

N° 12-606-X au catalogue  
ISBN 978-0-660-79935-3

## Trousse de la qualité des données

Date de diffusion : le 27 septembre 2017



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](http://www.statcan.gc.ca).

Vous pouvez également communiquer avec nous par :

**Courriel** à [infostats@statcan.gc.ca](mailto:infostats@statcan.gc.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 |

## 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](http://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 la ministre responsable de Statistique Canada

© Sa Majesté le Roi du chef du Canada, représenté par la ministre de l'Industrie, 2026

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.*

---

# Trousse de la qualité des données

## 1 Contexte

**Question :** Qu'obtient-on lorsqu'on combine « données » et « qualité »?

**Réponse :** La trousse de la qualité des données!

Eh bien, techniquement, on obtient la « qualité des données », mais dans le contexte de cette page Web, on peut envisager la trousse comme une prime très avantageuse.

Cela étant dit, une question demeure : en quoi consiste exactement la qualité des données?

Pour répondre à cette question, nous devons d'abord répondre à deux autres questions :

- Qu'est-ce que les données?
- Qu'est-ce que la qualité?

Les **données** sont constituées de chiffres, de lettres et de symboles. Elles deviennent de l'information lorsqu'on les organise sous forme d'ensembles, d'expressions ou de modèles. Nous utilisons l'information pour déterminer des besoins, pour mesurer des répercussions et pour prendre des décisions éclairées. Si les données à partir desquelles nous avons établi des conclusions ou pris des décisions sont incorrectes à certains égards, ces conclusions et ces décisions peuvent également être incorrectes ou trompeuses.

La **qualité** est mesurée en fonction de diverses caractéristiques, qui varient selon le point de vue choisi. Nous examinons ces caractéristiques dans la section qui suit.

Pour ce qui est du producteur de données, les mesures de la qualité comprennent la reproductibilité du processus, l'actualité et la ponctualité de la livraison des données et des métadonnées, la volonté et la capacité de soutenir les utilisateurs de données ainsi que l'impression d'autorité, d'impartialité et de confiance que dégage le producteur. Pour ce qui est des données et des métadonnées proprement dites, les mesures de la qualité comprennent la pertinence et l'utilité, la couverture, la granularité, l'exactitude et la fiabilité ainsi que la normalisation ou conformité.

Tous ces éléments nous amènent à la **qualité des données**, un concept modelé par les deux notions décrites ci-dessus, et qui fournit deux façons de déterminer si les données sont susceptibles d'être appropriées ou non :

- On peut décrire ce qui a été fait au cours de la collecte et du traitement des données en vue de s'assurer que les données sont correctes.
- On peut observer les caractéristiques mesurables des données.

Bien que le respect de bonnes pratiques en matière d'assurance de la qualité des données ne garantisse pas leur validité, il réduit la probabilité qu'il y ait des erreurs. Mener une évaluation de la qualité des données constitue un moyen de déterminer la mesure selon laquelle les données sont protégées des erreurs. En communiquant cette évaluation aux utilisateurs, on leur donne confiance en la qualité des données qu'ils utilisent.

## 2 Attributs de la qualité

### Attributs de la qualité liés au producteur de données<sup>1</sup>

**Pratiques d'assurance de la qualité :** Mesure selon laquelle les pratiques d'assurance de la qualité<sup>2</sup> visées et consignées ont été appliquées dans le cadre de la collecte et du traitement des données, aussi bien sur le plan de l'engagement du producteur de données à l'échelle organisationnelle, que sur le plan de la mise en œuvre de pratiques de suivi et de production de rapports à l'échelle opérationnelle.

**Reproductibilité du processus :** Mesure selon laquelle le processus de production de données peut être reproduit ou répété. Parmi les exemples de processus non reproductibles figurent les processus ponctuels ou encore les cas où des étapes intermédiaires ou des fichiers de données n'ont pas été archivés et ne peuvent être recréés.

**Actualité et ponctualité** : L'actualité correspond à l'intervalle entre la fin de la période de référence des données et le moment où elles sont mises à la disposition des utilisateurs. Idéalement, cet intervalle est court, et les données et les métadonnées sont diffusées au même moment. La ponctualité désigne la mesure selon laquelle les données et les métadonnées fiables sont diffusées au moment prévu, comme prévu ou tel que promis par le producteur de données.

**Caractère joignable** : Volonté et capacité du producteur de données à discuter des données avec des utilisateurs potentiels, voire à faciliter l'utilisation des données.

**Viabilité** : Mesure selon laquelle on peut prévoir que le producteur de données continuera de produire les données pendant une période de temps raisonnable.

**Impression d'autorité, d'impartialité et de confiance** : Mesure selon laquelle le producteur de données est perçu comme faisant autorité en la matière, comme étant à l'abri de toute influence excessive des intervenants ou d'autres organismes, et comme étant digne de confiance.

**Sécurité** : Mesure selon laquelle la sécurité des données est assurée dans tous les fonds de données et lors de toute communication, et selon laquelle l'accès aux données durant la production est limité aux personnes ayant reçu la formation et les autorisations appropriées. Notamment, l'accès est accordé selon le principe du besoin de savoir.

### Attributs de la qualité liés aux données et aux métadonnées<sup>3</sup>

**Pertinence et utilité** : Mesure selon laquelle les données concernent le phénomène voulu. Les données seront considérées comme étant moins pertinentes si elles datent de trop longtemps ou ne renseignent pas sur le sujet d'intérêt. L'utilité des métadonnées renvoie à la mesure selon laquelle les données sont décrites en fonction des méthodes, des concepts, des limites, des hypothèses et des pratiques d'assurance de la qualité adoptées.

**Couverture** : Mesure selon laquelle les données représentent l'ensemble du phénomène voulu. Cet attribut peut être évalué en fonction de la couverture temporelle ou géographique, ou de la couverture des unités de population (c.-à-d. personnes, ménages, entreprises). On utilise parfois le terme intégralité pour désigner la couverture (plus particulièrement dans le contexte des métadonnées).

**Granularité** : Désigne l'unité ou le niveau d'un enregistrement particulier de l'ensemble de données. Par exemple, un ensemble de données à un niveau très détaillé peut contenir les enregistrements pour des particuliers, des procédures médicales ou des lacs, tandis qu'un ensemble de données à un niveau moins détaillé peut contenir des enregistrements agrégés par province ou année. Plus un ensemble de données est détaillé ou local, plus la valeur perçue est élevée et, en contrepartie, plus le besoin de protéger les données contre une divulgation non autorisée est accru. Il est généralement facile d'agréger les données ou d'en réduire la granularité, mais il n'est habituellement pas possible d'accroître la granularité de données déjà agrégées.

**Exactitude et fiabilité** : Mesure selon laquelle les données décrivent correctement le phénomène qu'elles sont censées mesurer. La fiabilité désigne la mesure selon laquelle les données sont exactes au fil du temps. L'exactitude renvoie souvent à deux éléments : la **précision**, pour évaluer la similitude de mesures répétées d'une même situation, et le **biais**, pour mesurer tout écart systématique par rapport à la réalité des données. Les autres facteurs qui contribuent à l'exactitude et à la fiabilité des données sont la **validité**, soit la mesure selon laquelle les valeurs des variables de l'ensemble de données correspondent aux résultats attendus, et la cohérence, soit la mesure selon laquelle les données sont exemptes de contradictions.

**Normalisation ou conformité** : Mesure selon laquelle les données et les métadonnées respectent des normes reconnues en matière de présentation et de conventions d'appellation, ainsi que des normes de diffusion reconnues comme SDMX pour les produits statistiques. Parmi les autres aspects liés à la normalisation et à la conformité, mentionnons l'utilisation de logiciels et de formats de fichier conformes aux normes de l'industrie et le vocabulaire contrôlé pour les valeurs des données, au besoin.

**Protection de renseignements de nature délicate** : Il est interdit de divulguer des renseignements de nature délicate contenus dans des ensembles de données mis à la disposition d'utilisateurs autres que ceux autorisés à les consulter, à moins d'en avoir obtenu l'autorisation explicite. Parmi les exemples de renseignements de nature délicate, mentionnons les identificateurs qui permettent d'associer les données granulaires à une personne,

à un ménage ou à une entreprise, ou encore des données agrégées dont le niveau de détail est suffisant pour permettre de déduire les attributs d'une personne, d'un ménage ou d'une entreprise. Il existe différentes façons de protéger les données pour éviter la divulgation de renseignements délicats, en fonction de la nature et de la granularité des données. On peut notamment supprimer ces renseignements ou appliquer le concept de perturbation aléatoire aux valeurs de données. De nombreux algorithmes de contrôle des divulgations fournissent des diagnostics du niveau de protection atteint.

**Possibilité de combinaison ou de couplage :** Mesure selon laquelle il est possible d'intégrer deux sources de données ou plus. Par exemple, des identifiants uniques, comme le numéro d'assurance sociale, le numéro d'entreprise ou le numéro d'assurance-maladie, peuvent être appariés directement, tandis que des identificateurs non uniques, comme le nom, le sexe, la date de naissance et l'adresse, peuvent être couplés au moyen d'algorithmes de couplage statistiques fondés sur les probabilités. L'intégration des ensembles de données est améliorée lorsque le concept représenté par un seul enregistrement de chaque ensemble de données est aligné.

**Accessibilité :** Facilité avec laquelle l'utilisateur peut obtenir et utiliser les données et les métadonnées. Les données et les métadonnées facilement accessibles sont munies d'étiquettes, de mots clés et de balises pertinents et appropriés afin qu'elles puissent être repérées de façon électronique; elles sont présentées dans un format ou logiciel couramment utilisé; et elles sont téléchargeables ou accessibles au moyen de processus transparents et clairs. L'accessibilité consiste également à réduire les obstacles à l'accès, notamment les coûts.

**Possibilité de traitement et intelligibilité :** Facilité avec laquelle l'utilisateur peut manipuler, interpréter, explorer, analyser et utiliser les données et les métadonnées. Un autre élément important de cet attribut est la mesure selon laquelle les métadonnées et d'autres services de soutien des producteurs de données favorisent la bonne utilisation des données, par exemple, en y incluant les indicateurs de qualité appropriés.

**Impression de fiabilité et de crédibilité :** Mesure selon laquelle les données sont perçues comme étant fiables et les métadonnées comme étant crédibles.

### 3 Pratiques d'assurance de la qualité des données

Il existe un éventail de bonnes pratiques, et celles-ci peuvent être adoptées par toute organisation produisant des données. Les producteurs de données peuvent adapter ces pratiques à leur propre environnement. Dans ce cadre, nous les invitons à consigner leurs pratiques d'assurance de la qualité des données et à transmettre cette documentation à leurs utilisateurs de données. En effet, lorsque les utilisateurs sont informés des pratiques d'assurance de la qualité observées au cours de la production des données, leur confiance en la qualité des données augmente. Ces pratiques d'assurance de la qualité sont un sous ensemble de celles contenues dans le [Cadre d'assurance de la qualité](#) et les [Lignes directrices concernant la qualité](#).

#### Pratiques d'assurance de la qualité des données pour la production de registres et de bases de données

- Utiliser des identifiants uniques connus (p. ex. numéro d'assurance sociale, numéro d'entreprise, numéro d'assurance-maladie) avec des mesures de protection appropriées des renseignements délicats.
- Utiliser des chiffres de contrôle sur ces identifiants uniques connus pour garantir la validité des valeurs.
- Utiliser des menus déroulants, des tables de consultation ou des listes de référence pour les variables auxquelles correspond un ensemble de codes<sup>4</sup> déterminé.
- Utiliser des présentations standard reconnues dans la mesure du possible, par exemple, la norme ISO 8601 pour les dates (AAAAMMJJ) et l'heure (HH:MM), les abréviations standard des provinces (ON, MB, etc.).
- Inclure une validation intégrée permettant de signaler la saisie de données aberrantes ou de valeurs inattendues.
- Utiliser une règle logique et consignée d'attribution de noms pour les variables et les fichiers.
- Consigner les règles d'inclusion et d'exclusion, les procédures à suivre et les vérifications de la qualité.
- Produire des ensembles de données de sortie à des intervalles réguliers et prévisibles (c'est-à-dire le dernier jour de chaque mois ou le dernier jour de chaque année).
- Définir et mettre en œuvre une stratégie pour la sauvegarde, le stockage et la rétention des données.

### **Pratiques d'assurance de la qualité des données pour les données d'enquête (enquête-échantillon ou recensement)**

- Utiliser des méthodes d'échantillonnage, de pondération et d'estimation valables sur le plan statistique.
- S'assurer que toutes les méthodes sont consignées et reproductibles.
- S'assurer que la base de sondage est aussi à jour, complète et exacte que possible.
- Consigner la période, la couverture géographique et les unités de population de la base de sondage et de l'échantillon.
- Mettre à l'essai la fluidité et l'intelligibilité du questionnaire.
- Choisir une méthode de collecte appropriée pour la population cible et le domaine spécialisé, en tenant compte des facteurs liés aux coûts et d'autres facteurs.
- Utiliser une technique de contrôle de la qualité telle que le contrôle statistique du processus<sup>5</sup> pour vérifier l'exactitude des données recueillies.
- Tenter au moins une fois de prendre contact avec chaque unité échantillonnée et consigner chaque tentative.
- Utiliser les ressources de vérification de manière efficiente et efficace, c'est-à-dire faire en sorte que les données soient adaptées à leur utilisation prévue<sup>6</sup> sans chercher à les rendre « parfaites ».
- Valider les données agrégées ou totalisées par rapport à d'autres sources.

### **Pratiques d'assurance de la qualité des données pour la production de données numérisées, de données satellitaires et de données de compteurs**

- Utiliser une technique de contrôle de la qualité telle que l'échantillonnage pour acceptation<sup>7</sup> ou la vérification sur le terrain<sup>8</sup> pour garantir l'exactitude des données.
- S'assurer de communiquer les modifications à tous les utilisateurs, par exemple, lorsqu'un nouveau code universel des produits (CUP) est adopté.

### **Pratiques d'assurance de la qualité des données pour la combinaison de données provenant de différentes sources**

- S'assurer de la concordance des définitions des concepts, des populations d'intérêt, des unités d'observation et des périodes de référence.
- Mentionner toutes les sources de données et leur contribution au produit final.
- Analyser les données non appariées ou inutilisées pour comprendre les raisons pour lesquelles il n'y a pas eu de correspondance.
- S'assurer que toutes les méthodes sont consignées et reproductibles.
- Pratiques d'assurance de la qualité pour les métadonnées (documentation)
- Tenir compte des besoins en documentation dans la planification du projet et l'affectation des ressources.
- Consigner au fur et à mesure; ne pas laisser cette tâche pour la fin.
- Utiliser des modèles et des règles d'attribution de noms normalisées.
- Décrire tous les concepts, la population visée par les données, les limitations des données ou les exceptions éventuelles qu'elles contiennent et la période de référence.
- Décrire toutes les méthodes utilisées pour l'échantillonnage, la collecte, la saisie et la vérification des données ainsi que pour la combinaison de données provenant de différentes sources et pour la totalisation.
- Décrire les mesures en matière de sécurité des données.
- Décrire les méthodes d'assurance de la qualité suivies.
- Décrire les mesures prises pour prévenir la divulgation de renseignements de nature délicate.
- Fournir des statistiques sommaires portant sur les principales variables (moyenne, médiane, mode, intervalle et ensemble des valeurs valides).

- Fournir un dictionnaire de données ou un vocabulaire contrôlé pour les variables, au besoin.
- Utiliser des présentations standard reconnues dans la mesure du possible, par exemple, la norme ISO 8601 pour les dates (AAAAMMJJ) et l'heure (HH:MM), les abréviations standard des provinces (ON, MB, etc.).
- Mettre la documentation à la disposition des utilisateurs de données.
- Utiliser des étiquettes, des balises et des mots clés pertinents et appropriés pour que les données et les métadonnées associées puissent être repérées de façon électronique.
- Surveiller et consigner les mises à jour et les révisions.

#### **Pratiques d'assurance de la qualité des données pour la sécurité et l'accessibilité des données et la protection contre la divulgation de renseignements de nature délicate**

- Pendant le processus de production de données, limiter l'accès aux personnes ayant reçu la formation et les autorisations appropriées, et dont le besoin d'accéder aux renseignements a été établi (« besoin de savoir »).
- Pendant le processus de production de données, assurer la sécurité des données dans tous les fonds de données et lors de toute communication, au moyen du cryptage et d'autres techniques.
- Adopter la stratégie de la « source unique de vérité » en vue de minimiser la duplication des renseignements et des efforts, grâce notamment à des structures de base de données efficaces.
- Utiliser des présentations standard pour les noms, les dates, les adresses et les autres variables couramment utilisées (normes internationales, régionales et nationales au besoin, comme la norme ISO 8601 pour la date et l'heure : AAAAMMJJ HH:MM selon le système horaire de 24 heures).
- Utiliser des logiciels et formats de fichier standard pour les fichiers mis à la disposition des autres utilisateurs.
- Prévoir la possibilité de communiquer des ensembles de données au niveau de granularité (niveau de détail) le plus fin possible et se préparer en conséquence.
- Effectuer des sauvegardes régulières.
- Définir et mettre en œuvre une stratégie pour le stockage et la rétention des données.
- Assurer la protection contre la divulgation de renseignements délicats (l'identité ou les attributs d'une personne ou d'une entreprise) en masquant les valeurs<sup>9</sup> et au moyen d'autres techniques.

## **4 Listes de contrôle**

### **Liste de contrôle d'autoévaluation de la qualité des producteurs de données**

Les producteurs de données doivent remplir cette liste de contrôle d'autoévaluation. Cette autoévaluation offre aux producteurs de données la possibilité de réfléchir à leurs propres pratiques d'assurance de la qualité des données et de repérer les domaines où des améliorations pourraient être apportées. Lorsque l'autoévaluation est terminée et qu'elle comporte des éléments probants montrant la façon dont chacun des attributs de la qualité est assuré, elle peut être communiquée aux utilisateurs des données afin de les aider à prendre des décisions éclairées concernant l'adéquation des données à leur utilisation.

#### **Questions à propos de vous, le producteur de données**

**Q1** Êtes-vous un ministère ou un organisme gouvernemental, un établissement universitaire, un institut de recherche, une entreprise privée ou une organisation d'un autre type?

**Répondez par « oui » ou par « non » aux questions 2 à 7 et fournissez des hyperliens vers les renseignements, le cas échéant.**

**Q2** Pensez-vous que vous fournirez encore ces données dans un an?

**Q3** Êtes-vous disposé à offrir des services de consultation et de soutien aux personnes qui souhaitent utiliser ces données?

**Q4** Avez-vous appliqué les pratiques d'assurance de la qualité des données précisées pour la production de ces données?

**Q5** Toutes les méthodes et procédures utilisées sont-elles consignées?

**Q6** Toutes les méthodes et procédures utilisées sont-elles reproductibles?

**Q7** Y a-t-il des mécanismes en place pour prévenir les atteintes à la protection des données ou la perte de données dans tous les fonds et lors de toute communication?

#### **Questions sur les métadonnées**

**Répondez par « oui » ou par « non » à la question 8 et fournissez un ou plusieurs hyperliens vers les renseignements, le cas échéant.**

**Q8** Fournissez-vous aux personnes qui souhaitent utiliser vos données les métadonnées<sup>3</sup> connexes?

**Si la réponse à la question 8 est « oui », répondez à la question 9.**

**Q9** Dans l'affirmative, dans quel format les métadonnées sont-elles fournies (par exemple, Excel, Word, PDF, même format)?

**Répondez par « oui » ou par « non » aux questions 10 à 13 et fournissez des hyperliens vers les renseignements, le cas échéant.**

**Q10** Dans l'affirmative, les métadonnées sont-elles accessibles en même temps que les données?

**Q11** Dans l'affirmative, les métadonnées incluent-elles les concepts, les définitions et les descriptions des variables?

**Q12** Dans l'affirmative, les métadonnées incluent-elles la description des méthodes, des procédures et des pratiques d'assurance de la qualité adoptées dans le cadre de la production des données?

**Q13** Dans l'affirmative, les métadonnées sont-elles exactes, complètes, à jour et exemptes de contradictions?

#### **Questions sur les données**

**Q14** Dans quel délai, après la fin de la période de référence, ces données sont-elles mises à la disposition des utilisateurs (durée)?

**Q15** Quelle période ces données couvrent-elles (date de début et de fin)?

**Q16** Quelle région géographique ces données couvrent-elles (description géospatiale)?

**Q17** Quelles sont les unités de population représentées par ces données (décrire la population)?

**Q18** Ces données sont-elles mises à la disposition des utilisateurs au niveau de granularité (de détail) le plus bas possible (décrire le niveau de détail de l'unité)?

**Répondez par « oui » ou par « non » à la question 19 et fournissez un ou plusieurs hyperliens vers les renseignements, le cas échéant.**

**Q19** Fournissez-vous des indicateurs d'exactitude pour (au moins) les 5 à 6 variables les plus importantes de l'ensemble de données?

**Si la réponse à la question 19 est « oui », répondez aux questions 20 et 21 en remplissant le tableau conformément aux titres, et fournissez des hyperliens vers les renseignements, le cas échéant (la précision peut être exprimée sous forme de valeurs CV ou de limites supérieures et inférieures d'un intervalle de confiance).**

**Q20** Dans l'affirmative, veuillez fournir le nom de la variable ainsi qu'un indicateur ou une estimation de l'exactitude de chacune des 5 à 6 variables les plus importantes.

- Variable
- Indicateur

**Q21** Veuillez fournir le nom de la variable ainsi que le pourcentage de valeurs valides pour chacune des 5 à 6 variables les plus importantes.

**Q22** Les données des enregistrements individuels sont-elles logiques (par exemple, les vérifications effectuées sur les données)?

**Q23** Quelles sont les présentations standard utilisées et pour quelles variables sont-elles utilisées (par exemple, dates, adresses et autres variables suivant des présentations standard ou respectant des vocabulaires contrôlés)?

**Q24** Quelles sont les mesures prises pour prévenir la divulgation de renseignements de nature délicate (par exemple, méthodes de contrôle de la divulgation)?

**Répondez par « oui » ou par « non » aux questions 25 et 26 et fournissez des hyperliens vers les renseignements, le cas échéant.**

**Q25** L'ensemble de données comprend-il des identifiants uniques?

**Q26** Existe-t-il des variables d'appariement comme l'âge, le sexe, l'adresse, l'industrie et la profession?

**Q27** Dans quel format ces données sont-elles offertes (par exemple, Excel, fichier non hiérarchique)?

**Q28** Quelle est la taille moyenne d'un ensemble de données (par exemple, kilooctets [Ko] gigaoctets [Go])?

**Q29** Existe-t-il des exceptions ou des limitations relatives à ces données? Dans l'affirmative, quelles sont-elles?

### Liste de contrôle de l'évaluation de la qualité des utilisateurs de données

Cette liste de contrôle de l'évaluation doit être remplie par toute personne envisageant d'utiliser des données produites par une autre organisation. Les utilisateurs de données devraient avoir une idée du niveau de qualité requis et des attributs prioritaires s'ils connaissent l'utilisation envisagée des données.

Répondez à chaque question par « oui » ou par « non » et fournissez un ou plusieurs hyperliens vers les renseignements, le cas échéant.

#### Questions à propos du producteur de données

**Q1** Le producteur de données<sup>1</sup> est-il un ministère ou un organisme gouvernemental, un établissement universitaire, un institut de recherche, ou une entreprise privée? Ou bien les données proviennent-elles d'une autre source?

**Q2** Vous attendez-vous à ce que le producteur de données fournisse toujours ces données dans un an?

**Q3** Le producteur de données est-il disposé à offrir des services de consultation et de soutien relativement à l'utilisation de ces données?

**Q4** Les pratiques désignées d'assurance de la qualité ont-elles été appliquées dans le cadre de la production de ces données?

**Q5** Selon vous, le producteur de données est-il un spécialiste du domaine et est-il digne de confiance?

#### Questions sur les métadonnées

**Q6** Les métadonnées<sup>2</sup> sont-elles fournies en même temps que les données?

**Q7** Les métadonnées incluent-elles tous les concepts, toutes les définitions et toutes les descriptions de variables dont vous avez besoin?

**Q8** Les métadonnées incluent-elles des descriptions des méthodes, des procédures et des pratiques d'assurance de la qualité adoptées dans le cadre de la production des données?

**Q9** Les métadonnées sont-elles exactes, complètes, à jour et exemptes de contradictions?

**Q10** Selon vous, les métadonnées sont-elles crédibles?

#### Questions sur les données

**Q11** Les données sont-elles accessibles dans un délai raisonnable après leur production?

- Q12** Les données sont-elles accessibles ou reçues à la date prévue?
- Q13** Les données concernent-elles le phénomène souhaité?
- Q14** La période couverte par les données correspondelle à vos besoins?
- Q15** La couverture géographique des données correspondelle à vos besoins?
- Q16** La population visée par les données correspondelle à vos besoins?
- Q17** Les données sont-elles au niveau de granularité (niveau de détail) dont vous avez besoin?
- Q18** Existe-t-il des indicateurs d'exactitude pour les variables qui sont les plus importantes pour vous?
- Q19** Dans l'affirmative, le niveau de précision des variables les plus importantes pour vous correspond-il à vos besoins?
- Q20** L'absence de biais est-elle établie? (En l'absence d'indication de biais, veuillez répondre « Oui ».)
- Q21** Les variables les plus importantes pour vous ontelles suffisamment de valeurs valides pour vos besoins?
- Q22** Existetil suffisamment d'enregistrements ayant une cohérence interne pour vos besoins?
- Q23** Les variables respectent-elles des formats standard?
- Q24** Semble-t-il que les renseignements de nature délicate ont été protégés?
- Q25** L'ensemble de données comprend-il des identifiants uniques?
- Q26** Existetil des variables d'appariement comme l'âge, le sexe, l'adresse, l'industrie, la profession?
- Q27** Pouvez-vous facilement récupérer ou lire les données dans votre milieu de travail?
- Q28** Pouvezvous facilement traiter et manipuler les données dans votre milieu de travail?
- Q29** Selon vous, les données sont-elles fiables?

Merci d'avoir lu la trousse! Afin que nous puissions répondre encore mieux à vos besoins, veuillez prendre quelques minutes pour nous indiquer les parties que vous avez trouvées utiles, les renseignements qui pourraient être ajoutés et les aspects qui pourraient être améliorés. Si vous avez des questions, n'hésitez pas à nous en faire part, c'est avec plaisir que nous y répondrons. Veuillez envoyer un courriel au [Secrétariat de la qualité de Statistique Canada](#).

## Notes

- Producteur de données** : L'entité responsable de produire, de recueillir, de rassembler, de traiter et de manipuler les données, et de les mettre à la disposition des autres.
- Méthodes d'assurance de la qualité** : Utilisation de normes reconnues pour la date et l'heure, comme AAMMJJ HH : MM; données totalisées validées par rapport à d'autres sources.
- Métadonnées** : Les métadonnées sont des données qui décrivent d'autres données. Les métadonnées résument des renseignements de base sur les données, facilitant leur consultation et leur utilisation.
- D'ensemble de codes** : Ensemble de valeurs d'une variable, par exemple, l'ensemble de codes pour un feu de circulation est rouge, jaune et vert.
- Contrôle statistique du processus** : Type précis de contrôle de la qualité qui consiste à surveiller les variations du processus et à détecter les écarts par rapport à la norme, qui sont prévus lorsqu'on utilise des cartes de contrôle.
- D'« adaptation à l'utilisation »** : L'expression « adaptation à l'utilisation » ou « adéquation à l'utilisation » est utilisée pour rendre l'idée que le produit répond aux besoins du client ou de l'utilisateur, mais sans les excéder. Un exemple de voiture adaptée à l'utilisation serait une Honda ou une Mazda, non pas une Lamborghini.
- D'échantillonnage pour acceptation** : Méthode particulière de contrôle de la qualité où l'on utilise l'échantillonnage statistique pour déterminer si l'on doit accepter ou rejeter un ensemble de produits.

8. **Vérification sur le terrain** : Processus consistant à recueillir les données appropriées et objectives (démonstrables) plutôt que des renseignements obtenus par déduction.
9. **Masquage de valeurs** : Méthode simple pour prévenir la divulgation de renseignements délicats (p. ex. arrondissement, regroupement de valeurs, ajout d'un « bruit » aléatoire).