



Orientation sur la qualité des données

Publié : le 2024-01-01

© Sa Majesté le Roi du chef du Canada,
représenté par la présidente du Conseil du Trésor, 2024

Publié par le Secrétariat du Conseil du Trésor du Canada
90 rue Elgin, Ottawa, Ontario, K1A 0R5, Canada

No de catalogue BT48-42/2024F-PDF
ISBN: 978-0-660-69836-6

Ce document est disponible sur le site Web du gouvernement du Canada à l'adresse www.canada.ca

Ce document est disponible en médias substituts sur demande.

Nota : Pour ne pas alourdir le texte français, le masculin est utilisé
pour désigner tant les hommes que les femmes.

Also available in English under the title: Guidance on Data Quality

Orientation sur la qualité des données

Date de publication : 10 janvier 2024

Sur cette page

- [Préambule](#)
- [1. Concepts clés](#)
- [2. Dimensions de la qualité des données](#)
- [3. Pratiques exemplaires en matière de qualité des données](#)

Préambule

La présente orientation fournit aux fonctionnaires des ministères un vocabulaire commun pour comprendre la qualité des données et des conseils sur la manière d'aborder cette question dans la pratique, à l'appui de ce qui suit :

- les sections 4.3.1.1 et 4.3.2.1 de la *[Politique sur les services et le numérique](#)*;
- la section 4.3.1.3 de la *[Directive sur les services et le numérique](#)*;
- les sections 6.3.1, 6.3.3 et 6.3.4 de la *[Directive sur la prise de décisions automatisée](#)*.

1. Concepts clés

► Dans cette section

1.1 En quoi consiste la qualité des données?

La qualité des données est une caractéristique des données définie en fonction de neuf dimensions : accès, exactitude, cohérence, exhaustivité, constance, intelligibilité, pertinence, fiabilité et actualité.

Pourquoi la qualité des données est-elle importante?

La qualité des données détermine si les utilisateurs peuvent les trouver, les partager et les utiliser facilement lorsqu'ils en ont besoin. Une qualité élevée des données favorise la prise de décisions fondée sur des données probantes et l'utilisation de systèmes de décisions automatisés, et peut améliorer la conception et la mise en œuvre des politiques, programmes et services à l'échelle du gouvernement.

La qualité des données peut également aider les fonctionnaires des ministères à confirmer que les données répondent aux besoins et aux objectifs d'utilisateurs particuliers, c'est-à-dire qu'elles sont adaptées à leur but ultime. Toutefois, les données peuvent être considérées comme adaptées à leur usage sans pour autant répondre à toutes les dimensions de la qualité des données dans la même mesure. Bien que les fonctionnaires des ministères soient encouragés à considérer la qualité des données en fonction de l'objectif pour lequel les données sont utilisées, les dimensions de la qualité des données ne seront pas toutes pertinentes de la même façon dans tous les contextes. De même, la qualité des données doit être prise en compte tout au long du cycle de vie des données en question, car la pertinence de chaque dimension peut varier en fonction de l'étape à laquelle la qualité des données est examinée.

2. Dimensions de la qualité des données

Les neuf dimensions de la qualité des données fournissent aux ministères un vocabulaire commun pour définir et évaluer la qualité des données. Les dimensions peuvent aider les fonctionnaires à identifier et à articuler les différents problèmes de qualité des données, et à les traiter afin de garantir que les données soient utilisables et répondent à leurs besoins. Le respect de ces dimensions peut renforcer la gouvernance, le partage et la réutilisation des données à l'échelle du gouvernement.

Les dimensions peuvent se chevaucher et l'importance de chacune peut varier en fonction des besoins spécifiques de l'utilisateur. Dans certains cas, il peut être nécessaire de trouver un équilibre entre les deux. Les fonctionnaires peuvent apporter des précisions aux dimensions existantes ou envisager l'ajout de nouvelles dimensions pour s'adapter au type spécifique de données qu'ils traitent. Les fonctionnaires sont encouragés à consulter des experts (par exemple des gestionnaires de données, des conservateurs de données, des fournisseurs de données, des experts en la matière) qui possèdent les connaissances nécessaires pour répondre aux questions relatives à chaque dimension de la qualité des données.

Accès

L'accès désigne la facilité avec laquelle un utilisateur peut découvrir, traiter, manipuler et obtenir les données.

L'accès permet de savoir si les utilisateurs sont au courant de l'existence des données et s'ils ont l'autorisation d'y accéder. Même s'ils y ont accès, les utilisateurs n'ont pas toujours la capacité de les traiter ou de les manipuler pour répondre à leurs besoins en raison de contraintes techniques, de ressources insuffisantes, de l'absence d'information requise, ou de certaines politiques ou lois restreignant l'utilisation des données.

Exactitude

L'exactitude désigne la mesure dans laquelle les données décrivent les phénomènes du monde réel qu'elles sont censées représenter.

Les données sont exactes lorsqu'elles représentent un phénomène adéquatement. Les évaluations de l'exactitude varient selon le contexte, la méthodologie et la validité des hypothèses ou suppositions sous-jacents. Le maintien de l'exactitude dans les organisations du secteur public implique de garantir que les données recueillies pour administrer les services correspondent à ce que les clients ont partagé. Dans le cas des initiatives stratégiques et des programmes, pour garantir l'exactitude, les utilisateurs doivent souvent valider les données en consultant des sources fiables et en évaluant les méthodes ou les processus par lesquels les données ont été acquises.

Cohérence

La cohérence désigne la mesure dans laquelle les données provenant d'un ou de plusieurs sources sont comparables et peuvent être reliées entre elles.

Un jeu de données cohérent est conforme aux taxonomies ou nomenclatures d'architecture communes. Les utilisateurs peuvent améliorer la cohérence des données en adoptant des normes organisationnelles, fédérales, nationales ou internationales, et en particulier celles qui sont prescrites comme norme de référence pour les éléments de données du gouvernement du Canada. Lorsque les données sont plus cohérentes, elles peuvent facilement être réutilisées et combinées avec d'autres données, ce qui permet aux utilisateurs de les intégrer et de les comparer.

Exhaustivité

L'exhaustivité désigne la mesure dans laquelle les valeurs de données sont suffisamment remplies.

Les données peuvent être considérées comme complètes lorsqu'elles contiennent les entrées nécessaires pour que les utilisateurs les utilisent de façon appropriée. L'information contextuelle et de fond permet aux utilisateurs de comprendre un jeu de données dans leurs secteurs d'activité respectifs.

Constance

La constance désigne la mesure dans laquelle les données sont non contradictoires en interne.

La cohérence permet de garantir la validité logique d'un jeu de données. Un jeu de données est cohérent si les relations entre ses composantes sont déterminées comme étant logiques.

Intelligibilité

L'intelligibilité désigne la mesure dans laquelle les données peuvent être comprises dans leur contexte approprié.

Un jeu de données peut être interprété si un utilisateur (humain ou machine) est en mesure de comprendre ses entrées, de déterminer pourquoi et comment il a été recueilli ou créé, et de juger de sa pertinence pour une politique, un programme, un service ou une autre initiative gouvernementale.

Pertinence

La pertinence désigne la mesure dans laquelle les données sont jugées convenables pour appuyer un objectif.

La pertinence des données dépend de la valeur informative ou analytique qu'elles apportent à l'objectif de l'utilisateur. L'évaluation de la pertinence dépend du contexte et des besoins de l'utilisateur. Ainsi, les mêmes données peuvent être pertinentes dans un contexte, mais non pertinentes dans un autre contexte.

Fiabilité

La fiabilité désigne la mesure dans laquelle la variabilité des données peut être expliquée.

La fiabilité porte sur le fait que les données répondent aux attentes des utilisateurs au fil du temps. Un jeu de données est fiable si les utilisateurs peuvent expliquer comment il évolue ou change au fil du temps. La fiabilité consiste également à veiller à ce que les données demeurent intactes et ne soient pas modifiées, ou alors qu'elles soient modifiées uniquement de manière documentée grâce à des mesures de contrôle de l'intégrité des données.

Actualité

L'actualité désigne la durée entre la fin de la période à laquelle les données se rapportent et la période à laquelle ces données sont disponibles pour répondre aux besoins des utilisateurs.

L'actualité est une mesure du délai entre deux moments : la période à laquelle les données se rapportent et la période où les utilisateurs peuvent effectivement utiliser ces données. L'actualité décrit la mesure dans laquelle les utilisateurs ont accès aux données lorsqu'ils en ont besoin.

3. Pratiques exemplaires en matière de qualité des données

La présente orientation associe les neuf dimensions aux pratiques exemplaires correspondantes afin de fournir une approche commune pour l'évaluation et la gestion de la qualité des données. La liste des pratiques recommandées n'est pas exhaustive, mais elle permet aux fonctionnaires d'interpréter et d'appliquer les neuf dimensions de manière cohérente. Ces pratiques peuvent également être utilisées pour orienter les approches en matière d'évaluation, de maintien ou d'amélioration de la qualité des données. Les dimensions et les pratiques exemplaires peuvent être utilisées pour tous les types de données, et les fonctionnaires peuvent les adapter à leurs besoins spécifiques. Les pratiques ne sont pas pertinentes dans tous les cas ou n'ont pas la même signification d'un contexte à l'autre : les fonctionnaires sont donc libres de décider s'il y a lieu de les appliquer, et de déterminer dans quels cas et de quelle façon.

Accès

- Élaborer un répertoire ou un catalogue des jeux de données afin d'appuyer les politiques, les programmes ou les services.
- Appliquer les normes de référence des métadonnées (c'est-à-dire les normes de référence des métadonnées prescrites à l'échelle de l'organisation ou du ministère) pour décrire les concepts, les variables ou les nomenclatures de vos actifs de données, conformément à la Norme pour la gestion des métadonnées et à la Norme sur les données géospatiales.

- Établir des processus concernant la manière dont votre organisation documente, conserve, publie, archive et élimine les données qu'elle recueille ou crée.
- Attribuer des catégories de sécurité aux actifs de données, comme l'exige la Directive sur la gestion de la sécurité.
- Définir les droits d'accès, les privilèges et les restrictions en ce qui concerne les actifs de données, conformément à la Directive sur la gestion de la sécurité et à la Directive sur la gestion de l'identité.
- Veiller à ce qu'il existe des processus et procédures pour appuyer la production de données en réponse aux demandes de renseignements en vertu de la Loi sur l'accès à l'information et de la Loi sur la protection des renseignements personnels.
- Utiliser un langage simple (voir le Guide de rédaction du contenu du site Canada.ca) et des formats lisibles par machine (par exemple les fichiers CSV, les fichiers de langage XML, les fichiers JSON) pour faciliter le partage, le traitement, l'utilisation, la publication et l'archivage des données, y compris dans les métadonnées requises.
- Fournir aux utilisateurs de multiples méthodes d'accès aux données et d'extraction de celles-ci, notamment en offrant les données dans des formats multiples et au moyen d'interfaces de protocole d'application (API) accessibles élaborées conformément aux Normes du gouvernement du Canada sur les API.
- Investir dans les infrastructures de données afin de fournir un accès facile et sécurisé aux données, conformément à l'approche « en nuage d'abord » prévue dans la Directive sur les services et le numérique. Les données sensibles (Protégé B, Protégé C ou Classifié) doivent être conservées dans des systèmes situés à l'intérieur des frontières géographiques du Canada ou dans les organisations du gouvernement du Canada à l'étranger (voir l'Orientation sur l'utilisation sécurisée des services commerciaux d'informatique en nuage : Avis de mise en œuvre de la Politique sur la sécurité [AMOPS] et le Profil des mesures de sécurité pour les services du gouvernement du Canada fondés sur l'informatique en nuage).

- Travailler à l'ouverture par défaut et publier des données dans le Portail du gouvernement ouvert, conformément à la Directive sur le gouvernement ouvert et dans le respect des cadres fédéraux applicables en matière de protection des renseignements personnels, de sécurité et de propriété intellectuelle.
- Réaliser des sondages pour identifier les obstacles à la recherche, à l'accès et à l'utilisation des données au sein de votre organisation.
- Signaler tout accès aux données ou utilisation des données non autorisés aux responsables de la sécurité désignés et, en cas de renseignements personnels, aux responsables de la protection des renseignements personnels du ministère, ainsi qu'au Secrétariat du Conseil du Trésor du Canada et au Commissariat à la protection de la vie privée du Canada, comme l'exige la Directive sur les pratiques relatives à la protection de la vie privée.

Exactitude

- Consulter les sources de données fiables pour identifier les sources d'erreur et déterminer les mesures requises pour corriger ces erreurs, le cas échéant, vérifier le contenu et comprendre le contexte entourant les données.
- Veiller à ce que les données soient décrites conformément aux normes de référence prescrites pour les métadonnées (c'est-à-dire au niveau de l'organisation ou du ministère) afin de permettre aux utilisateurs de déterminer leur exactitude. Les métadonnées pertinentes pourraient inclure des renseignements sur la source, le but et la méthode de collecte, le traitement, les révisions, la couverture, le modèle de données, ainsi que sur les hypothèses connexes.
- Veiller à ce que les données soient adéquatement représentatives de tous les domaines (par exemple les zones géographiques, les populations) qu'elles contiennent.

- Assurer des plages de valeurs adéquates, le cas échéant. Des explications concernant les valeurs aberrantes doivent être fournies aux utilisateurs de données.
- Élaborer des règles permettant de veiller à ce que les données ne comportent pas d'erreurs, y compris le dédoublement dans un jeu de données. Appliquer les règles tout au long du cycle de vie des données, en particulier lors de la collecte et du partage des données.
- Veiller à ce que les méthodes utilisées tout au long du cycle de vie des données minimisent les biais et les erreurs statistiques (par exemple les erreurs d'échantillonnage). (Voir la méthode fondée sur l'erreur d'enquête totale et Analyse comparative entre les sexes Plus [ACS Plus]).
- Veiller à ce qu'il existe une source faisant autorité pour les données, dans la mesure du possible.
- Élaborer des procédures permettant de corriger ou de mettre à jour les renseignements personnels sur demande (voir la Directive sur les pratiques relatives à la protection de la vie privée).
- Valider les concepts et les hypothèses utilisés avec des experts en la matière, afin de déterminer dans quelle mesure les données correspondent à ce que l'utilisation recherche.
- Fournir des renseignements sur le niveau d'erreur ou d'incertitude dans le cas des données (par exemple une erreur standard, des intervalles de confiance), le cas échéant.
- Veiller à ce que les résultats des systèmes d'intelligence artificielle (IA) (par exemple l'IA générative utilisée ou déployée par un ministère) soient évalués en termes d'exactitude, notamment au moyen de tests de partialité.

Cohérence

- Le cas échéant, adopter les Normes référentielles pangouvernementales relatives aux données du gouvernement du

Canada, notamment pour le codage et le formatage des données.

- En l'absence d'une norme référentielle pangouvernementale relative aux données du gouvernement du Canada, adopter ou adapter les normes relatives aux données ministérielles, nationales et/ou internationales existantes et documenter les différences de pratiques, en particulier lors du partage de données avec d'autres organisations ou de la publication de données sur le Portail du gouvernement ouvert. Les normes pertinentes pourraient être propres à un domaine, conçues pour des types particuliers de données (par exemple les données statistiques ou géospatiales).
- Consigner de manière cohérente les normes référentielles relatives aux données utilisées dans un répertoire ou un catalogue de données ou dans des ententes de partage de données. Si de nouvelles normes référentielles relatives aux données sont élaborées, documenter les raisons pour lesquelles on n'utilise pas les normes gouvernementales ou ministérielles existantes et applicables ne sont pas utilisées.
- Veiller à ce que les éléments de données soient définis, classifiés et représentés selon les architectures de données communes, conformément au Cadre de l'architecture intégrée du gouvernement du Canada.
- Veiller à ce que les concepts, les définitions et les nomenclatures soient compatibles dans les jeux de données et entre eux, afin de permettre la comparaison et l'intégration des données, tant en interne qu'entre le gouvernement du Canada et les organisations externes.
- Utiliser des tableaux de concordance permettant d'illustrer les écarts et les transitions entre les normes utilisées pour toutes les sources de données.
- Réduire la duplication des données entre les jeux de données afin d'améliorer l'intégrité des données et de garantir leur unicité.

Exhaustivité

- Veiller à ce qu'aucune entrée, colonne ou ligne essentielle ne soit manquante ou incomplète.
- Tenir à jour les valeurs, les concepts, les définitions, les nomenclatures et les méthodologies.
- Attribuer des étiquettes obligatoires et facultatives aux colonnes ou aux lignes d'un jeu de données afin de faciliter les évaluations de l'exhaustivité.
- Compléter les données au moyen des métadonnées appropriées qui précisent le contexte et le but de leur acquisition. Les métadonnées pourraient également préciser les facteurs relatifs à la vie privée, à la confidentialité ou à l'exactitude ayant une incidence sur l'exhaustivité.

Constance

- Élaborer des règles pour valider les relations logiques codées dans un jeu de données. Cela pourrait inclure des règles officialisant la relation entre deux variables interreliées.
- Valider régulièrement la constance des jeux de données. Les processus de validation devraient être normalisés et automatisés pour favoriser l'efficience.
- Tenir un registre des problèmes de constance décelés au moyen des procédures de validation des données et revoir périodiquement les règles de validation afin de garantir leur pertinence et leur efficacité.
- Obtenir les métadonnées appropriées auprès du fournisseur de données pour en apprendre davantage sur les classes d'entités d'un jeu de données, les valeurs qu'elles sont censées permettre et les relations qui existent entre elles.

Intelligibilité

- Adopter, adapter ou développer des vocabulaires contrôlés afin de veiller à ce que les concepts clés soient nommés et définis de façon

uniforme dans un jeu de données. Voir les métadonnées ou les normes référentielles prescrites relativement aux données. Respecter les normes référentielles prescrites régissant les valeurs admissibles des éléments d'un jeu de données (par exemple les données de référence, les données de base).

- Conformément à la Norme pour la gestion des métadonnées, appliquer les normes référentielles prescrites en matière de métadonnées pour les définitions et les procédures afin de préciser pourquoi et comment les données ont été recueillies, ainsi que leur catégorisation en matière de sécurité, en tenant compte des besoins des publics cibles.
- Consigner l'information nécessaire pour interpréter les données de façon significative, notamment en ce qui concerne le but initial des méthodes de collecte et de calcul des données, et maintenir des liens entre cette information et les données tout au long de leur cycle de vie.
- Veiller à ce que les utilisateurs soient conscients des limites des données.

Pertinence

- Établir des processus permettant de consulter les intervenants au sujet de leurs besoins en matière de données. Il pourrait s'agir de tirer parti des répertoires ou des catalogues de données pour déterminer les données existantes et réduire au minimum la collecte de données redondantes (voir la Ligne directrice sur les services et le numérique pour obtenir des directives sur la collecte de renseignements et de données).
- Déterminer les besoins en matière de données et les sources de données en fonction des objectifs opérationnels et des besoins des utilisateurs.
- Évaluer et documenter la façon dont les actifs de données satisfont aux exigences en matière de données afin d'évaluer leur pertinence. Cela pourrait comprendre le suivi sur la manière dont les actifs de données sont utilisés et réutilisés.

- Utiliser les résultats des évaluations de la pertinence pour orienter l'acquisition future de données et les activités connexes de gestion du cycle de vie et de gouvernance.
- Établir des critères permettant de garantir que les efforts d'acquisition de données établissent un équilibre approprié entre les besoins opérationnels et les risques en matière de protection des renseignements personnels et de sécurité (voir les Principes de nécessité et de proportionnalité de Statistique Canada).
- Veiller à ce que l'institution dispose du pouvoir législatif nécessaire pour recueillir ou créer des données concernant une personne identifiable et que la collecte soit directement liée à un programme ou à une activité opérationnelle au sein de l'institution.
- Préserver les données et les métadonnées connexes qui ont une valeur historique ou archivistique conformément à la Loi sur la Bibliothèque et les Archives du Canada et aux instruments de politique connexes.

Fiabilité

- Veiller à ce que les méthodes de collecte et d'analyse des données soient clairement articulées afin de faciliter la validation par un tiers et de maintenir l'intégrité du processus de production des données.
- Identifier et consigner les sources susceptibles de modifier directement ou indirectement un jeu de données. Les sources de changement pourraient inclure ce que représentent les données, les méthodes de collecte de données, les technologies de saisie et de stockage de données, les plateformes de traitement de données, les mesures législatives ou réglementaires, les exigences de politique et les cyberattaques.
- Tester les instruments de collecte ou de création de données avant de les déployer, et documenter les étalonnages et tenir de la variation des résultats.

- Tenir un registre des modifications apportées à vos actifs de données afin que les utilisateurs puissent déterminer leur provenance et suivre l'évolution depuis leur création (c'est-à-dire documenter au moyen de métadonnées).
- Déterminer et consigner les dépendances entre les actifs de données liés dans une architecture de données ou dans le contexte de l'analyse des données.
- Appuyer la compatibilité des concepts, des définitions et des nomenclatures au fil du temps. Préciser et expliquer les écarts quant à la façon dont ces éléments sont conservés au fil du temps.
- Protéger les actifs de données contre les activités frauduleuses ou non autorisées qui pourraient nuire à leur crédibilité. Cela comprend la définition, la mise en œuvre et la tenue à jour des mesures de sécurité permettant de répondre aux exigences en matière de sécurité de la technologie de l'information (TI), conformément à la Directive sur la gestion de la sécurité et à la Directive sur les pratiques relatives à la protection de la vie privée.
- Utiliser des approches de conservation numérique pour surveiller et prévenir la détérioration des actifs de données tout au long de leur cycle de vie. Cela comprend des vérifications régulières de l'intégrité des données (par exemple au moyen du hachage ou en utilisation des sommes de contrôle) et la documentation de toute preuve de détérioration conformément à la Loi sur la Bibliothèque et les Archives du Canada et aux instruments de politique connexes.
- Signaler l'altération ou la destruction non autorisée des actifs de données aux agents de sécurité désignés.
- Veiller à ce que les données aient une source faisant autorité, dans la mesure du possible.

Actualité

- Déterminer les besoins actuels et futurs des utilisateurs en matière de données, y compris les considérations de temps (par exemple les périodes de référence, les exigences législatives ou stratégiques, les normes relatives aux services).
- Consulter les fournisseurs de données pour évaluer si les besoins en données peuvent être satisfaits sans délai et informer les utilisateurs de données de tout problème prévu. Cela pourrait comprendre la confirmation de la capacité du fournisseur de données à respecter les délais prévus dans les ententes de partage de données.
- Veiller à ce que les fournisseurs de données disposent d'un calendrier de diffusion des données indiquant les étapes du processus de production des données et tenant compte des écarts et des retards (par exemple au moyen de la planification d'urgence).
- Publier des données préliminaires sur le Portail du gouvernement ouvert, le cas échéant, et conformément à la Directive sur le gouvernement ouvert.

© Sa Majesté le Roi du chef du Canada, représentée par le président du Conseil du Trésor, 2024,

ISBN : 978-0-660-69836-6

Date de modification :

2024-02-01