

No 12-539-X au catalogue
ISSN 1708-6264



Statistique Canada : Lignes directrices concernant la qualité

Sixième édition – Décembre 2019

Date de diffusion : le 4 décembre 2019

 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 STATCAN.infostats-infostats.STATCAN@canada.ca

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

- | | |
|---|----------------|
| • Service de renseignements statistiques | 1-800-263-1136 |
| • Service national d'appareils de télécommunications pour les malentendants | 1-800-363-7629 |
| • Télécopieur | 1-514-283-9350 |

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 « Contactez-nous » > « [Normes de service à la clientèle](#) ».

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.

Publication autorisée par le ministre responsable de Statistique Canada

© Sa Majesté la Reine du chef du Canada, représentée par le ministre de l'Industrie 2019

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

| | |
|--|-----------|
| Préface | 5 |
| Abréviations | 6 |
| Définitions importantes | 6 |
| Sommaire exécutif | 7 |
| Introduction | 8 |
| Renseignements statistiques et crédibilité des producteurs de données..... | 8 |
| Gestion de la qualité dans le système statistique canadien | 8 |
| But et portée des lignes directrices..... | 8 |
| Lignes directrices pour assurer la qualité des données | 10 |
| Pertinence | 11 |
| Définition, importance et objectifs | 11 |
| Lignes directrices pour assurer la pertinence des données | 11 |
| Indicateurs de la qualité de la pertinence | 11 |
| Actualité et ponctualité..... | 11 |
| Définition, importance et objectifs | 11 |
| Lignes directrices pour assurer l'actualité des données..... | 12 |
| Indicateurs de la qualité de l'actualité..... | 12 |
| Exactitude et fiabilité | 12 |
| Définition, importance et objectifs | 12 |
| Lignes directrices pour assurer l'exactitude des données..... | 13 |
| Indicateurs de la qualité de l'exactitude | 14 |
| Accessibilité et clarté..... | 14 |
| Définition, importance et objectifs | 14 |
| Lignes directrices pour assurer l'accessibilité des données..... | 14 |
| Indicateurs de la qualité de l'accessibilité..... | 15 |
| Intelligibilité..... | 15 |
| Définition, importance et objectifs | 15 |
| Lignes directrices pour assurer l'intelligibilité des données..... | 15 |
| Indicateurs de la qualité de l'intelligibilité..... | 15 |
| Cohérence et comparabilité | 16 |
| Définition, importance et objectifs | 16 |
| Lignes directrices pour assurer la cohérence des données..... | 16 |
| Indicateurs de la qualité de la cohérence | 16 |

| | |
|--|-----------|
| Lignes directrices concernant la qualité des processus | 17 |
| But et portée des lignes directrices sur la qualité des processus..... | 17 |
| Qualité et étapes du processus de production statistique..... | 17 |
| 1. Détermination des besoins | 19 |
| 2. Conception..... | 22 |
| 3. Construction..... | 26 |
| 4. Acquisition des données | 33 |
| 5. Profilage et préparation des données | 38 |
| 6. Intégration, estimation et compilation..... | 42 |
| 7. Analyse..... | 46 |
| 8. Diffusion | 50 |
| 9. Évaluation..... | 52 |
| Bibliographie | 54 |

Préface

La population et le gouvernement du Canada disposent d'un système statistique national (SSN) performant mené par Statistique Canada sur lequel ils peuvent compter pour avoir accès à une source de renseignements statistiques de qualité. La crédibilité du système et la confiance accordée à l'information qu'il produit ne peuvent être établies que si cette dernière correspond aux besoins du pays et représente bien l'environnement dans lequel nous vivons et qu'elle tente de décrire, le tout en s'assurant de ne pas nuire à la vie privée des citoyens. En d'autres mots, l'information doit être pertinente, de grande qualité et obtenue éthiquement.

La qualité est donc un élément de culture essentiel à l'exécution du mandat confié au système statistique canadien, à savoir produire de l'information. La qualité comporte d'ailleurs une caractéristique qu'il est important de bien comprendre : elle se détériore automatiquement en l'absence de mesures proactives. C'est la raison pour laquelle les méthodes et les outils employés pour maintenir une production de données de qualité doivent évoluer au fur et à mesure que notre environnement évolue.

Une telle évolution a résulté en des changements rapides des méthodes d'acquisition de données, en plus d'une conscience accrue des risques au respect de la vie privée associés à ce nouvel environnement. Le système statistique Canadien est sous pressions croissantes de justifier ces activités dans un contexte d'éthique et de protection de vie privée. Conséquemment, il est nécessaire que le SSN se conforme à une approche transparente et cohérente pour aborder ces problèmes qui se rattachent au processus de production statistique.

Le système statistique canadien doit composer avec ces réalités et, sous la direction de Statistique Canada, il fournit depuis longtemps une orientation quant aux principes à appliquer dans les différents processus de production statistique. Ainsi, les expériences et les conclusions à l'égard de ce que constituent des « pratiques exemplaires » ont été regroupées dans un ensemble de lignes directrices concernant la qualité. La première édition des lignes directrices a été publiée en 1985, puis elle a été révisée à quatre reprises, soit en 1987, 1998, 2003 et 2009. La présente édition s'harmonise avec la nécessité d'une actualisation continue des *Lignes directrices concernant la qualité* et elle est basée sur trois motifs : a) présenter un document de référence à tous les autres producteurs de données du système statistique canadien; b) être davantage adaptée à la nouvelle réalité des données administrative en traitant des principaux processus de production statistique ayant cours dans le système; c) faciliter le respect des méthodes actuelles d'assurance de la qualité.

Il est important de comprendre que les présentes lignes directrices ne s'appliquent pas toutes uniformément à chacun des processus de production de données. Il faut donc juger minutieusement leur pertinence et leur importance en fonction des exigences et des contraintes propres à chaque programme et à chaque étape.

Les présentes lignes directrices ne sauraient remplacer l'expertise et le jugement du personnel chargé de produire des données. Cependant, le souci de la qualité qui en émane doit toujours se manifester dans l'ensemble des activités de tout processus de production. Il incombe à tout le personnel intervenant dans les activités statistiques de veiller à ce que la qualité soit considérée comme hautement prioritaire durant la conception et la mise en œuvre des méthodes et des procédures statistiques.

Il convient de remercier et de féliciter les nombreux experts qui ont collaboré à la préparation de cette sixième édition des *Lignes directrices concernant la qualité*, ainsi que le Secrétariat de la qualité, qui a mené et coordonné le projet.



Anil Arora

Statisticien en chef

Abréviations

| | |
|---------|--|
| BMDI | Base de métadonnées intégrée |
| CAQ | Cadre d'assurance de la qualité |
| CCIIM | Centre de collaboration internationale et d'innovation en méthodologie |
| CEE-ONU | Commission économique des Nations Unies pour l'Europe |
| FMGD | Fichier de microdonnées à grande diffusion |
| GSBPM | Modèle générique du processus de production statistique |
| INS | Institut national de statistique |
| OCDE | Organisation de coopération et de développement économiques |
| ONU | Organisation des Nations Unies |
| RDA | Répertoire des données administratives |
| SCDDDE | Système de contrôle de la divulgation et de la diffusion des données économiques |
| SSN | Système statistique national |

Définitions importantes

| Concept | Définition |
|---------------------------------------|---|
| Processus de production statistique | Toute activité visant à acquérir, à recueillir ou à manipuler des renseignements en vue de produire des données statistiques |
| Approche participative | Utilisation de réseaux spécialisés ou recours au grand public pour acquérir des données à des échelles géographiques qui seraient inaccessibles à des chercheurs qui ne sont pas suffisamment nombreux ou qui ne peuvent pas être présents à plusieurs endroits en même temps. |
| <i>Web Mapping</i> | Technique d'extraction du contenu des sites Web en vue de le transformer en données utiles dans un tout autre contexte. |
| Profilage de données | Processus consistant à examiner les données disponibles dans une source existante (base de données, fichier administratif, site Web, etc.) pour en recueillir des statistiques (moyenne, médiane, fréquence, variation, etc.) et des renseignements sur ses caractéristiques (structure, contenu, format, classifications, etc.). |
| Confidentialité | Protection contre la divulgation de renseignements personnels identifiables concernant une personne, une entreprise ou une organisation. |
| Politique | Consignes officielles qui imposent des responsabilités précises aux ministères. La politique a une portée obligatoire. |
| Directive | Instructions officielles qui obligent les ministères à instaurer (ou à éviter) une mesure particulière. La directive a une portée obligatoire. |
| Norme | Ensemble de mesures, de procédures ou de pratiques opérationnelles ou techniques qui s'appliquent à l'échelle de l'administration publique. La norme a une portée obligatoire. |
| Ligne directrice | Document fournissant orientation, conseils et explications aux gestionnaires ou aux spécialistes fonctionnels. La ligne directrice est une recommandation qu'il est fortement suggéré de suivre. |
| Système statistique national canadien | Système constitué dans la pratique de l'ensemble des ministères et des organismes fédéraux et provinciaux qui produisent et diffusent des données statistiques. Il n'a pas d'existence légale dans la mesure où la <i>Loi sur la statistique</i> du Canada ne fait mention que de Statistique Canada à l'article 3. |
| Vie privée | Droit de se retirer, de ne pas être sujet à une quelconque forme de surveillance, et de préserver l'anonymat de son information personnelle. Dans ce document, le terme vie privée réfère à des renseignements concernant les particuliers. |
| Besoin | Lacune en produits statistiques nécessitant d'être comblée. |

Sommaire exécutif

L'évolution du monde de la statistique et des données au cours des dernières années a engendré le besoin de transformer à la fois la façon d'acquérir et aussi de traiter les données à des fins de production de statistiques officielles. Par souci d'entretenir le niveau élevé de qualité dans les produits et processus statistiques, il est devenu nécessaire de se pencher sur les méthodes et procédures existantes afin d'évaluer leur pertinence, et d'y apporter les modifications nécessaires afin d'assurer le maintien cette pertinence. En effet, la récente et rapide révolution des données a nécessité la création de nouvelles techniques et méthodes afin de préserver la capacité de prendre des décisions éclairées appuyées par ces données. À ce titre, cette nouvelle version des lignes directrices propose des principes directeurs et pratiques exemplaires sur des sujets tels que l'utilisation de données alternatives aux enquêtes traditionnelles ainsi que l'intégration de données.

Bien sûr, les techniques de collecte et de traitement traditionnelles des enquêtes probabilistes demeurent en vigueur et bien représentées dans les nouvelles Lignes directrices concernant la qualité. Avec l'addition de principes directeurs et pratiques exemplaires qui touchent les données alternatives ainsi qu'une plus grande emphase sur les pratiques éthiques, la protection de la vie privée et la proportionnalité, la sixième édition offre maintenant une meilleure représentation de la réalité d'aujourd'hui. À titre d'exemple, plutôt que de parler de collecte de données, il est maintenant question du concept plus vaste d'acquisition de données.

Statistique Canada a également saisi l'opportunité durant cette mise-à-jour de présenter l'information en adoptant une structure similaire au Modèle générique du processus de production statistique (GSBPM). Cette nouvelle façon de présenter l'information guide le lecteur avec plus d'aisance à travers chacune des étapes et sous-étapes du processus de production statistique.

Introduction

Renseignements statistiques et crédibilité des producteurs de données

Les renseignements statistiques sont un élément essentiel au fonctionnement de toute société humaine organisée. L'absence de données de qualité mettrait gravement en péril les processus décisionnels, la répartition des ressources et la capacité des gouvernements, des entreprises, des établissements et du grand public à comprendre les réalités sociale et économique du pays. Un SSN joue un rôle crucial dans la production et la diffusion de renseignements statistiques.

La crédibilité d'un organisme producteur de statistiques officielles repose sur plusieurs facteurs, dont voici les plus importants : la production de données statistiques de qualité, un bon rapport coût-efficacité, la protection des renseignements personnels, la confidentialité, la transparence et un effectif compétent, éthique et motivé qui a une grande expertise dans le domaine des méthodes statistiques.

La qualité prise en compte en fonction de dimensions, en particulier la dimension de la pertinence de l'information qui est produite, revêt une importance fondamentale pour un organisme producteur de statistiques. Si ce dernier était incapable de produire des données de qualité, la confiance des utilisateurs et des fournisseurs s'éroderait rapidement, ce qui rendrait sa mission plus difficile à réaliser. Ainsi, la mise en œuvre d'un ensemble de mécanismes rigoureux permettant de gérer tout ce qui a trait à la qualité s'avère nécessaire dans un SSN.

Gestion de la qualité dans le système statistique canadien

Dans le système statistique canadien, c'est dans le [Cadre d'assurance de la qualité \(CAQ\)](#) que sont décrits les stratégies et mécanismes mis en œuvre pour faciliter et assurer une gestion efficace de la qualité dans tous les programmes statistiques et toutes les initiatives organisationnelles. On y trouve une description de la structure de gestion, les politiques et les lignes directrices, les mécanismes de consultation ainsi que l'approche de réalisation et de gestion des projets se fondant sur l'approche scientifique.

L'efficacité du CAQ ne dépend pas d'un seul processus ou mécanisme, mais bien de l'effet combiné de nombreuses mesures interdépendantes qui s'appuient sur les intérêts professionnels et la motivation du personnel et qui se consolident mutuellement avec comme objectif la satisfaction des besoins des clients. Ces mesures mettent l'accent sur le professionnalisme du personnel et illustrent le souci manifesté à l'égard de la qualité des données. L'un des traits dominants de la stratégie est la synergie attribuable au fait que les nombreux intervenants des programmes travaillent dans un cadre où sont privilégiées la cohérence des processus et l'uniformité des messages.

Les présentes lignes directrices s'inscrivent dans ce cadre. Elles se veulent un document d'accompagnement dans lequel sont décrites les pratiques exemplaires jalonnant toutes les étapes du processus de production statistique. Elles sont destinées aux membres des équipes chargées de l'élaboration et de la mise en œuvre des activités statistiques.

But et portée des lignes directrices

Le principal objectif des *Lignes directrices concernant la qualité* consiste à fournir une liste plus ou moins exhaustive de principes directeurs et de pratiques exemplaires à appliquer au cours d'un processus de production statistique.

Le document traite principalement de la façon d'assurer la qualité grâce à la conception ou à la mise en œuvre efficace et adéquate d'un projet ou d'un programme statistique, du début du projet à la diffusion des données et l'évaluation du processus. Les lignes directrices sont fondées sur les connaissances et l'expérience collective d'un grand nombre de personnes ayant travaillé dans le SSN canadien, en particulier à Statistique Canada.

Le document comporte deux sections principales qui traitent respectivement des questions relatives à la qualité du produit et des questions relatives à la qualité des processus dans une activité de production statistique. Toutefois, il importe de se rappeler que le contexte dans lequel on prépare une activité de production statistique impose certaines contraintes. Chaque nouvelle activité statistique, tout en visant à répondre à certains besoins immédiats en matière d'information, ajoute également des renseignements à une base de données statistiques

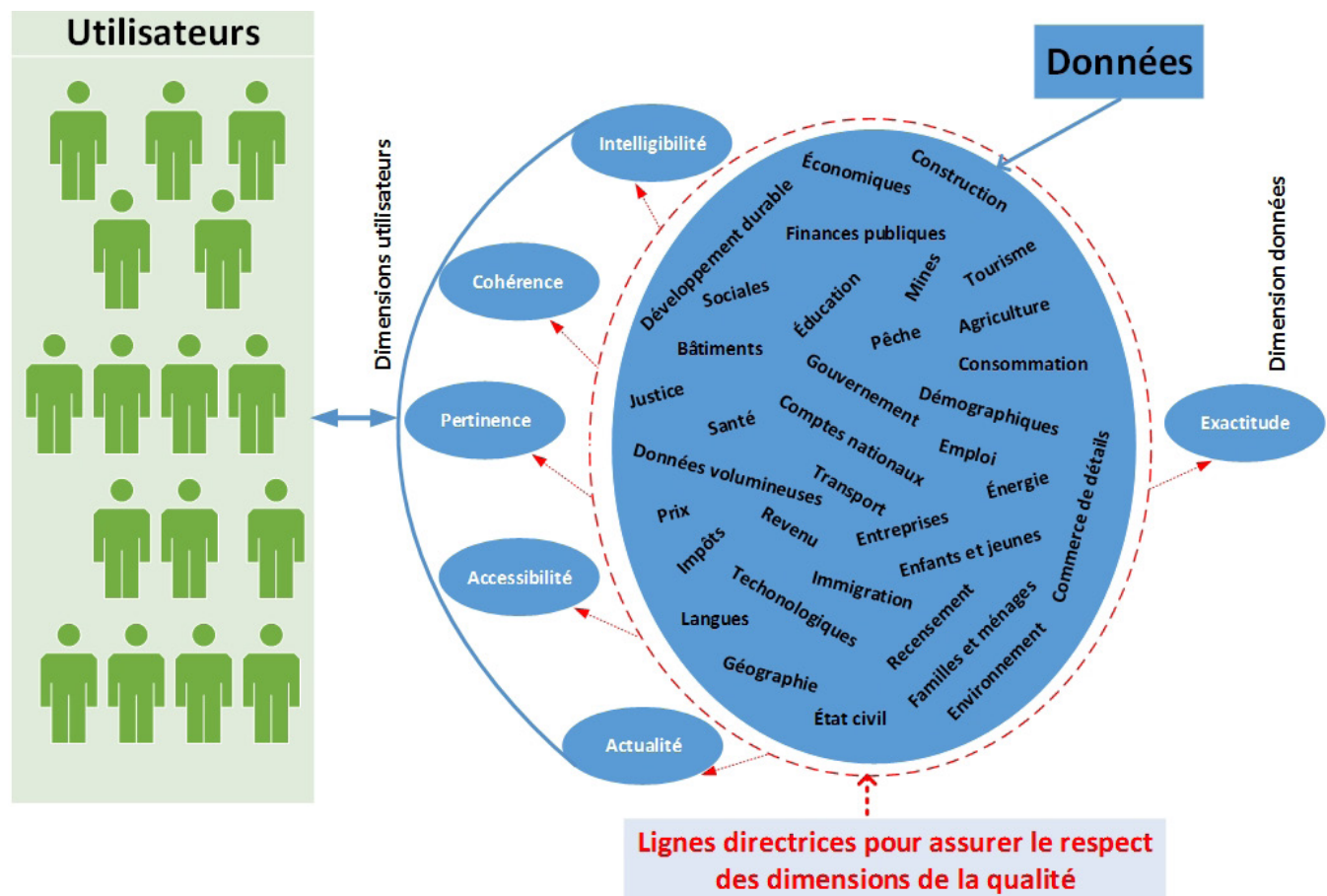
qu'on peut utiliser à des fins qui sont beaucoup plus nombreuses que celles établies au moment de sa conception. Il est donc important de s'assurer que le résultat de chaque activité peut, dans la mesure du possible, être intégré à des données sur des sujets connexes tirées d'autres activités et être utilisé conjointement avec ces données.

Cela démontre la nécessité d'examiner et de respecter les normes statistiques sur le contenu et les domaines qui ont été mis en œuvre pour offrir une cohérence et une harmonie des données dans le SSN. Parmi ces normes, mentionnons les cadres statistiques (p. ex. le Système de comptabilité nationale), les systèmes de classification statistique (p. ex. ceux concernant les industries ou la géographie) ainsi que d'autres concepts et définitions qui précisent les variables statistiques à mesurer. L'utilité des nouvelles données statistiques est accrue lorsqu'on peut utiliser ces dernières conjointement avec des données existantes.

Lignes directrices pour assurer la qualité des données

Dans le système statistique canadien, la qualité des données produites est définie au moyen des six dimensions de la qualité : la pertinence, l'exactitude, l'actualité, l'accessibilité, l'intelligibilité et la cohérence. Ces dimensions sont mentionnées dans l'un des huit principes directeurs du CAQ, qui décrit les stratégies mises en œuvre pour faciliter et assurer une gestion efficace de la qualité dans tous les programmes statistiques et toutes les initiatives organisationnelles. Ces six dimensions sont abordées dans le CAQ de façon à décrire ce que les producteurs de statistiques officielles doivent faire pour assurer la qualité à l'échelle de leur institution. Dans les présentes lignes directrices, elles sont abordées de telle sorte qu'elles permettent d'évaluer l'adéquation des données produites. La figure ci-après illustre l'étendue des lignes directrices et montre que, quel que soit le type de données à produire, leur mise en application sert à assurer le respect des dimensions de la qualité.

Figure 1



Les sections suivantes décrivent les façons, dans le système statistique canadien, dont la qualité des données produites est assurée, à savoir la prise en charge et la gestion efficace des six dimensions. Chaque section porte sur une dimension. La première sous-section présente une définition, suivie d'une courte explication sur l'importance de la dimension (le pourquoi) et les objectifs visés. La deuxième sous-section fournit des lignes directrices visant à guider le personnel des organismes producteurs de données dans l'accomplissement de leurs tâches. Finalement, puisque la gestion des dimensions de la qualité a une incidence directe sur la qualité des produits statistiques, la troisième sous-section propose des indicateurs de la qualité permettant non seulement de mesurer la dimension en question, mais également de l'évaluer par rapport aux objectifs à atteindre.

Pertinence

Définition, importance et objectifs

Par « **pertinence** des données statistiques », on entend la mesure dans laquelle les besoins réels des utilisateurs sont satisfaits. Pour être qualifiées de pertinentes, les données doivent éclairer les utilisateurs sur les questions les plus importantes à leurs yeux. C'est par cette dimension que les instituts nationaux de statistique (INS) ajustent la direction de leurs programmes lorsque cela s'avère nécessaire. Toutefois, son évaluation est subjective étant donné qu'elle dépend des divers besoins des utilisateurs.

C'est dans cette dimension qu'on tient généralement compte des ressources humaines, matérielles et financières qui doivent être utilisées pour chaque programme. Ainsi, le grand défi de tout INS est de jauger les besoins conflictuels des utilisateurs actuels et potentiels afin de concevoir un programme qui répond le mieux possible aux principaux besoins compte tenu des contraintes en matière de ressources. Les mécanismes qui servent à assurer la pertinence permettent non seulement d'exercer un contrôle de base sur les autres aspects de la qualité, mais également d'évaluer les besoins des utilisateurs par rapport aux autres dimensions.

Il faut donc jauger l'importance relative des besoins en données et explorer les différentes alternatives. Cette notion est importante puisqu'elle permet de déterminer l'équilibre optimal entre le besoin en données des utilisateurs et le besoin de protection de la vie privée des répondants. Bien qu'il puisse être ardu d'évaluer cet aspect, il est primordial d'y consacrer un effort approprié.

Lignes directrices pour assurer la pertinence des données

- ▶ Établir et tenir à jour la liste des utilisateurs actuels et potentiels de tous les produits statistiques de l'organisme producteur de données;
- ▶ Maintenir des liaisons bilatérales et multilatérales avec les principaux utilisateurs;
- ▶ Procéder à des évaluations périodiques des programmes statistiques;
- ▶ Confirmer et réajuster les priorités des programmes dans le cadre d'un processus de planification stratégique en examinant les besoins en matière d'information émergente et la pertinence des programmes courants;
- ▶ Tenir compte de l'incidence qu'auront sur d'autres programmes connexes les rajustements et les modifications apportés à un programme statistique;
- ▶ Énoncer clairement les buts et les objectifs des activités statistiques;
- ▶ Mesurer les utilisations des données et le niveau de satisfaction des utilisateurs;
- ▶ Déterminer les paramètres en vue d'établir l'équilibre optimal entre le besoin en données et l'impact sur la vie privée des canadiens, et choisir la meilleure option parmi celles considérées.

Indicateurs de la qualité de la pertinence

- ▶ Existence d'une liste actualisée des utilisateurs des données du programme;
- ▶ Niveau de conformité des données du programme avec les normes régionales et internationales;
- ▶ Nombre de consultations en ligne des produits du programme;
- ▶ Taux de satisfaction des utilisateurs des produits du programme;
- ▶ Pourcentage d'utilisateurs qui utilisent réellement le produit statistique.

Actualité et ponctualité

Définition, importance et objectifs

Par « **actualité** des données statistiques », on entend le délai entre le point de référence (ou la fin de la période de référence) auquel se rapporte l'information et la date à laquelle les données sont disponibles. La ponctualité, quant à elle, se définit comme la différence entre le moment planifié pour sa disponibilité et sa disponibilité réelle.

Une gestion efficace de cette dimension permet aux données produites et diffusées d'être en phase avec la réalité présente. L'actualité de l'information doit être planifiée dès l'étape de la conception de l'activité statistique. Sa planification est très souvent basée sur des compromis entre la précision recherchée, le coût que l'organisme producteur consent à assumer et la disponibilité des données. L'actualité est donc une caractéristique importante. On doit assurer un suivi régulier du calendrier et du déroulement des activités afin de prévenir certaines détériorations et de détecter des manquements aux bonnes pratiques pouvant mener à des cas extrêmes de non-ponctualité.

Lignes directrices pour assurer l'actualité des données

- ▶ En premier lieu, s'assurer qu'il n'y a pas de mandat législatif rattaché à la diffusion des données. S'il y en a un, cela peut diminuer la flexibilité du calendrier de mise en œuvre et la capacité de le modifier;
- ▶ Planifier et élaborer un calendrier pour chacune des étapes du processus de production et avoir des mécanismes de suivi réguliers de leur mise en œuvre;
- ▶ Négocier des ententes de partage avec les fournisseurs de données qui permettent un transfert à récurrence planifiée et ponctuelle;
- ▶ S'entendre avec les fournisseurs de données sur un calendrier de disponibilité de leurs bases de données;
- ▶ Tenir compte des facteurs externes dans la planification et l'élaboration du calendrier de diffusion;
- ▶ Annoncer, au préalable, le calendrier de diffusion et les changements de dates prévus à long terme;
- ▶ Adopter, dans le cas de certains programmes, la stratégie de diffusion des données préliminaires, suivie de la diffusion de chiffres révisés et définitifs qui sont plus actuels;
- ▶ Utiliser les nouvelles technologies pour raccourcir les délais de diffusion;
- ▶ Évaluer la faisabilité de diffuser des données d'abord provisoires, puis révisées;
- ▶ Informer les utilisateurs au sujet des contraintes associées à l'actualité;
- ▶ Dans le cas d'un décalage du calendrier prévu, donner des explications et fixer à l'avance une nouvelle date de parution;
- ▶ Évaluer l'importance des besoins des utilisateurs en fonction du moment auquel ces besoins doivent être comblés.

Indicateurs de la qualité de l'actualité

- ▶ Existence d'une date de début et de fin pour chaque étape du processus de production;
- ▶ Existence d'un calendrier de diffusion;
- ▶ Existence d'un plan de suivi et de contingence;
- ▶ Existence de documentation sur l'évaluation des besoins;
- ▶ Degré de ponctualité selon le calendrier de diffusion prévu (exprimé en nombre de jours, de mois ou d'années de retard par rapport à la date de diffusion initiale);
- ▶ Période entre la date de référence et la date de diffusion;
- ▶ Période entre la date de diffusion provisoire et la date de diffusion de la version définitive.

Exactitude et fiabilité

Définition, importance et objectifs

Par « **exactitude** des données statistiques », on entend la mesure dans laquelle l'information décrit correctement le phénomène qu'elle doit mesurer. L'exactitude est habituellement caractérisée en fonction de l'erreur dans les estimations statistiques et elle est traditionnellement décomposée en biais (erreur systématique) et en variance (erreur aléatoire). On peut aussi la définir par rapport aux sources principales d'erreurs susceptibles de mener à des données imprécises (p. ex. la couverture, l'échantillonnage, la non-réponse). Liée à la notion d'exactitude, la fiabilité correspond à la mesure dans laquelle l'information statistique décrit correctement, et uniformément dans le temps, le phénomène qu'elle doit servir à mesurer.

Le principal objectif dans cette dimension est que les données produites rendent compte de la réalité avec le plus de justesse possible. Ainsi, la gestion de l'exactitude nécessite une attention particulière durant les phases de conception, de mise en œuvre et d'évaluation de l'activité statistique. Le niveau d'exactitude atteint est donc fonction des méthodes et des processus adoptés pour détecter et contrôler les erreurs potentielles dans les diverses étapes de la mise en œuvre de l'activité statistique.

En ce qui a trait au degré d'exactitude acceptable, sa détermination et la justification de son choix découlent de l'équilibre optimal entre les besoins des utilisateurs et le niveau de protection de la vie privée nécessaire afin de satisfaire ces besoins, des contraintes en matière de ressources budgétaires, matérielles et humaines, ainsi que de la prise en considération des facteurs contextuels relatifs à l'activité statistique à entreprendre.

Lignes directrices pour assurer l'exactitude des données

- ▶ Appliquer les méthodes appropriées de contrôle de qualité et mettre en œuvre des processus d'assurance de la qualité pour détecter et contrôler les erreurs éventuelles dans les diverses étapes du processus de production;
- ▶ Consulter les producteurs de données administratives pour valider les concepts et pour comprendre le traitement des données dans le cadre d'affaire d'où elles proviennent (utilisation, couverture, cycle de collecte et de réconciliation, valeurs manquantes ou de faible qualité) et faire les ajustements requis;
- ▶ Faire des suivis auprès des producteurs de données si des variables ou des données semblent manquantes ou erronées;
- ▶ Mesurer la qualité de l'appariement des enregistrements provenant de différents fichiers (p. ex. l'appariement de données administratives aux registres statistiques);
- ▶ Minimiser et évaluer les erreurs de couverture et de classement des unités statistiques dans la base utilisée;
- ▶ Effectuer des prétests et des mises à l'essai de toutes les versions des instruments d'acquisition des données (questionnaire, plateforme de saisie, module de transfert, programmes ou logiciel d'imputation et d'estimation, etc.);
- ▶ Privilégier les méthodes d'échantillonnage sans biais ou avec biais minimales, comme les méthodes d'échantillonnage probabilistes;
- ▶ Considérer des méthodes non probabilistes uniquement s'il est possible de le faire dans un cadre d'inférence valide (Beaumont, 2018) ou s'il n'y a aucune autre possibilité;
- ▶ Sélectionner un plan de sondage qui permet d'optimiser l'exactitude des données produites;
- ▶ Vérifier l'échantillon finalement enquêté par rapport au plan initial;
- ▶ Vérifier la représentativité des données produites;
- ▶ Avoir un plan de suivi quant à la collecte pour améliorer les taux de réponse et la qualité des données;
- ▶ Minimiser et évaluer les erreurs de traitement (erreurs de saisie, erreurs de vérification et d'imputation des données, erreurs de manipulation des fichiers, etc.);
- ▶ Vérifier que le calcul des poids correspond à un total connu;
- ▶ Déterminer et décrire les raisons principales de la non-réponse (aucune communication, refus, inadmissibilité, etc.);
- ▶ Évaluer l'ampleur des erreurs dues à la non-réponse;
- ▶ Donner une estimation de la variance qui tient compte du plan d'échantillonnage et des diverses manipulations effectuées sur les données d'enquête (ajustement de la non-réponse, imputation, calibration, etc.);
- ▶ Valider les modèles utilisés et les hypothèses de départ;
- ▶ Utiliser d'autres sources pour comparer, évaluer et valider les estimations calculées;
- ▶ Utiliser les nouvelles technologies pour automatiser le plus possible les procédures et ainsi minimiser les erreurs introduites par les manipulations et l'incohérence des données;
- ▶ Évaluer l'importance relative des besoins des utilisateurs des données et celui en protection de la vie privée des répondants afin d'en déterminer l'équilibre optimal.

Indicateurs de la qualité de l'exactitude

Des sept indicateurs précisés ci-après, les cinq premiers sont généralement liés à l'exactitude de données issues des processus d'enquêtes par échantillonnage. Statistique Canada réalise actuellement des travaux de recherche afin d'élaborer des indicateurs pour les données provenant des autres processus de production statistique. Ces travaux permettront également de vérifier la pertinence d'avoir recours à certains indicateurs utilisés pour mesurer les données d'enquêtes. Les résultats de ces travaux feront partie de la prochaine édition des lignes directrices.

- ▶ Taux de couverture;
- ▶ Taux de réponse;
- ▶ Variance (coefficient de variation);
- ▶ Taux et ratio d'imputation;
- ▶ Variance due à l'imputation;
- ▶ Erreurs de mesure;
- ▶ Erreurs de précocité ou de fin de période (pour les estimations de population).

Accessibilité et clarté

Définition, importance et objectifs

Par « **accessibilité et clarté** des données statistiques », on entend la facilité avec laquelle les utilisateurs peuvent connaître l'existence de l'information (incluant les métadonnées), la localiser, la visualiser et l'importer dans leur propre environnement de travail. Il s'agit également du caractère approprié de la présentation de l'information ou du média au moyen duquel on peut avoir accès aux données. Pour certains utilisateurs, le coût peut également être un aspect de l'accessibilité.

C'est au moyen de cette dimension que les utilisateurs peuvent prendre connaissance du contenu de leurs demandes d'information et savoir si cette information répond bien à leurs besoins. L'accessibilité est très liée à la pertinence dans la mesure où la satisfaction des utilisateurs à l'égard d'un produit statistique doit être mesurée non seulement sur son contenu, mais également sur son mode de diffusion.

L'objectif dans le système statistique canadien en matière de diffusion est d'optimiser l'utilisation de l'information qu'il produit tout en s'assurant que les coûts de diffusion ne réduisent pas sa capacité à recueillir et à traiter les données. Les politiques de diffusion des organismes producteurs et les systèmes de livraison déterminent la plupart des aspects de l'accessibilité.

Lignes directrices pour assurer l'accessibilité des données

- ▶ Mettre en œuvre un système de documentation et d'archivage des données statistiques;
- ▶ Concevoir les produits statistiques en s'assurant qu'ils sont bien intégrés aux systèmes de catalogage de l'organisme producteur;
- ▶ Vérifier périodiquement les dossiers archivés pour s'assurer de leur accessibilité aux utilisateurs;
- ▶ Établir des procédures dès le début du cycle de vie des enquêtes à caractère répétitif pour s'assurer que tous les fichiers importants sont conservés;
- ▶ Produire et mettre en œuvre des procédures pour distribuer des fichiers à usage restreint, s'il y a lieu;
- ▶ Appliquer une méthodologie de contrôle de la divulgation qui minimise les suppressions;
- ▶ Proposer de multiples méthodes d'accès et d'extraction;
- ▶ Mettre à profit les nouvelles technologies de visualisation des données;
- ▶ Rendre les fichiers de données accessibles en prévision des activités analytiques futures;
- ▶ S'assurer que les produits sont conformes aux exigences des politiques et des normes établies par l'organisme en matière d'accessibilité. Sinon, se référer aux politiques en vigueur à Statistique Canada (Politique sur l'accès à l'information, Politique sur l'accès aux microdonnées et Politique sur les faits saillants des publications);

- ▶ S'assurer que les produits sont conformes aux exigences de l'organisme. Sinon, consulter celles mentionnées dans la [Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie](#) de Statistique Canada;
- ▶ Informer les utilisateurs sur les publications disponibles, leur fréquence ainsi que leurs principaux modes de diffusion;
- ▶ Organiser des séances de diffusion officielles de résultats des processus statistiques;
- ▶ Rédiger et diffuser des communiqués clairs et concis qui renvoient au produit statistique;
- ▶ Effectuer des enquêtes sur la satisfaction des utilisateurs concernant la facilité d'accès et les modes de diffusion.

Indicateurs de la qualité de l'accessibilité

- ▶ Existence d'un système d'archivage et d'entreposage du produit statistique;
- ▶ Disponibilité du produit statistique dans plusieurs formats;
- ▶ Existence d'une référence au produit statistique au sein du système d'archivage et d'entreposage;
- ▶ Organisation de séances de diffusion officielles du produit statistique;
- ▶ Communiqués sur le produit statistique;
- ▶ Mesures disponibles quant à la satisfaction des utilisateurs concernant la facilité d'accès et les modes de diffusion.

Intelligibilité

Définition, importance et objectifs

Par « **intelligibilité** des données statistiques », on entend la disponibilité de renseignements supplémentaires et de métadonnées nécessaires à l'interprétation et à l'utilisation appropriée des données. Il s'agit en général de renseignements sur les variables, les classifications et les concepts sous-jacents utilisés, sur les méthodes de collecte et de traitement des données et sur les indicateurs de la qualité des données statistiques.

Le système statistique canadien a notamment pour responsabilité d'offrir suffisamment d'information pour permettre aux utilisateurs d'interpréter efficacement les renseignements statistiques. La gestion de l'intelligibilité porte principalement sur la communication des métadonnées ou des renseignements sur l'information.

Lignes directrices pour assurer l'intelligibilité des données

- ▶ Consulter la politique de l'organisme en matière d'intelligibilité des données. Sinon, se référer à la [Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie](#) ainsi qu'à la Directive sur la documentation des métadonnées statistiques en vigueur à Statistique Canada;
- ▶ S'assurer que la base de métadonnées de l'organisme est à jour. Un gabarit permet de compléter ou de mettre à jour les enregistrements de la Base de métadonnées intégrée (BMDI) de Statistique Canada;
- ▶ Permettre l'accès aux renseignements sur les utilisations des données et les objectifs auxquels elles sont censées satisfaire pour assurer la transparence;
- ▶ S'assurer que des indicateurs de la qualité sont disponibles pour que les utilisateurs soient pleinement informés et puissent connaître les forces et les limites des données.

Indicateurs de la qualité de l'intelligibilité

- ▶ Une copie des documents méthodologiques concernant les données est toujours à la disposition des utilisateurs;
- ▶ Une base de métadonnées à jour est mise à la disposition des utilisateurs. Elle inclut des indicateurs de la qualité;
- ▶ Un rapport d'évaluation du processus statistique est disponible à l'interne.

Cohérence et comparabilité

Définition, importance et objectifs

Par « **cohérence et comparabilité** des données statistiques », on entend la mesure dans laquelle celles-ci peuvent être jumelées et comparées à d'autres renseignements statistiques dans un vaste cadre analytique au fil du temps. L'utilisation de concepts, de classifications et de populations cibles types favorise la cohérence, tout comme l'utilisation de méthodes d'enquêtes communes. Par « cohérence », on n'entend pas nécessairement la concordance numérique parfaite.

Lignes directrices pour assurer la cohérence des données

- ▶ Tenir à jour les concepts, les définitions, les nomenclatures et les méthodologies;
- ▶ Tenir à jour un document des méthodologies et un canevas d'acquisition des données administratives et des données provenant des opérations de terrain;
- ▶ Utiliser une base commune pour toutes les enquêtes de même type;
- ▶ Utiliser des questions communément formulées;
- ▶ Documenter les répercussions de l'introduction de nouvelles classifications et de nouveaux concepts sur les séries statistiques;
- ▶ Se conformer aux normes établies. Sinon, documenter clairement les causes de leur non-utilisation et les répercussions sur la comparabilité et la cohérence du produit;
- ▶ Préciser, au besoin, les différences entre les concepts et les méthodes utilisés à différentes périodes pour les mêmes variables;
- ▶ Évaluer et documenter les effets dus aux différences entre les méthodes utilisées à des périodes différentes;
- ▶ Mesurer les différences entre les pratiques locales, les pratiques régionales et les pratiques internationales, s'il y a lieu.

Indicateurs de la qualité de la cohérence

- ▶ Conformité des concepts, des définitions et des classifications utilisés aux normes établies;
- ▶ Souscription aux normes régionales et internationales en matière de méthodes statistiques;
- ▶ En cas d'incohérence qui ne peut être expliquée entièrement par l'exactitude, présentation d'une description de cette incohérence.

Lignes directrices concernant la qualité des processus

But et portée des lignes directrices sur la qualité des processus

Le présent chapitre réunit des lignes directrices et des listes de contrôle liées à de nombreuses questions dont il faut tenir compte dans la poursuite des objectifs de qualité que sous-tend l'exécution des activités statistiques. Il traite principalement de la façon d'assurer la qualité des processus de production des données, du début du projet à l'évaluation, la documentation ou la diffusion. Ces lignes directrices sont fondées sur les connaissances et l'expérience collective d'une grande partie du personnel de Statistique Canada et de celui d'autres organismes publics producteurs de données.

Le principal objectif des lignes directrices consiste à fournir une liste exhaustive de principes directeurs et de pratiques exemplaires à appliquer durant le processus de production des données. Pour mieux apprécier la portée des présentes lignes directrices, il importe de définir le sens donné au terme « processus de production statistique » dans le système statistique canadien.

Le terme « processus de production statistique » est un terme générique qui désigne toutes les activités visant l'acquisition et la manipulation de renseignements en vue de produire des données statistiques, à savoir :

1. l'acquisition des données provenant des dossiers administratifs, où les données sont généralement créées ou tirées de documents initialement conservés à des fins non statistiques;
2. le recensement et l'enquête par sondage, qui consistent à recueillir des données respectivement sur toute la population ou sur un échantillon (habituellement aléatoire) de celle-ci;
3. le couplage d'enregistrements, qui consiste à trouver les enregistrements qui sont associés à une même personne ou entité dans un ou plusieurs fichiers ou registres statistiques;
4. les activités statistiques dérivées, qui consistent à faire des estimations par intégration des données provenant de plusieurs sources différentes;
5. l'approche participative, ou *Crowdsourcing*, et le *Web Mapping*, qui sont d'autres types de processus de production de données en pleine expansion, mais qui ne sont pas abordés dans ce document.

Contrairement aux versions antérieures des *Lignes directrices concernant la qualité*, qui étaient beaucoup plus axées sur le recensement et l'enquête par sondage, la présente version concerne les quatre premiers types de processus préalablement décrits, le cinquième étant encore à l'état embryonnaire.

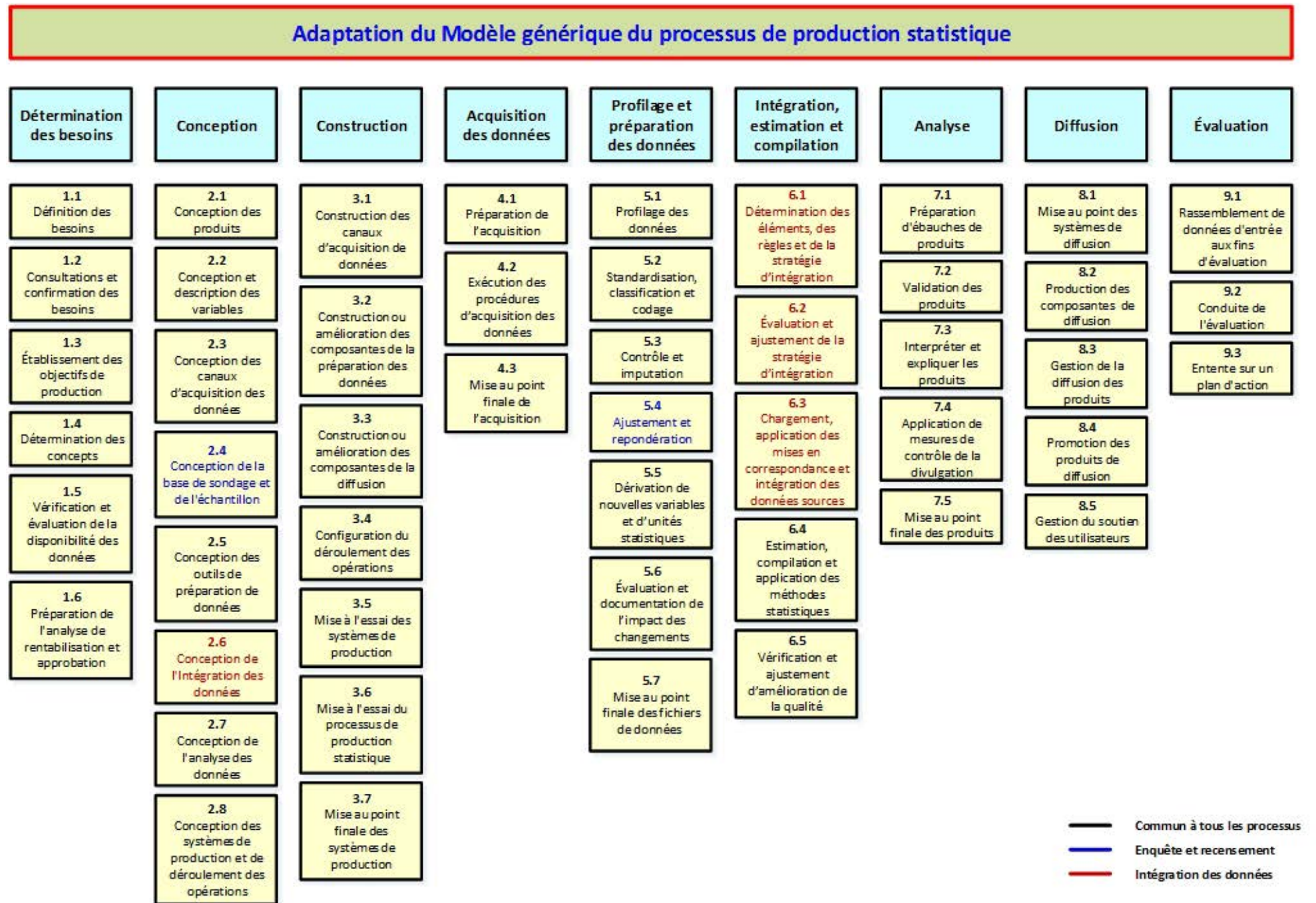
En outre, l'utilisation de lignes directrices dans le processus de production statistique permet de normaliser les pratiques de production, ce qui garantit l'aspect répétitif de leur exécution et de leur continuité dans le temps quant à la mobilité du personnel et qui permet leur analyse, leur évaluation et leur amélioration.

Qualité et étapes du processus de production statistique

Parmi les principes de base à suivre dans tout processus de production figure la nécessité de considérer la qualité à toutes les étapes. Le domaine statistique n'y échappe pas, ce qui pousse les organismes statistiques à modéliser leur processus statistique pour en faciliter le bon déroulement et pour être en mesure d'intégrer et d'appliquer les notions de qualité à chacune de ses étapes. C'est dans ce contexte que le SSN canadien utilise, depuis l'édition d'avril 1987 des *Lignes directrices concernant la qualité*, un diagramme schématique afin d'en représenter le processus. Plus récemment, le SSN canadien a adopté le Modèle générique du processus de production statistique (GSBPM) afin de normaliser ses processus de production statistique. Le GSBPM est une norme internationale provenant d'un groupe de travail conjoint de la Commission économique des Nations Unies pour l'Europe (CEE-ONU), d'Eurostat et de l'Organisation de coopération et de développement économiques (OCDE). Adopté et adapté par de nombreux instituts nationaux de statistique et organismes internationaux, le GSBPM, qui en est à sa cinquième version, propose un modèle de structure pour le processus de production statistique.

Ce modèle de production a été adapté de sorte que non seulement il reflète le contexte de production ayant cours dans le système statistique canadien, mais il est le plus exhaustif possible, dans la mesure où cette adaptation prend en compte certaines spécificités des différents types de processus statistique énoncés précédemment.

Figure 2
Étapes du processus de production statistique



L'adaptation du modèle conserve la proposition initiale du GSBPM, laquelle divise le processus statistique en un certain nombre de phases ou d'étapes. Ainsi, les principes selon lesquels la qualité est de nature multidimensionnelle, doit être intégrée à chaque étape et communiquée de façon transparente mènent logiquement à la conceptualisation du processus de gestion de l'assurance de la qualité sous la forme d'un modèle défini par les aspects liés, d'une part, aux dimensions de la qualité et, d'autre part, aux phases ou aux étapes du processus statistique. Une approche de gestion de la qualité globale exige de prendre en compte tous les éléments de ce modèle.

Le présent chapitre comporte neuf sections correspondant aux neuf étapes que les présentes lignes directrices proposent pour mener à bien tout processus de production statistique dans le système statistique canadien. Chaque section est divisée en sous-sections structurées de même façon : elles décrivent tout d'abord l'étape et les sous-étapes, puis elles présentent les lignes directrices pour assurer la qualité à cette étape et, enfin, elles précisent les dimensions et les indicateurs de la qualité associés à l'étape. En outre, étant donné que chaque processus comporte certaines particularités qui lui sont propres, des lignes directrices sont présentées de manière spécifique pour illustrer les cas d'exception.

1. Détermination des besoins

Description de l'étape et des sous-étapes

La première étape de tout processus de production statistique consiste à en déterminer les principaux objectifs. L'étape de la détermination des besoins s'amorce lorsqu'un besoin de nouvelles statistiques est exprimé ou lorsque des rétroactions sont reçues des parties prenantes concernant des statistiques actuelles.

La détermination et l'énoncé des objectifs doivent inclure les éléments suivants : les exigences particulières à l'égard des données et leurs utilisations, les principales attentes en matière de qualité, les enjeux relatifs à la protection de la vie privée, les contraintes budgétaires, les analyses de rentabilisation et les dates de livraison prévues. C'est aussi à cette étape que sont définis les concepts pertinents, l'unité d'analyse et la population cible. Ainsi, les utilisateurs visés, et même les utilisateurs potentiels, peuvent savoir si les résultats prévus par le projet répondent à leurs besoins et dans quelle mesure ils y répondent.

Généralement, la responsabilité de déterminer les besoins revient à certains services ou certaines divisions spécialisées des organismes producteurs de statistiques officielles, qui travaillent avec d'autres intervenants du SSN touchant à leur domaine d'expertise.

L'étape de la détermination des besoins comporte les six sous-étapes suivantes :

| | | | | | |
|--------------------------------------|---|---|--|--|--|
| 1.1 Définition des besoins | 1.2 Consultations et confirmation des besoins | 1.3 Établissement des objectifs de production | 1.4 Détermination des concepts | 1.5 Vérification et évaluation de la disponibilité des données | 1.6 Préparation de l'analyse de rentabilisation et approbation |
|--------------------------------------|---|---|--|--|--|

1.1. Définition des besoins : Cette sous-étape consiste en la détermination et l'examen des données nécessaires de même que des besoins auxquels elles doivent répondre. Les pratiques ayant cours dans d'autres organismes statistiques régionaux et internationaux produisant des données semblables sont également examinées, notamment en ce qui a trait aux normes et aux méthodes utilisées. La prise en compte des besoins de groupes particuliers d'utilisateurs (personnes handicapées, groupes ethniques, etc.), ainsi que la sensibilité de l'information nécessaire pour combler les besoins, est fortement recommandée à la phase de planification.

1.2. Consultations et confirmation des besoins : De larges consultations avec les différentes parties prenantes de l'activité statistique sont nécessaires à cette sous-étape. Ces consultations permettent non seulement de clarifier et de confirmer leurs besoins dans les moindres détails, mais également de savoir quand, comment et pourquoi il faut produire les données. Dans le cadre des processus de production nécessitant plusieurs types de données d'entrée, tels que les comptes macroéconomiques ou les estimations de population, il est primordial d'élargir la portée des consultations à d'autres programmes similaires ou connexes pour répondre à certaines exigences législatives nationales et internationales en la matière.

- 1.3. Établissement des objectifs de production :** Cette sous-étape porte sur la détermination des produits statistiques devant répondre aux besoins exprimés par les utilisateurs. Durant l'établissement des objectifs de production, il faut généralement prendre en compte l'équilibre entre les mesures de qualité recherchée et les ressources dont dispose l'organisme statistique pour réaliser cette activité.
- 1.4. Détermination des concepts :** C'est à cette sous-étape que prennent forme les concepts devant être mesurés par l'activité statistique. De plus, il n'est pas obligatoire, à la phase de planification, d'avoir des concepts alignés sur les normes statistiques existantes. Cet alignement ainsi que le choix et la définition des concepts et des variables statistiques à utiliser se déroulent à la sous-étape 2.2.
- 1.5. Vérification et évaluation de la disponibilité des données :** Cette sous-étape consiste à vérifier la disponibilité de sources de données pouvant répondre aux besoins des utilisateurs. Avant toute prise de décision, il est obligatoire de procéder à une évaluation des données disponibles pour savoir si elles conviennent à des fins statistiques et si elles répondent aux besoins et aux objectifs exprimés. L'évaluation du cadre légal permettant la collecte et l'utilisation de données provenant de sources de rechange afin d'assurer que les règles d'éthiques et la protection de la vie privée sont respectés est tout aussi nécessaire.
- 1.6. Préparation de l'analyse de rentabilisation et approbation :** Cette sous-étape consiste à documenter et à analyser tous les types possibles de processus de production de ces données statistiques et à déterminer, en fonction des ressources disponibles, celui qui est le plus apte à répondre aux besoins des utilisateurs ou des clients. Généralement présenté sous forme de document d'analyse de rentabilisation, le processus statistique choisi devra satisfaire à plusieurs normes préétablies et être approuvé par la haute gestion avant d'être mis en œuvre.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

- Déterminer et analyser les besoins en matière d'information des utilisateurs internes et externes par rapport à la nouvelle demande d'information ou au changement environnemental requis;
- Comparer les opérations statistiques similaires dans d'autres organismes statistiques régionaux et internationaux, notamment en ce qui a trait aux normes et aux méthodes utilisées;
- Déterminer les besoins particuliers de divers groupes d'utilisateurs, comme les personnes handicapées, les groupes ethniques, etc.;
- Examiner les besoins en matière d'information selon une recherche sur les solutions les plus rationnelles à court terme et à long terme;
- Mener systématiquement de vastes consultations auprès des utilisateurs pour s'entendre sur le contenu et pour solliciter l'appui des partenaires à l'égard du projet;
- Établir et entretenir des relations avec les utilisateurs de données dans tous les secteurs pour rehausser la pertinence de l'information appelée à être produite et pour améliorer la diffusion des produits et des services;
- Déterminer et définir les contraintes opérationnelles telles que la période de référence, les coûts, les ressources et les méthodes d'acquisition des données;
- Établir les objectifs de production conjointement avec les utilisateurs et les principaux intervenants;
- Intégrer à l'énoncé des objectifs les aspects mesurables de la qualité qui doivent être satisfaits;
- Tenir compte des objectifs et des besoins des activités statistiques subséquentes ou parallèles au moment de la détermination des objectifs de production;
- Élaborer les concepts devant être mesurés par l'information statistique qui sera produite (population cible, unité statistique, etc.). Il n'est pas obligatoire, à la phase de planification, d'avoir des concepts alignés sur les normes statistiques existantes;
- Analyser les données disponibles et accessibles dans le secteur en question en fonction de la pertinence, de la fréquence, de la qualité, de l'actualité, etc.;

- Vérifier si le couplage d'enregistrements est une option viable en commençant le repérage des ensembles de données qui pourraient être couplés;
- Évaluer le cadre juridique de toutes les sources possibles en ce qui a trait à leur acquisition et leur utilisation.

► **Processus statistique avec des données administratives**

- Consulter le répertoire de données de l'organisme pour vérifier si les données sont déjà disponibles. Statistique Canada dispose d'un vaste répertoire de données administratives. Les renseignements concernant leur accessibilité se trouvent dans le Répertoire des données administratives (RDA);
- Joindre d'autres organismes publics et privés dans le domaine en question pour s'enquérir de l'existence et de la disponibilité de données administratives pouvant être utilisées dans la production de statistiques. À Statistique Canada, les orientations concernant l'utilisation des données administratives figurent dans la [Politique sur l'utilisation de données administratives obtenues en vertu de la Loi sur la statistique](#) et dans le Manuel sur les données administratives;
- Effectuer une évaluation exploratoire des données avant leur acquisition pour s'assurer qu'elles répondent aux besoins exprimés en matière de concepts et de qualité. Spécifiquement, s'assurer que l'acquisition de ces données est nécessaire (c'est-à-dire que les autres options évaluées ne conviennent pas à satisfaire adéquatement les besoins) et proportionnelle à l'intensité des besoins exprimés. Un document à cet effet intitulé « Guide d'évaluation des données administratives – Phases d'exploration et d'acquisition initiale » a été produit par Statistique Canada;
- Analyser le contexte dans lequel les renseignements administratifs ont été créés (p. ex. législation, objectifs et besoins);
- Évaluer le pouvoir discriminant des variables de couplage disponibles dans les données repérées pour déterminer la faisabilité de produire l'information statistique par couplage d'enregistrements.

► **Processus statistique par intégration de données**

- Pour une production statistique par couplage d'enregistrements :
 - consulter les règles en vigueur dans l'organisme en matière de couplage d'enregistrements. À Statistique Canada, elles sont définies dans la [Directive sur le couplage de microdonnées](#);
 - évaluer la faisabilité de répondre aux besoins en matière d'information en jumelant les données administratives et les données d'enquêtes, tout en tenant compte des contraintes méthodologiques;
 - repérer l'ensemble des données qui peuvent être couplées et l'unité de couplage;
 - examiner la disponibilité et la qualité des variables d'identification communes à tous les ensembles de données sources pouvant faciliter le couplage;
 - évaluer le pouvoir discriminant des variables disponibles pour juger de la faisabilité du couplage;
 - évaluer les besoins en matière d'ajustements si les variables obtenues ne répondent pas complètement aux besoins.
- Pour la production des comptes nationaux :
 - consulter la documentation en rapport aux exigences législatives nationales et internationales dans les domaines des finances publiques;
 - consulter les manuels internationaux de comptabilité macroéconomique et les guides internationaux sur la compilation pour assurer un meilleur encadrement du processus;
 - vérifier la disponibilité des données requises pour la création des comptes, les conditions dans lesquelles ces données seront disponibles, y compris toute restriction concernant leur utilisation.
- Pour la production des estimations et projections de population :

- consulter la documentation en rapport aux exigences législatives nationales et internationales dans les domaines des estimations de la population;
- vérifier la disponibilité des données requises (enquêtes, couplages et fichiers administratifs) pour la création des estimations de population, les conditions dans lesquelles ces données seront disponibles, y compris toute restriction concernant leur utilisation.

► **Processus statistique dans le cadre d'une nouvelle enquête**

- Déterminer dans quelle mesure l'enquête répond aux besoins des utilisateurs et trouver un équilibre raisonnable entre ces besoins et les questions liées au fardeau de réponse et à la protection des renseignements personnels des répondants;
- Examiner la relation d'équilibre entre les statistiques déjà disponibles pouvant répondre aux besoins des utilisateurs, le coût, le temps requis et la valeur ajoutée de la mise en œuvre d'une nouvelle enquête.
- Préparer une analyse de rentabilisation du processus de production aux fins d'approbation. Cette analyse doit comporter :
 - a. la description du processus opérationnel, y compris les lacunes et les problèmes à résoudre;
 - b. les solutions précisant la façon dont le processus de production statistique sera élaboré pour produire de nouvelles statistiques ou des statistiques révisées;
 - c. une évaluation des coûts et des avantages ainsi que les contraintes externes.

Dimension et indicateurs de la qualité associés à cette étape

a. **Dimension de la qualité** : Pertinence.

b. **Indicateurs de la qualité** :

- Existence d'une description des besoins des utilisateurs et de la façon dont ils comptent utiliser les données produites;
- Existence de plans d'analyse qui comprennent un descriptif des tableaux à diffuser;
- Existence d'une analyse de rentabilisation du processus établissant l'écart existant entre les besoins des utilisateurs et les résultats visés.

2. Conception

Description de l'étape et des sous-étapes

L'étape de la conception dans le processus de production statistique consiste à décrire les activités d'élaboration, de création, de développement ainsi que les travaux de recherche nécessaires pour parvenir aux produits statistiques, aux concepts, aux variables, aux méthodologies, aux canaux d'acquisition des données et aux processus opérationnels. On y trouve tous les éléments nécessaires pour définir ou améliorer les produits ou les services statistiques précisés dans l'analyse de rentabilisation. C'est à l'étape de la conception que sont définies les métadonnées essentielles à la poursuite du processus de production et à l'interprétation des produits statistiques. C'est également à cette étape qu'une évaluation attentive des besoins permet de confronter la nécessité d'information avec la sensibilité des données nécessaire afin de combler ces besoins. Ce n'est qu'après une analyse approfondie que cet équilibre optimal est trouvé, ce qui assure que l'effort consacré à la collecte des données est proportionnel aux besoins exprimés.

Les normes nationales, régionales et internationales sont largement utilisées dans les activités de conception. Cela permet non seulement de produire des données comparables, mais aussi de réduire la durée et le coût du processus de production. On encourage donc les organismes producteurs de statistiques à réutiliser ou à adapter les éléments de conception à partir des processus existants.

L'étape de la conception comporte huit sous-étapes, qui peuvent être réalisées de manière séquentielle ou parallèle et qui peuvent être, pour la plupart, itératives. Les sous-étapes 2.4 et 2.6 sont spécifiques et elles se rapportent respectivement au processus par voie d'enquêtes et à celui par intégration de données. Les huit sous-étapes sont les suivantes :

| | | | | | | | |
|---------------------------------------|---|---|--|--|---|---|--|
| 2.1 Conception des produits | 2.2 Conception et description des variables | 2.3 Conception des canaux d'acquisition des données | 2.4 Conception de la base de sondage et de l'échantillon | 2.5 Conception des outils de préparation des données | 2.6 Conception de l'intégration des données | 2.7 Conception de l'analyse des données | 2.8 Conception des systèmes de production et de déroulement des opérations |
|---------------------------------------|---|---|--|--|---|---|--|

- 2.1. Conception des produits :** Cette sous-étape consiste à décrire dans les moindres détails des données statistiques à produire, de même que les services connexes. Elle consiste également à définir et à concevoir la stratégie et la méthode d'intégration des données, s'il y a lieu. Cette sous-étape comprend aussi la description détaillée de la préparation des systèmes et des outils qui seront utilisés à l'étape de la production et de la diffusion de données, de même que celle des méthodes de contrôle de la divulgation et des processus régissant l'accès aux données confidentielles.
- 2.2. Conception et description des variables :** Cette sous-étape vise à définir les variables statistiques qui décrivent le phénomène pour lequel les données sources seront acquises de même que celles qui en seront dérivées (sous-étape 5.5., à savoir la dérivation de nouvelles variables et d'unités statistiques). Elle vise également à déterminer la façon dont ces données sources seront transformées pour être conformes aux concepts et aux conventions liées au type d'activité statistique en question. Il est fortement recommandé, à cette sous-étape, d'utiliser autant que possible les normes et les classifications nationales, régionales et internationales dans la définition des variables statistiques.
- 2.3. Conception des canaux d'acquisition des données :** Cette sous-étape consiste en la détermination et la description des canaux d'acquisition et des méthodes permettant de recueillir les données. Les activités menées dans le cadre de cette sous-étape peuvent varier selon le type de processus statistique. Elles comprennent les interviews assistées par ordinateur, les questionnaires papier, l'auto-dénombrement, les interfaces de transfert de données administratives, l'Internet, etc. Cette sous-étape comprend également la conception de toute entente formelle se rapportant à l'acquisition de données, comme les protocoles d'entente et la confirmation du fondement juridique pour l'acquisition des données.
- 2.4. Conception de la base de sondage et de l'échantillon :** Cette sous-étape s'applique uniquement aux processus de production qui requièrent la collecte des données par échantillonnage. Elle consiste à déterminer et à délimiter la population d'intérêt, à élaborer un plan d'échantillonnage et à déterminer la méthode et les critères d'échantillonnage les plus appropriés au phénomène que l'on veut mesurer. C'est à cette étape qu'a lieu l'analyse pour vérifier si la population cible est couverte par la ou les bases de sondage choisies.
- 2.5. Conception des outils de préparation des données :** Cette sous-étape consiste à déterminer les méthodes statistiques les plus appropriées à appliquer pour le profilage et la préparation des données. Elle comprend généralement les spécifications des routines de vérification de conformité, de détection et de correction d'erreurs, d'imputation, de désaisonnalisation, de modélisation, de déflation ou d'étalonnage, de validation et de mise au point des données finales.
- 2.6. Conception de l'intégration des données :** Cette sous-étape s'applique uniquement aux processus qui requièrent l'intégration de sources multiples de données. Elle consiste à concevoir la stratégie et la méthode d'intégration, l'évaluation de la qualité, la détermination de contraintes et d'identités et les exigences en matière d'accès. Dans le cas de couplage d'enregistrements, la méthode la plus appropriée est déterminée en fonction de la disponibilité de certaines variables d'intérêt.
- En outre, si les données sources n'existent pas, la conception de modèles peut être nécessaire lorsqu'il s'agit d'un processus statistique appelé à produire des comptes macroéconomiques, des estimations de population ou des données par microsimulation.
- 2.7. Conception de l'analyse des données :** Cette sous-étape consiste à déterminer les méthodes statistiques les plus appropriées à appliquer durant l'étape de l'analyse.
- 2.8. Conception des systèmes de production et du déroulement des opérations :** C'est à cette sous-étape que sont conçus les systèmes de production et du déroulement de toutes les opérations du processus, de l'acquisition des données à la diffusion des résultats. Il est important de vérifier la compatibilité des systèmes

existants avant de se lancer dans la conception de nouveaux systèmes. L'exécution de cette sous-étape doit tenir compte de l'ensemble des étapes du processus de production pour garantir leur agencement et éviter les activités redondantes.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

- Décrire de façon détaillée les données statistiques à produire de même que les produits connexes;
- Décrire de manière détaillée les systèmes et les outils qui seront utilisés à l'étape de la diffusion des données;
- Définir et décrire les métadonnées qui accompagneront la diffusion des données et les différents formats sous lesquels ces dernières seront accessibles;
- Définir et décrire les méthodes de contrôle de la divulgation qui seront utilisées et les différents processus qui régiront l'accès aux données confidentielles;
- Définir la population cible, y compris l'ensemble des unités statistiques sur lesquelles on souhaite obtenir des renseignements;
- Énoncer clairement les concepts et les variables du phénomène à mesurer et la façon dont ils seront utilisés;
- Définir les concepts, les variables, les classifications, les unités statistiques et les populations en utilisant, au besoin, les définitions normalisées énoncées dans la [Politique concernant les normes de Statistique Canada](#);
- S'assurer de recourir à la version la plus récente de toutes les variables utilisées pour éviter le phénomène de désuétude;
- Privilégier d'abord la liste officielle des concordances entre certaines classifications internationales normalisées et celles du système statistique canadien, à moins que le terme recherché n'y soit pas consigné;
- Tenir compte, dans le choix des règles d'affectation des noms, des écarts entre la norme et l'usage. Autrement dit, réserver les titres normalisés aux éléments définis dans les systèmes de nomenclatures utilisées;
- Recourir aux concepts, aux variables et aux classifications employés dans le cadre de productions statistiques connexes en l'absence de normes officielles. Consulter la Division des normes de Statistique Canada au besoin;
- Définir les variables dérivées et les rendre opérationnelles;
- Agréger les données à un niveau plus élevé pour répondre aux besoins particuliers de l'analyse ou pour respecter les contraintes de confidentialité ou de fiabilité des données;
- Adopter, si possible, les classes ou les agrégations du niveau supérieur dictées par la norme. Sinon, opter pour une stratégie de regroupement commune, puis documenter les écarts entre la norme et les niveaux de classification ou d'agrégation choisis;
- Recourir à des classifications qui reflètent à la fois les niveaux détaillés et agrégés. Toujours expliquer aux utilisateurs de quelle façon les classifications sont intégrées au niveau supérieur;
- Définir et déterminer le ou les canaux les plus appropriés pour acquérir les données et les métadonnées;
- Déterminer et concevoir la méthode de traitement statistique et d'analyse des données à acquérir. Il s'agit de spécifier les routines de codage, d'imputation, d'estimation, de modélisation, de désaisonnalisation, de déflation, de validation et de mise au point de l'ensemble des résultats;
- Concevoir, acheter ou adopter des systèmes de production et établir le déroulement de toutes les opérations du processus, de la collecte des données à la diffusion des résultats;
- Vérifier la compatibilité des systèmes existants avant de se lancer dans la conception de nouveaux systèmes;

- Tenir compte de l'ensemble des étapes du processus de production pour garantir leur agencement et éviter les activités redondantes.

► **Processus statistique avec des données administratives**

- Concevoir, acheter ou adopter, de concert avec l'organisme fournisseur, une plateforme de saisie de même qu'un manuel de codage lorsque les données existent uniquement sur papier. À Statistique Canada, les principales orientations à cet égard se trouvent dans la [Politique sur l'utilisation de données administratives obtenues en vertu de la Loi sur la statistique](#);
- Concevoir, acheter ou adopter le mode de transmission des données le plus approprié lorsque les données existent en format électronique. Il faut toujours se référer aux directives de l'organisme en matière de conception ou d'adoption du mode de transmission de données. À Statistique Canada, les principales orientations à cet égard se trouvent dans la Directive sur la transmission des renseignements protégés;
- S'assurer d'avoir un document définissant clairement chacune des variables contenues dans le fichier;
- Maintenir une communication avec le fournisseur pour être au courant de tout changement apporté au fichier;
- Collaborer avec les concepteurs de l'organisme fournisseur chargés de remanier les systèmes administratifs ou d'en concevoir de nouveaux pour favoriser l'intégration des exigences statistiques aux systèmes dès le début du projet.

► **Processus statistique par intégration de données**

- Concevoir la stratégie la plus appropriée d'intégration des données et celle de l'évaluation de la qualité pour veiller à ce que l'ensemble des données intégrées soit adapté à l'utilisation prévue;
- Concevoir la stratégie d'intégration en fonction de la disponibilité des variables connues et classées comme identificateurs potentiels d'intégration;
- Prendre en compte les objectifs du projet, l'utilisation finale des données intégrées et les ressources disponibles durant la conception de la stratégie d'intégration;
- Examiner les méthodes et les processus utilisés dans des projets similaires au moment de déterminer la stratégie d'intégration;
- Déterminer les mesures de validation interne et externe pour évaluer respectivement la qualité du processus d'intégration et l'exactitude des données intégrées;
- Concevoir des ajustements visant à améliorer la qualité globale de l'ensemble des données intégrées;
- Concevoir des modèles de création de données au cas où certaines données sources n'existent pas.

► **Processus statistique dans le cadre d'une nouvelle enquête**

- Définir et déterminer les instruments et la ou les méthodes de collecte les plus appropriés dans le cadre de l'enquête à mener;
- Étudier et évaluer soigneusement différents modes de collecte des données;
- Consulter les principaux utilisateurs des données pendant la conception du questionnaire pour bien comprendre la façon dont ils comptent utiliser les données;
- Concevoir le processus de collecte de façon à alléger le fardeau du répondant, à réduire les coûts, à obtenir des données les plus exactes possibles et à accélérer l'obtention de ces données;
- Envisager de jumeler la collecte et la saisie autant que possible (p. ex. au moyen d'une collecte de données en format électronique);
- Utiliser des outils et des méthodes de collecte normalisés (p. ex. écrans standard et questions normalisées) pour faciliter le travail et limiter les risques d'erreurs à la saisie;

- Effectuer une revue des bibliothèques de questions et des outils d'élaboration des questionnaires existants;
- Établir le plan d'échantillonnage et déterminer la méthode et les critères de sélection d'échantillonnage les plus appropriés au phénomène que l'on veut mesurer. Penser à utiliser toute information auxiliaire disponible permettant d'améliorer l'efficacité du plan d'échantillonnage;
- Vérifier et confirmer que la population cible est couverte par la base de sondage choisie ou conçue;
- Choisir une approche méthodologique qui repose sur des bases de sondage multiples (combinaison de deux bases de sondage ou plus) si aucune base de sondage ne peut assurer, à elle seule, la couverture requise de la population cible;
- S'assurer que la base de sondage correspond le plus possible, du point de vue de ses mises à jour, à la période de référence de l'enquête;
- Envisager de recourir à des méthodes d'échantillonnage à plusieurs degrés ou des méthodes d'échantillonnage indirect lorsqu'aucune base de sondage ne permet de traiter la population d'unités visée par l'enquête sans engendrer des coûts trop élevés;
- Utiliser les données complémentaires tirées d'autres sources pour remédier à l'erreur de couverture des bases de sondage;
- Instaurer des procédures pour repérer et corriger d'éventuelles erreurs de couverture dans la base de sondage;
- Utiliser la même base de sondage dans les enquêtes portant sur une même population cible pour augmenter la cohérence, éviter les contradictions, faciliter la combinaison des estimations tirées des enquêtes et réduire les coûts relatifs à la mise à jour et à l'évaluation des bases de sondage. Penser également à utiliser comme base de sondage les registres statistiques déjà existants dans l'organisme;
- Utiliser autant que possible les systèmes de localisation GPS lorsque vient le temps de créer des sous-ensembles géographiques dans la base de sondage;
- Conserver et stocker l'information sur l'échantillonnage et la collecte des données pour qu'il soit possible de coordonner les enquêtes et de mieux gérer les relations avec les répondants ainsi que le fardeau de réponse.

Dimensions et indicateurs de la qualité associés à cette étape

a. **Dimensions de la qualité** : Pertinence, exactitude, cohérence et intelligibilité.

b. **Indicateurs de la qualité** :

- Description détaillée des principaux concepts statistiques, notamment les mesures statistiques, la population, les variables, les unités, les domaines et la période de référence;
- Références exactes concernant les concepts, les variables et les classifications normalisés utilisés;
- Rapport de vérification d'une couverture adéquate de la population cible;
- Description détaillée des instruments et des méthodes d'acquisition, de préparation et d'analyse des données. Dans le cas d'une enquête ou de l'intégration de données, une description détaillée de la méthodologie utilisée est aussi disponible.

3. Construction

Description de l'étape et des sous-étapes

L'étape de la construction permet d'élaborer et de mettre à l'essai l'environnement opérationnel complet de production jusqu'à ce qu'il soit prêt à être utilisé en milieu « réel ». Les produits conçus à l'étape de la conception orientent l'assemblage et la configuration des différentes composantes de cette étape pour créer l'environnement opérationnel. L'élaboration d'une nouvelle composante doit être une exception et doit uniquement viser à combler des lacunes dans le catalogue existant parce qu'elle doit avoir comme objectif la réutilisation des composantes

à grande échelle. Dans le cas des productions récurrentes, cette étape a lieu la première fois, puis à la suite d'un examen ou d'un changement dans la méthodologie ou la technologie.

Cette étape comporte sept sous-étapes, qui se déroulent généralement en ordre séquentiel, mais qui peuvent aussi être exécutées de manière parallèle. Ces sous-étapes sont les suivantes :

| | | | | | | |
|--|--|--|---|---|---|---|
| 3.1 Construction des canaux d'acquisition de données | 3.2 Construction ou amélioration des composantes de la préparation des données | 3.3 Construction ou amélioration des composantes de la diffusion | 3.4 Configuration du déroulement des opérations | 3.5 Mise à l'essai des systèmes de production | 3.6 Mise à l'essai du processus de production statistique | 3.7 Mise au point finale des systèmes de production |
|--|--|--|---|---|---|---|

3.1. Construction ou amélioration des canaux d'acquisition des données : Cette sous-étape consiste à préparer ou à construire les canaux par lesquels les données seront acquises. Cela doit être fait conformément aux spécifications établies à l'étape de la conception concernant le ou les modes d'acquisition retenus. L'exécution de cette sous-étape comporte également la mise à l'essai du contenu et du fonctionnement de l'instrument développé. Les canaux d'acquisition des données prennent plusieurs formes. Il peut s'agir des routines d'extraction ou des plateformes de transmission de données utilisées pour recueillir des ensembles existants de données statistiques ou administratives, ou des interviews sur place ou au téléphone, des questionnaires papier, électroniques ou en ligne, dans le cas d'enquêtes ou de recensements.

Création des instruments de collecte : Dans le cas d'un processus statistique mené au moyen d'une enquête par échantillonnage, cette sous-étape consiste à élaborer le questionnaire, à créer ou à mettre à jour la base de sondage et à sélectionner l'échantillon. En ce qui concerne cette dernière activité, elle consiste à tirer un sous-ensemble d'unités représentatives de la population cible dans la base de sondage. Elle exige une bonne coordination avec d'autres processus de production utilisant la ou les mêmes bases de sondage pour contrôler le chevauchement au besoin. Ceci permet de bien répartir le fardeau des répondants. L'assurance de la qualité de la base de sondage créée ou mise à jour fait également partie des activités exécutées à cette sous-étape du processus.

3.2. Construction ou amélioration des composantes de la préparation des données : Cette sous-étape consiste à déterminer les mécanismes, les outils et les méthodes permettant d'examiner et de recueillir des renseignements sur les données acquises, et de détecter et de corriger les problèmes actuels et potentiels. Il s'agit donc de trouver, en fonction du type de processus statistique, les outils et les méthodes appropriés pour réaliser le profilage, le nettoyage et la transformation des données.

Lorsqu'il s'agit de processus d'intégration de données, cette sous-étape établit et codifie les règles logiques requises pour transformer l'information source obtenue à la sous-étape 3.1 pour respecter les concepts et les présentations requises. Elle décrit également les activités et les outils requis pour mettre en correspondance les données sources transformées.

3.3. Construction ou amélioration des composantes de la diffusion : Cette sous-étape consiste à construire et à développer les composantes et les nouveaux services ou à améliorer ceux existants qui sont nécessaires à la diffusion des résultats obtenus dans le cadre de l'activité statistique. Les composantes et les services de la diffusion doivent prendre en compte la dimension de l'accessibilité en offrant aux utilisateurs plusieurs options d'accès aux produits de données ouvertes de même qu'aux microdonnées.

3.4. Configuration du déroulement des opérations : Cette sous-étape consiste à configurer le déroulement des opérations et les systèmes utilisés tout au long du processus de production statistique, de l'acquisition des données à l'évaluation du processus proprement dit. Cette sous-étape est importante, parce qu'elle permet de matérialiser le déroulement des opérations établi à la sous-étape 2.8.

3.5. Mise à l'essai des systèmes de production : Cette sous-étape comprend la mise à l'essai technique et l'approbation de nouveaux programmes et de nouvelles routines informatiques. Elle permet aussi de confirmer que des routines utilisées dans d'autres processus de production statistique conviennent dans le cas de l'activité en question. La mise à l'essai des services et opérations configurés au point précédent fait également partie des activités exécutées à cette sous-étape.

L'architecture opérationnelle dans de nombreux INS, dont Statistique Canada, offre l'avantage d'être toujours opérationnelle et adaptable à différents processus de production. La mise à l'essai des systèmes de production est donc l'une des sous-étapes les plus simples à exécuter après certains ajustements.

- 3.6. Mise à l'essai du processus de production statistique :** Cette sous-étape consiste à mener un essai pilote du processus de production statistique. Elle comporte généralement l'utilisation d'un petit échantillon des données acquises (20 à 100 enregistrements ou plus) pour mettre à l'essai les instruments et les mécanismes d'acquisition des données, de suivi du traitement et d'analyse des données acquises et ainsi s'assurer que le processus de production statistique fonctionne comme prévu.
- 3.7. Mise au point finale des systèmes de production :** Cette sous-étape comporte toutes les activités de mise en mode de production des outils, des mécanismes et des services développés, y compris ceux qui ont été modifiés, de manière à ce qu'ils soient prêts à être utilisés au moment opportun. Bien que la mise au point finale de certaines composantes du processus de production fasse partie de l'étape précédente, il est important de tester la performance de l'ensemble des composantes à l'intérieur de l'environnement de production pour s'assurer qu'elles fonctionnent comme prévu dans cet environnement.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

• Couverture des bases de données

- Évaluer la fiabilité et la pertinence de différentes bases de données disponibles et accessibles à l'étape de la planification avant de fixer son choix;
- Négocier les changements requis avec les gestionnaires des bases de données choisies pour les activités statistiques dérivées où les changements de couverture peuvent échapper au contrôle du gestionnaire immédiat;
- Apporter des rajustements aux données des bases choisies ou utiliser des données supplémentaires tirées d'autres sources pour compenser l'erreur de couverture;
- Intégrer, dans la documentation de l'activité statistique, la description des populations cibles et observées, les différences entre la population cible et la population observée, ainsi que la description des bases de données et des erreurs de couverture.

• Saisie et codage de données

- Considérer l'utilisation d'un système automatisé ou de techniques d'apprentissage automatisé pour attribuer des codes aux descriptions. Statistique Canada a développé un système automatisé, appelé G-Code, qui attribue des codes aux descriptions;
- Construire une plateforme de saisie des données ou adapter celle existante aux besoins de l'activité statistique de façon à réduire les coûts afférents, à accélérer l'obtention des données et à en garantir la qualité;
- Intégrer des règles de contrôle au système de collecte électronique et en ligne pour empêcher la numérisation des données non valides;
- Intégrer des algorithmes et leurs paramètres dans les systèmes automatisés de saisie des données pour réduire les taux d'erreur;
- Privilégier des processus de saisie et de codage flexibles pour pouvoir faire des changements si les normes en matière d'efficacité le requièrent;
- Utiliser le codage manuel ou automatique s'il est impossible de recourir au précodage, par exemple dans le cas où les données acquises incluent des questions ouvertes;
- Préparer des documents et des activités de formation à l'intention des préposés à la saisie et au codage;
- Mettre à l'essai les systèmes automatisés de saisie des données qui sont basés sur la reconnaissance intelligente des caractères à partir d'images numérisées avant de les utiliser;
- Établir un lien direct avec les métadonnées et les données à saisir;

- Employer des méthodes de contrôle de la qualité statistique pour évaluer et améliorer la qualité des opérations d'acquisition, de saisie et de codage des données;
 - S'assurer que le système automatisé de codage a été testé, modifié au besoin et validé et qu'il est prêt à être utilisé en production.
- **Profilage de données**
 - Déterminer, en fonction du type de processus statistique, les méthodes, les techniques et les outils les plus appropriés pour le profilage des données;
 - Utiliser des logiciels généralisés et réutilisables pour le profilage des données, bien qu'une intervention manuelle soit parfois nécessaire dans certaines situations;
 - Éviter autant que possible d'augmenter le volume des contrôles si ces derniers n'ont que peu d'incidence sur les résultats finaux.
- **Stratégie d'imputation**
 - Inclure dans la composante du système de préparation des algorithmes permettant d'imputer les données;
 - Tenir compte des systèmes généralisés disponibles dans l'organisme lorsqu'on élabore les méthodologies d'imputation parce qu'ils offrent une variété de méthodes déjà programmées. Sinon, il est possible de consulter le système généralisé de vérification et d'imputation BANFF de Statistique Canada, qui offre une variété de méthodes pour des données continues et catégoriques;
 - Consulter LogiPlus, un logiciel intégré dans Microsoft Windows qui permet de créer, d'éditer et de vérifier la logique et de traiter et d'imputer par tables de décision;
 - Déterminer les variables susceptibles d'agir comme variables auxiliaires pour l'imputation des données manquantes en explorant diverses sources (p. ex. données d'enquêtes en cours, données historiques, données administratives, paradonnées);
 - Évaluer la qualité et la pertinence des variables disponibles pour déterminer lesquelles utiliser comme variables auxiliaires;
 - Tenir compte du type de caractéristique à estimer dans le choix des variables auxiliaires et de la stratégie d'imputation pour maintenir les relations d'intérêt entre les variables;
 - Si un système généralisé est utilisé, s'assurer que les méthodes et les paramètres choisis ont été testés, modifiés au besoin et validés et que le système est prêt à être utilisé en production.
- **Règles de confidentialité**
 - Consulter les ressources disponibles dans l'organisme en matière de confidentialité avant de prendre des décisions techniques. Autrement, se référer à la Politique sur la protection des renseignements personnels et la confidentialité de Statistique Canada;
 - Consulter les ressources disponibles dans l'organisme en matière de contrôle de divulgation. Sinon, joindre le Centre de ressources sur le contrôle de la divulgation de Statistique Canada;
 - Utiliser un logiciel généralisé de contrôle de la divulgation bien établi, tel que G-CONFID, plutôt que des systèmes personnalisés;
 - Évaluer la faisabilité d'utiliser les plateformes corporatives telles que le Système de contrôle de la divulgation et de la diffusion des données économiques (SCDDDE).
- **Analyse des données**
 - Déterminer la méthode analytique appropriée aux données avant d'examiner les logiciels offerts pour l'appliquer;
 - Utiliser un logiciel commercial ou non commercial qui est approuvé par l'organisme et qui convient aux types d'analyse choisis;
 - Vérifier s'il est nécessaire de reformater les données pour utiliser le logiciel choisi;

- Consulter, au besoin, le Centre de ressources en analyse de données (CRAD); il s'agit d'une équipe de statisticiens-conseils et de chercheurs au sein de la Direction de la méthodologie de Statistique Canada;
- Consulter, au besoin, le Centre de recherche et d'analyse en séries chronologiques (CRASC), qui regroupe des spécialistes en séries chronologiques qui assurent, à Statistique Canada, l'exactitude des méthodes d'analyse utilisées dans le domaine des séries chronologiques et leur conformité avec les plus récents développements.
- **Diffusion des données**
 - Consulter les spécialistes en communications et en diffusion de l'organisme pour toute question relative à ce sujet;
 - Consulter, au besoin, la politique et la directive de Statistique Canada en ce qui a trait aux composantes de la diffusion des données et aux normes : a) Politique sur la diffusion officielle; b) Directive sur la diffusion des fichiers de microdonnées.
- **Opérations**
 - Créer un comité multidisciplinaire pour approuver le déroulement des opérations;
 - Élaborer un plan de communication entre les membres de l'équipe;
 - Définir et partager les responsabilités selon l'expertise, l'expérience et les ensembles particuliers d'activités ou de composantes;
 - Examiner les plans des composantes élaborées pour déterminer les entrées et les sorties de chacune d'elles ainsi que leurs dépendances;
 - Établir et maintenir un calendrier des livrables et des dépendances entre les différents services corporatifs;
 - Établir des mécanismes de suivis et des mécanismes de mises à jour des activités ou des échéanciers;
 - Formuler la version définitive du déroulement des opérations et du plan pour la mise à l'essai.
- **Mise à l'essai**
 - Mettre à l'essai et approuver les nouveaux programmes et les nouvelles routines élaborés aux étapes précédentes;
 - Confirmer que les routines existantes utilisées dans d'autres processus de production statistique conviennent à celui en cours;
 - Mettre à l'essai les interactions entre les services assemblés et configurés et la prise de mesures pour s'assurer que la solution de production fonctionne en tant qu'ensemble cohérent de processus, de renseignements et de services;
 - Effectuer une acquisition (collecte, transfert, etc.) des données sur une petite échelle pour mettre à l'essai les canaux, le traitement et l'analyse des données et ainsi s'assurer que le processus de production statistique fonctionne comme prévu;
 - Revenir, au besoin, à une étape précédente pour apporter des rajustements aux instruments, aux systèmes ou aux composantes à la suite des essais pilotes;
 - Produire la documentation voulue au sujet des composantes du processus, y compris la documentation technique et les manuels des utilisateurs;
 - Transférer les composantes du processus dans l'environnement de production et s'assurer qu'elles fonctionnent comme prévu dans cet environnement.
- ▶ **Processus statistique avec des données administratives**
 - Consulter la politique, la directive et les lignes directrices de l'organisme en matière d'acquisition de données administratives. Sinon, se référer à celles en vigueur à Statistique Canada :

1. [Politique sur l'utilisation de données administratives obtenues en vertu de la *Loi sur la statistique*](#);
2. Directive sur l'obtention des données administrative en vertu de la *Loi sur la statistique*;
3. Directive sur la sécurité des renseignements statistiques de nature délicate;
4. Lignes directrices sur les données accessibles au public.

► **Processus statistique par intégration de données**

- Consulter les directives de l'organisme en matière d'intégration de données administratives;
- Consulter, au besoin, le Centre de ressources en couplage d'enregistrements (CRCE) de Statistique Canada, lequel fournit des services en matière de couplage d'enregistrements et de prétraitement des enregistrements en vue d'un tel couplage, et G-Coup, un système probabiliste de couplage de Statistique Canada qui a été conçu pour résoudre principalement les problèmes de couplage de fichiers lorsqu'il n'y a pas d'identificateurs uniques.

► **Processus statistique dans le cadre d'une nouvelle enquête**

- **Conception et remaniement de questionnaire**
 - Communiquer avec le Centre de ressources en conception de questionnaires (CRCQ) de Statistique Canada pour toute question relative à l'élaboration et la conception des questionnaires. Le CRCQ est le centre d'expertise et d'expérience dans le domaine de la conception et l'évaluation de questionnaires. Statistique Canada dispose également d'une Politique sur l'élaboration des questionnaires, qui peut être consultée au besoin;
 - Être en communication permanente avec les autres parties prenantes et utilisateurs des données de l'enquête durant tout le processus de conception ou de remaniement du questionnaire pour assurer une compréhension des utilisations future des données;
 - Consulter l'annexe B de la Directive sur l'information des répondants aux enquêtes de Statistique Canada pour s'assurer que les meilleures pratiques ont été adoptées en ce qui a trait aux exigences minimales en matière de communication de l'information aux répondants;
 - S'assurer que le questionnaire ne contient que des questions pertinentes à l'atteinte des objectifs mentionnés;
 - S'assurer que les questions préliminaires d'ouverture s'appliquent à tous les répondants, qu'elles sont faciles à comprendre et qu'elles permettent de savoir si le répondant fait partie de la population observée;
 - Utiliser des mots et des concepts qui ont la même signification pour les répondants et pour les concepteurs du questionnaire. Dans le cas des entreprises, choisir des questions, des périodes de référence et des catégories de réponse qui sont compatibles avec les pratiques de tenue de livres de l'établissement;
 - Utiliser une formulation et une terminologie qui encouragent les répondants à répondre aux questions le plus exactement possible. Le questionnaire doit être axé sur le sujet de l'enquête, être le plus bref possible, comporter des questions qui suivent un ordre logique pour faciliter le rappel des répondants et diriger ces derniers vers la source d'information appropriée;
 - Dans la mesure du possible, utiliser les concepts et la terminologie déjà en usage ou le contenu harmonisé. Au besoin, utiliser les questions déjà élaborées pour assurer la cohérence et la comparabilité par rapport aux résultats d'enquêtes similaires qui ont déjà mesurées adéquatement les mêmes concepts;
 - Vérifier la concordance des versions française et anglaise du questionnaire et, le cas échéant, celle des versions dans d'autres langues. En effet, certaines enquêtes, comme celles concernant les peuples autochtones, sont parfois menées dans d'autres langues;
 - Établir un lien direct entre le questionnaire et le système de métadonnées statistiques de manière à faciliter la saisie des métadonnées durant la phase de collecte;

- Concevoir des questionnaires professionnels, attrayants et faciles à administrer ou à remplir pour le répondant. Si le questionnaire doit être administré par un enquêteur, il doit être convivial pour ce dernier;
 - Donner des titres ou des entêtes à chaque section du questionnaire. Intégrer des directives et des espaces de réponse qui permettent d'obtenir une réponse exacte aux questions;
 - Utiliser des symboles et des polices de caractère pour attirer l'attention des répondants ou des enquêteurs, pour les guider dans les parties du questionnaire qui doivent être lues et pour indiquer où doivent figurer les réponses;
 - S'assurer que les instructions à l'intention des répondants ou des enquêteurs sont courtes, précises et faciles à trouver. Fournir les définitions requises à l'endroit approprié pour que le répondant soit en mesure de répondre correctement aux questions;
 - Utiliser le caractère gras pour mettre en évidence les éléments importants dans les questions. S'assurer que les périodes de référence (date ou période) et les unités de réponse sont claires pour le répondant;
 - Laisser un espace à la fin du questionnaire pour les commentaires additionnels des répondants et intégrer une formule d'appréciation destinée aux répondants;
 - Choisir parmi une multitude de méthodes pour tester et évaluer le questionnaire avant l'implémentation en production. Par exemple, des études qualitatives et quantitatives tel que des groupes de consultation, des entretiens cognitifs ou des enquêtes pilote.
- **Couverture et création de la base de sondage**
 - Tenir compte des systèmes généralisés disponibles dans l'organisme. Consulter, au besoin, le G-Éch, un système généralisé de Statistique Canada qui fournit des fonctionnalités pour l'échantillonnage probabiliste;
 - Utiliser une même base de sondage dans les enquêtes ayant la même population cible pour éviter les incohérences, faciliter la combinaison des estimations tirées des enquêtes et réduire les coûts de la mise à jour et de l'évaluation de la base de sondage;
 - Considérer utiliser les registres déjà disponibles et maintenus dans l'organisme;
 - Tenter d'obtenir le plus de renseignements possible sur la base qui permettra d'améliorer l'efficacité dans les différentes étapes de l'enquête, que ce soit durant l'échantillonnage, la collecte, la vérification, l'imputation ou l'estimation. Au besoin, jumeler la base et d'autres sources pour acquérir davantage de renseignements pertinents;
 - Conserver et stocker les renseignements sur l'échantillonnage, la rotation et la collecte des données pour procéder à une coordination des enquêtes et mieux gérer les relations avec les répondants et le fardeau de réponse. Par exemple, il faut indiquer la fréquence à laquelle chaque unité est sélectionnée par chacune des enquêtes utilisant la même base de sondage;
 - Avoir des procédures de vérification et de validation en vigueur pour diminuer le nombre d'erreurs dans la base;
 - En ce qui concerne les bases aréolaires, utiliser les données géospatiales et les statistiques associées pour créer des aires géographiques raisonnables par rapport aux besoins des enquêtes et au moment de la conception de l'échantillonnage;
 - Vérifier si les renseignements sur le plan de sondage peuvent être intégrés à l'analyse et, le cas échéant, déterminer la procédure à suivre;
 - Utiliser un logiciel conçu particulièrement pour l'analyse de données d'enquête qui tient compte des estimations ponctuelles pondérées et de la variance pour des estimations pondérées par les poids de sondage.

Dimensions et indicateurs de la qualité associés à cette étape

- a. **Dimensions de la qualité** : Pertinence, exactitude, cohérence, actualité, accessibilité et intelligibilité (soit les six dimensions de la qualité du processus de production d'une enquête).

b. Indicateurs de la qualité :

- Mécanismes et outils d'acquisition des données et composantes du processus de production et de diffusion fonctionnels et prêts à être utilisés comme prévu;
- Existence d'un tableau de bord consignait le déroulement des opérations, et les systèmes et les transformations utilisés au sein des processus de production statistique, de la collecte des données à la diffusion.

4. Acquisition des données**Description de l'étape et des sous-étapes**

L'étape de l'acquisition dans le processus de production statistique consiste à acquérir ou à faciliter l'acquisition de tous les renseignements nécessaires (données et métadonnées) au moyen de différents canaux et à les charger ensuite dans un environnement approprié en vue de leur préparation à l'utilisation et de leur analyse. L'acquisition des données est plus qu'une source d'information, c'est aussi le principal lien entre l'organisme producteur et le grand public, qu'il faut convaincre de participer dans le cas des processus d'enquête, et le principal lien avec les autres organismes producteurs, avec lesquels il faut négocier dans le cas de bases de données statistiques ou administratives, de registres ou d'autres bases de données non statistiques.

L'étape de l'acquisition comporte trois sous-étapes, qui se déroulent généralement en ordre séquentiel. Ces sous-étapes sont les suivantes :

| 4.1 Préparation de l'acquisition | 4.2 Exécution des procédures d'acquisition des données | 4.3 Mise au point finale de l'acquisition |
|-------------------------------------|---|--|
|-------------------------------------|---|--|

4.1. Préparation de l'acquisition : Les activités exécutées à cette sous-étape ont pour but d'assurer que le personnel, les mécanismes et les outils sont fin prêts pour l'acquisition des données et des métadonnées selon la stratégie établie. Lorsque le processus de production est récurrent, certaines des activités ne sont pas requises explicitement.

Lorsqu'il s'agit d'un processus d'enquête, cette sous-étape comprend les activités suivantes : élaboration d'une stratégie de collecte; formation du personnel; vérification de la disponibilité des ressources; configuration des systèmes de demande et de réception des renseignements; préparation des instruments de collecte; et préparation du matériel pour le répondant (selon la stratégie de collecte : lettre aux répondants, dépliant de l'enquête, rappel en cas de non-réponse ou refus, information pour les participants aux enquêtes).

Pour les processus de production statistique autres que les enquêtes, cette sous-étape consiste, entre autres, à veiller à ce que les processus, les systèmes et les procédures de confidentialité requis soient en vigueur pour recevoir ou extraire les renseignements nécessaires de la source principale.

4.2. Exécution des procédures d'acquisition : Cette sous-étape consiste en la mise en œuvre de l'acquisition, au moyen des différents instruments conçus et élaborés aux étapes deux et trois du processus, pour recueillir les renseignements prévus. C'est à cette sous-étape que s'établissent les relations avec les fournisseurs en vue de l'acquisition effective des données. Elle comprend également la gestion de la communication avec les fournisseurs de données pour s'assurer que les relations avec l'organisme statistique demeurent positives et constructives.

4.3. Mise au point finale de l'acquisition : Cette sous-étape consiste à télécharger ou à saisir les données et les paradonnées acquises dans un environnement électronique conçu ou modifié à cet effet en vue de la prochaine étape de traitement.

Lignes directrices pour assurer la qualité à cette étape**► Ensemble des processus statistiques**

- **Préparation et exécution des procédures de l'acquisition**

- Utiliser les systèmes intégrés protégés dans la mesure du possible pour obtenir des renseignements de sources externes (répondants, utilisateurs, destinataires de données ou fournisseurs de données);
- S'assurer que des systèmes intégrés sont utilisés pour les transmissions sortantes de renseignements protégés. Toute exemption doit être approuvée par l'unité responsable de la sécurité de l'information dans l'organisme. Consulter la directive de l'organisme en la matière. Sinon, se référer à celle en vigueur à Statistique Canada : Directive sur la transmission des renseignements protégés;
- Signaler à l'unité responsable de la sécurité de l'information dans l'organisme tout problème de sécurité informatique durant la transmission ou la réception des données.

- **Saisie des données**

- S'assurer que les agents de saisie des données reçoivent la formation appropriée et qu'ils disposent des outils de travail adéquats, notamment les manuels d'instructions et tout le matériel individuel nécessaire;
- Exploiter la technologie accessible pour rehausser l'efficacité et la qualité du processus de saisie des données. Par exemple, l'utilisation de la saisie optique des données et l'utilisation d'un scanner sont autant de possibilités fournies par les nouvelles technologies;
- Mettre en œuvre des systèmes de contrôle efficaces pour assurer la sécurité de la saisie, de la transmission et de la manipulation des données;
- S'assurer que les procédures de saisie sont suivies de manière constante dans toutes les unités pour éviter, le plus possible, de commettre des erreurs;
- Intégrer des vérifications automatiques pour les erreurs que l'opérateur de saisie des données peut corriger (c.-à-d. des vérifications qui mettront en évidence les erreurs d'introduction sur clavier). Enregistrer ces cas pour une analyse et un examen ultérieurs;
- Centraliser la saisie pour réduire les coûts et pour faciliter l'apport des connaissances spécialisées disponibles;
- Examiner et analyser les mesures et les résultats du contrôle de la qualité de façon à faciliter la détermination des principales causes fondamentales des erreurs;
- Réquisitionner des préposés à la saisie des données pour qu'ils vérifient l'exactitude de la saisie automatique à partir d'un échantillon. Les résultats de cette évaluation permettront d'améliorer le processus;
- Développer des procédures encadrant le processus de destruction des données qui ne sont plus nécessaires;
- Effectuer une évaluation rétrospective de toutes les opérations de saisie des données et documenter les résultats en prévision d'une utilisation ultérieure.

- ▶ **Processus statistique avec des données administratives**

- Utiliser le protocole d'entente signé avec le ou les fournisseurs des données pour confirmer les modalités de transmission des données et des métadonnées associées;
- Utiliser les mécanismes et les outils existants pour vérifier que toutes les données ont été transmises et qu'elles répondent aux normes préalablement établies durant les négociations avec le ou les fournisseurs;
- Vérifier que les données transmises sont stockées dans le répertoire sécurisé qui a été créé à cet effet;
- S'assurer que l'accès d'autres personnes dans l'organisme aux données administratives respecte le principe du besoin de savoir;
- Communiquer rapidement avec le ou les fournisseurs de données pour tout problème relevé durant la transmission des données.

► Processus statistique dans le cadre d'une nouvelle enquête**• Échantillonnage**

- Utiliser les variables auxiliaires disponibles sur la base de sondage qui sont les plus corrélées aux variables d'intérêt de l'enquête pour construire un plan de sondage efficace (p. ex. au moyen d'une stratification de la base ou d'une attribution judicieuse de l'échantillon);
- Former des strates les plus homogènes possibles par rapport aux variables d'intérêt de l'enquête pour augmenter la précision des estimations;
- Tenir compte autant que possible des domaines d'intérêt durant la construction des strates et la répartition de l'échantillon dans les strates (Singh, Gambino et Mantel, 1994) pour garantir une taille d'échantillon adéquate dans les domaines d'intérêt et ainsi réduire la nécessité de recourir à des méthodes d'estimation pour petits domaines;
- Réduire autant que possible le nombre de degrés d'échantillonnage et la sélection de grappes pour réduire l'effet du plan sur les estimations et pour diminuer la possibilité d'observer des domaines d'intérêt vides;
- Choisir les strates en fonction des objectifs de l'enquête, de la disponibilité des variables de la base de sondage, de la distribution de la variable d'intérêt et du niveau de précision visé pour les estimations;
- Pour les enquêtes longitudinales, choisir des variables de stratification qui correspondent à des caractéristiques qui demeurent stables au fil du temps;
- Créer une strate de grandes unités dont l'inclusion dans l'enquête est certaine pour les populations dont la distribution est fortement asymétrique;
- Pour déterminer la taille d'échantillon et la répartition dans les strates, tenir compte des taux d'erreur de classification prévus dans les unités, de la non-réponse et d'autres anomalies de la base;
- Vérifier, pour chaque type d'unité, les éléments suivants : la disponibilité d'une base adéquate ou la possibilité de créer une base, la facilité de la communication et de la collecte et de la mesure des données, la qualité des données fournies par l'unité et le coût de la collecte;
- Pour déterminer la taille de l'échantillon, tenir compte :
 - o des niveaux de précision nécessaires à la production des estimations de l'enquête,
 - o du type de plan et d'estimateur à utiliser,
 - o de l'accessibilité à l'information auxiliaire,
 - o des contraintes budgétaires,
 - o des facteurs d'échantillonnage (p. ex. mise en grappe, stratification) et des facteurs hors échantillonnage (p. ex. non-réponse, présence d'unités hors du champ de l'enquête, attrition dans les enquêtes longitudinales);
- Utiliser les résultats d'enquêtes antérieures ou similaires pour calculer l'effet de plan de sondage et ainsi déterminer la taille de l'échantillon, dans le cas de certains plans d'échantillonnage complexes exigeant ce calcul. Consulter Gambino (2001), Kish (1965) et Gabler et coll. (2006).

• Enquête périodique

- Tenir compte des naissances et des décès prévus liés aux unités au sein de la population observée en évolution;
- Élaborer des procédures pour surveiller la qualité du plan d'échantillonnage au fil du temps;
- Élaborer une méthode visant à maintenir la taille de l'échantillon et, par conséquent, les coûts de collecte stables lorsque la taille de l'échantillon augmente parallèlement à la taille de la population;

- S'assurer que le plan des enquêtes est le plus souple possible pour pouvoir faire face aux changements futurs, comme l'augmentation ou la réduction de la taille de l'échantillon, la restratification, le ré-échantillonnage et l'actualisation des probabilités de sélection;
 - Mettre en œuvre une stratégie d'actualisation pour le remaniement sélectif des strates qui ont subi une grave détérioration;
 - Former les strates en combinant de petites unités stables liées aux domaines établis lorsque des estimations sont requises pour des domaines d'intérêt précis (p. ex. petites régions géographiques);
 - À l'étape de la mise en œuvre, comparer :
 - o la taille et les caractéristiques de l'échantillon réel par rapport aux attentes,
 - o la précision des estimations par rapport aux objectifs prévus.
- **Enquête longitudinale**
 - Dans le cas d'une enquête longitudinale par panel, déterminer la durée du panel (sa durée dans l'échantillon) en tentant de maintenir un juste équilibre entre, d'une part, la satisfaction des besoins de l'enquête (données relatives à la durée) et, d'autre part, les effets d'attrition et de conditionnement de l'échantillon;
 - Utiliser un plan fondé sur des panels chevauchants (c.-à-d. ayant une durée en chevauchement) lorsqu'il est nécessaire de produire des estimations transversales de pair avec les estimations longitudinales.
- **Préparation et exécution de la collecte**
 - Déterminer les rôles et les responsabilités vis-à-vis tous les aspects de la collecte, y compris la stratégie de communication, la mise en œuvre, l'évaluation, la surveillance, la planification de mesures d'urgence et la sécurité;
 - Concevoir le processus de collecte de façon à alléger le fardeau du répondant, à réduire les coûts et à accélérer l'obtention des données les plus exactes possibles;
 - Mettre en œuvre des mesures en vue de garantir la confidentialité des données à collecter;
 - Exploiter la technologie accessible pour rehausser l'efficacité et la qualité des processus de collecte (p. ex. la collecte électronique des données par Internet et la possibilité d'obtenir des renseignements à partir de différents supports, tels que les tablettes, les cellulaires ou tout autre type d'assistant numérique personnel);
 - Établir des procédures et des mesures de contrôle des échantillons pour chaque étape de la collecte de données (p. ex. livraison et renvoi des questionnaires papier ou électronique, suivi des lacunes ou incohérences et suivi des cas de non-réponse);
 - Élaborer une stratégie de communication avec les répondants pour maximiser les taux de réponse, notamment faire la promotion de l'enquête, informer à l'avance les répondants qu'ils ont été sélectionnés pour prendre part à l'enquête, fournir un courriel et un numéro de téléphone sans frais pour toute question, publier des statistiques clés pouvant inciter les répondants à participer, faciliter la communication des renseignements destinés au public (p.ex. sites Web, guide d'utilisation du questionnaire ou ligne d'information) et remercier les répondants de leur participation;
 - S'assurer que les intervieweurs reçoivent la formation appropriée et qu'ils disposent des outils de travail adéquats, notamment les manuels d'instruction et tout le matériel individuel nécessaire;
 - Développer une méthode de communication permettant aux intervieweurs de communiquer avec le bureau central pendant la collecte;
 - Contrôler attentivement les opérations de distribution et de récupération des questionnaires papier ou en ligne au cours des enquêtes en veillant à ce que chaque unité qui a été sélectionnée dans l'enquête reçoive le questionnaire qui convient;

- Fournir des plans et des outils permettant de gérer activement la collecte de données en même temps qu'elle progresse en utilisant des mesures de productivité et des indicateurs des coûts;
- Choisir le moment opportun pour communiquer avec le répondant ou la personne désignée du ménage ou de l'organisme répondant en consultant les parodonnées produites durant des cycles précédents de l'enquête ou celles d'une enquête similaire;
- Communiquer avec les répondants au moment qui leur convient le mieux et veiller à ce que le nombre d'appels, de visites ou de rappels n'excède pas les limites acceptables;
- Établir un ordre de priorité des unités sélectionnées pour entrer en communication et faire les suivis de sorte à atteindre un bon équilibre entre la qualité souhaitée et les coûts de collecte;
- Mener des recherches pour dépister les répondants dont les coordonnées ne semblent pas à jour;
- Établir des mécanismes permettant aux répondants de faire la mise à jour de leurs coordonnées entre les cycles de l'enquête;
- Laisser le répondant communiquer les données selon la méthode et dans le format qui lui conviennent;
- Assouplir les modalités de déclaration pour alléger le fardeau du répondant et faciliter la collecte de données (p.ex. on pourrait suggérer des modalités particulières aux entreprises qui participent à plusieurs enquêtes à la fois pour qu'elles ne fassent qu'une seule déclaration pour l'ensemble des enquêtes);
- Lorsque le questionnaire est rempli, vérifier l'exactitude des renseignements de couverture et la qualité des données fournies. Idéalement, des vérifications et des validations doivent être intégrées au questionnaire pour permettre des corrections pendant la collecte. Sinon, il peut être nécessaire d'effectuer une contre-vérification pour s'assurer de la qualité des réponses fournies;
- Lorsque le questionnaire n'est pas reçu, il faut chercher à savoir pourquoi (refus, disparition du répondant, fermeture, etc.) et faire un suivi auprès des non-répondants au téléphone, en personne ou par courriel;
- Prendre des rendez-vous au besoin avec le répondant pour le convaincre de participer à l'enquête ou pour faire une interview;
- S'assurer d'informer rapidement le personnel de collecte de l'enregistrement des questionnaires retournés pour éviter les suivis inutiles;
- À la fin de la collecte, communiquer avec un sous-échantillon ou avec l'ensemble des unités non répondantes pour vérifier qu'elles sont bien admissibles à l'enquête. Dans l'affirmative, il est recommandé d'obtenir quelques renseignements essentiels pouvant servir à faire des ajustements pour la non-réponse. Consulter les données administratives courantes sur l'ensemble des unités non répondantes pour obtenir ces renseignements;
- Effectuer une évaluation rétrospective de toutes les opérations de collecte et documenter les résultats en prévision d'une utilisation ultérieure;
- Évaluer les méthodes employées et tirer des leçons pour améliorer chacune des composantes;
- Veiller à recueillir des renseignements pour savoir si les outils et les procédures de collecte exigent des améliorations sur le plan de la qualité en prévision des cycles d'enquête futurs.

Dimension et indicateurs de la qualité associés à cette étape

a. **Dimension de la qualité** : Exactitude.

b. **Indicateurs de la qualité** :

- Taux de couverture des bases de données utilisées;
- Taux de déclaration par personne interposée;

- Taux de réponse totale et partielle;
- Taux de refus;
- Incidence des stratégies de suivis;
- Erreurs imputables à l’admissibilité ou à l’inadmissibilité à l’enquête;
- Répartition des interviews selon leur durée moyenne;
- Incidence du mode de collecte (suggestion, effet de mode);
- Taux de rejet à la vérification;
- Taux d’erreurs à la saisie ou au codage.

5. Profilage et préparation des données

Description de l’étape et des sous-étapes

L’étape du profilage et de la préparation des données consiste à modifier les renseignements sources originaux au moyen d’un processus par lequel des données incomplètes, incorrectes, inexactes ou non pertinentes sont détectées par ce qu’on appelle le profilage et sont ensuite modifiées, corrigées ou supprimées. Elle peut également impliquer la suppression d’erreurs typographiques et la correction de valeurs selon une liste d’entités connues. Une autre pratique courante de préparation des données consiste en leur amélioration par l’ajout de renseignements connexes ou complémentaires. L’objectif de cette étape est de s’assurer que les renseignements sources acquis sont non seulement cohérents, mais aussi prêts à passer aux prochaines étapes.

L’étape du profilage et de la préparation des données comporte sept sous-étapes, qui peuvent se dérouler en ordre séquentiel ou de manière parallèle et qui peuvent également être itératives. Avant toute autre chose, il est essentiel d’effectuer le profilage des données. Les sept sous-étapes sont les suivantes :

| | | | | | | |
|-------------------------------------|---|--------------------------------------|---|--|--|--|
| 5.1 Profilage des données | 5.2 Standardisation, classification et codage | 5.3 Contrôle et imputation | 5.4 Ajustement et repondération | 5.5 Dérivation de nouvelles variables et d’unités statistiques | 5.6 Évaluation et documentation de l’incidence des changements | 5.7 Mise au point finale des fichiers de données |
|-------------------------------------|---|--------------------------------------|---|--|--|--|

5.1. Profilage des données : Cette sous-étape du processus de la préparation consiste à examiner et à récolter le plus de renseignements possible sur les données qui ont été acquises pour découvrir leurs différentes caractéristiques et en dresser un profil. Les principaux objectifs des renseignements recueillis sont les suivants :

- a. vérifier si les renseignements contenus dans les données peuvent être facilement utilisée à d’autres fins;
- b. déterminer les sous-ensembles d’enregistrement et de variables provenant de sources administratives et confirmer lesquels sont dans le champ du programme statistique;
- c. avoir des mesures sur la qualité des données et sur leur conformité par rapport aux normes de l’organisme producteur;
- d. évaluer les risques engendrés par l’intégration de ces données à d’autres sources, y compris les défis liés au couplage d’enregistrements;
- e. vérifier si les métadonnées décrivent correctement les renseignements contenus dans la source de données;
- f. évaluer les modèles de distribution des valeurs et les dépendances fonctionnelles des renseignements contenus dans la source de données;
- g. avoir une bonne compréhension de l’enjeu des renseignements sur les projets à venir pour éviter les mauvaises surprises;
- h. avoir une vue globale des renseignements quant aux utilisations telles que la gestion des bases de données ou l’intendance des données pour améliorer leur qualité.

- 5.2. Standardisation, classification et codage :** Cette sous-étape consiste en la normalisation de la structure, du format et des ensembles de codes et d'abréviations des variables relatives aux données sources pour garantir leur uniformisation, leur comparabilité et leur stabilité dans tous les fichiers. Cette sous-étape comprend aussi les activités permettant d'attribuer une valeur numérique ou alphanumérique à une réponse texte selon un système de classification préétablie. Bien qu'elles soient fréquemment automatisées, les activités exécutées à cette sous-étape requièrent parfois une intervention manuelle parce qu'elles peuvent exiger une prise de décision complexe. La classification et le codage des données permettent de produire des données formatées qui seront utilisées aux étapes subséquentes.
- 5.3. Contrôle et imputation :** Cette sous-étape est une suite logique de la sous-étape du profilage des données. Elle a lieu lorsque, durant le profilage, les données sont déclarées incorrectes, manquantes ou peu fiables. L'imputation est donc le processus qui permet d'attribuer des valeurs de remplacement à des données manquantes, invalides ou incohérentes qui ont été rejetées à l'étape de la validation des données. Ce processus doit suivre certaines règles et il comprend les phases suivantes :
- la détermination de la pertinence d'ajouter ou de modifier des données;
 - le choix de la méthode à utiliser;
 - l'ajout ou la modification des valeurs des données à imputer;
 - la production de métadonnées sur le processus d'imputation.
- 5.4. Ajustement et pondération :** Cette sous-étape sert en particulier au rajustement des poids de sondage et a lieu essentiellement dans le cas des processus statistiques d'enquêtes par échantillonnage lorsque l'échantillon observé est plus petit que celui initialement sélectionné. Ce phénomène est observé en cas de non-réponse totale ou à cause des unités hors champ sélectionnées. Les facteurs d'ajustement doivent être fondés sur la probabilité de réponse de chaque unité échantillonnée dans le cas où la non-réponse est reliée aux variables mesurées.
- 5.5. Dérivation de nouvelles variables et d'unités statistiques :** Cette sous-étape consiste à créer de nouvelles variables et unités statistiques à partir des valeurs de celles obtenues de l'ensemble des données sources. Ces variables et unités sont le plus souvent nécessaires à la création du produit statistique souhaité. La dérivation est réalisée au moyen de l'application de formules mathématiques à l'une ou à plusieurs des variables présentes dans l'ensemble de données ou de l'application de différentes hypothèses de modélisation. Il est parfois nécessaire de procéder à la dérivation de manière itérative, puisque certaines variables qui servent à dériver d'autres variables peuvent elles-mêmes avoir été créées par dérivation. Il importe donc de veiller à ce que les variables soient dérivées dans le bon ordre.
- 5.6. Évaluation et documentation de l'incidence des changements :** Cette sous-étape consiste à évaluer et à documenter l'incidence des différentes transformations opérées tout au long des cinq sous-étapes précédentes. Il s'agit d'un exercice important parce qu'il permet non seulement de dresser un portrait global des améliorations apportées aux données sources, mais également de préciser l'incidence qu'ont eue les transformations sur la qualité des données.
- 5.7. Mise au point finale des fichiers de données :** C'est à cette sous-étape que l'ensemble des résultats des six autres sous-étapes du traitement sont mis ensemble pour produire le fichier de données devant servir à l'étape suivante.
- Cette sous-étape consiste à créer le fichier de traitement final, qui constitue la base de diffusion de tous les fichiers de données, à savoir le fichier principal, le fichier de microdonnées à grande diffusion (FMGD) et le fichier partagé.
- Les fichiers de diffusion sont les fichiers qui seront utilisés à l'intérieur et à l'extérieur de Statistique Canada pour analyser les données de l'enquête. Le FMGD est une version abrégée du fichier principal qui a été minutieusement évalué par les statisticiens pour éliminer tous les risques raisonnables d'identification des répondants pouvant présenter des caractéristiques passablement uniques. Le fichier partagé, quant à lui, ne contient que les enregistrements des répondants qui ont consenti à ce que leurs enregistrements soient partagés.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

• Profilage des données

- Déterminer la ou les méthodes de profilage en fonction de l'utilisation qui doit être faite des données (optimisation des requêtes, nettoyage, intégration, analyse, gestion scientifique, etc.) aux étapes subséquentes du processus de production statistique;
- Déterminer le logiciel approprié en fonction du type de profilage à effectuer;
- Procéder à l'analyse individuelle des différentes colonnes de la base de données en mettant l'accent sur les éléments suivants : cardinalité, modèles des valeurs et types de données, distribution et répartition des valeurs, etc.;
- Procéder à l'analyse multiple des colonnes de la base de données en mettant l'accent sur les éléments suivants : unicité (découverte de clés, conditionnelle, partielle), dépendances d'inclusion (découverte de clés étrangères, conditionnelles, partielles), dépendances fonctionnelles (conditionnelles, partielles);
- Procéder à des analyses de chevauchement lorsqu'on utilise plusieurs sources de données;
- Élaborer un rapport comportant les résultats du profilage et utiliser ces derniers pour mettre à jour les métadonnées, s'il y a lieu.

• Codage des données

- Centraliser les opérations de codage pour réduire les coûts et pouvoir profiter plus facilement des connaissances des experts;
- S'assurer que les préposés au codage possèdent la formation et les outils adéquats au succès des opérations de codage;
- S'assurer de suivre les procédures de codage de manière constante dans toutes les unités du fichier pour éviter, le plus possible, de commettre des erreurs;
- Utiliser des méthodes de contrôle de la qualité pour vérifier le niveau d'exactitude des données codées par le système ou les préposés par rapport aux critères préétablis;
- Recourir aux services d'une équipe de préposés au codage des données pour traiter des cas spéciaux échappant au codage automatisé;
- Créer des fichiers de référence et en faire la mise à jour pour maximiser le nombre de phrases reconnues par le système et limiter les erreurs;
- Procéder à l'évaluation d'un échantillon et vérifier l'exactitude des données codées automatiquement;
- Utiliser les résultats de cette évaluation pour augmenter et améliorer le contenu des fichiers de référence utilisés pour le codage des données.

• Examen, vérification et nettoyage des données

- Offrir une formation adéquate à tout le personnel concerné par l'examen et la vérification des données acquises;
- Fournir un guide des règles de vérification;
- Effectuer un suivi des travaux pour s'assurer que l'interprétation des règles de vérification est identique et cohérente pour l'ensemble des vérificateurs;
- Appliquer de nouveau les vérifications aux unités qui ont subi des corrections pour s'assurer qu'aucune autre erreur n'a été introduite directement ou indirectement par le processus de correction;
- Déterminer le modèle de non-réponse après l'analyse des caractéristiques des répondants et des non-répondants puis, faire le choix de la méthode ou de l'approche appropriée pour compenser les données manquantes (imputation ou repondération);

- Supprimer les erreurs typographiques relevées à la sous-étape du profilage et corriger certaines valeurs par rapport à une liste d'entités connue, s'il y a lieu;
 - Utiliser des outils de mise en forme pour harmoniser et normaliser les variables pour faciliter leur compilation et leur classification.
- **Vérification et imputation**
 - Déterminer les variables susceptibles d'agir comme variables auxiliaires pour l'imputation des données manquantes en explorant diverses sources de données;
 - Évaluer la qualité et la pertinence des variables disponibles pour déterminer celles pouvant être utilisées comme variables auxiliaires ou pour établir des classes d'imputation;
 - Déterminer les variables auxiliaires qui peuvent expliquer le ou les mécanismes de non-réponse et les utiliser pour enrichir la méthode d'imputation;
 - Tenir compte du type de caractéristiques à estimer dans le choix des variables auxiliaires et de la stratégie d'imputation pour maintenir les relations d'intérêt;
 - Certaines enquêtes exigent l'utilisation de plusieurs méthodes d'imputation selon la disponibilité de l'information auxiliaire. Dans un tel cas, il faut :
 - établir un nombre limité d'échelons hiérarchiques des méthodes en utilisant des règles prédéfinies,
 - élaborer et tester les méthodes associées à chaque échelon hiérarchique,
 - élaborer et tester les méthodes d'imputation en tenant compte des nouvelles classes dans les situations où les classes d'imputation ont été regroupées;
 - Dans le cas d'une imputation par donneur :
 - imputer toutes les données manquantes d'un enregistrement à partir d'un seul donneur, si possible, pour assurer la cohérence entre les données et maintenir les relations entre les variables,
 - Considérer la méthode du plus proche voisin pour trouver un donneur qui soit le plus près possible de l'enregistrement à imputer et choisir des variables appropriées pour déterminer le plus proche voisin,
 - Choisir un donneur au hasard dans l'ensemble des donneurs potentiels qui sont à la même distance (ou à une distance similaire) du receveur à imputer;
 - Les enregistrements imputés doivent ressembler étroitement à l'enregistrement rejeté à l'étape du contrôle et de la vérification;
 - Considérer exclure certaines unités dans le calcul des valeurs imputées si ces dernières sont très différentes des unités à imputer;
 - Déterminer les valeurs imputées et désigner clairement les méthodes et les sources d'imputation;
 - Conserver les valeurs imputées et non imputées des champs de l'enregistrement pour les besoins de l'évaluation;
 - Évaluer le degré et les effets de l'imputation;
 - Mesurer la variance ajoutée introduite par imputation.
 - **Ajustement et repondération**
 - Rajuster le poids de sondage pour réduire les erreurs de non-réponse, de couverture et d'échantillonnage ou pour assurer la cohérence avec d'autres sources de données;
 - Envisager le calage dans le cas où les données auxiliaires sont corrélées aux variables d'intérêt. Il faut cependant que les données auxiliaires soient disponibles pour les unités échantillonnées et que les totaux de population correspondants soient connus ou estimés avec précision;

- Utiliser les méthodes de contrôle de l'étendue des poids pour éviter les poids extrêmes ou négatifs. Voir, par exemple, Deville et Särndal (1992);
 - Rajuster les poids longitudinaux et les poids transversaux pour tenir compte de l'érosion de l'échantillon lorsqu'il s'agit d'analyse longitudinale;
 - Utiliser, si possible, l'information auxiliaire si elle est de qualité appropriée et corrélée avec les principales variables de l'enquête, ce qui permet d'améliorer l'uniformité et la précision des estimations;
 - Tenir compte de la non-réponse en choisissant des variables auxiliaires qui sont à la fois reliées à la probabilité de répondre et aux variables principales de l'enquête;
 - Pondérer selon la non-réponse en modélisant les indicateurs de réponse et en validant le modèle au moyen de diagnostics;
 - Former des classes de pondération pour la non-réponse pour obtenir une certaine robustesse par rapport à une défaillance du modèle de non-réponse et pour éviter des poids extrêmes;
 - Mentionner les classes de pondération, le facteur d'ajustement de la régression généralisée, les unités à inclure et exclure dans le calcul des facteurs d'ajustement, et les facteurs d'ajustement fondés sur la probabilité de réponse dans le cas où la non-réponse est reliée aux variables mesurées.
- **Évaluation et documentation**
 - Documenter chacun des processus (profilage, codage, examen et nettoyage des données, vérification et imputation, et ajustement et pondération), y compris les méthodes et les outils utilisés de même que les résultats trouvés, les répercussions sur la qualité des données et les recommandations formulées;
 - S'assurer de l'accessibilité de toute la documentation produite.
 - **Traitement final**
 - Ajouter les poids de l'enquête aux enregistrements individuels, y compris les poids à utiliser avec le fichier partagé, s'il y a lieu;
 - Créer des codes d'identification d'enregistrements pour le fichier principal, le FMGD et le fichier partagé;
 - Ajouter les nouvelles variables qui auraient pu avoir été créées à la suite des suppressions requises pour la création d'un FMGD ou d'un fichier partagé.

Dimensions et indicateurs de la qualité associés à cette étape

a. **Dimensions de la qualité** : Exactitude, intelligibilité et cohérence.

b. **Indicateurs de la qualité** :

- Une description et une justification de la méthodologie utilisée pour chaque étape du profilage et de la préparation des données sont disponibles, avec résultats à l'appui;
- Des indicateurs de la qualité, des mesures d'exactitude et des mesures d'assurance de la qualité sont disponibles pour les différentes étapes;
- Dans le cas où des modèles ont été utilisés, une description des hypothèses des modèles et une évaluation de leurs effets probables sur la qualité des données sont disponibles;
- Les systèmes généralisés et les paramètres utilisés ont été testés, modifiés au besoin et validés;
- Au besoin, un dictionnaire des données et un guide d'utilisateur sont disponibles.

6. Intégration, estimation et compilation

Description de l'étape et des sous-étapes

L'intégration des données regroupe un ensemble d'activités qui consistent à extraire les données provenant de différentes sources (bases de données, fichiers, applications, services Web, courriels, etc.), à leur faire subir

certaines transformations (jointures, déduplication, concaténation, calculs, etc.) pour finalement les consolider dans une seule et même base de données.

La compilation consiste à créer des données agrégées et des chiffres de population à partir des microdonnées ou d'agrégats de niveau inférieur. Dans le cadre des processus de production statistique impliquant plusieurs sources de données d'entrée, cette étape consiste à mettre en correspondance plusieurs types de statistiques, concepts, classifications et conventions au moyen d'un processus itératif conduisant à des données agrégées. Dans ce cas, l'intégration et la compilation se réalisent de manière simultanée.

Dans le cas des processus d'enquêtes-échantillons, si des données auxiliaires sont disponibles, on peut améliorer la précision des estimations. L'intégration de données auxiliaires dans ces processus d'estimation est appelée calage.

L'étape de l'intégration et de la compilation comporte cinq sous-étapes, qui peuvent se dérouler en ordre séquentiel ou de manière parallèle et qui peuvent être itératives. Ces sous-étapes sont les suivantes :

| 6.1 | 6.2 | 6.3 | 6.4 | 6.5 |
|---|--|--|--|---|
| Détermination des éléments, des règles et de la stratégie d'intégration | Évaluation et ajustement de la stratégie d'intégration | Chargement, application des mises en correspondance et intégration des données sources | Estimation, compilation et application des méthodes statistiques | Vérification et ajustement d'amélioration de la qualité |

6.1. Détermination des éléments, des règles et de la stratégie d'intégration : Cette sous-étape consiste à déterminer les éléments d'intégration. Il s'agit essentiellement d'une variable ou d'une combinaison de variables qui rendent faisable l'intégration des données de sources multiples. Ce sont les identifiants de base de données ou clés. Certains fichiers de données comportent des clés uniques pour chaque enregistrement, ce qui peut faciliter leur intégration lorsque ces clés sont les mêmes pour tous les fichiers de l'intégration.

Dans le cas d'un processus statistique par couplage d'enregistrements, cette sous-étape comprend la détermination d'un ensemble de paires d'enregistrements correspondant parfaitement à une clé donnée ou à un critère particulier et représentant des paires potentiellement jumelées, que l'on appelle également paires potentielles.

Cette sous-étape consiste aussi à élaborer et à définir les règles et la stratégie qui seront utilisées pour l'intégration des données. Dans le cadre d'un processus de couplage d'enregistrements, cette sous-étape commence par la comparaison des champs et des enregistrements où, pour chaque paire, les attributs des enregistrements couplés sont comparés. La comparaison d'attributs est en outre fondée sur des fonctions de comparaison qui diffèrent selon la manière dont elles traitent les valeurs manquantes. Les résultats de la comparaison servent ensuite à la prise de décision en ce qui a trait à la stratégie de couplage. Cette sous-étape se termine généralement par la production des clés de couplage préliminaires.

Dans les cas des processus impliquant plusieurs types de données d'entrée tels que les comptes macroéconomiques et les estimations de population, il arrive souvent que cette sous-étape se limite à confirmer qu'il n'y a rien à modifier. Cependant, dans certains cas, de nouveaux types de transactions peuvent émerger pour lesquels aucune structure ni règle n'a encore été créée ou élaborée. Par conséquent, des règles et un traitement devront être déterminés avant que les composantes ou les comptes soient compilés.

6.2. Évaluation et ajustement de la stratégie d'intégration : Cette sous-étape consiste à évaluer la qualité des éléments d'intégration.

Dans le cas d'un processus par couplage d'enregistrements, elle porte sur l'évaluation de la qualité du couplage (validation interne) et de l'adéquation à l'utilisation de l'ensemble de données couplées (validation externe). Le but de cette étape est de veiller à ce que l'ensemble de données couplées présente un niveau de qualité adéquat pour l'utilisation prévue. Toute limite des données couplées relevées au cours de cette sous-étape devrait contribuer à la décision relative à leur utilisation.

Cette sous-étape permet également de perfectionner, au besoin, la stratégie de couplage d'enregistrements en fonction des résultats de la validation interne et externe. Cela peut consister à perfectionner les règles de couplage ou les seuils utilisés pour vérifier si les paires correspondent ou non. Cette sous-étape se termine généralement par la production des clés de couplage finales.

En ce qui a trait aux processus macroéconomiques, cette sous-étape suppose l'analyse des comptes pour en vérifier la cohérence et la nature raisonnable. Différentes perspectives ou dimensions sont examinées, comme la dimension relative à la cohérence temporelle et la dimension relative à la structure. Les restrictions comptables et la mise en équilibre sont également examinées dans le cadre de ce sous-processus pour assurer une cohérence globale. À la suite de cette étape, il est possible qu'on revienne à la phase de conception, puisque les connaissances acquises au cours de cette étape peuvent entraîner une modification de la conception du compte. Une fois les ajustements apportés à la conception, il faut revenir à la sous-étape 6.1 pour mettre à jour les structures comptables, les règles et la stratégie d'intégration.

Pour ce qui est des estimations et des projections de population, cette sous-étape suppose l'analyse des tendances ainsi que de l'accroissement démographique pour en vérifier la cohérence et la nature raisonnable. La structure selon l'âge et le sexe ainsi que la distribution géographique feront également partie de l'analyse.

6.3. Chargement, application des mises en correspondance et intégration des données sources : Cette sous-étape consiste en l'utilisation d'identifiants de base pour créer de nouvelles données à partir de deux ensembles de données sources ou plus.

Dans les cas de processus statistique comme les comptes macroéconomiques, les estimations ou les projections de population, cette sous-étape consiste à charger les données sources acquises et les transformer de manière à ce qu'elles soient conformes aux concepts et aux mises en correspondance de la comptabilité macroéconomique ou de la démographie. Tout renseignement acquis au cours des premières étapes est également pris en considération, puisqu'il s'agit de renseignements importants dans le processus général de création de comptes macroéconomiques ou des composantes démographiques.

En ce qui a trait au processus statistique par couplage d'enregistrements, l'intégration comprend généralement le couplage systématique, l'utilisation de clés de couplage pour créer un ensemble de données couplées et le rapprochement des variables lorsque deux ensembles de données sources ou plus contiennent les mêmes variables. Pour examiner le processus de fusion, il convient d'effectuer des vérifications de la fréquence ou des enregistrements sur chaque ensemble de données sources avant le processus de fusion, puis à nouveau après la fusion. On compare ensuite le nombre d'enregistrements afin de déceler tout écart.

Pour ce qui est des processus d'enquêtes-échantillons, l'intégration a lieu avec des données auxiliaires et est appelée calage (voir la sous-étape 5.4, à savoir l'ajustement et la repondération). Elle consiste à rajuster les poids initialement calculés pour que les estimations de la ou des variables auxiliaires correspondent à des totaux connus.

6.4. Estimation, compilation et application des méthodes statistiques : Cette sous-étape consiste à créer des données agrégées et des chiffres de population à partir de microdonnées ou d'agrégats de faible niveau. Elle comprend la sommation des données pour les enregistrements qui ont certaines caractéristiques en commun, la détermination des mesures de la moyenne et de la dispersion, et l'application des poids de la sous-étape 5.4 (ajustement et repondération) pour calculer les totaux appropriés. Dans le cas des enquêtes-échantillons, les erreurs d'échantillonnage peuvent également être calculées dans cette sous-étape, et associées aux agrégats pertinents. La compilation est un processus itératif qui peut être une bonne méthode pour évaluer la qualité des données parce qu'elle permet de faire ressortir certaines incohérences. C'est également à cette sous-étape que les méthodes statistiques, telles que la désaisonnalisation, la déflation ou l'étalonnage, sont appliquées.

6.5. Vérification et ajustement d'amélioration de la qualité : Cette sous-étape consiste à vérifier que les données intégrées ou compilées répondent aux objectifs de production et peuvent être utilisées de manière fiable. Les activités de vérification consisteront notamment à confronter le produit statistique à d'autres données pertinentes (internes et externes), à confronter les statistiques aux attentes et aux connaissances du domaine ou à étudier les incohérences présentes dans les statistiques.

Dans le cadre des processus de couplage d'enregistrements, cette sous-étape est importante parce qu'elle fournit des renseignements supplémentaires sur l'adéquation à l'utilisation de l'ensemble de données couplées. Les erreurs ou les limites décelées à cette sous-étape peuvent exiger de plus amples ajustements de la stratégie de couplage d'enregistrements, de l'évaluation de la qualité ou des ajustements d'amélioration de la qualité exigeant la répétition de sous-étapes antérieures.

Lignes directrices pour assurer la qualité à cette étape

► Intégration de données par couplage d'enregistrements

- Utiliser le processus d'indexage ou de blocage pour générer et évaluer l'ensemble des paires potentielles pour réduire le nombre de paires possibles que génèrent les produits des tableaux croisés, lorsque les ensembles de données sources sont volumineux;
- Faire le choix de l'une des deux principales méthodes connues (probabiliste et déterministe) pour comparer les paires d'enregistrements selon des règles basées sur une évaluation du contenu des différentes régions des paires;
- Effectuer un examen minutieux (pouvant inclure de la vérification manuelle) des paires jumelées sélectionnées, des enregistrements non jumelés ainsi que du taux de couplage global, et apporter des ajustements nécessaires à la stratégie initiale de couplage jusqu'à l'obtention d'une stratégie optimale. Ce processus peut être itératif jusqu'à la production des clés de couplage préliminaires;
- Générer des estimations d'erreurs pour l'ensemble de données couplées, comme les taux de liens faux positifs et faux négatifs, la spécificité et la sensibilité;
- Procéder à une validation interne en comparant les taux de couplage globaux à des niveaux attendus qui découlent de l'expérience acquise ou à des sources externes, ou encore en analysant les taux de couplage pour les sous-groupes ou populations pour déceler d'éventuels biais ou confirmer les tendances prévues;
- Procéder à une validation externe en comparant certains résultats issus des données couplées à des données externes. La validation doit être réalisée avec la participation de spécialistes du domaine, s'il y a lieu;
- Ajuster les règles et la stratégie finale de couplage en fonction des résultats de la validation interne et externe en vue de la production des clés de couplage finales;
- Rédiger un rapport sur le processus de couplage qui précise les limites des données et dont le format permet aux utilisateurs des données de comprendre les notions de base de la stratégie de couplage ainsi que les résultats de l'évaluation de la qualité.

► Intégration de données de sources multiples

- Établir, mettre à jour ou valider certaines règles ou hypothèses concernant des composantes ou des comptes à intégrer. Ce point peut se résumer à confirmer qu'il n'y a rien à modifier dans le cadre d'un processus récurrent;
- S'assurer que des transformations sont effectuées pour toutes les données d'entrée pour qu'elles soient conformes aux concepts et aux mises en correspondance de l'activité statistique en question (comptabilité nationale, estimations ou projections de population, etc.);
- Examiner sous différentes perspectives et dimensions les comptes ou les composantes de l'activité statistique en question, telles que la dimension relative à la cohérence temporelle ou celle relative à la structure;
- Examiner les restrictions comptables et la mise en équilibre des comptes pour assurer une cohérence globale lorsqu'il s'agit d'un processus relatif aux comptes macroéconomiques;
- Effectuer les modifications et les ajustements nécessaires au cas où des incohérences ou des déséquilibres sont constatés aux points précédents;
- Mettre à jour la structure des comptes, les règles ou les données sources lorsqu'il s'agit d'un processus relatif aux comptes macroéconomiques.

► Estimation et compilation

- Utiliser les poids d'estimation (poids de sondage ajustés) pour calculer les statistiques descriptives et analytiques des domaines d'intérêt;
- Utiliser des méthodes spéciales pour estimer les petits domaines au cas où les exigences relatives à ceux-ci n'ont pas été prises en compte dans le plan d'échantillonnage;

- Calculer pour chacune des estimations de l'enquête, une estimation de son erreur d'échantillonnage, sous forme de variance d'échantillonnage, d'erreur-type, de coefficient de variation, de marge d'erreur ou d'intervalle de confiance;
- Utiliser l'estimation composite pour les enquêtes périodiques comportant un chevauchement d'échantillons importants entre les cycles (Gambino, Kennedy et Singh 2001);
- Utiliser, s'il y a lieu, un logiciel généralisé d'estimation disponible dans l'organisme. Sinon, il est possible de consulter le système généralisé d'estimation G-EST de Statistique Canada, un système généralisé modulaire pour l'estimation par domaine dans les enquêtes sur échantillon.

► **Vérification et ajustement d'amélioration de la qualité**

- Vérifier, dans le cas des processus impliquant plusieurs types de données d'entrée, la dimension relative à la cohérence temporelle et celle relative à la structure des composantes;
- Appliquer les ajustements d'amélioration de la qualité selon le plan élaboré au cours de la sous-étape 2.6 et les révisions effectuées à la sous-étape 6.2.

Dimensions et indicateurs de la qualité associés à cette étape

a. **Dimensions de la qualité** : Pertinence, exactitude, intelligibilité, cohérence et accessibilité.

b. **Indicateurs de la qualité** :

- Existence d'un rapport de validation des données intégrées;
- Existence d'un dictionnaire des données;
- Existence d'indicateurs de la qualité sur les données intégrées.

7. Analyse

Description de l'étape et des sous-étapes

L'étape de l'analyse consiste, d'une part, à examiner, à interpréter et à préparer les données aux fins de diffusion et, d'autre part, à élaborer des réponses à certaines questions. Les activités exécutées aux différentes sous-étapes de l'analyse permettent aux analystes en statistique de comprendre les statistiques produites. Elles consistent à cerner les sujets d'analyse, à déterminer la disponibilité de données appropriées, à décider des méthodes qu'il y a lieu d'utiliser pour répondre aux questions d'intérêt, à appliquer les méthodes et à évaluer, résumer et communiquer les résultats.

Cette étape joue également un rôle clé dans l'évaluation de la qualité des données grâce à la mise en lumière de certains problèmes spécifiques et elle constitue de ce fait une des sources pouvant influencer sur les améliorations futures au processus.

Généralement, les résultats de l'analyse des données sont souvent publiés ou résumés dans les diffusions officielles de l'organisme producteur. De plus, certains programmes dépendent des résultats analytiques à titre de principal produit de données parce que, pour des raisons de confidentialité, il est impossible de diffuser les microdonnées.

L'étape de l'analyse comporte cinq sous-étapes, qui se déroulent généralement en ordre séquentiel, mais qui peuvent aussi se dérouler de manière parallèle, et qui peuvent être itératives. Ces sous-étapes sont les suivantes :

| | | | | |
|--|---------------------------------------|---|--|---|
| 7.1 Préparation d'ébauches de produits | 7.2 Validation des produits | 7.3 Interpréter et expliquer les produits | 7.4 Application de mesures de contrôle de la divulgation | 7.5 Mise au point finale des produits |
|--|---------------------------------------|---|--|---|

7.1. Préparation d'ébauches de produits : Cette sous-étape de l'analyse consiste à transformer les données recueillies et traitées durant l'étape précédente en produits statistiques. Elle comporte également des activités visant à produire d'autres mesures connexes.

7.2. Validation des produits : Cette sous-étape consiste à valider la qualité des données produites par rapport au CAQ et aux objectifs de production initialement définis pour le processus statistique. Elle consiste également à rassembler le plus de renseignements possible sur le phénomène à l'étude et à les comparer

ensuite aux données acquises pour repérer toute divergence par rapport aux attentes et pour permettre des analyses basées sur des données probantes.

- 7.3. Interpréter et expliquer les produits :** Les activités de cette sous-étape permettent aux analystes de comprendre, par l'interprétation des données, les questions pertinentes actuelles et émergentes et de déterminer la façon de présenter les résultats au grand public. Les études analytiques effectuées à cette sous-étape permettent également d'expliquer le comportement de certaines caractéristiques et les éventuelles relations qui peuvent exister entre elles.
- 7.4. Application de mesures de contrôle de divulgation :** L'objectif de cette sous-étape consiste à s'assurer que les dispositions régissant la protection de la confidentialité sont respectées tout en préservant le plus possible l'utilité des données produites. Elle comporte des activités de vérification pouvant déceler la divulgation primaire et secondaire ainsi que l'application de méthodes de suppression de données ou de techniques de perturbation.
- Les principes qui sous-tendent les activités de contrôle de la divulgation sont presque exclusivement régis par les dispositions de la *Loi sur la statistique* [1970, S.R.C. 1985, c. S19, plus précisément l'alinéa 17(1)b]. Le programme vigilant de contrôle de la divulgation et de protection de la confidentialité dans le système statistique au Canada permet de maintenir la confiance du public, ce qui est essentiel pour l'obtention de données, pertinentes et de qualité, autant chez les répondants que chez les fournisseurs de données.
- 7.5. Mise au point finale des produits :** Cette sous-étape consiste à s'assurer que les statistiques produites, les métadonnées qui les accompagnent et les services connexes atteignent le niveau de qualité requis et peuvent être utilisés.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

• Validation des produits

- Consulter la Directive sur la validation des produits statistiques et les Lignes directrices sur la validation des produits statistiques de Statistique Canada;
- Vérifier la cohérence des résultats par rapport à des sources de données internes et externes similaires, comme des enquêtes, d'autres éditions ou des données administratives;
- Vérifier la cohérence interne en calculant des ratios qui sont censés se situer à l'intérieur de certaines limites connues (proportions hommes-femmes, valeur moyenne des biens, etc.);
- Examiner la contribution individuelle des grandes unités aux totaux (généralement appliquées aux enquêtes-entreprises);
- Comparer les indicateurs de la qualité des données calculés aux étapes précédentes aux objectifs de production;
- Vérifier les tableaux croisés pour assurer l'uniformité et l'exactitude des variables clés et des domaines importants;
- Organiser des rencontres de rétroaction avec le personnel affecté à l'acquisition et au traitement des données;
- Mandater des spécialistes de l'externe qui sont familiarisés avec le domaine en question, pour qu'ils vérifient si ses résultats sont plausibles et qu'ils rédigent un rapport sur les travaux en cours, avant la diffusion des résultats;
- Évaluer, entre autres, les erreurs de couverture et d'échantillonnage, les erreurs attribuables à la non-réponse, ainsi que les erreurs de mesure et de traitement, à la lumière des analyses d'autres étapes du processus de production;
- Analyser la cohérence des données sur la base des événements récents dans l'actualité.

• Interprétation et explication des résultats

- Choisir une approche analytique qui convient à la question examinée et aux données à analyser;

- Vérifier si les sources sont cohérentes et, si plus d'une source de données est utilisée dans l'analyse, trouver une façon efficace de les combiner;
 - Vérifier si les valeurs imputées doivent être incluses dans l'analyse et, le cas échéant, déterminer la façon dont elles doivent être traitées;
 - Préciser dans l'analyse la façon dont la non-réponse totale ou partielle a été traitée et prendre en compte l'importance des données manquantes et les types de données manquantes dans les sources de données utilisées;
 - Déterminer les autres méthodes qui peuvent être utilisées pour rendre compte correctement de l'effet de la non-réponse dans l'analyse, si les valeurs imputées ne sont pas utilisées;
 - Faire toutes les mises en garde nécessaires sur la façon dont les méthodes utilisées pour traiter les données manquantes peuvent influencer sur les résultats;
 - Éviter de tirer des conclusions en ce qui concerne la causalité;
 - Durant l'analyse de tendances à court terme, ne pas oublier de tenir compte des tendances à moyen et à long terme. Les tendances à court terme ne représentent souvent que de légères fluctuations d'une tendance plus importante à moyen ou à long terme;
 - Éviter les points de référence arbitraires dans le temps, comme le changement entre l'année précédente et l'année en cours;
 - Tenir compte des points de référence comportant une plus grande signification tels que le dernier point décisif pour les données économiques, les différences intergénérationnelles pour les statistiques démographiques et les changements législatifs pour les statistiques sociales;
 - Consulter des experts au sujet des domaines spécialisés et des méthodes statistiques;
 - Expliquer les pratiques ou les méthodes d'arrondissement;
 - Prendre soin d'établir la distinction entre le taux de variation et la variation en points de pourcentage lorsqu'on présente des détails sur les taux;
 - Définir la base de calcul des taux et déterminer la base conceptuelle des mesures (par exemple, s'il s'agit de dollars constants ou d'un index, indiquer l'année de référence).
- **Contrôle de la divulgation**
- Consulter les directives de l'organisme en matière de contrôle de divulgation. Autrement, se référer aux Lignes directrices sur la gestion des fichiers de microdonnées statistiques et des statistiques agrégées de Statistique Canada pour déterminer les méthodes de contrôle de divulgation les plus appropriées pour les types de données à traiter;
 - Distinguer le type de données à traiter pour faire le bon choix des méthodes de contrôle de divulgation. Chaque type de données a des méthodes qui lui sont propres;
 - Éviter de dévoiler les paramètres et les règles utilisées pour contrôler la divulgation. La connaissance de ces paramètres peut aider à mieux préciser la valeur de certains répondants;
 - Déterminer le degré de confidentialité de chaque cellule par rapport aux règles établies dans l'organisme;
 - Éviter de diffuser un tableau de données quantitatives si ce dernier fournit des valeurs rattachées à des cellules jugées confidentielles;
 - Tenir compte du risque de divulgation résiduelle qui est la possibilité d'estimer les données confidentielles par un recoupement de l'information diffusée avec d'autres renseignements accessibles, y compris les diffusions antérieures de l'organisme;
 - Vérifier si les cellules à fréquence nulle posent problème. Ces dernières peuvent révéler des renseignements confidentiels dans les tableaux de données quantitatives;
 - Supprimer les cellules confidentielles dans les tableaux;
 - Vérifier si les catégories et les hiérarchies utilisées dans les tableaux se chevauchent;

- Utiliser, s'il y a lieu, l'une des techniques d'arrondissement des valeurs des cellules pour protéger adéquatement la confidentialité;
 - En ce qui a trait à la diffusion de microdonnées :
 - o évaluer le risque de divulgation dans les fichiers de microdonnées,
 - o appliquer l'une des deux méthodes des données à diffusion restreinte généralement utilisée pour contrôler le risque de divulgation,
 - o s'assurer que la population est assez grande dans le cas de certains groupes identifiables,
 - o élargir les catégories de variables,
 - o regrouper les valeurs extrêmes supérieures et inférieures,
 - o supprimer certaines variables provenant de certains répondants,
 - o supprimer les répondants du fichier, s'il y a lieu;
 - Consulter le paragraphe 17(2) de la *Loi sur la statistique*, qui prévoit que certains types de renseignements confidentiels peuvent être diffusés à la discrétion du statisticien en chef et en vertu d'une ordonnance;
 - Consulter les ressources disponibles dans l'organisme en matière de confidentialité. Autrement, joindre les ressources ci-après à Statistique Canada :
 - o la Division de l'accès aux microdonnées, qui offre des avis et conseils à propos des politiques liées à la confidentialité des renseignements recueillis,
 - o le Comité de la confidentialité et des mesures législatives et ses sous-comités, le Comité de gestion sur l'accès aux microdonnées et le Comité de la diffusion des microdonnées, lesquels offrent des stratégies et des pratiques de contrôle de la divulgation,
 - o le Centre de ressources sur le contrôle de la divulgation au sein du Centre de collaboration internationale et d'innovation en méthodologie (CCIIM), qui offre l'aide technique, ainsi que l'équipe de soutien des systèmes généralisés pour le logiciel G-Confid.
- **Présentation des résultats**
- Mettre l'accent sur les variables et les sujets importants dans le texte de présentation des résultats;
 - Structurer les idées de façon logique, en fonction de leur pertinence ou de leur importance;
 - Recourir à des titres, des sous-titres et des encadrés pour renforcer la structure du texte de présentation des résultats;
 - Rédiger le texte en langage aussi simple que le sujet le permet. Selon le public cible, il est parfois souhaitable de perdre un peu en précision pour rendre le texte plus compréhensible;
 - Insérer des graphiques et des tableaux en complément du texte de présentation des résultats pour faciliter la transmission du message;
 - Essayer autant que possible de privilégier les titres qui véhiculent un message (p. ex. « Les revenus des femmes demeurent inférieurs à ceux des hommes »), plutôt que des titres de graphique classiques (p. ex. « Revenus selon l'âge et le sexe »);
 - Commenter l'information fournie dans les tableaux et les graphiques pour permettre au lecteur de mieux la comprendre;
 - Veiller à ce que la présentation générale des tableaux insérés dans le texte contribue d'une part à la clarté des données qu'ils contiennent et, d'autre part, prévient les erreurs d'interprétation. Cela comprend l'espacement, la formulation, l'emplacement et l'apparence des titres, les titres de lignes et de colonnes et autres étiquetages;
 - Expliquer les pratiques ou les méthodes d'arrondissement;

- Satisfaire aux exigences en matière de confidentialité imposées par le processus statistique dont les données font l'objet;
- Fournir des renseignements sur les méthodes analytiques et les outils utilisés. Inclure soit une section portant sur les méthodes, soit un renvoi indiquant au lecteur où obtenir les détails;
- Inclure des renseignements sur la qualité des résultats. Les erreurs types, les intervalles de confiance ou les coefficients de variation fournissent au lecteur des renseignements importants sur la qualité des données. Le choix de l'indicateur peut varier selon l'endroit où l'article est publié;
- S'assurer que toutes les références sont exactes et uniformes et qu'elles font l'objet de renvois dans le texte;
- S'assurer qu'il n'y a pas d'erreurs dans l'article. Vérifier les détails, par exemple la cohérence des chiffres, les tableaux et les graphiques, et l'exactitude des données externes et des calculs arithmétiques simples;
- S'assurer que ce qui est annoncé dans l'introduction est effectivement exprimé dans le reste de l'article. S'assurer que les conclusions sont cohérentes avec les résultats de l'analyse;
- Faire réviser l'article par d'autres personnes pour en vérifier la pertinence, l'exactitude et l'intelligibilité, peu importe où il doit être diffusé;
- S'assurer que le texte est conforme aux normes d'édition en vigueur dans l'organisme ou se référer aux Lignes directrices sur l'édition de Statistique Canada. Ces normes sont applicables aux graphiques, aux tableaux et au style, entre autres;
- S'assurer de la bonne qualité dans l'une ou l'autre des langues officielles de toutes les composantes à diffuser et que les deux versions concordent en ce qui a trait aux données et au texte.

Dimensions et indicateurs de la qualité associés à cette étape

a. **Dimensions de la qualité** : Pertinence, exactitude, accessibilité, intelligibilité et cohérence.

b. **Indicateurs de la qualité** :

- Disponibilité d'un rapport de validation des données;
- Disponibilité d'un rapport d'analyse;
- Disponibilité d'un dictionnaire des données;
- Disponibilité d'indicateurs de la qualité;
- Renseignements disponibles sur les méthodes utilisées;
- Mesures en vigueur pour le contrôle de la divulgation;
- Révision des données et des textes qui seront diffusés.

8. Diffusion

Description de l'étape et des sous-étapes

L'étape de la diffusion consiste à mettre les données recueillies et traitées dans le cadre d'une activité statistique à la disposition des utilisateurs par divers moyens. Il importe aussi de communiquer efficacement les données à leurs utilisateurs et de faire savoir que les données sont disponibles. Les activités de diffusion et de communication des résultats de tout processus statistique doivent être connexes et avoir comme objectif d'optimiser l'utilisation des données en répondant aux besoins des utilisateurs et en leur offrant un large accès à l'information.

L'étape de la diffusion comporte cinq sous-étapes, qui se déroulent généralement en ordre séquentiel, mais qui peuvent aussi se dérouler de manière parallèle, et qui peuvent être itératives. Ces sous-étapes sont les suivantes :

| 8.1 Mise au point des systèmes de diffusion | 8.2 Production des composantes de diffusion | 8.3 Gestion de la diffusion des produits | 8.4 Promotion des produits de diffusion | 8.5 Gestion du soutien des utilisateurs |
|---|---|--|---|---|
|---|---|--|---|---|

- 8.1. Mise au point des systèmes de diffusion :** Cette sous-étape consiste à faire les dernières mises au point des systèmes de diffusion avec les données et métadonnées qui vont être publiées. Elle comprend les activités de formatage des données et des métadonnées de manière à ce qu'elles soient prêtes à être chargées dans le module de sortie, et de la vérification de l'existence d'un lien entre les données et les métadonnées pertinentes.
- 8.2. Production des composantes de diffusion :** Cette sous-étape consiste à finaliser la production des données selon la manière qu'elles ont été conçues à la sous-étape 2.1 du processus. Elle comporte les activités de préparation des composantes du produit, telles que le texte explicatif, les tableaux et graphiques ou les énoncés sur la qualité. Elle inclut aussi l'assemblage des composantes préparées et une vérification finale pour s'assurer que le produit répond aux objectifs de départ et qu'il est conforme aux normes de diffusion.
- 8.3. Gestion de la diffusion des produits :** Cette sous-étape consiste à mettre en œuvre tous les éléments en vue de la diffusion, y compris ceux permettant de gérer la période durant laquelle se déroulera la diffusion. Parmi les activités faisant partie de cette sous-étape, mentionnons les séances d'information pour les médias et d'autres intervenants intéressés au sujet traité, des présentations à certains ministres du gouvernement qui doivent répondre aux questions les concernant, la gestion de l'accès aux données confidentielles pour des utilisateurs autorisés, etc. Les activités de présentation qui ont lieu avant la diffusion proprement dite sont toujours soumises à un embargo.
- 8.4. Promotion des produits de diffusion :** Cette sous-étape permet de promouvoir les produits résultant d'un processus de production statistique particulier pour leur donner plus de visibilité médiatique et les aider à atteindre le plus vaste auditoire possible.
- 8.5. Gestion du soutien des utilisateurs :** Cette sous-étape consiste à s'assurer que les demandes concernant le produit diffusé sont enregistrées et acheminées au bon endroit et que des réponses sont fournies dans les meilleurs délais, qu'il s'agisse de demandes de renseignements ou de services des clients, de demandes d'accès aux microdonnées, etc.

Lignes directrices pour assurer la qualité à cette étape

► Ensemble des processus statistiques

- Consulter les directives de l'organisme en matière de diffusion. Autrement, se référer à la Politique sur la diffusion officielle de Statistique Canada;
- Consulter les directives de l'organisme en matière de diffusion des microdonnées. Autrement, se référer à la Directive sur la diffusion de fichiers de microdonnées;
- Consulter les directives de l'organisme sur les faits saillants des publications. Autrement, se référer à la Politique sur les faits saillants des publications de Statistique Canada;
- Consulter les directives de l'organisme en matière d'édition de documents. Autrement, se référer aux Lignes directrices sur l'édition de Statistique Canada;
- Éviter de préparer des produits (ébauches préliminaires) lorsque le traitement des données est en cours;
- S'assurer que les données à diffuser correspondent aux données de source obtenues;
- Examiner à fond toutes les données (y compris les produits sous-jacents) préalablement à la diffusion pour s'assurer que les données sont exactes, que l'analyse est rigoureuse, que le traitement est adéquat, que la publication est pertinente pour l'organisme et que la communication est efficace;
- Recourir, s'il y a lieu, à des outils automatisés, par exemple le système Édition électronique intelligente ou un comparateur de texte, pour réduire les risques d'erreurs humaines;

- Tester les produits électroniques (tableaux de données et autres liens) avant leur diffusion pour s'assurer qu'ils fonctionnent comme prévu;
- Veiller à ce que les produits écrits soient examinés par une personne n'ayant pas pris part à l'activité statistique;
- Vérifier soigneusement les chiffres, les périodes de référence (p. ex. au cours du dernier semestre ou du dernier trimestre) et les mots qui décrivent les tendances (p. ex. à la hausse, à la baisse) dans les articles et les publications pour s'assurer qu'ils sont exacts;
- Valider les chiffres cités dans les articles et les publications en les comparant aux chiffres d'autres produits tabulaires;
- Veiller à ce que le texte soit de grande qualité dans les deux langues officielles et que les deux versions concordent en ce qui a trait aux données et au texte;
- S'assurer que les mesures de la qualité des données ou, si possible, les outils nécessaires à leur calcul (p. ex. les tables de validation des coefficients de variation, les programmes de variance d'échantillonnage) accompagnent le produit diffusé;
- S'assurer que la documentation décrivant la qualité et la méthodologie utilisée est prête pour accompagner le produit diffusé;
- S'assurer que les données sont diffusées avec le nom d'une ou de plusieurs personnes-ressources, d'un numéro de téléphone et d'une adresse de courrier électronique pour toute information complémentaire;
- S'assurer que la ou les personnes-ressources sont prêtes à répondre aux demandes officielles d'entrevue présentées par les médias, à fournir des commentaires et à interpréter les données;
- Évaluer l'incidence de toute déclaration erronée ou erreur d'interprétation véhiculée par les médias et déterminer le meilleur moyen d'y donner suite.

Dimensions et indicateurs de la qualité associés à cette étape

a. **Dimensions de la qualité** : Pertinence, accessibilité, actualité, exactitude, cohérence et intelligibilité (soit les six dimensions de la qualité du processus de production d'une enquête).

b. **Indicateurs de la qualité** :

- Disponibilité d'un calendrier de diffusion et d'une stratégie de suivi;
- Stratégie de communication en vigueur et documentée;
- Délai raisonnable entre la date ou période de référence et celle de la diffusion du produit;
- Délai raisonnable entre la date de diffusion prévue et la celle de la diffusion réelle;
- Disponibilité de la documentation des erreurs détectées avant et après la diffusion et évaluation des répercussions;
- Disponibilité de la fréquence à laquelle les utilisateurs accèdent au produit d'information au fil du temps et évaluation de l'utilité du produit.

9. Évaluation

Description de l'étape et des sous-étapes

L'étape de l'évaluation consiste à organiser une ou plusieurs séances de remue-méninges sur les étapes du processus de production statistique. Cette étape, la dernière de toute activité de production statistique, est généralement effectuée par l'ensemble de l'équipe y ayant participé. Elle est basée sur l'analyse de tous les intrants quantitatifs et qualitatifs recueillis tout au long des huit étapes précédentes. Elle permet à l'équipe de cerner les forces et les faiblesses du processus et de faire des recommandations sur les améliorations potentielles dans l'avenir.

L'étape de l'évaluation comporte trois sous-étapes, qui se déroulent généralement en ordre séquentiel. Ces sous-étapes sont les suivantes :

| 9.1 | 9.2 | 9.3 |
|---|--------------------------|------------------------------|
| Rassemblement de données d'entrée aux fins d'évaluation | Conduite de l'évaluation | Entente sur un plan d'action |

- 9.1. Rassemblement des données d'entrées :** Cette sous-étape consiste à rassembler toute la documentation produite durant les différentes étapes du processus de production statistique et la mettre ensuite à la disposition de l'équipe chargée de l'évaluation. Cette documentation se trouve sous plusieurs formes, y compris les commentaires des utilisateurs, les métadonnées et paradonnées sur les processus, les mesures des systèmes, et les suggestions du personnel. Les rapports de suivi du plan d'action du projet doivent également servir de données d'entrée à l'évaluation.
- 9.2. Conduite de l'évaluation :** Cette sous-étape consiste à analyser les données recueillies à la sous-étape précédente et d'en faire la synthèse dans un rapport d'évaluation. Ce dernier devra faire état de tous les problèmes rencontrés au cours du processus, comment ils ont été gérés et les résultats engendrés, et recommander des changements au besoin. Ces recommandations peuvent être sur les changements à apporter à une étape ou une sous-étape quelconque du processus ou peuvent consister à proposer que le processus statistique en entier ne soit pas reconduit. Il est également important de noter les processus qui ont bien fonctionné dans l'évaluation.
- 9.3. Entente sur un plan d'action :** Cette sous-étape implique de chercher le consensus auprès de la haute gestion sur la nécessité d'élaborer un plan d'action basé sur les recommandations du rapport et sur les moyens de sa mise en œuvre. Ce plan doit inclure certains mécanismes de surveillance des répercussions des mesures prévues, lesquels peuvent, à leur tour, servir de base aux évaluations futures du processus.

Lignes directrices pour assurer la qualité à cette étape

- Choisir la personne ou l'équipe qui sera chargée de mener l'évaluation du programme (processus de production récurrente) ou du projet (processus de production spéciale);
- S'assurer que le responsable de chacune des étapes du processus rassemble toutes les données d'entrées et les mette à la disposition de la personne ou de l'équipe d'évaluation. Il s'agit essentiellement :
 - des commentaires des utilisateurs et des membres des comités ayant participé aux consultations,
 - des métadonnées et paradonnées sur les étapes et sous-étapes,
 - des mesures des systèmes,
 - des suggestions du personnel,
 - des rapports de suivi par rapport à un plan d'action;
- S'assurer que le rapport d'évaluation est une synthèse de l'analyse des données d'entrées, qu'il mentionne tout problème de qualité particulier relevé au cours du processus de production statistique et qu'il recommande des changements au besoin;
- Rassembler les pouvoirs décisionnaires nécessaires pour établir un plan d'action fondé sur les recommandations du rapport d'évaluation et s'entendre sur ce plan.

Dimensions et indicateur de la qualité associés à cette étape

- a. **Dimensions de la qualité :** Intelligibilité et pertinence.
- b. **Indicateur de la qualité :** Rapport d'évaluation du processus de production statistique.

Bibliographie

- AITKEN, Alexis, et coll. 2004. *Handbook on improving quality by analysis of process variables*, European Commission, Eurostat, 156 p.
- BERGDAHL, Mats, et coll. 2007. *Handbook on Data Quality Assessment Methods and Tools*, Wiesbaden, European Commission, Eurostat, 141 p.
- CAI, Li et Yangyong ZHU. 2015. « The Challenges of Data Quality and Data Quality Assessment in the Big Data Era », *Data Science Journal*, vol. 14, n° 2, p. 1 à 10.
- EUROPEAN STATISTICAL SYSTEM. 2015. *Quality Assurance Framework of the European Statistical System*, Version 1.2, European Statistical System, 52 p.
- EUROSTAT. 2012. *Guide de la statistique de la coopération au développement de la commission européenne*, version 2.0, Luxembourg, Union européenne, 509 p., « Methodologies & Working papers ».
- INSTITUTO NACIONAL DE ESTADISTICA. 2015. *Directrices de calidad del INE*, Unidad de Calidad, 63 p.
- LOHR, Sharon L. et Trivellore E. Raghunathan. 2017. « Combining survey data with other data sources », *Statistical Science*, vol. 32, n°2, p. 293 à 312.
- NAUMANN, Felix. 2013. *An Introduction to Data Profiling*, Hasso Plattner Institut, IT Systems Engineering, Universität Potsdam.
- NAUMANN, Felix. 2013. « Data Profiling Revisited », *SIGMOD Record*, vol. 42, n° 4, p. 40 à 49.
- SCHOMM, Fabian. 2016. « Data Profiling as a Process: Bridging the Gap Between Academia and Practitioners », *GVD*.
- STATISTIQUE CANADA. 2000. [Politique visant à informer les utilisateurs sur la qualité des données et la méthodologie](https://www.statcan.gc.ca/fra/apercu/politique/info-usager), <https://www.statcan.gc.ca/fra/apercu/politique/info-usager>.
- STATISTIQUE CANADA. 2003. *Méthodes et pratiques d'enquête*, produit n° 12-587-XPF au catalogue de Statistique Canada, Ottawa (Ontario).
- STATISTIQUE CANADA. 2004. [Politique concernant les normes](https://www.statcan.gc.ca/fra/apercu/politique/normes), <https://www.statcan.gc.ca/fra/apercu/politique/normes>.
- STATISTIQUE CANADA. 2009. *Lignes directrices concernant la qualité*, Cinquième édition, produit n° 12-539-X au catalogue de Statistique Canada, Ottawa (Ontario).
- STATISTIQUE CANADA. 2012. *Politique sur l'accès aux microdonnées*.
- STATISTIQUE CANADA. 2012. *Politique sur la protection des renseignements personnels et la confidentialité*.
- STATISTIQUE CANADA. 2015. *Directive sur l'obtention des données administratives en vertu de la Loi sur la statistique*.
- STATISTIQUE CANADA. 2015. *Directive sur la validation des produits statistiques*, version 1.4,
- STATISTIQUE CANADA. 2015. [Politique sur l'utilisation de données administratives obtenues en vertu de la Loi sur la statistique](https://www.statcan.gc.ca/fra/apercu/politique/donnees_admin), https://www.statcan.gc.ca/fra/apercu/politique/donnees_admin.
- STATISTIQUE CANADA. 2016. *Directive sur la transmission des renseignements protégés*.
- STATISTIQUE CANADA. 2016. *Politique sur la diffusion officielle*, version mise à jour le 27 septembre.

STATISTIQUE CANADA. 2017. *Directive sur le couplage de microdonnées*.

STATISTIQUE CANADA. 2017. *Directives sur les corrections apportées aux communiqués du Quotidiens et aux produits statistiques*, version mise à jour le 1^{er} novembre.

STATISTIQUE CANADA. 2017. *Le cadre d'assurance de la qualité de Statistique Canada*, Troisième édition, produit n° 12-586-X au catalogue de Statistique Canada, Ottawa (Ontario).

STATISTIQUE CANADA. 2018. *Directive sur la diffusion de fichiers de microdonnées*.

STATISTIQUE CANADA. 2018. *Directive sur la gestion des fichiers de microdonnées statistiques*, version mise à jour le 30 septembre.

STATISTIQUE CANADA. 2018. *Directive sur la gestion des statistiques agrégées*, version mise à jour le 30 septembre.

UNECE. 2013. [Modèle générique du processus de production statistique \(GSBPM\)](http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0), version 5.0, Décembre 2013, <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>

UNECE BIG DATA QUALITY TASK TEAM. 2014. *A Suggested Framework for the Quality of Big Data*, version 4.1, CEE ONU, 32 p.

YU, Frank. 2003. [Défis reliés à la réalisation l'enquêtes pour la prochaine décennie : Mise en évidence de la qualité des données dans la pratique](https://www150.statcan.gc.ca/n1/fr/pub/11-522-x/2003001/session14/7715-fra.pdf?st=Qkp8-Uos), produit n° 11-522-XIF au catalogue de Statistique Canada, Ottawa (Ontario). Présentation au Symposium international sur les questions de méthodologie de Statistique Canada. <https://www150.statcan.gc.ca/n1/fr/pub/11-522-x/2003001/session14/7715-fra.pdf?st=Qkp8-Uos>