

AIDE-MÉMOIRE SUR LA QUALITÉ DES DONNÉES

*Pour des données
exploitables et fiables*



Octobre 2024

TABLE DES MATIÈRES

Métadonnées	1
Étape 1 - Décrire le cas d'utilisation des données	2
Étape 2 - Identifier des sources de données potentiellement utiles	3
Étapes 3 à 5 - Définir les besoins, évaluer la qualité des données et adopter des mesures	4
Conseils	6
Conseils pour l'étape 1	6
Conseils pour l'étape 2	6
Définitions des dimensions de la qualité des données	8
À propos de Nord Ouvert	9

MÉTADONNÉES

Nom de l'organisation ou du projet		
Numéro de version du document	Dernière mise à jour (yyyy-mm-dd)	Personne responsable de la qualité des données
Collaborateur(s)	Actions prioritaires	

ÉTAPE 1

Décrire le cas d'utilisation des données ¹

Description
Rétroaction

(1) Voir la section [Conseils](#) ci-dessous pour des conseils utiles sur la manière de réaliser cette étape.

ÉTAPE 2

Identifier des sources de données potentiellement utiles ²

Sources de données potentielles	Pertinence

(2) Voir la section Conseils ci-dessous pour des conseils utiles sur la manière de réaliser cette étape.

ÉTAPES 3 À 5

Définir les besoins, évaluer la qualité des données et adopter des mesures

Dimension de la qualité des données *	Étape 3 - Exigences Traduisez la dimension en exigences spécifiques pour votre cas d'utilisation. Fournissez une justification si vous pensez que la dimension ne s'applique pas. **	Étape 4 - Évaluation À quel niveau vos données répondent-elles à ces exigences ?	Étape 5 - Plan d'action Si les exigences ne sont pas satisfaites de manière adéquate, que pouvez-vous faire pour les améliorer ?
Bien documentées	<ul style="list-style-type: none"> • Définir les champs de métadonnées à remplir obligatoirement. • Les descriptions des métadonnées doivent être écrites dans un langage clair et simple. 	Faible Moyen Bon	
Collectées en temps voulu	Pourcentage de données suffisamment récentes et opportunes pour une tâche donnée. Note : vous pouvez déterminer des intervalles, par exemple : <ul style="list-style-type: none"> • 75-100% = good • 50-75% = fair • below 50% = poor 	Faible Moyen Bon	
Complètes	<ul style="list-style-type: none"> • Définir le nombre ou le pourcentage attendu de valeurs complètes pour réaliser une tâche ou une analyse. Vous pouvez également définir des intervalles pour les seuils d'exhaustivité. 	Faible Moyen Bon	
Fiables	<ul style="list-style-type: none"> • Seuil d'erreur acceptable à 1 mètre dans le cas d'une mesure de distance entre deux objets ou personnes (exactitude). • Aucun doublon (unicité) • 100 % des dates sont formatées conformément à la norme ISO 8601 (cohérence). 	Faible Moyen Bon	

Pensez aux types de mesures suivants :

Technique (outils et logiciels, automatisation, fonctions de détection des doublons, tableaux de bord sur la qualité des données).

Administratif (processus, protocoles, politiques, formation).

Gouvernance (structures de prise de décision).



* Définitions fournies en dernière page

** Des exemples sont proposés pour chaque dimension

(suite)

Dimension de la qualité des données *	Étape 3 - Exigences **	Étape 4 - Évaluation	Étape 5 - Plan d'action
Réutilisables	<ul style="list-style-type: none"> Tous les ensembles de données doivent être stockés dans un format lisible par machine. La structure des données et les conventions de nommage sont normalisées conformément aux normes du domaine. 	Faible Moyen Bon	
Accessibles	<ul style="list-style-type: none"> Tous les membres du personnel ayant la formation et l'autorité nécessaires peuvent accéder aux données. 	Faible Moyen Bon	
Sécurisées	<ul style="list-style-type: none"> Les renseignements personnels ont été identifiés et signalés en vue d'une analyse de risque. Des mesures de protection techniques, administratives et physiques appropriées ont été prévues et mises en œuvre. 	Faible Moyen Bon	
Pertinentes	<ul style="list-style-type: none"> Des parties prenantes concernées par le cas d'usage ont été consultées en vue de combler les lacunes en matière de perspectives. Des stratégies de suivi et d'évaluation sont mises en œuvre pour déterminer si les données ont permis d'atteindre les objectifs. Les données de l'échantillon sociodémographique sont représentatives de la population cible. 	Faible Moyen Bon	
<i>(Dimension supplémentaire)</i>		Faible Moyen Bon	

* Définitions fournies en dernière page

** Des exemples sont proposés pour chaque dimension

CONSEILS

Conseils pour l'étape 1

Soyez aussi précis que possible lorsque vous décrivez les différentes composantes de votre cas d'utilisation des données :

- L'énoncé du problème.
- Le contexte dans lequel le problème émerge.
- Les parties prenantes concernées.
- Les résultat(s) et objectif(s) souhaité(s).
- L'énoncé de votre hypothèse sur la manière dont l'utilisation des données permettra d'atteindre le(s) objectif(s) fixé(s).
- Toute autre information pertinente, telle que les conditions de réussite.

Conseils pour l'étape 2

Voici quelques conseils pour vous aider à identifier les données adaptées à votre cas d'utilisation.

1. Définir le champ d'application des données

Pour réfléchir à la qualité des données, il faut d'abord se questionner sur leur pertinence, c'est-à-dire trouver des données utiles et adaptées à la tâche à accomplir. Voici quelques exemples de questions :

quoi	Qu'est-ce que je veux observer et mieux comprendre (les individus, l'environnement bâti, la nature, les services, etc.) Comment cet élément a-t-il été mesuré dans le passé ? Quelles sont les forces et les limites de ces méthodes ?
où	Les données sont-elles liées à des frontières géographiques spécifiques (pays, région, continent) et à des juridictions (fédérales, étatiques, provinciales ou locales) ?
quand	Existe-t-il une période de temps pertinente pendant laquelle les données doivent être collectées ?
thème	Les données sont-elles liées à un sujet ou à un domaine spécifique, tel que les soins de santé, la mobilité, le logement, etc. Le sujet est-il multidisciplinaire ?
fréquence	Les données doivent-elles être mises à jour à une fréquence spécifique ?
format	Les données doivent-elles être disponibles dans un format spécifique ?

2. Identifier les sources de données potentiellement pertinentes

Données internes	Votre organisation ou votre équipe collecte-t-elle actuellement ces données ou a-t-elle la capacité de le faire à l'avenir ?
Données ouvertes	Les données sont-elles publiées de manière ouverte ? Dans un souci de transparence de la part des pouvoirs publics, les organismes publics proposent de plus en plus souvent des plateformes de données ouvertes sur lesquelles vous pouvez filtrer vos recherches pour trouver des ensembles de données ouvertes. En fonction de vos contraintes géographiques, vous pouvez consulter les portails de données ouvertes au niveau fédéral, de l'état ou de la province, ou au niveau local ³ . En outre, il existe de nombreux référentiels de données qui publient régulièrement des ensembles de données ouvertes à l'échelle mondiale (voir les exemples ci-dessous).
Shared data	Recherchez des plateformes commerciales où acheter des ensembles de données sur mesure (par exemple auprès de Statistique Canada) ou, alternativement, recherchez des partenariats de données existants dans lesquels vous pouvez rejoindre des organisations partageant les mêmes idées afin de collaborer, d'échanger, et de mutualiser des données dans le but d'atteindre des objectifs communs. Le cours «Introduction aux Partenariats de Données» de Nord Ouvert (Code D104F), disponible sur notre centre de formation en ligne gratuit , propose des exemples de partenariats de données existants dans divers domaines tels que le climat ou la santé.

(3) Exemples : [Portail fédéral canadien de gouvernement ouvert](#) | [Portail de données provinciales du Québec](#) | [Portails de données municipales pour les villes de Montréal et de Toronto](#).

Exemples de référentiels de données à l'échelle mondiale				
Dépôts certifiés Core - CoreTrustSeal	Re3data	FAIRsharing	Google Public Data Explorer	Données de l'ONU
Données de la Banque mondiale	Open Knowledge Hub	Our World in Data	Observatoire de la santé mondiale	OpenStreetMap

DÉFINITIONS DES DIMENSIONS DE LA QUALITÉ DES DONNÉES⁴

Bien documentées	Les données bien documentées sont accompagnées de métadonnées complètes et adéquates.
Collectées en temps voulu	Les données sont suffisamment à jour pour la tâche à accomplir.
Complètes	Les données complètes ne contiennent aucune valeur manquante ou un nombre acceptable de valeurs manquantes pour la tâche ou l'analyse en question.
Fiables	Des données fiables sont le résultat de processus de collecte de données cohérents et stables qui peuvent être répétés avec la certitude d'obtenir des résultats similaires.
Réutilisables	Les données réutilisables ont le potentiel de servir à d'autres fins que celles pour lesquelles elles ont été conçues à l'origine.
Accessibles	Les données accessibles peuvent être facilement utilisées et consultées par les utilisateurs prévus.
Sécurisées	Les données sécurisées sont protégées et l'accès est limité aux personnes possédant la formation et les autorisations nécessaires.
Pertinentes	Les données pertinentes sont utiles et applicables à la tâche à accomplir.

(4) Pour plus de détails, voir le guide de [Nord Ouvert : La qualité des données en pratique](#)

À Propos de Nord Ouvert

Nord Ouvert est une organisation à but non lucratif dédiée à l'avancement du bien commun. Nous travaillons avec les gouvernements, les organisations à vocation citoyenne et les entreprises à mission pour créer et mettre en œuvre des solutions pratiques grâce à des stratégies numériques transformatrices et à des cadres de gouvernance des données.

L'équipe de Nord Ouvert est composée de professionnels possédant un large champ d'expertise, notamment en matière de gouvernance, de technologie de l'information, de planification stratégique et opérationnelle, d'urbanisme, de développement communautaire, de recherche appliquée, de développement international et d'élaboration de politiques publiques. Grâce à la pluralité de leurs compétences, les membres de l'équipe d'Open North apportent des perspectives et une expérience précieuses à tous les projets.

Nord Ouvert fait partie de Montréal en commun, un projet piloté par la Ville de Montréal dans le cadre du Défi des villes intelligentes et réalisé grâce au soutien financier du Gouvernement du Canada.

opennorth.ca/fr

À propos du Défi des villes intelligentes et de Montréal en commun

Montréal en commun est une communauté d'innovation pilotée par la Ville de Montréal dont les partenaires expérimentent des solutions en accès à l'alimentation, en mobilité et en réglementation municipale dans un désir de repenser la ville. Les projets de Montréal en commun sont mis en œuvre grâce au prix octroyé à la Ville de Montréal par le Gouvernement du Canada dans le cadre du Défi des villes intelligentes.

Auteur: Samuel Kohn

Ce travail est protégé par le droit d'auteur de Nord Ouvert sous licence [Creative Commons 4.0 International](#) (CC BY 4.0), à l'exception des photographies, des images, des logos, de la marque et des autres marques de commerce de Nord Ouvert.





Aide-mémoire sur la qualité des données