délais;
- réduire les coûts récurrents liés aux aspects opérationnels des enquêtes afin de réaliser des économies.

Pour atteindre ces objectifs, les programmes d'enquête ont dû s'adapter aux exigences particulières du modèle du PISE mais, en même temps, le modèle devait être conçu de manière suffisamment souple pour remplir les exigences propres aux programmes. Le plus grand défi à relever lors de la mise en œuvre du PISE a été de trouver un équilibre entre l'élaboration d'un modèle normalisé et cohérent et le maintien de la souplesse nécessaire pour répondre aux exigences propres aux programmes.

Les caractéristiques communes des enquêtes du PISE qui favorisent l'efficacité opérationnelle comprennent :

- l'utilisation pleine et entière du Registre des entreprises comme base de sondage;
- le recours à des questionnaires électroniques comme principal mode de collecte;
- la gestion active de la collecte en fonction des indicateurs de qualité;
- la mise en œuvre d'une stratégie de vérification commune;
- l'application de métadonnées normalisées à la conception des questionnaires, à l'échantillonnage, à la vérification et l'imputation, de même qu'aux processus de répartition et d'estimation;
- la gouvernance partagée dans l'ensemble des divisions opérationnelles et spécialisées, y compris la gestion du changement.

Les enquêtes du PISE font appel à ces caractéristiques communes mais il existe de nombreuses variantes qui répondent aux exigences propres à chaque enquête. Par exemple, le modèle est conçu de manière suffisamment souple pour traiter des enquêtes de fréquence variable, notamment chaque mois, tous les trimestres et une fois l'an, et ayant une couverture différente, comme des enquêtes à l'échelle de l'économie, des enquêtes axées sur l'industrie et des enquêtes fondées sur les activités.

3. Les pierres angulaires d'un système d'infrastructure intégré

Depuis le début, la vision du PISE consiste à créer une infrastructure d'enquêtes entreprises souple et efficace. Pour la souplesse, le système a été conçu de manière à intégrer une foule d'enquêtes différentes tout en réduisant les contraintes liées au traitement. Par exemple, le système peut traiter simultanément les données de nombreuses enquêtes, ce qui n'était pas possible dans l'EUE.

Par ailleurs, le système a été conçu de telle sorte que les modifications apportées aux exigences, y compris l'ajout de nouvelles enquêtes ou le changement des spécifications des enquêtes actuelles, puissent être instaurées sans exercer de contraintes excessives sur le personnel affecté au traitement. On est parvenu à la variabilité dimensionnelle et à l'efficacité en concevant un système fondé sur des métadonnées. Dans le système de métadonnées intégré du PISE, on peut exécuter systématiquement tous les processus sans intervention manuelle afin de transformer les fichiers à mesure qu'une enquête passe d'un processus à l'autre.

a) Modèle fondé sur des métadonnées

Depuis longtemps, Statistique Canada utilise des métadonnées afin de gérer ses publications, ses services et ses fonds de données statistiques. Toutefois, relativement peu de programmes d'enquête comportent des métadonnées bien élaborées pour gérer les opérations d'enquête. Le programme de l'EUE a effectivement utilisé un système de métadonnées qui comprenait des règles de vérification ainsi que des numéros et des descriptions de cellules[2]. Pour le PISE, ce cadre de métadonnées a été élargi afin d'englober tous les aspects du traitement des enquêtes. On augmente ainsi l'efficacité, la robustesse et l'agilité pendant le traitement des programmes du PISE.

Dans le modèle du PISE, les métadonnées sont stockées dans des tableaux faciles à modifier qui servent à alimenter les systèmes. Voilà une déviation par rapport au modèle de l'EUE dans lequel un code fixe était souvent attribué aux métadonnées dans les programmes. Les systèmes du PISE ont simplement accès à l'information à partir des tableaux de métadonnées afin d'en orienter l'exécution.

L'un des principaux avantages du système fondé sur les métadonnées du PISE vient du fait que l'on peut réagir aux changements qui s'imposent à mesure que les besoins du programme évoluent en modifiant les métadonnées plutôt qu'en réécrivant le code système. Cela permet à l'équipe affectée au traitement d'exercer un meilleur contrôle et aux utilisateurs d'avoir une plus grande marge de manœuvre.

Selon les lignes directrices en matière de gestion des métadonnées, ces dernières doivent être actives, créées pour une raison et utilisées dans des processus en aval. Bien que le principe « aucune donnée sans métadonnées » soit souvent appliqué à la production de données finales sous forme de métadonnées descriptives, il s'applique aussi au traitement, particulièrement dans le PISE. Lorsqu'on crée une variable, elle est étiquetée par des éléments descriptifs, comme un nom et une origine, mais les métadonnées indiquent aussi comment faire la validation, la vérification et l'imputation, et permettent de suivre l'évolution de la variable dans les diverses étapes de traitement.

Les utilisateurs disposeront d'un point d'entrée unique dans le PISE puisque les métadonnées seront intégrées dans toutes les étapes du traitement et que leur gestion ainsi que les processus qu'elles dirigent formeront naturellement les composantes du même portail homogène. Cette intégration permettra au système de vérifier instantanément la satisfaction des conditions de fonctionnement. À titre d'exemple, si un utilisateur choisit d'exécuter un processus, l'interface peut lui demander d'entrer les métadonnées nécessaires et vérifier que d'autres préalables soient en place. Les intrants peuvent ensuite être validés automatiquement et, après cette étape, le système peut soit cesser ses avertissements, soit permettre à l'utilisateur de passer aux étapes suivantes. Selon la méthodologie des interfaces de métadonnées, l'utilisateur n'a pas à s'occuper de nombreuses applications et n'a pas besoin de connaître l'ordre des diverses étapes nécessaires pour exécuter le traitement.

Bien que l'objectif d'efficience d'un système fondé sur des métadonnées consiste à réduire au minimum la reprise du travail et à faciliter la réutilisation, une meilleure qualité et une plus grande cohérence sont des résultats tout aussi importants. L'intégration des métadonnées dans les opérations de traitement facilite les vérifications

2. Voir le document de Hostetter dans les références.

automatiques de cohérence. Les règles d'intégrité des données sont appliquées dans la base de données du système pour assurer la qualité des intrants. En outre, les métadonnées génèrent de précieux renseignements sur la gestion qui aident à surveiller les progrès, ce qui améliore la qualité générale du traitement des enquêtes.

b) Création d'une infrastructure intégrée

Les principales composantes de l'infrastructure du PISE reposent sur une série de services génériques pour exécuter l'échantillonnage (SGECH), la vérification et l'imputation (BANFF), l'estimation (SGE), la confidentialité (G CONFID) et la diffusion (G EXPORT). En outre, plusieurs dépôts des différents types de données, paradonnées et métadonnées qui sont utilisés tout au long du processus servent de centres de service de données.

Les principaux centres de service de données sont les suivants : le Registre des entreprises, qui est la base de sondage; l'entrepôt fiscal, qui renferme tous les dossiers fiscaux; la Base de métadonnées intégrée (BMDI), qui contient les métadonnées liées au contenu; le dépôt de données du PISE, où sont conservés tous les fichiers de traitement des données; et l'entrepôt de données de collecte, où se trouvent les données brutes des répondants.

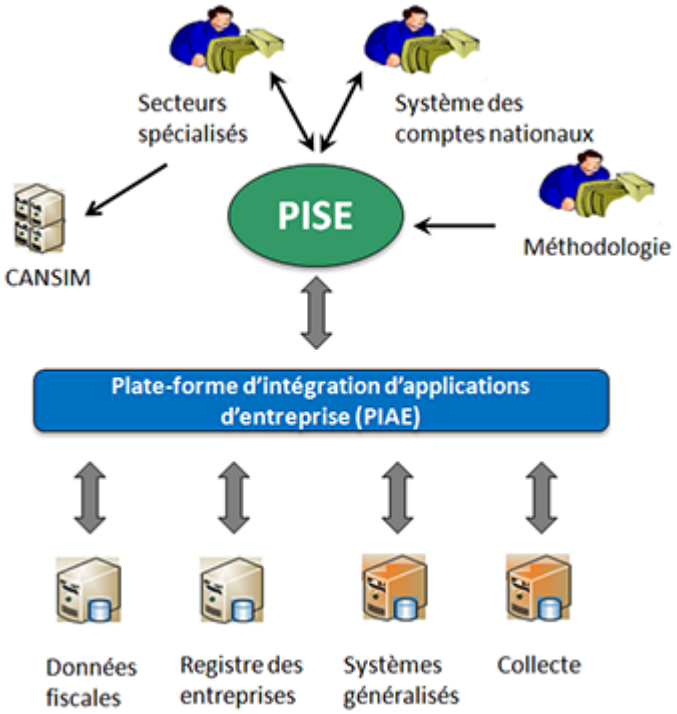
On peut intégrer les services génériques à l'aide de la Plate-forme d'intégration d'applications d'entreprises (PIAE), qui permet de faire des transformations de données homogènes entre les principaux services sur les entreprises, notamment :

- les services de données du Registre des entreprises, les services de données fiscales, les métadonnées de la BMDI et, à l'avenir, les données sur la collecte;
- les services génériques servant au traitement des données (SGECH, BANFF, SGE).

L'un des grands avantages de l'adoption de la PIAE vient du fait que, lorsqu'un centre de service de données modifie un service de données, le PISE n'a pas à adapter ses interfaces pour tenir compte du changement, ce qui favorise la stabilité à l'intérieur du système et améliore la fiabilité.

Le diagramme (figure 1) montre que les services de données sont directement liés à la PIAE. La technologie permet alors à toutes les parties connectées d'avoir accès aux fichiers de données ou aux services de traitement requis.

Figure 1
Infrastructure du PISE



L'architecture des technologies du PISE (AT PISE) soutient l'infrastructure intégrée. Elle repose sur les capacités matérielles et logicielles logiques nécessaires pour assurer le déploiement des services opérationnels, de données et d'application, notamment l'infrastructure de la TI, les intergiciels, les réseaux, les communications, le traitement et les normes.

L'AT PISE est une description officielle du système du PISE, de la structure des composantes, de leurs relations ainsi que des principes et lignes directrices régissant leur conception et leur évolution dans le temps. En outre, elle comprend les principales composantes de l'intégration avec lesquelles le nouveau système doit interagir et précise comment mettre en place l'architecture pour tenir compte de ces composantes externes importantes.

L'AT PISE fournit un système fondé sur des métadonnées fortement intégré dans les services de données de base de Statistique Canada. Voici certaines de ses caractéristiques clés :

- Modularité / maintenabilité / extensibilité
  - architecture en couches
  - architecture axée sur le service et PIAE
  - architecture orientée objet
- Sécurité
  - groupes de sécurité fondés sur les rôles
  - authentification au moyen de la connexion au compte d'utilisateur
  - mécanisme de vérification pour suivre les activités des utilisateurs
  - pas d'accès direct aux tables des bases de données
- Convivialité
  - architecture itérative
  - architecture axée sur le service
- Rendement / disponibilité / fiabilité
  - architecture intégrée, PIAE
  - services d'infrastructure organisationnelle
  - services d'hébergement généraux

c) Fonctionnalité de gestion de l'information

La saine gestion des données statistiques est une composante intégrale des travaux de Statistique Canada visant à donner à la population canadienne accès à une source d'information fiable. Dans le cas du PISE, l'infrastructure intégrée est conçue de manière à faciliter la gestion des données en fonction des exigences législatives et des directives sur la gestion de l'information de Statistique Canada.

L'un des principaux aspects de la gestion de l'information consiste à conserver la documentation fournissant le contexte des fichiers de microdonnées et de données agrégées, ce qui comprend les dictionnaires de données, la définition des concepts et autres documents dont les utilisateurs ont besoin pour comprendre les forces, les limites et les utilisations prévues de l'information. Le niveau de la documentation peut varier selon la valeur à long terme attribuée aux renseignements. Aux fins du PISE, la majeure partie de la documentation requise pour atteindre les objectifs en matière de gestion de l'information sera créée, tenue à jour et conservée soit dans le cadre de métadonnées, soit dans le dépôt de données du PISE où se trouvent les fichiers. Les utilisateurs pourront avoir accès aux fichiers de données et aux métadonnées connexes à l'aide d'outils normalisés.

Un autre aspect important de la gestion de l'information est l'application des règles de conservation aux données, ainsi que la suppression de toute information n'ayant aucune valeur opérationnelle. Dans le cas du PISE, des métadonnées indiquant des périodes de conservation seront intégrées dans les fichiers de données pour en faciliter la gestion. Les fichiers qui ne seront plus nécessaires seront supprimés au moyen de processus automatisés.

4. Une base de sondage commune : le Registre des entreprises

Le Registre des entreprises (RE) est la base de sondage commune de toutes les enquêtes qui suivent le modèle du PISE. Le RE est un centre de service de données mis à jour par l'entremise de plusieurs sources, notamment les fichiers de données administratives, la rétroaction de la conduite des enquêtes entreprises de Statistique Canada, de même que l'établissement de profils, y compris le contact direct avec les entreprises pour obtenir des renseignements sur leurs activités et les résultats de recherche sur Internet. L'utilisation du RE assure la qualité, tout en évitant les chevauchements entre les enquêtes et en allégeant le fardeau de réponse le plus possible.

Le RE contient toutes les entreprises exploitées au Canada et toutes les entreprises étrangères ayant des liens avec des sociétés canadiennes. Il renferme de l'information sur le mode d'organisation des entreprises, la classification par industrie, leur taille (revenus gagnés et nombre d'employés) et leur emplacement.

La vaste majorité des entreprises qui figurent dans le RE sont dites des entreprises simples, en ce sens qu'elles possèdent une seule unité d'exploitation et qu'elles sont classées dans une seule industrie et une seule région géographique. Les entreprises simples représentent 99 % de toutes les entreprises du RE. Les entreprises complexes, c.-à-d. celles qui ont plus d'une unité d'exploitation, sont en petit nombre (1 % du total) mais elles constituent environ 50 % des revenus générés. La structure des entreprises complexes est surtout mise à jour grâce à l'établissement de profils et à la rétroaction d'enquête. Des données administratives sont également utilisées mais elles ne sont souvent pas disponibles au niveau nécessaire pour mettre à jour la structure des entreprises complexes.

Quatre attributs statistiques sont dérivés pour chaque entreprise qui se trouve dans le RE. Le niveau le plus élevé est l'entreprise statistique, qui correspond souvent à une unité légale mais, pour certaines entreprises très complexes, il peut s'agir d'une famille d'unités légales. Le niveau suivant de la structure est la compagnie statistique, qui est le niveau le plus bas où l'on peut mesurer les investissements. On peut également calculer le bénéfice d'exploitation à ce niveau et mesurer l'actif et le passif.

Sous la compagnie se trouve l'établissement statistique qui, dans la plupart des cas, correspond à un centre de profit. À ce niveau, on peut mesurer la valeur de la production, le coût des intrants et de la main-d'œuvre. Enfin, l'emplacement statistique est l'entité du niveau le plus bas. Les données sur l'emploi et/ou les revenus sont disponibles à ce niveau.

Les enquêtes du PISE cibleront différents niveaux de la structure des entreprises, selon leurs besoins. Par exemple, pour les enquêtes mesurant l'actif et le passif, on ciblera l'entreprise ou la compagnie statistique. Par opposition, de nombreuses enquêtes sur la production cibleront l'établissement statistique ou l'emplacement statistique puisque ces niveaux fournissent le point de vue le plus détaillé qui soit sur la répartition géographique et par industrie de l'activité économique au Canada.

a) Source des données aux fins de la répartition

Pour produire des données économiques au niveau le plus détaillé de la structure d'une entreprise, ce qui est obligatoire pour compiler les comptes nationaux, il faut parfois avoir recours à un processus de répartition. En général, dans ce processus, on utilise les données d'enquête ou administratives qui ne sont disponibles qu'au niveau de la compagnie ou de l'entreprise statistique, et on les répartit entre les divers établissements et emplacements de l'entreprise, de sorte que les répartitions statistiques provinciales ou territoriales puissent être créées. Avec le temps, les programmes ont conçu leur propre méthode de répartition en réponse à des besoins précis et les processus de répartition se sont diversifiés.

Afin de normaliser les processus et méthodes dans le cadre du PISE, on a examiné les stratégies de répartition et on a élaboré et instauré une méthode commune. Les avantages liés à l'adoption d'une seule méthodologie de répartition comprennent : une meilleure cohérence des données administratives et d'enquête et l'intégration plus simple de la méthode de répartition dans le traitement des enquêtes.

On a décidé que le RE serait le meilleur endroit où stocker les variables de répartition, qui sont appelées facteurs de répartition, puisqu'il s'agit d'un outil centralisé, accessible et fréquemment utilisé dans l'organisation.

L'information servant à obtenir les facteurs de répartition provient généralement de deux sources principales : les variables des profils dans le RE et les données administratives provenant des fichiers supplémentaires T4 de l'Agence du revenu du Canada, qui présentent la répartition des salaires par province et par territoire. Ensemble, ces données sont utilisées pour établir deux ensembles de facteurs : les facteurs relatifs aux revenus et les facteurs d'emploi. Ils indiquent respectivement la part des revenus totaux ou celle de tous les emplois se rapportant à chacune des entités opérationnelles de l'entreprise.

On détermine d'abord les facteurs de répartition en utilisant les données T4, à condition qu'elles remplissent les trois critères suivants :

- elles doivent être disponibles pour l'entreprise visée, selon le numéro d'entreprise (NE);
- elles doivent porter sur l'ensemble des provinces et des territoires où l'entité juridique a au moins une entité d'exploitation;
- elles doivent être plus actuelles que les variables des profils qui figurent dans le RE.

Dans ces conditions, les valeurs de répartition de l'entreprise sont produites de façon à ce que la répartition de ses emplois et de ses revenus corresponde à la répartition provinciale ou territoriale indiquée par les données T4 sur les salaires et les traitements.

Si les données T4 ne remplissent pas les critères susmentionnés ou si toutes les entités d'exploitation sont situées dans la même province ou le même territoire, on calcule les facteurs de répartition directement à partir des variables des profils qui sont disponibles dans le RE. L'établissement des profils est le processus par lequel des analystes spécialisés mènent de façon régulière des enquêtes auprès des entreprises afin de veiller à ce que le RE contienne une liste à jour des entités opérationnelles pour une entreprise donnée. Dans le cadre de cet exercice, on obtient une estimation du total des emplois et des revenus pour chacune de ces entités. Pour les entreprises les plus complexes en activité au Canada, les données du profil sont fréquemment mises à jour et sont, par conséquent, plus susceptibles d'être utilisées pour créer les facteurs de répartition dans le RE.

Pour la plupart des enquêtes du PISE, le facteur de répartition des revenus dans le RE est utilisé pour répartir les données sur les finances, les biens et services et les caractéristiques dans les cas où les données recueillies englobent les opérations nationales des entreprises complexes. Il y a cependant une exception : la variable des salaires et des traitements versés, pour laquelle le facteur de répartition des emplois dans le RE est appliqué.

Dans le cadre du processus d'assurance de la qualité de Statistique Canada, ce sont les analystes des domaines spécialisés qui examinent et qui valident les facteurs de répartition. Au cours de cet examen, des modifications peuvent être apportées : celles-ci sont souvent fondées sur des données améliorées mises à la disposition des analystes, comme l'information provenant de sources de confrontation des données. Une fois validés par les analystes, les facteurs de répartition sont appliqués et des estimations sont totalisées. Pour les programmes du PISE, les estimations sont généralement produites à un niveau détaillé de l'industrie pour chaque province ou territoire canadien.

La portée de la répartition diffère selon les programmes d'enquêtes. Elle dépendra : de la structure d'une industrie donnée, et particulièrement du nombre d'entreprises complexes dans la population; de l'ampleur des données déclarées par les répondants; et de la stratégie de collecte utilisée. Par exemple, Statistique Canada peut recueillir auprès d'entreprises des données sur leurs opérations nationales ou provinciales, ou sur tous leurs emplacements. Lorsque des données sur leurs opérations nationales sont recueillies, les répercussions de la répartition seront plus importantes. En revanche, si les données sur les opérations d'une entreprise sont déclarées séparément pour chaque province ou territoire, il n'est alors pas nécessaire d'appliquer les facteurs de répartition pour désagréger davantage les données.

Chaque programme choisit une stratégie de collecte en fonction des exigences en matière d'extrant, comme le degré de précision requis pour les différentes estimations produites. L'information concernant la stratégie de collecte pour une industrie donnée est mise à la disposition des utilisateurs lorsque les données sont diffusées publiquement.

b) Source de l'information sur les biens, services et activités

Afin d'améliorer la qualité des estimations sur les biens et services et de sélectionner de manière plus efficace des échantillons, les responsables des programmes d'enquête ont déterminé qu'il fallait cibler les entreprises qui produisent des biens et services particuliers et celles qui s'occupent d'activités précises. Par exemple, au moment d'effectuer une enquête sur les dépenses en immobilisations, on peut choisir un échantillon de manière plus efficace si l'on dispose d'information sur la base de sondage afin de signaler quelles entreprises ont des dépenses en immobilisations.

Le RE peut maintenant conserver l'information sur le bien, le service ou l'activité au moyen du champ « Production et mesures additionnelles ». Les divisions spécialisées sont tenues de remplir et de mettre à jour ce champ. Une fois l'information disponible, on peut l'utiliser pendant la sélection de l'échantillon pour cibler les entreprises qui auront une incidence importante sur la qualité des estimations sur le bien, le service ou l'activité. Aux fins du PISE, plusieurs programmes utiliseront cette fonctionnalité. En fait, l'échantillon de la phase 1 du PISE, que nous décrivons plus loin, servira à recueillir de l'information sur le bien, le service et l'activité, qui sera alors utilisée pour bien choisir les entreprises qui recevront un questionnaire plus détaillé de l'enquête de la phase 2 (voir la section 9).

5. Élaboration d'un modèle de contenu harmonisé

Évidemment, la création d'une infrastructure intégrée permet d'obtenir des gains d'efficacité opérationnelle mais on peut ne profiter de tous les avantages de l'approche intégrée que si l'on instaure un cadre conceptuel harmonisé. Dans le cas du PISE, cela a débuté par l'application de normes. Toutes les enquêtes du PISE doivent appliquer des normes statistiques, notamment :

- le Système de classification des industries de l'Amérique du Nord (SCIAN) afin de classer la population cible par industrie;
- le Système de classification des produits de l'Amérique du Nord (SCPAN) pour classer et recueillir des données sur les intrants et les extrants des entreprises;
- le Plan comptable (PC)[3], qui est la taxonomie de référence servant à organiser l'information financière sur les entreprises (p. ex., les revenus, les dépenses, l'actif et le passif).

Plusieurs variables financières sont communes à de nombreuses enquêtes économiques. En harmonisant les définitions de ces variables et en appliquant des normes de manière systématique, on a conçu du contenu commun et on l'a instauré dans tous les programmes.

Le modèle de contenu du PISE repose sur une série de modules génériques englobant des variables communes et est appliqué aux enquêtes sans que l'on doive apporter de modifications d'une enquête à l'autre. Cette façon de faire joue un rôle capital dans la création de la cohérence dans l'ensemble des programmes et la réduction des efforts nécessaires pour concevoir, tester et mettre en œuvre le contenu des enquêtes.

Essentiellement, les modules normalisés sont une série de questions d'enquêtes-entreprises utilisées pour recueillir des renseignements afin de répondre aux besoins des programmes. Il existe des modules normalisés pour les données sur l'état des résultats (revenus et dépenses), les données relatives aux ventes selon le type de client, les données relatives aux ventes selon l'emplacement du client, et les intrants sur les achats de service.

L'objectif, qui consistait à utiliser les données fiscales de manière optimale, a orienté l'élaboration du contenu des questionnaires. Tout spécialement, les variables sur les revenus et dépenses du PISE ont été intégrées directement dans les données disponibles dans les dossiers fiscaux. Ce lien direct élimine la nécessité de recueillir de l'information financière auprès des petites et moyennes entreprises, puisqu'il est facile d'avoir accès aux données sur ces dernières à partir de sources administratives.

L'un des principaux problèmes qu'il a fallu régler pour élaborer le contenu des données financières a consisté à veiller à ce que les besoins conceptuels du Système de comptabilité nationale du Canada soient comblés grâce à l'utilisation des données administratives. Le PC fait le pont entre ces deux ensembles de concepts. Afin d'élaborer le modèle de contenu du PISE, on a examiné et révisé le PC pour vérifier que ses variables, qui sont directement liées aux concepts fiscaux, remplissent les besoins d'information des comptables nationaux.

Dans le but d'accroître la souplesse et de répondre aux besoins d'enquêtes particulières, les employés des secteurs spécialisés peuvent adapter certains modules qui font partie de leurs questionnaires du PISE. Par exemple, les produits apparaissant dans les questionnaires sur la fabrication sont différents de ceux qui se trouvent dans les questionnaires sur les industries des services. Il se peut également que certains modules normalisés ne soient pas nécessaires parce qu'ils ne s'appliquent pas à l'industrie. Pendant l'élaboration des questionnaires d'enquête du PISE, le personnel doit simplement choisir les modules de contenu standardisés pertinents et se concentrer sur l'élaboration du contenu propre à l'industrie, au besoin. Cela réduit grandement le temps nécessaire pour élaborer, mettre en œuvre et tester de nouveaux questionnaires.

3. Voir le document de Martineau (2012) dans les références.

6. Gestion du fardeau de réponse

Statistique Canada s'enorgueillit d'un passé riche en utilisation de données administratives pour alléger le fardeau de réponse. En réalité, dans le cadre du modèle de l'enquête-entreprise de l'EUE, les données fiscales ont été utilisées pour remplacer directement un sous-échantillon d'unités échantillonnées et pour imputer les enregistrements de non-réponse.

Au fil du temps, grâce à leur utilisation, les méthodes d'imputation des données fiscales se sont améliorées et la qualité de l'information s'est accrue, ce qui a renforcé encore plus la fiabilité des données fiscales comme principale source d'information. Pendant la transition vers le modèle du PISE, les méthodes ont été adaptées afin de profiter pleinement de la disponibilité des données fiscales, ce qui a réduit d'autant plus le fardeau de réponse dans les programmes d'enquête.

a) Stratégie de remplacement des données fiscales

Les données fiscales constituent le pilier des estimations des données financières pour les enquêtes du PISE et la majorité des entreprises de l'échantillon n'auront pas besoin de fournir de données pour les modules sur les états des revenus et dépenses[4]. En utilisant les données administratives comme leurs principales sources d'information financière, les questionnaires du PISE pourront se concentrer sur les biens et services produits, les pratiques des entreprises et d'autres caractéristiques, notamment les activités de recherche et développement, et les types de dépenses en immobilisations.

L'utilisation des données administratives est pleinement intégrée dans le plan d'échantillonnage du PISE. Pour chaque enquête du PISE, la population d'intérêt se divise en deux parties :

- les entreprises très complexes (il y en a environ 2 000 dans le RE), pour lesquelles on ne peut pas facilement utiliser les données fiscales;
- les petites et moyennes entreprises, pour lesquelles les données fiscales sont utilisées directement pour la majorité des variables financières.

Dans la plupart des cas, on effectuera des enquêtes auprès des grandes entreprises complexes afin de recueillir tous les renseignements nécessaires, y compris les données financières et caractéristiques. Cela est nécessaire puisque Statistique Canada a pour mandat de publier de l'information financière par industrie et région géographique. L'information administrative n'est souvent pas disponible à ce niveau de détail pour les entreprises complexes qui ont des activités importantes dans plus d'une industrie et/ou province.

Néanmoins, pour alléger le fardeau de réponse de certaines entreprises très complexes, on a conçu un outil qui leur permet de produire une déclaration une seule fois grâce à un questionnaire combiné pour toutes leurs activités. L'outil de déclaration des grandes entreprises complexes est décrit à la section 7. Par ailleurs, les données fiscales de certaines entreprises complexes seront remplacées par la stratégie de remplacement intelligente du PISE décrite ci après.

Dans le cas des entreprises simples, on instaurera des méthodes de renouvellement de l'échantillon et de coordination pour veiller à ce que le fardeau de réponse soit bien géré.

Le plan d'échantillonnage s'accompagne d'estimateurs et d'indicateurs de qualité connexes qui tiennent compte de la conception afin de produire les meilleures estimations qui soient à partir des données disponibles.

4. La collecte des données sur l'état des résultats d'enquête sera toujours nécessaire pour certaines grandes entreprises complexes et lorsque les microdonnées sont partagées avec d'autres ministères fédéraux en raison de la législation fiscale.

b) Stratégie de remplacement intelligente

Par le passé, toutes les grandes entreprises complexes étaient exemptées des stratégies de remplacement des données fiscales.

Dans le cadre de la stratégie de remplacement intelligente du PISE, il a été établi que les données administratives pouvaient être utilisées pour remplacer les données d'enquête de certaines entreprises complexes si les données administratives remplissaient les critères, notamment la disponibilité, l'exactitude et la cohérence[5], et si l'entreprise avait un certain degré de stabilité opérationnelle et structurelle pendant les trois dernières années.

Il s'agit là d'un important changement de paradigme. Les enquêtes du PISE recueilleront uniquement de l'information financière auprès des grandes entreprises complexes si les conditions susmentionnées ne sont pas remplies. Pour mettre en œuvre cette stratégie, il faut examiner séparément chaque entreprise complexe afin de déterminer si elle est une bonne candidate au remplacement intelligent.

Les données administratives sont évaluées et comparées aux données d'enquête pour les trois cycles d'enquête les plus récents. Cet exercice comprend la comparaison de l'exactitude des données, l'examen de la stabilité de la source des données administratives et l'examen de l'ampleur des données disponibles à partir de la source administrative. L'équipe chargée de la cohérence dans la Division de la statistique des entreprises (DSE) est tenue de préparer les cas à examiner, de présenter les candidates au remplacement intelligent aux spécialistes du domaine et d'instaurer la stratégie dans le cas d'entreprises sélectionnées.

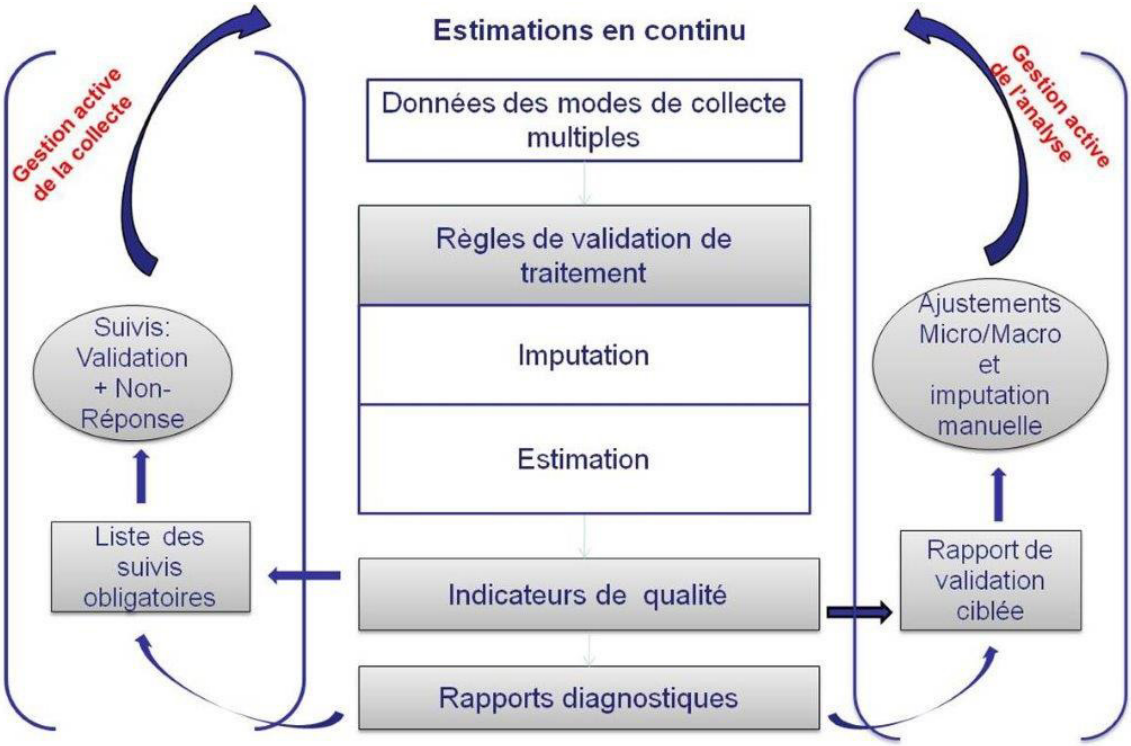
c) Gestion active de la collecte

La gestion active de la collecte est une expression qui signifie que les efforts de collecte sont ajustés de manière dynamique en fonction des données déjà reçues.

La pierre angulaire de la gestion active de la collecte est le modèle des estimations en continu (EC), selon lequel on produit des estimations de manière itérative jusqu'à ce qu'on ait atteint un niveau de qualité acceptable[6]. Les indicateurs de qualité de variables clés par domaine d'estimation et mesures de l'impact de chaque unité fournissent l'information nécessaire pour déterminer quelles autres unités requièrent un suivi. En plus de la gestion de la collecte, les estimations en continu jouent un rôle capital dans la stratégie d'analyse du PISE, qui est décrite plus loin dans le présent document.

5. Voir le document de Martineau (2013) dans les références.
6. Voir le document de Mills et al. dans les références.

Figure 2
Processus de gestion active de la collecte



Si tous les objectifs de qualité sont atteints pour une enquête particulière, la collecte active est fermée et le suivi peut prendre fin; sinon, les ressources chargées du suivi ou de la modification sont affectées aux unités considérées avoir une incidence sur les estimations clés et leur qualité. Les autres unités passeront par des processus automatiques afin de produire un fichier de microdonnées uniforme et utilisable.

Grâce au modèle des estimations en continu et à la gestion active de la collecte, on s'attend à une réduction globale des activités de suivi de la collecte, ce qui allégera le fardeau de réponse et améliorera la cohérence de la stratégie afin de cibler la qualité dans les différents domaines d'estimation.

7. Intégration d'un cadre d'analyse de cohérence

La vaste majorité des extrants du PISE seront intégrés par le Système de comptabilité nationale dans leurs divers cadres (p. ex., tableaux d'entrées-sorties) et comparés les uns aux autres par d'autres intervenants. On ne peut douter que l'harmonisation des concepts dans l'ensemble des enquêtes-entreprises aide à rendre les données des divers programmes plus cohérentes mais, en soi, cela ne suffit pas. Il est aussi extrêmement important de mener une analyse de cohérence tôt dans le cycle de traitement de l'enquête pour déterminer les problèmes de données, y compris les lacunes entre les données des niveaux de l'établissement et de l'entreprise, le chevauchement entre les différents programmes d'enquête et les cas où les concepts n'ont pas été déclarés de manière uniforme.

Dans le modèle de l'EUE, l'analyse de cohérence était généralement menée à la fin du cycle d'enquête et souvent une fois que les données étaient déjà diffusées. Le processus consistait à comparer les données déclarées pour des entreprises sélectionnées dans plusieurs enquêtes et aux sources de données administratives. L'équipe chargée de la cohérence de Statistique Canada effectuait une analyse préliminaire et en transmettait les résultats au personnel des divisions spécialisées et du Système de comptabilité nationale (SCN).

Ces travaux se sont avérés extrêmement utiles pour déterminer les problèmes statistiques et de déclaration ainsi que pour trouver des solutions. Or, étant donné le moment où le processus avait lieu, les divisions spécialisées ont eu de la difficulté à intégrer les changements dans leurs données statistiques. Souvent, les données avaient déjà été publiées et le personnel des domaines spécialisés était obligé d'attendre les dates prévues de révision des données avant d'intégrer les résultats de l'analyse de cohérence.

Selon le modèle du PISE, les données sont disponibles dans les systèmes de traitement presque instantanément, ce qui permet d'effectuer des vérifications de cohérence au début du processus d'enquête, c.-à-d. pendant la collecte active. En repérant les problèmes tôt, on a amplement l'occasion de confirmer l'information auprès des répondants et d'améliorer l'exactitude des données déclarées, ce qui réduit le risque de devoir apporter des révisions majeures.

Pour accélérer ces travaux, on a conçu un prototype d'afficheur d'analyse de cohérence qui renferme des données de sources administratives et d'enquêtes auxiliaires. Le fait d'intégrer toute l'information disponible sur une entreprise donnée dans un seul endroit facile d'accès permet d'avoir un processus d'analyse de cohérence plus approfondi et des données de meilleure qualité.

8. Les grandes entreprises complexes : un cas spécial

La majorité des enquêtes-entreprises et des enquêtes agricoles menées par Statistique Canada reposent sur des échantillons, ce qui signifie que seul un sous-ensemble d'unités de la population cible est sélectionné en vue de remplir un questionnaire. Toutefois, comme les échantillons sont conçus pour donner des estimations de grande qualité tout en allégeant le fardeau de réponse, on sélectionne toujours les plus grandes unités pour ce qui est des revenus générés. Voilà pourquoi les grandes entreprises et exploitations agricoles ont une incidence importante sur les estimations statistiques. En obtenant des données de ces unités, on peut atteindre des cibles de qualité tout en envoyant relativement peu de questionnaires aux petites et moyennes entreprises et exploitations agricoles.

Pour que cette stratégie remporte du succès, il est impératif que les plus grandes entreprises déclarent leurs renseignements de manière uniforme, ce qui est tout un défi puisqu'elles sont priées de participer à de nombreuses enquêtes. Afin d'assurer une bonne déclaration et d'améliorer les relations avec ces entreprises, Statistique Canada a conçu et mis en place deux programmes.

a) Les programmes GPE et SGEA

Le premier est le Programme de gestion de portefeuille d'entreprises (GPE), dont le mandat consiste à collaborer avec les entreprises les plus grandes et les plus complexes du Canada afin de tenir à jour le profil de leurs opérations, de négocier des modalités de déclaration, d'assurer la cohérence dans la déclaration des données

entre toutes les enquêtes et de fournir un seul point d'entrée pour réagir rapidement aux inquiétudes ou aux problèmes soulevés par ces entreprises ou par les analystes de l'industrie.

Le second programme est celui des Statistiques sur les grandes entreprises agricoles (SGEA). Comme le programme GPE, l'équipe chargée du programme SGEA doit veiller à ce que le profil des plus grandes entreprises agricoles au pays soit mis à jour et représentatif de toutes les activités agricoles courantes. En outre, l'équipe est tenue de recueillir des données auprès de ces exploitations agricoles, notamment sur l'élaboration de modalités de déclaration de données personnalisées.

Ces deux programmes ont remporté un franc succès. Le programme GPE porte sur 330 groupes d'entreprises qui, ensemble, possèdent plusieurs milliers d'activités définies par les établissements dans le Registre des entreprises. Les taux de réponse représentent régulièrement 90 % des entreprises visées par ce programme.

Le programme SGEA porte sur 370 entreprises se composant d'environ 800 exploitations agricoles qui exercent leurs activités sur 8 000 parcelles de terrain. En général, le taux de réponse est de 90 %.

b) Collecte personnalisée

Depuis plusieurs années, les équipes des programmes GPE et SGEA s'affairent à élaborer des modalités de déclaration personnalisées pour plusieurs des unités visées par leur programme respectif. Ces modalités ont été conçues pour réagir aux préoccupations des répondants selon lesquelles le besoin d'information statistique constituait un fardeau beaucoup trop lourd et il y avait un chevauchement entre plusieurs programmes.

Puisque de nombreuses grandes entreprises et exploitations agricoles étaient sélectionnées pour des enquêtes mensuelles, trimestrielles et annuelles, il y avait certainement lieu de faire valoir le bien fondé du partage des données entre les programmes. Pour aboutir à un compromis raisonnable entre les données qui devraient être recueillies pour chaque programme d'enquête et celles qui pourraient être partagées entre les programmes, le gestionnaire du programme GPE ou SGEA rencontre en général les analystes du domaine spécialisé afin d'élaborer une stratégie.

Souvent, les stratégies conçues comprenaient la création d'un chiffrier de déclaration intégrant les variables essentielles à recueillir. La méthode du chiffrier a aidé à atteindre le but, soit améliorer les relations avec les répondants en allégeant le fardeau de déclaration et en éliminant les chevauchements. Cependant, de telles stratégies de déclaration ont également entraîné un travail manuel considérable afin d'intégrer les données déclarées dans les systèmes de traitement. Parce que ce travail exigeant en ressources n'est pas intégré dans tous les programmes, on ne peut pas effectuer d'analyse de cohérence tôt dans le cycle d'enquête. Il est uniquement possible de le faire une fois que tous les programmes ont incorporé les données dans leurs systèmes.

Dans le cadre du PISE, ces problèmes sont réglés grâce à la mise en œuvre de l'outil de déclaration personnalisable nouvellement conçu. Ce dernier ressemble à l'outil du chiffrier qui est actuellement employé pour les modalités de déclaration d'entreprises complexes sélectionnées. Toutefois, plusieurs améliorations majeures y sont apportées. L'outil a été conçu à l'aide de métadonnées qui permettent de transférer de façon transparente l'information déclarée à l'infrastructure de collecte.

Non seulement cette façon de faire épargne-t-elle du temps, mais encore elle réduit la possibilité de commettre des erreurs pendant la saisie des données. Les vérifications de la collecte sont appliquées en temps opportun, pour permettre au personnel chargé des programmes GPE et SGEA de faire un suivi auprès de leurs répondants peu après l'arrivée des données. L'équipe chargée de la cohérence est également en mesure de cerner et de régler les problèmes bien plus tôt dans le cycle d'enquête, ce qui améliore la qualité des estimations des enquêtes.

9. Approche méthodologique

L'approche méthodologique du PISE tient compte des objectifs de base du programme, notamment alléger le fardeau de réponse, maintenir la qualité et optimiser l'utilisation des données administratives. Deux des principales méthodes instaurées pour atteindre ces objectifs sont l'échantillonnage à deux phases et le calage des estimations en fonction des sources de données fiscales.

a) Échantillonnage à deux phases

Le plan d'échantillonnage à deux phases[7] vise à améliorer la qualité tout en allégeant le fardeau de réponse. Les enquêtes du PISE visent, entre autres, des variables telles que les types de biens et services produits ou des activités particulières, comme la recherche et le développement. Dans le modèle de l'EUE, parce que les unités incluses dans l'échantillon étaient choisies uniquement en fonction des variables financières et du SCIAN, il était difficile de contrôler la qualité des estimations des biens ou services. Il s'agissait là d'une lacune du modèle puisque les estimations des biens ou services sont des intrants essentiels à la compilation des comptes nationaux.

L'approche de l'échantillonnage à deux phases comprend l'élargissement de l'étape du contact préliminaire qui est courante pour les enquêtes économiques. Tout spécialement, l'idée consiste à ajouter des questions à l'application de collecte afin d'obtenir de l'information que l'on peut utiliser plus tard, pendant la sélection de l'échantillon. En voici des exemples : obtenir de l'information sur les dépenses en immobilisations d'une entreprise, ses activités de recherche et développement, ainsi que les types de biens et services qu'elle produit.

Le but consiste à produire de meilleures estimations à moindre coût pour les variables ou les sous-populations qu'il n'est pas facile de repérer dans le Registre des entreprises. Grâce à la souplesse de l'application de collecte de la phase 1, on peut inclure d'autres types de questions dans les enquêtes actuelles ou à frais recouvrables selon les besoins des utilisateurs, ce qui constitue un élément clé pour accroître la souplesse du modèle du PISE.

Pour chaque enquête, il faut décider si l'approche à deux phases est indiquée. Les enquêtes annuelles sur les industries manufacturières, les dépenses en immobilisations, la recherche et le développement, le commerce de détail hors magasin, l'industrie du commerce de gros, l'énergie et l'agriculture pourraient profiter de cette approche afin de réagir rapidement à certains événements. Par exemple, seulement de 40 % à 60 % des 30 000 unités sélectionnées dans l'échantillon historique de l'Enquête sur les dépenses d'immobilisations déclarent effectivement des dépenses en immobilisations. L'approche à deux phases aide à alléger le fardeau de réponse et à réduire les coûts en précisant quelles entreprises il faut cibler au moyen du questionnaire détaillé sur les dépenses en immobilisations de la phase 2.

Une fois la phase 1 terminée, l'information recueillie est traitée en vue de la sélection de l'échantillon de la phase 2. Tout au long de la collecte de la phase 1, les analystes du domaine spécialisé peuvent déterminer les mises à jour nécessaires au moyen du Registre des entreprises. La DMEE intègre l'information mise à jour afin de générer l'échantillon de la phase 2.

À cette étape, la collecte des données d'enquête débute et les questionnaires électroniques en sont le principal instrument. Les mises à jour du SCIAN provenant de la collecte sont directement disponibles dans le RE, tandis que les données recueillies à mêmes les suppléments des enquêtes, couvrant les emplacements des établissements, sont disponibles par l'intermédiaire de l'entrepôt du PISE. Une fois validée, l'information mise à jour à partir des suppléments est rentrée dans le RE par le Programme de gestion de portefeuille d'entreprises ou le programme des profils du RE.

b) Calage

La stratégie d'estimation du PISE tient compte du plan d'échantillonnage à deux phases et intègre les estimations par calage afin de produire des estimations de programme de haute qualité.

Dans le plan d'échantillonnage à deux phases, il y a deux occurrences de sélection d'échantillon. Une unité peut être choisie pendant la sélection de la phase 1 et, si elle est retenue, elle peut être sélectionnée pendant la

7. Voir le document deTurmelle et al. (2013) et de Turmelle et al. (2012) dans les références.

sélection de la phase 2. Par conséquent, il y a deux poids d'échantillonnage, w_{1k} pour la sélection de la phase 1 et w_{2k} pour la phase 2. Néanmoins, pendant l'estimation, les deux poids sont combinés pour donner un seul poids, exprimé comme w_k, pour chaque unité de la phase 2.

Dans le PISE, on utilise alors l'estimateur de calage pour étalonner les estimations à l'aide du total des données fiscales. Le calage se sert des données auxiliaires pour produire un estimateur ayant une plus petite variance que les estimateurs non calés et qui s'étalonne avec le total des données auxiliaires. Pour ce faire, on utilise les données auxiliaires afin de modifier le poids d'échantillonnage pour en faire ce qu'on appelle des poids de calage. Il est essentiel que le calage produise un poids calé, $\tilde{w}_k$, qui remplace le poids d'échantillonnage, w_k, à l'étape de l'estimation.

Ces poids calés ne sont calculés que pour des unités à tirage partiel. Les unités à tirage complet conservent un poids de « un ». Pour les unités à tirage nul, les données fiscales ou les valeurs estimées à l'aide des données fiscales, sont agrégées et ajoutées aux composantes à tirage complet et à tirage partiel afin de compiler l'estimation de tout l'univers.

10. Stratégie d'analyse et de diffusion

Le modèle du PISE a été conçu pour favoriser l'approche descendante pendant l'analyse et la validation des données. Pour de nombreux programmes, cette approche contraste avec les méthodes employées auparavant. L'idée d'une approche descendante consiste à produire et à examiner des estimations à une étape très préliminaire du cycle de traitement de l'enquête puis à orienter les ressources de manière à entreprendre le travail et à améliorer grandement la qualité des estimations. L'approche descendante du PISE sera appliquée à mesure que les données seront recueillies grâce à la gestion active de la collecte, décrite dans la section 6 c), et à la stratégie d'analyse commune.

a) Stratégie d'analyse

La clé de la nouvelle approche analytique est la production d'estimations à une étape préliminaire du traitement des enquêtes. Par le passé, le premier ensemble d'estimations de bon nombre de programmes annuels était produit seulement quelques mois avant la diffusion des données. C'est uniquement à ce moment que les analystes avaient une vue d'ensemble des estimations de leurs industries et étaient en mesure de comparer les données d'autres sources dans le cadre du processus de validation. Avant que les estimations ne soient produites, les analystes consacraient souvent énormément de temps et d'effort aux corrections manuelles des microdonnées. Même si l'on avait l'impression que ce travail manuel donnait des estimations de bien meilleure qualité, des études ont démontré qu'une partie de cet effort n'avait pas d'incidence sur les estimations[8].

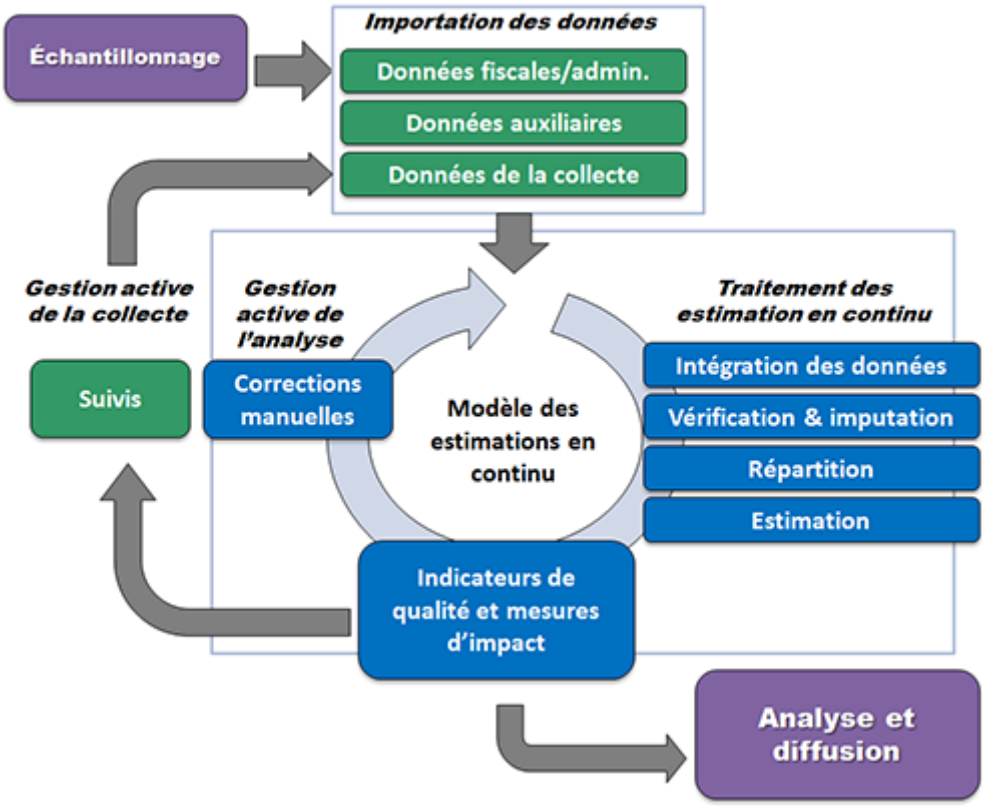
En produisant des estimations bien plus tôt dans le processus, les analystes pourront déterminer plus facilement les domaines d'estimation qui ne sont pas cohérents et concentrer leurs efforts sur ces domaines. Cette approche centrée de la vérification aura certainement un impact quantifiable sur les estimations. Elle pourrait également se traduire par des gains de ponctualité puisque certains domaines ne nécessiteront plus beaucoup d'interventions manuelles pour atteindre les cibles de qualité.

Dans le but de produire des estimations en continu, le système du PISE aura accès à toutes les données disponibles pour chaque unité échantillonnée, notamment les données déclarées, les données historiques et les données administratives. Les enregistrements de non-réponse seront imputés afin de compléter les ensembles de microdonnées. La répartition et l'estimation seront alors exécutées sans intervention manuelle. Il sera possible de consulter les estimations grâce à l'outil d'analyse du PISE.

Les employés des secteurs spécialisés compareront les estimations agrégées aux données auxiliaires et aux valeurs attendues selon les connaissances de l'industrie, pour déterminer si un examen plus poussé et une correction sont requis et à quel endroit les effectuer. Un ensemble de rapports sera disponible pour déterminer les enregistrements ayant la plus grande incidence dans les domaines qui nécessitent un examen plus poussé.

8. Voir le document de Saint-Pierre et al. (2011) et de Godbout et al. (2011) dans les références.

Figure 3
Processus des estimations en continu



b) Intégration avec le Système de comptabilité nationale (SCN)

Les membres du personnel qui s'occupent du SCN sont des utilisateurs clés des données du PISE. Un bon nombre de processus, méthodes et concepts du PISE ont été mis en place pour qu'il soit facile d'intégrer les données de ce programme dans les comptes.

Par exemple, l'adoption du contenu harmonisé fondé sur l'industrie, le bien ou le service et les normes financières donne de l'information plus cohérente, conçue pour s'intégrer au cadre conceptuel du SCN. Il ne sera donc pas nécessaire que le personnel chargé du SCN transforme autant de données.

Les outils d'analyse et d'accès aux données du PISE ont été conçus de manière à ce que les employés des secteurs spécialisés et le personnel chargé du SCN puissent utiliser les mêmes interfaces pour accéder aux données figurant dans le dépôt de données du PISE et pour les analyser. Cela facilitera la communication entre les deux groupes puisqu'ils ont tous les deux la même fenêtre sur les données et le même accès à toute l'information pertinente pour expliquer les données, comme les commentaires des répondants ou les notes explicatives ajoutées par les employés des secteurs spécialisés.

11. Modèle de gouvernance

La gouvernance du PISE s'inspire de la structure éprouvée qui a été utilisée avec succès pour régir l'EUE pendant de nombreuses années. Cela dit, des ajustements ont été apportés pour tenir compte de l'introduction de nouveaux programmes d'enquête. En raison du grand nombre de nouveaux programmes, plusieurs autres participants ont dû être ajoutés aux réunions.

L'objectif de la structure de gouvernance du PISE est d'assurer une cohérence et une transparence des décisions par une série de réunions axées sur l'action. Cet objectif ne peut être atteint que si le personnel qui assiste aux réunions a le pouvoir de représenter ses programmes et de prendre des décisions. Il faut également établir le nombre maximum de représentants au-delà duquel le bon fonctionnement de ces comités est compromis.

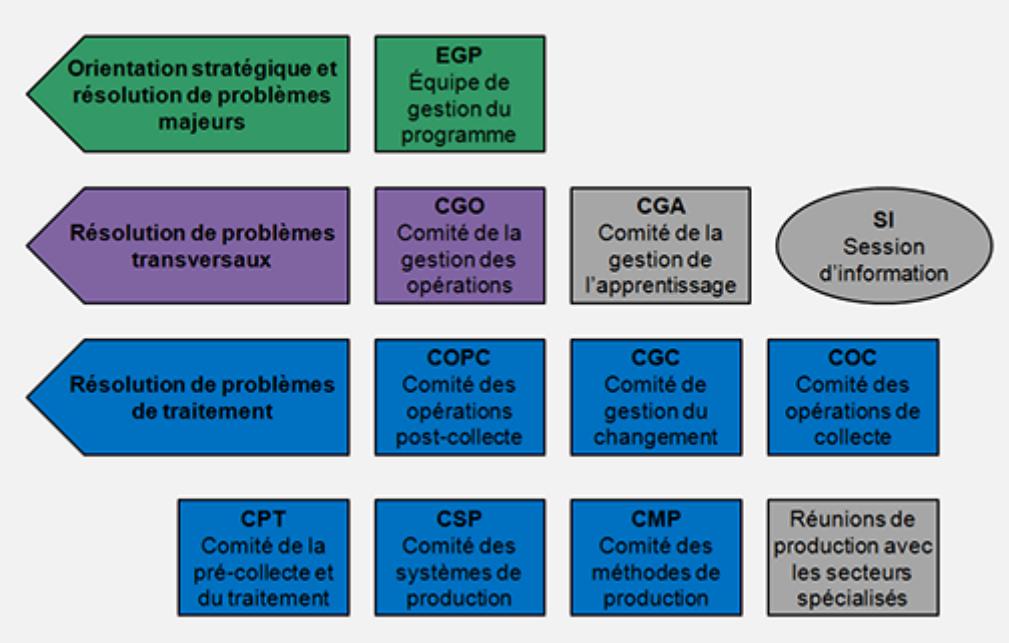
a) Structure des comités

La figure 4 illustre la hiérarchie des comités du PISE. Au total, huit comités interreliés régissent ensemble toutes les activités opérationnelles du PISE. Chacun des comités a un mandat précis et un pouvoir décisionnel particulier. Au sommet de la hiérarchie, l'équipe de gestion des programmes est responsable de toutes les décisions stratégiques. Elle se compose des cadres supérieurs qui représentent de nombreuses divisions opérationnelles et spécialisées.

De bonnes communications entre les comités sont essentielles pour que le PISE fonctionne sans anicroches et le comité de gestion du changement est un maillon essentiel dans cette chaîne. C'est grâce à la gestion du changement que les principaux intervenants sont avisés des problèmes et des risques et qu'ils sont priés de prendre part à la résolution des problèmes.

Des communications claires sont tout spécialement importantes lorsqu'il faut apporter des changements au processus. L'outil organisationnel JIRA, qui a été instauré dans le processus de gestion du changement du PISE, est le mécanisme qui permet de tenir les partenaires au courant des problèmes et risques liés au projet. Or, il ne s'agit pas simplement d'un outil qui permet de transmettre de l'information. Le système de gestion du changement et des problèmes JIRA permet également au personnel d'attribuer des tâches, de documenter des problèmes et des solutions, et même de faire remonter les problèmes jusqu'au personnel supérieur, au besoin.

Figure 4
Structure de gouvernance



b) Système d'information de gestion

Les enquêtes s'intégrant dans le PISE profiteront de l'accès à une foule de rapports normalisés pour évaluer l'état d'avancement des processus et déterminer si les activités opérationnelles sont sur la bonne voie.

Le personnel de la Division de la statistique des entreprises compile et met à jour un calendrier intégré des activités indiquant les principaux jalons, à partir des consultations avec les partenaires opérationnels et des domaines spécialisés. Ce calendrier est une composante importante du système d'information de gestion. Il sert à vérifier que les activités sont bien harmonisées et à repérer les éventuels goulots d'étranglement lorsque des activités ne sont pas complétées comme prévu.

En plus du calendrier, il existe une série de rapports d'information, notamment les suivants :

- les rapports de collecte, qui affichent les progrès en ce qui concerne les taux de réponse et l'état des vérifications;
- les rapports d'évaluation de la qualité, qui veillent à ce que les données circulant entre les processus soient complètes et qu'elles respectent les niveaux de qualité exigés;
- les rapports de vérification et d'imputation, qui fournissent aux analystes des domaines spécialisés de l'information sur la qualité des données après le processus de vérification et d'imputation;
- une série de rapports d'analyse, utilisés pour comparer les données au fil du temps et entre les itérations de traitement;
- les rapports servant à évaluer les processus d'échantillonnage et d'estimation ainsi que la justesse des estimations en continu;
- les rapports d'indicateurs de qualité, qui aident les analystes à se concentrer sur les enregistrements ayant une grande incidence sur les estimations.

Le système de gestion de l'information du PISE est également conçu de manière suffisamment souple pour permettre l'ajout ou la modification des rapports en fonction des besoins des programmes. Toutes les divisions partenaires pourront y avoir accès afin que chacune dispose des derniers renseignements disponibles pour prendre des décisions éclairées.

12. Conclusion

Le projet du PISE a été un défi à bien des égards, mais pas aussi important que l'effort nécessaire pour négocier l'application de stratégies génériques communes aux programmes qui utilisaient depuis toujours des solutions personnalisées. La combinaison d'une forte gouvernance, de la transparence dans la prise de décisions et de la participation des partenaires à l'élaboration des solutions s'est traduite par une approche fructueuse. Des communications fréquentes à tous les niveaux et par de nombreux moyens ont tenu les intervenants bien informés et engagés envers le projet.

Le PISE a instauré de nombreux changements dans toutes les activités liées à la collecte des données, au traitement des données et à la compilation des estimations. Tout, depuis l'examen du contenu jusqu'à la façon dont les différents secteurs qui s'occupaient du travail interagissaient les uns avec les autres, a été modifié. Nul doute qu'au cours des prochaines années, il faudra constamment l'améliorer, à mesure que nous en apprendrons plus sur ce qui fonctionne bien et ce qui doit être ajusté. Comme dans le cas du programme de l'EUE, il faudra plusieurs années avant que le PISE n'arrive à maturité et n'atteigne un équilibre. En ce sens, l'engagement continu de tous les intervenants sera la clé, tout comme le soutien de la haute direction.

Le PISE procurera de nombreux avantages. Les méthodes et processus d'enquête mis à jour profiteront des toutes dernières technologies des systèmes disponibles. Les données seront plus cohérentes dans l'ensemble des programmes grâce à l'application de concepts harmonisés et de contenu de questionnaires normalisé, ce qui facilitera l'intégration des extrants dans le Système de comptabilité nationale. On allégera le fardeau de réponse en utilisant des données administratives de façon plus intensive. Les répondants disposeront d'une option de production de rapports électroniques et il ne sera pas nécessaire de faire autant de suivi de la collecte. L'analyse sera plus facile grâce à l'introduction d'estimations en continu et à l'application d'outils communs pour valider les données. Et de nombreux autres programmes profiteront de l'intégration dans le modèle du PISE.

13. Références

Brodeur M., Koumanakos, P., Leduc, J., Rancourt, E., Wilson, K. *(2006)*. The integrated approach to economic surveys in Canada, Statistics Canada. Catalogue 68-514.

Enterprise Statistics Division. (2010). Integrated Business Statistics Program Blueprint, Statistics Canada internal document.

Godbout, S., Beaucage, Y., Turmelle, C. (2011). Achieving Quality and Efficiency Using a Top-Down Approach in the Canadian Integrated Business Statistics Program, Paper for the Conference of European Statistics, Slovenia, 2011.

Hostetter, E. (2013). IBSP Conceptual Framework, Statistics Canada internal document.

Martineau, P. (2013). Smart Tax Replacement, Statistics Canada internal document.

Martineau, P. (2012). Use of the Chart of Accounts in determining the content of Statistics Canada business surveys, Statistics Canada internal document.

Mills, F., Godbout, S., Bosa, K., Turmelle, C. (2013). Mulivariate Selective Editing in the Integrated Business Statistics Program, Joint Statistical Meeting.

Saint-Pierre, É., Bricault, M. (2011). The Common Editing Strategy and Data Processing of Business Statistics Surveys. Paper for the Conference of European Statistics, Slovenia.

Sambell, C., Maffett, P., Felx, P. (2012). The Processing of Rolling Estimates, Findings and Issues, Statistics Canada internal document.

Turmelle, C., Beaucage, Y. (2013). Le Programme intégré de la statistique des entreprises, Utilisation d'un plan de sondage à deux phases pour produire des estimations fiables, Recueil du Symposium 2013 de Statistique Canada.

Turmelle, C., Godbout, S., Bosa, K. (2012). Methodological Challenges in the Development of Statistics Canada's New Integrated Business Statistics Program, International Conference of Establishment Surveys, 2012.