

Solution de données alternatives – Expérience de la Division des prix à la production

Gaétan Garneau and Mary Beth Garneau¹

Résumé

Au cours de la dernière décennie, la Division des prix à la production de Statistique Canada a élargi son programme d'Indices des prix à la production pour les services, tout en continuant d'améliorer son programme d'Indices des prix à la production de la construction et des biens. Même si la majorité des indices des prix sont fondés sur des méthodes d'enquêtes traditionnelles, des efforts ont été déployés pour utiliser davantage de données administratives et d'autres sources de données, afin de réduire le fardeau pour nos répondants.

Ce manuscrit qui est axé principalement sur les programmes des prix à la production fournit également des renseignements sur l'importance grandissante des sources de données alternatives à Statistique Canada. Il présente aussi les défis opérationnels et les risques auxquels les bureaux statistiques pourraient faire face s'ils comptaient de plus en plus sur les produits de tiers. Finalement, il présente les outils en cours d'élaboration afin d'intégrer les données alternatives, tout en recueillant des métadonnées.

Mots Clés : données; alternatives; administratives; défis; risques; prix; production; indices.

1. Introduction

1.1 Description

Au cours de la dernière décennie, la Division des prix à la production de Statistique Canada a élargi son programme d'Indices des prix à la production pour les services (IPPS), tout en continuant d'améliorer son programme d'indices des prix de la construction et des biens. Même si la majorité des indices des prix sont fondés sur des méthodes d'enquêtes traditionnelles, des efforts ont été déployés pour utiliser davantage de données alternatives qui comprennent les données administratives et les autres sources de données. Cette recherche pour des sources de données qui ne sont pas des données d'enquêtes traditionnelles cadre bien avec l'effort du Bureau pour diminuer le fardeau pour nos répondants ainsi que stabiliser ou diminuer les coûts de collecte pour l'organisme. Ces nouvelles sources de données prennent de nombreuses formes, par exemple, les prix de catalogue obtenus via l'Internet, les fichiers de données obtenus de tierces parties et les microdonnées des enquêtes sur les chiffres d'affaires.

La deuxième section résume les diverses sources de données alternatives accessibles par les Bureaux de statistique nationaux (BSN) ainsi que Statistique Canada, avec les défis que posent l'utilisation de ces données et les solutions possibles. Cette section présente également quelques exemples de la façon dont Statistique Canada utilise ces données alternatives pour produire ses indices de prix dans les secteurs des transports, de la finance, de la construction, des biens et des services professionnels. La troisième section décrit comment Statistique Canada et la Division des prix à la production travaillent à se donner des outils pour intégrer un ensemble varié de données. Enfin, la section 4 conclut le document en soulevant d'autres idées à étudier. Finalement, cet article est une adaptation et une mise à jour d'un document écrit dans le cadre du groupe de travail de Voorburg sur les services (Garneau, 2015).

¹ Gaétan Garneau, Statistique Canada, Pré Tunney, Ottawa, Canada, K1A 0T6 (gaetan.garneau@canada.ca);
Mary Beth Garneau, Statistique Canada, Pré Tunney, Ottawa, Canada, K1A0T6 (marybeth.garneau@canada.ca)

2. Sources de données alternatives

Un grand nombre de BSN, incluant Statistique Canada utilisent une vaste gamme de sources de données qui ne sont pas directement collectées à partir d'enquêtes gérées par le Bureau. Il est d'ailleurs de plus en plus encouragé de regarder pour d'autres sources de données pour diminuer le fardeau de réponse des répondants et les coûts tout en gardant un niveau de qualité exemplaire pour Statistique Canada. En fait très souvent, la qualité des programmes est améliorée par l'utilisation de ces données alternatives directement dans la production ou pour la validation des données obtenues via des enquêtes dites traditionnelles. D'ailleurs, le Statisticien en chef de Statistique Canada a réaffirmé dans son allocution annuelle de mars 2016 toute l'importance des données alternatives (administratives et autres sources de données) dans les programmes du Bureau.

En général, une agence nationale de statistique a un grand contrôle des données régulières d'enquêtes qu'elle compile et qu'elle traite, mais dans le cas des sources différentes de celles obtenues via des enquêtes régulières, il est normal de découvrir que chaque source présente des défis et des risques qui lui sont propres. Dans cette section, nous verrons certains de ces enjeux, nous allons décrire la façon d'atténuer les risques et finalement, nous présenterons quelques exemples.

2.1 Principales sources de données alternatives

Les cinq principales sources de données utilisées à Statistique Canada et plus particulièrement dans le programme des indices de prix à la production du Canada sont :

- prix affichés dans les sites Web
- prix affichés dans des catalogues ou dans des journaux
- fichiers de données de tierces parties du secteur privé
- données administratives ou données d'autres ministères ou d'organismes de réglementation
- fichiers de données d'autres enquêtes

La *Loi sur la statistique* confère à Statistique Canada le pouvoir d'obtenir les données maintenues par les ministères fédéraux, provinciaux ou territoriaux ou les bureaux municipaux, personnes morales, entreprises ou organismes aux fins de l'application de la loi. La loi impose par ailleurs le maintien rigoureux de la confidentialité des données acquise par la Loi sur la statistique. À titre complémentaire, Statistique Canada utilise aussi des renseignements accessibles par le public, sous licence ou pas, avec ou sans frais d'utilisation.

2.2 Prix affichés dans les sites Web

Les prix affichés dans les sites Web des entreprises des échantillons des programmes d'indices de prix peuvent être une source de données précieuses surtout si ces prix (de catalogue) sont semblables ou très peu différents des prix effectifs de transaction. Ainsi, Statistique Canada et la Division des prix à la production recueillent par exemple des renseignements sur les prix ou les tarifs sur plusieurs sites Web comme les prix des services de messagerie. Dans ce cas-ci, les tarifs ou les tableaux de prix des entreprises individuelles sont facilement disponibles. Des outils de recherche de tarifs ou des calculateurs de factures en ligne peuvent aussi servir à obtenir des estimations de prix. Finalement, les prix sont recueillis mensuellement pour un ensemble détaillé de spécifications de prix couvrant la géographie (p. ex., origine et destination du service), le type de colis ou le type de service (express ou non-express).

Pour l'instant, ces données sont recueillies manuellement dans les sites Web. Outre les limitations connues de l'utilisation des prix de catalogue à la place des prix effectifs des transactions, il y a plusieurs facteurs à prendre en compte dans l'utilisation du Web comme source de données. Dans certaines industries, les prix peuvent varier selon le nombre de visionnements ou le dispositif utilisé pour accéder au site Web. Par exemple, des interrogations répétées sur les sites de voyages peuvent être comptées comme une augmentation de la demande, ce qui pourrait amener des hausses de prix. Également, l'utilisation d'un dispositif ou du média pour accéder au site Web pourrait avoir un impact sur le prix. En effet, certains voyageurs ont signalé que les prix donnés sur les appareils mobiles sont parfois plus élevés que les prix donnés via un ordinateur de bureau. Aussi, il faut s'assurer que le site Web est fiable et qu'il sera

accessible pour la collecte continue des prix. Sa disparition pourrait donner bien des maux de tête pour les programmes d'indices de prix comptant sur la disponibilité des prix mois après mois.

La collecte de données sur le Web pourrait éventuellement se faire par des moyens automatisés tels que le moissonnage Web (« web-scraping ») à Statistique Canada. De fait, non seulement le moissonnage Web pourrait réduire les frais de collecte, mais encore il permettrait de recueillir beaucoup plus de données dans un laps de temps plus court. Jusqu'à présent, Statistique Canada n'utilise pas le moissonnage Web pour la collecte régulière de données pour ses programmes. Un projet pilote qui permettra d'en savoir plus sur le sujet est en cours. Depuis le début de cette étude, outre les avantages mentionnés ci-haut, le projet pilote a permis d'identifier certaines difficultés potentielles quant à l'utilisation de cette méthode de collecte.

En premier lieu, sur certaines pages ou sur certains sites Web, le mode de communication des prix utilisé rend difficile ou impossible pour un logiciel de moissonnage Web de trouver les prix, du moins par les méthodes standards connus de Statistique Canada. Par exemple, pour la collecte des prix sur le Web de l'Indice des prix des services ferroviaires de transport, Statistique Canada saisit les prix qui sont affichés dans des images de prix. Pour la collecte de données sur le Web pour les services de messagerie, certains sites n'affichent le prix que lorsqu'on passe la souris sur une zone bien précise de la page. Aussi, pour la collecte de prix pour les services de messagerie, les changements fréquents au site Web sont aussi un obstacle éventuel au moissonnage Web.

En outre, les conditions de certains sites Web interdisent le moissonnage Web automatisé. Une des raisons pour lesquelles les sociétés interdisent le moissonnage Web est que cette pratique pourrait générer trop de trafic et paralyser un site Web. En l'occurrence, cela entacherait la réputation du BSN responsable. Sans compter que le BSN pourrait être tenu responsable des dommages causés.

Enfin, il y a des préoccupations au sujet de la perception publique de l'utilisation que fait le BSN du moissonnage Web. Certains logiciels de moissonnage Web (non utilisés par Statistique Canada) utilisent des méthodes intrusives qui pourraient être perçues comme des atteintes à la sécurité des sites Web. Cette technologie a parfois été associée à des activités illégales ou des attaques malicieuses.

Pour limiter les risques associés au moissonnage Web, une bonne idée serait que les BSN communiquent d'abord avec les propriétaires des sites Web avant le moissonnage. Le moissonnage Web pratiqué ne devrait pas être intrusif ni créer de risque de sécurité pour le BSN. Enfin, la pratique du moissonnage devrait être très limitée afin de ne pas surcharger le serveur du site Web. Statistique Canada suit ces approches pour son étude de faisabilité quant à l'utilisation du moissonnage Web comme méthode de collecte d'information/prix. Finalement, une nouvelle façon est d'obtenir l'accès direct à la base de données qui est à la base du site Web. C'est de celle-ci que les prix sont générés lorsque les utilisateurs entrent des paramètres de recherche pour des prix de produits. Cette interface de programmation applicative (API en anglais pour Application Programming Interface) est une solution qui décrit comment les programmes voulant avoir accès à la base de données peuvent y avoir accès. Avec l'accord du site Web, cette solution est de plus en plus mise de l'avant pour offrir l'accès aux bases de données des sites en question et par le fait même éviter les problèmes reliés au moissonnage du Web.

2.3 Prix affichés dans des catalogues ou dans des journaux

Il s'agit de l'information collectée dans des catalogues de produits ou des journaux. Ces prix peuvent représenter par exemple, des prix de location d'outils demandés ou des moyennes de transactions obtenus de produits ou services sur des marchés locaux ou internationaux. L'accès à ces catalogues et à ces journaux peut dépendre d'une inscription payante ou non. Pour les entreprises sélectionnées, pour un ou des produits sélectionnés dans un échantillon, la sélection de ces prix de rechange représentatifs des prix de transaction sera une façon de diminuer le fardeau de réponse de l'entreprise en question. Nous collectons en ce moment un grand nombre de prix dans les journaux comme le prix de l'or et de l'argent et celui de plusieurs métaux. Pour un ensemble de produits défini, si le prix moyen de transaction obtenu via un marché représente bien l'ensemble des répondants et si le produit est homogène, il peut être utilisé comme une approximation du prix de tous les répondants pour ce même produit. Ainsi, ces données de remplacement épargnent temps et fardeau de la part des répondants. Dans tous les cas, la collecte est souvent faite d'une façon

manuelle et un système de collecte et de capture devra être mis en place pour intégrer ces prix aux systèmes de traitement des BSN (voir la solution identifiée pour la Division des prix à la production dans la section 3).

2.4 Fichiers de données de tierces parties du secteur privé

Les entreprises du secteur privé accèdent à une quantité croissante de données qu'elles achètent elles-mêmes ou qu'elles collectent. Il y a un argumentaire à faire pour l'utilisation de ces données lorsque la collecte d'un ensemble comparable de données coûte plus cher pour le BSN que l'achat ou l'acquisition de ces données. Comme pour l'utilisation des renseignements de toute source de données, le BSN doit évaluer si les données s'alignent sur les classifications et les concepts du programme et si elles satisfont aux critères de qualité des données. Advenant que des données satisfassent aux critères et qu'il est économiquement viable de les obtenir, il faudra aussi évaluer les risques.

Le recours à une tierce partie pour l'obtention de données peut ajouter des risques d'interruptions de communication si l'entreprise met fin au service ou se retire des affaires. Lorsque plus d'une entreprise offre un service semblable, le temps qu'il faut pour acquérir des données de remplacement pourrait être long. L'acquisition pourrait entraîner des bris dans la série lorsqu'on passe d'une source à une autre. Le BSN peut ne pas avoir le contrôle des coûts lorsqu'elle compte sur la valeur marchande pour acquérir des données de tierces parties via des requêtes de services. Par exemple, la Division des prix à la production a eu à faire face à ce genre de risques, dans le cadre d'un processus d'appel d'offres pour l'Indice de prix des ordinateurs et des périphériques ainsi que pour l'Indice de prix des logiciels commerciaux.

Idéalement, un grand nombre des risques cernés plus haut peuvent être atténués dans le processus d'acquisition. Il y a moyen de négocier des contrats avec options de renouvellement annuel pour plusieurs années. Cela assure la stabilité des coûts et de la fourniture des données pour de nombreuses années et garde au minimum le nombre éventuel de solutions de rechange pour faire face à un bris de continuité d'une série. Finalement, un processus de vérification des données sur place pourrait être établi dans le contrat pour s'assurer de la qualité du produit ainsi que l'obtention d'un ensemble complet d'information à propos de ces données (métadonnées).

2.5 Données administratives ou données d'autres ministères ou d'organismes de réglementation

Les ministères et organismes de réglementation sont une source importante de données pour les BSN. Ils peuvent fournir des renseignements détaillés sur de grandes populations à coût minime, ce qui permet aux bureaux de statistiques d'élaborer des produits statistiques sans alourdir le fardeau de réponse. De nombreux pays utilisent les données administratives de l'impôt des sociétés pour les estimations du chiffre d'affaires. Il y a aussi des occasions d'utiliser des données administratives pour les indices de prix à la production.

Pour l'acquisition des données, les gestionnaires de programme de Statistique Canada suivent une directive applicable à l'échelle de l'organisme, la « Directive sur l'obtention de données administratives en vertu de la *Loi sur la statistique* ». Cette directive comprend l'obligation de documenter les dispositions et les conditions d'accès pour veiller à ce que Statistique Canada respecte toutes les exigences de la loi et maintienne la confiance du public dans ses communications sur l'utilisation de données administratives via son site Web. Statistique Canada favorise l'utilisation d'accords formels pour l'acquisition de données, car ils permettent aux deux organismes de comprendre parfaitement les conditions de l'accord².

Certains fournisseurs de données pourraient vouloir inclure une mention particulière sur l'utilisation de leurs données administratives. Les accords devraient fixer des conditions pour la disponibilité future des données afin d'assurer la continuité du programme de données pour le cas où le fournisseur n'aurait plus à collecter les données à des fins administratives. Ainsi, le protocole d'entente entre Statistique Canada et un autre organisme précise explicitement que les données doivent servir uniquement aux fins de la *Loi sur la statistique*. Il énonce que Statistique Canada utilisera les données pour produire des statistiques agrégées, et en particulier le nom de l'indice. Il signale en outre que, si elle

² Directive sur l'obtention de données administratives en vertu de la *Loi sur la statistique*, date de révision : 1^{er} avril 2015

devait mettre fin à la collecte du questionnaire particulier, l'organisme s'engage à travailler à transférer la responsabilité de la collecte à Statistique Canada, si Statistique Canada le demande. De plus, l'organisme en question pourra informer ses répondants de la nouvelle utilisation qui est faite de leurs données.

La directive énonce aussi l'obligation de documenter les données en vue de maximiser la source des données administratives. Le secteur de programme qui repère une nouvelle source possible de données vérifie si les données sont déjà disponibles et utilisées ailleurs à Statistique Canada. Lorsqu'un programme a repéré un autre ensemble de données pouvant servir à d'autres programmes, il y a un processus de consultation pour déterminer les autres utilisateurs éventuels au sein de l'organisme, et leurs besoins.

2.6 Fichiers de données d'autres enquêtes

Enfin, certains indices de prix à la production peuvent être construits avec des données d'enquête existantes. L'ajout de questions ou la modification d'enquêtes existantes présente une solution de rechange efficace et viable au développement de nouvelles enquêtes de prix. L'évaluation de la faisabilité de profiter d'autres enquêtes est une étape naturelle de la recherche d'une solution pour combler les lacunes statistiques.

Dès le départ de la construction d'un indice de prix à la production, on analyse les industries pour connaître leurs principales activités, on analyse le degré d'homogénéité de ses activités ou des mouvements de prix entre activités, et les principales sources de données servant à estimer les produits finaux. Après avoir établi que des données d'une autre enquête sont choisies, et sachant quelle est leur place dans le calcul du prix du produit final, on peut analyser les méthodes à utiliser pour compléter les données afin de calculer les prix. Ainsi, on pourrait désagréger la collecte des données existantes pour produire des niveaux de détail plus fins afin de refléter l'homogénéité des prix et des mouvements de prix des principales activités dans une industrie. Une autre possibilité, lorsqu'une enquête sert à recueillir des données détaillées sur les chiffres d'affaires, consiste à compléter les renseignements par le nombre ou la valeur des transactions afin de calculer un prix unitaire.

L'utilisation des données d'autres enquêtes est parfois possible et efficace, mais il existe diverses limitations. Les données de ventes sont agrégées au lieu d'être au niveau de la transaction. Le couplage des renseignements avec d'autres données, même si les données sont à un niveau d'agrégation détaillée, produit des prix et des indices de valeurs unitaires. On reproche souvent aux indices de valeurs unitaires leur incapacité de refléter les rajustements pour la qualité ou les changements de composition du produit.

Bien que ces caractéristiques ne soient pas l'idéal, la collecte des renseignements nécessaires au calcul des prix de valeurs unitaires a été maintenue, en raison de la possibilité de meilleurs taux de réponse et d'une plus grande facilité de déclaration, et compte tenu des contraintes de ressources. La réduction du besoin de détails allège la déclaration, et est plus favorable pour certains répondants qui pourraient autrement ne pas donner de renseignements sur les prix. En outre, l'exploitation des sources existantes coûte moins cher que la tenue d'une nouvelle enquête faisant appel à des ressources consacrées à cette seule fin. En outre, les données détaillées sur les chiffres d'affaires qui recueillent les prix par transaction constituent une bonne source de renseignements sur les facteurs de pondération au niveau du produit. Cette désagrégation pourrait aider à atténuer les problèmes de biais de valeurs unitaires décrits plus haut (Diewert, W. E. et Peter Von der Lippe (2010)).

À l'heure actuelle, Statistique Canada fait l'expérience de l'élaboration d'indices de prix à l'aide de données d'autres sources. Par exemple, les données de l'Enquête trimestrielle des télécommunications servent à construire les IPPS pour les télécommunications. De même, l'ajout de quelques nouvelles variables à une enquête existante, l'Enquête trimestrielle sur les caisses de retraite en fiducie, permettra probablement de produire un IPPS pour cette industrie.

3. Une solution pour l'utilisation de diverses sources données

La croissance de l'utilisation des données alternatives comparativement aux données de collecte régulière ou traditionnelle et le fait qu'elles proviennent de diverses sources rendent difficile l'intégration de celles-ci dans les systèmes de production. À l'exception des données collectées d'une façon régulière, la Division des prix à la production a relevé plus de 250 sources de données différentes qui alimentent tous ses indices de prix à la production

de services, de la construction et des biens. En ce moment, Statistique Canada travaille à élaborer une solution générique à ce défi pour l'ensemble de ses programmes, car la Division des prix à la production n'est pas la seule division comptant sur la collecte des données alternatives. En attendant ce produit pour l'ensemble du Bureau, la Division des prix à la production a développé une solution qui servira possiblement de modèle pour un futur système corporatif. Pour la solution de la Division des prix à la production, une hiérarchie conceptuelle a été mise de l'avant pour bâtir les composantes du système et pour s'assurer d'une bonne connexion entre les modules de données.

Le *fournisseur de données* est le niveau le plus élevé et représente l'organisme avec lequel nous obtenons les données, c'est le point de départ. Pour un fournisseur de données, nous pouvons avoir plusieurs façons d'obtenir l'information, par exemple, sur plusieurs pages Web ou dans plusieurs publications. Nous appellerons ce deuxième niveau un *article*. Finalement, pour un article en particulier, nous pouvons collecter plusieurs prix pour différents éléments d'intérêt que nous appellerons des *produits*. Voici un exemple pour bien comprendre ce concept:

Fournisseur de données: Agriculture et Agroalimentaire Canada

Article 1 : Page Web principale

Produit 1: Prix d'une douzaine d'œufs, calibre gros, Québec
Produit 2 : Prix d'une douzaine d'œufs, calibre gros, Ontario

Article 2 : Magazine Agriculture

Produit 1: Prix d'un bovin, Québec

En gardant en tête cette hiérarchie, nous avons développé deux modules de données et deux méthodes différentes de saisie de données selon la façon dont les données sont organisées:

Modules de données:

- Registre des sources de données
- Entrepôt de données

Méthodes de saisie:

- Saisie/capture des données de sources non standard
- Adaptateurs personnalisés pour le chargement des fichiers de données structurées

3.1 Registre des sources de données

Le registre des données renferme les métadonnées concernant la source des données, les coûts, les licences et dates de renouvellement, s'il y a lieu. Il inclut également les descriptions et toutes les autres données pertinentes pour la gestion de l'offre des données. Un grand nombre de ces métadonnées répond aux besoins de Statistique Canada pour faciliter la documentation des données administratives requises dans la « Directive sur l'obtention de données administratives en vertu de la *Loi sur la statistique* », mentionnée à la section 2.5.

3.2 Entrepôt de données

L'entrepôt de données permet l'entreposage central de toutes les données provenant des différentes sources alternatives de données pour en faciliter la réutilisation par d'autres programmes de la Division.

3.3 Saisie/capture des données d'une source non standard

La Division des prix à la production a élaboré deux modèles standards ou gabarits pour recueillir les données à l'aide d'outils de collecte existants de l'organisme. Statistique Canada a investi dans un système de questionnaires électroniques pouvant servir aux déclarations en ligne des répondants ou aux interviews téléphoniques assistées par ordinateur. Cette plateforme est maintenant utilisée pour saisir les données de sources non structurées, comme Internet ou les catalogues. Il est ainsi possible d'utiliser un ensemble de variables définies dans le registre et dans l'entrepôt de données pour personnaliser un de deux gabarits pour la capture des données des produits incluant des caractéristiques et des prix.

La section du haut du questionnaire (voir Figure 1 ci-dessous) contient ainsi de l'information sur le fournisseur de données, sur l'enquête qui utilisera ces données, un hyperlien avec la page Web où la personne faisant la saisie pourra trouver de l'information supplémentaire sur le fournisseur de données ou l'article sélectionné. La section du milieu comprend un ensemble de variables qui définissent l'unicité du produit collecté et qui sont utilisées comme variables clés pour lier le produit à l'article pour lequel nous faisons la collecte. Des variables aussi sont disponibles pour identifier des caractéristiques pour un même produit ainsi qu'un encadré donnant plus d'information sur le produit collecté. Juste sous cette section, nous retrouvons les variables identifiant le mois et le produit. C'est à cet endroit que les prix sont entrés avec la possibilité de réviser des prix pour les mois précédents. D'ailleurs, les prix pour les périodes précédentes sont inscrits pour s'assurer que le produit est bien identifié au travers les mois. Ainsi, nous évitons beaucoup d'erreurs de saisie ou d'identification qui auraient pu passées inaperçues advenant que seul le prix de la période courante soit demandé et inscrit (p. ex. en voyant que le prix de la période passée était de 1,20 \$, il y a peu de chance que la personne entrant l'information pour la période courante entre 12,50 \$ au lieu de 1,25 \$.) Finalement, il y a un endroit pour entrer des commentaires pour une vérification supplémentaire par les spécialistes du sujet-matière ou les spécialistes de prix. Il est possible aussi de cocher une case pour attirer l'attention de ces mêmes experts. Cet écran de saisie peut être utilisé pour la capture par des gens spécialisés dans la collecte ou par les spécialistes de prix selon la complexité de la capture. Cet écran de capture est très polyvalent et peut s'adapter à un grand nombre de produits à collecter pour un seul article jusqu'à un maximum de 260 cellules de prix. Tout ça a été construit pour éviter d'avoir à concevoir de nouvel outil de capture à chaque fois.

Figure 1 : Gabarit de l'écran de capture

The screenshot shows a web interface for data collection. At the top, there are navigation links: 'Ressources utiles', 'Contactez-nous', and 'Paramètres du compte'. Below this, the page title is 'Données Alternatives' with a progress indicator at 0%. The main content area is divided into several sections:

- Article Information:** ID Article : 110, Nom de l'article : Gabarit 2, # de données : 8, Origine du fournisseur de données : Externe, Type de source de collecte : Site Internet. A link 'Hyperlien de la source' is provided.
- Remarque pour la collecte:** Données recueillies mensuellement. Nom du répondant-Contact du Répondant.
- Instruction pour la collecte:** A link to 'select for detailed collection instructions'.
- Nom de l'enquête:** Indice des prix des services de messageries et des services de messagers (IPSMMSM).
- Information sur l'enquête:** A link to 'select for more information on the survey'.

The form includes several input fields:

- ID Produit ADI:** 1
- Code Postal Orig.:** QC - H2Y 1C6
- Code Postal Dest.:** NB - E3B 4Y7
- Type de Produit:** Lettre
- Type de service:** Rapide
- Nom du produit:** Prioritaire
- Origine/Adresse de ramassage:** (empty)
- Destination/Adresse de livraison:** (empty)

A text area for 'Description du produit' contains the text: "Prépayé prioritaire" service, Lettre Expresse, Lettre Document, Enveloppe Régulière, sélectionner le poids, de 400 à 500 grammes. Below this, there are two rows of pricing information:

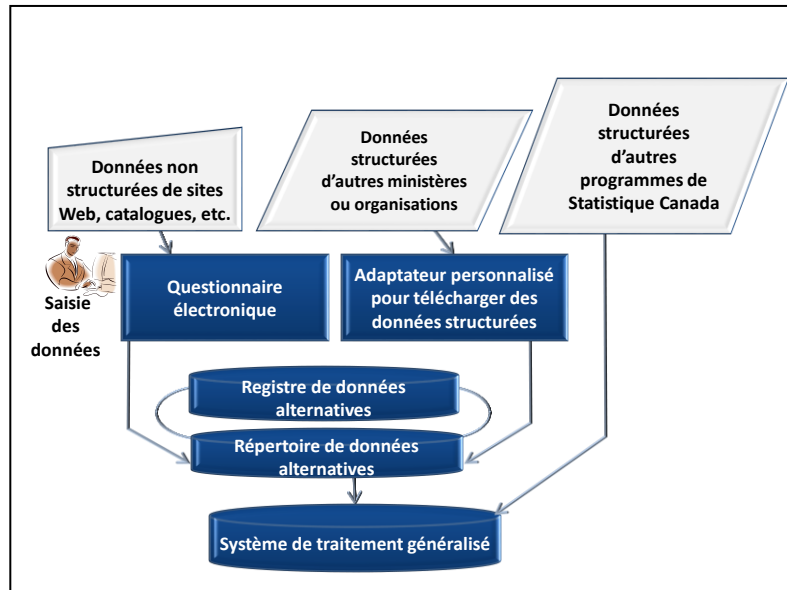
- Jan 2016 Taux de Base:** (empty)
- Jan 2016 Taux de supplément carburant(%):** (empty)
- Dec 2015 Taux de Base:** 10,25
- Dec 2015 Taux de supplément carburant(%):** 2,25

At the bottom, there is a checkbox for 'Révision requise par Sujet Matière' and a text area for 'Commentaires sur la collecte' with a 200-character limit.

3.4 Adaptateurs particularisés pour le chargement des fichiers de données structurées

Là où des fichiers de données structurées sont disponibles de façon récurrente, des adaptateurs particularisés sont créés pour charger les données directement dans l'entrepôt de données. Ces chargeurs permettront d'appliquer des calculs complexes et de créer les variables dérivées que nécessite le système de traitement. L'utilisation de chargeurs est une solution qui représente moins de 10 % des sources d'autres données utilisées dans les prix.

Figure 2 : Modèle opérationnel



4. Conclusion

Malgré les défis et les complexités que représente l'utilisation de sources de données alternatives, les possibilités de combler les lacunes statistiques tout en limitant les coûts et en allégeant le fardeau de réponse sont indéniables. Pour aider à gérer les divers défis associés à ces sources, Statistique Canada a divisé les rôles et responsabilités de divers aspects entre ses divisions spécialisées. Les divisions de programme sont responsables de la qualité de leurs programmes. Il appartient aux secteurs de programme de déterminer, d'évaluer et de mettre en œuvre l'utilisation des données alternatives pour remplacer la collecte directe ou la compléter là où ces données allégeraient le fardeau de réponse et réduiraient les coûts de la collecte, combleraient des lacunes statistiques ou donneraient des données de meilleure qualité que celles obtenues directement des répondants par voie d'enquêtes et de recensements.

La Division de la gestion de l'information est responsable de l'appui et de l'accessibilité de la gestion de l'information à Statistique Canada et doit voir à faire respecter les lois, les politiques et les directives applicables. Elle voit à ce que l'accessibilité, l'utilisation, l'entreposage et la gestion des données par Statistique Canada respectent toutes les exigences de la loi, et notamment les dispositions de confidentialité des lois, et en particulier de la *Loi sur la statistique*, de même que les politiques et directives de Statistique Canada et du gouvernement fédéral.

La Division des données administratives est responsable de l'acquisition, du traitement commun, de l'entreposage et de l'accessibilité des sources de données administratives qui ont une vaste portée, comme les données fiscales. Elle est en outre chargée d'élaborer et d'appliquer les stratégies de l'organisme pour faciliter l'acquisition, l'utilisation, la gestion et l'élimination des données administratives de l'organisme.

Statistique Canada a des dispositions dans sa *Loi sur la statistique* qui lui donnent accès aux données d'autres organismes à des fins statistiques, mais rien n'oblige les ministères fédéraux ni les autres organismes, à consulter Statistique Canada lorsqu'ils créent ou refont des systèmes administratifs.

Un grand défi pour les BSN est la capacité d'exercer une influence sur le type de données qui sont réunies, particulièrement par les ministères, les organismes de réglementation et le secteur privé. L'établissement de relations positives avec les dépositaires de ces données et un historique de production d'indices de qualité pertinents à partir de diverses sources est le premier pas dans cette direction. Avec un accord de concertation, le BSN serait en mesure de formuler des recommandations pour améliorer la convivialité des données alternatives à des fins statistiques.

Pendant qu'il poursuit l'élaboration d'indices de prix à la production pour les services, la construction et les biens, Statistique Canada continuera de rechercher de nouvelles sources de données pour produire des statistiques de grande qualité avec la plus grande efficacité possible.

Bibliographie

Articles de revue

Diewert, W. E. et Peter von der Lippe (2010), "Notes on Unit Value Index Bias", Discussion Paper No. 10-08, Department of Economics, University of British Columbia, Vancouver (Canada).
http://econ.sites.olt.ubc.ca/files/2013/06/pdf_paper_erwin-diewert-10-8-notes-unit-value.pdf.

Garneau, Mary Beth (2015), « Use of alternative data sources in Canadian SPPIs », 30th Voorburg Group Meeting, Sydney, Australia, September 21-25, 2015. <http://www.voorburggroup.org/Documents/2015%20Sydney/4004.pdf>

Loi sur la statistique

Document non publié

Directive sur l'obtention de données administratives en vertu de la Loi sur la statistique, date : 1^{er} avril 2015