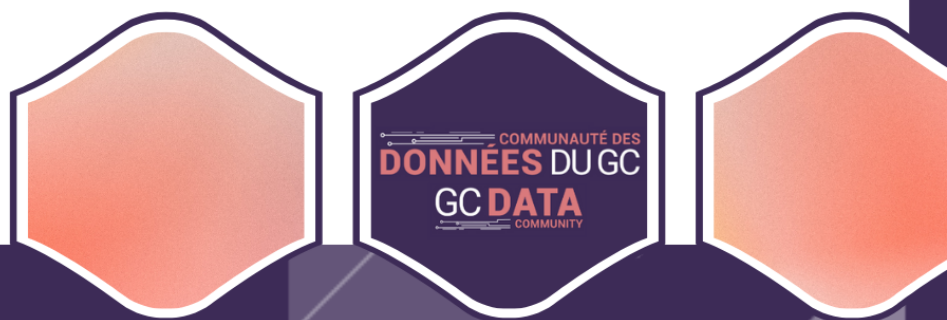


CADRE DE COMPÉTENCE EN MATIÈRE DE DONNÉES

DU GOUVERNEMENT DU CANADA



COMMUNAUTÉ DES DONNÉES DU GC

EN PARTENARIAT AVEC :

GROUPE DE TRAVAIL SUR LES PERSONNES ET LA CULTURE,
COMMUNAUTÉ DE PRATIQUE SUR LES DONNÉES INTÉGRÉES DU GC
ACADÉMIE DU NUMÉRIQUE, ÉCOLE DE LA FONCTION PUBLIQUE DU
CANADA



Vanessa Vermette

Vice-présidente
Direction générale de l'innovation et du
perfectionnement des compétences
École de la fonction publique du Canada



Christopher Valiquet

Directeur
Communauté des données du GC
École de la fonction publique du Canada

Avant-propos

Comme les données font partie intégrante de notre quotidien, elles sont essentielles à l'établissement d'une fonction publique numérique axée sur les citoyens et les citoyennes. On croirait à tort que les compétences en matière de données demeurent des connaissances spécialisées réservées à un groupe restreint de personnes. Effectivement, dans un monde qui évolue rapidement où les données forment la base de l'information, l'amélioration de ces compétences s'impose dans toutes les professions.

Les connaissances et le savoir-faire à l'égard des données sont au cœur des mesures de modernisation. Le cadre de compétences en matière de données est conçu pour soutenir les conversations et faire progresser les connaissances au sujet des données en permettant aux fonctionnaires du gouvernement fédéral d'avoir une compréhension commune des compétences liées aux données, ainsi qu'un langage commun en la matière. Cette compréhension peut servir à soutenir les mesures destinées à accroître la littératie en matière de données, comme l'élaboration de parcours d'apprentissage et d'évaluations connexes, ainsi que les initiatives de gestion des talents dans le domaine des données. En somme, elle nous permettra de mieux comprendre et d'accroître la valeur des données comme actif organisationnel.

Compte tenu des différences entre les ministères et agences en ce qui concerne leur contexte organisationnel et leurs secteurs d'activité, le cadre se veut un guide adaptable à leurs besoins.

Nous remercions les organisations et nos collègues du gouvernement du Canada qui ont contribué à l'élaboration du cadre.

Enfin, nous souhaitons que le cadre favorise l'unification et le renforcement des connaissances et du savoir-faire en matière de données à l'échelle de la fonction publique, dans l'espoir de constituer un effectif dont les données feront la force.

Table des matières

Contexte	4
Remerciements	4
Introduction	5
Comment le cadre peut-il être utilisé ?	5
Le cadre en un coup d’œil	7
Niveaux de compétence	8
1. Concepts et culture des données	10
1.1 Connaissances des données, du numérique et de l’organisation	10
1.2 Éthique et confidentialité relativement aux données	12
1.3 Prise de décisions fondées sur des données probantes	14
2. Gouvernance, collecte et intendance des données	16
2.1 Gouvernance, intendance et normes relatives aux données	16
2.2 Collecte des données	18
2.3 Qualité et valeur des données, et confiance à l’égard de celles-ci	20
2.4 Accès, sécurité et interopérabilité	22
3. Analyse et évaluation	24
3.1 Questions et définition des difficultés	24
3.2 Analyses et science des données	25
3.3 Narration et visualisation	28
3.4 Évaluation des résultats	30
4. Systèmes de données et architecture	32
4.1 Architecture des données intégrées	32
4.2 Systèmes de données	34
Annexe A :	
Définitions des termes clés	36
Annexe B :	
Ressources supplémentaires	40

Contexte

Remerciements

Nous tenons à souligner l'importante contribution de nos nombreux partenaires :

- l'Académie du numérique et l'École de la fonction publique du Canada ;
- le Groupe de travail sur les personnes et la culture de la Communauté de pratique sur les données intégrées du GC, avec la contribution directe de plus d'une douzaine de ministères ;
- le Secrétariat du Conseil du Trésor du Canada ;
- le Bureau du Conseil privé ;
- la Division de l'analyse stratégique, des publications et de la formation de Statistique Canada ;
- le Réseau de la science des données ;
- la Communauté de la gestion de l'information ;
- le Bureau de la dirigeante principale des données d'Emploi et Développement social Canada.

Une reconnaissance particulière est accordée à Emploi et Développement social Canada (EDSC) pour l'opérationnalisation du cadre de compétences en matière de données en tant que scénario d'utilisation opérationnelle et pour leur partage ouvert afin de promouvoir la collaboration au sein de la communauté des données et au-delà.

Introduction

La fonction publique fédérale se doit d'être compétente en matière de données pour favoriser la modernisation et demeurer efficace pour les personnes qu'elle sert. Le développement de la littératie en matière de données nécessite des définitions et une compréhension communes des compétences nécessaires permettant aux fonctionnaires de travailler ensemble pour traiter les données en tant que bien organisationnel. Le cadre de compétences en matière de données du gouvernement du Canada permet d'orienter les efforts en matière de littératie des données, qui concernent l'ensemble des fonctionnaires qui travaillent avec des données ou qui les utilisent.

L'élaboration de ce cadre a été un effort de la communauté qui a commencé après la publication en 2018 de la Feuille de route de la stratégie en matière de données pour la fonction publique fédérale. Dirigée par l'Académie numérique de l'École de la fonction publique du Canada (EFPC) et Statistique Canada, en collaboration avec l'Université Dalhousie, la première étape fut l'élaboration d'un fondement à l'aide d'une analyse documentaire sur la littératie en matière de données. En 2021, la Communauté de données du GC et le Groupe de travail sur les personnes et la culture de la Communauté de pratique des données d'entreprise ont mené des consultations avec des partenaires issus des communautés des données, du numérique, de l'information et des RH pour mettre à jour le cadre afin de refléter les divers contextes organisationnels et les besoins du gouvernement du Canada. En 2022, la Communauté des données du GC a poursuivi la mise en correspondance des compétences avec les niveaux de compétence, en s'appuyant sur le travail effectué par le Bureau principal des données d'Emploi et Développement social Canada. Les commentaires sont les bienvenus et seront utilisés pour les itérations futures de ce cadre. Les questions ou les commentaires peuvent être envoyés à la communauté des données du GC à l'EFPC (gcdc-cdgc@cspc-efpc.gc.ca).

Comment le cadre peut-il être utilisé ?

Le cadre a été développé pour établir une compréhension commune des compétences en matière de données de base, intermédiaires et avancées requises dans l'ensemble du gouvernement fédéral. Il répertorie l'ensemble des connaissances et des compétences requises tout au long du cycle de vie des données pour permettre une prise de décision efficace et rigoureuse fondée sur des données probantes.

Le cadre a trois utilisations principales.

- 1) Aider les **employés et employées** de l'ensemble de la fonction publique à mieux comprendre le paysage des données et à renforcer leurs connaissances en la matière.

- 2) Aider les **ministères et les dirigeantes et dirigeants** à comprendre l'étendue des compétences en matière de données requises dans l'ensemble d'une organisation pour permettre une prise de décision efficace fondée sur les données :
- gérer et exploiter efficacement les données en tant que biens stratégiques et organisationnels ;
 - assurer une bonne gouvernance selon des normes éthiques et opérationnelles ;
 - concevoir et mettre en place une infrastructure durable et sûre pour la collecte, l'accès, l'utilisation, la gestion, l'interopérabilité et la conservation des données ;
 - soutenir un effectif compétent en matière de données ;
 - accorder la priorité aux besoins, aux attentes et à la confiance des personnes desservies par le gouvernement du Canada lorsqu'il s'agit de politiques, de programmes et de prestation de services.
- 3) Guider les ministères et les agences dans la réalisation d'**évaluations de la littératie des données** afin de cerner les compétences disponibles et les lacunes.

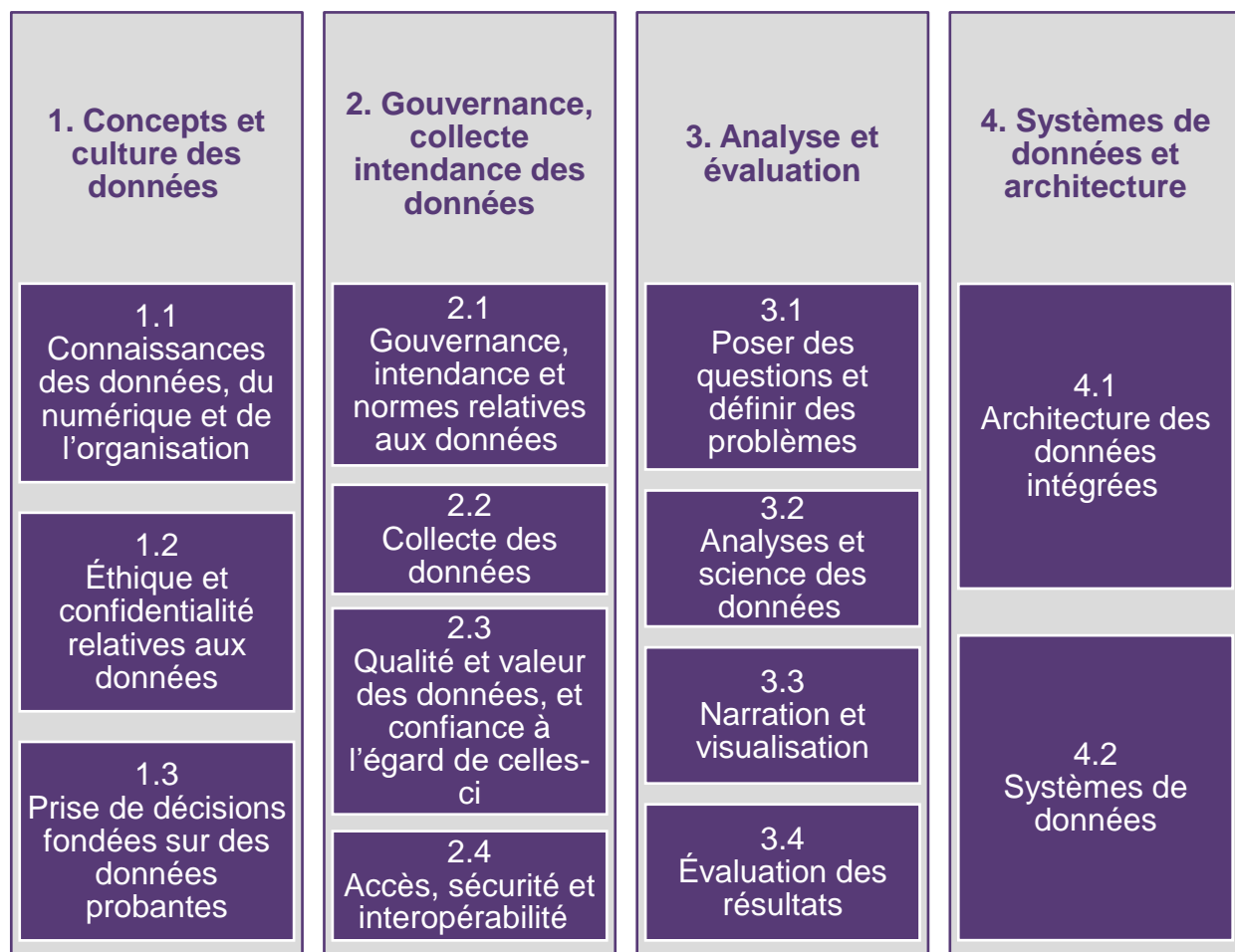
Il est important de noter qu'étant donné que le cadre n'a pas été mis en correspondance avec les classifications et les descriptions de postes, il n'est pas destiné à être utilisé comme un outil des RH formel ou pour effectuer des évaluations du rendement. Cela dit, il peut être utilisé pour orienter et informer les conversations sur les besoins en compétences en matière de données.

Comment le cadre est-il organisé ?

Le cadre comporte quatre catégories qui jouent chacune un rôle essentiel dans le cycle de vie des données :

- 1) **concepts et culture des données** ;
- 2) **gouvernance, collecte et gestion des données** ;
- 3) **analyse et évaluation** ;
- 4) **systèmes de données et architecture**.

Le cadre en un coup d'œil



Ensemble, ces catégories comprennent un total de **13 compétences**, qui sont réparties en indicateurs qui, eux, sont divisés en niveaux fondamental, intermédiaire et avancé. Un glossaire des termes est présenté à l'[annexe A](#).

Niveaux de compétence

Niveau 1 — Principes fondamentaux Définir le niveau de base de compréhension et de conscience

- Trouver, saisir, lire, comprendre et utiliser des données, des diagrammes et des graphiques.
- Comprendre les termes et concepts simples relatifs aux données et à l'IA.
- Comprendre les étapes du travail avec les données.
- Effectuer des tâches simples (par exemple, effectuer des calculs Excel, créer des tableaux Excel, produire des rapports).

Niveau 2 — Principes intermédiaires Mettre la théorie en pratique

- Maîtriser parfaitement les connaissances de base en matière de données.
- Savoir comment accéder aux données qui sont les mieux adaptées à l'usage prévu, comment les utiliser et comment découvrir les lacunes.
- Réaliser avec succès diverses tâches liées aux données (par exemple, utiliser différentes méthodes statistiques et techniques d'analyse, nettoyer et traiter les données).
- Utiliser des termes et des concepts liés aux données lors de conversations.
- Accéder à des données provenant de différentes sources et les traiter avec assurance.
- Analyser des données quantitatives et qualitatives et tirer des conclusions à partir de celles-ci.
- Tenir compte de la source des données et comprendre que la raison pour laquelle elles ont été recueillies ainsi que la manière dont elles ont été recueillies peuvent influencer leur qualité et leur utilité.

Niveau 3 — Principes avancés Applications avancées et habiliter les autres

- Posséder une connaissance approfondie des données ainsi que de la manière d'y accéder et de les traduire en connaissances exploitables.
- Combiner des données qualitatives et quantitatives pour comprendre la complexité d'une question ou d'un problème sous différents angles et à différents niveaux.
- Cerner, analyser et résoudre des problèmes complexes liés aux données.
- Trouver des moyens créatifs d'utiliser les données pour résoudre des problèmes ou orienter les politiques.
- Donner l'exemple, aider les autres et leur enseigner.

Le contexte est essentiel pour cerner les compétences et les niveaux de maîtrise nécessaires. Si certaines compétences, telles que la sensibilisation aux données, sont nécessaires au niveau de base pour l'ensemble des membres du personnel, d'autres ne sont applicables qu'à des rôles particuliers, par exemple les systèmes de données. En outre, les indicateurs sont censés fournir une gamme d'aptitudes pour une compétence donnée. Ainsi, les combinaisons appropriées de compétences, d'aptitudes et de niveaux de maîtrise requis dépendront des besoins spécifiques d'un rôle, d'une opération, d'une activité ou d'un projet. Il existe de nombreuses situations dans lesquelles l'ensemble des compétences d'une compétence donnée n'est pas nécessaire ; par exemple, un projet de recherche par sondage typique ne nécessitera pas les indicateurs liés à l'intelligence artificielle énumérés dans la compétence Analyse et science des données.

1. Concepts et culture des données

1.1 Connaissances des données, du numérique et de l'organisation

Connaissances sur la façon dont les données façonnent le gouvernement d'aujourd'hui, et comment les utiliser efficacement. Cela comprend les compétences qui sous-tendent une organisation axée sur les données.

Principes fondamentaux

1.1.1 Connaître les termes, normes, politiques, documents et communautés clés en matière de données et de numérique, notamment :

- Termes relatifs aux données
- Rôle des bureaux principaux de données fédéraux et ministériels
- Stratégies fédérales et ministérielles en matière de données
- [Stratégie relative aux données de 2023-2026 pour la fonction publique fédérale](#)
- [Loi sur la statistique](#)
- [Politique sur les services et le numérique](#)
- [Ambition numérique du Canada](#) et [normes relatives au numérique](#)

1.1.2 Comprendre [ce que sont les données](#), le cycle de vie des données (planifier, recueillir, traiter, utiliser et partager) et les nombreux [types de données](#) qui existent.

1.1.3 Comprendre le concept d'information et les termes liés à la valeur des données, à l'information et à l'analyse.

1.1.4. Comprendre la valeur des données en tant qu'atout stratégique et l'importance de la maîtrise des données pour l'organisation en vue de soutenir la prise de décision, la recherche, l'apprentissage et le développement, la prestation de services et la mesure des résultats.

1.1.5 Comprendre les rôles, les responsabilités et les obligations en matière de données.

1.1.6 Connaître les rôles, les politiques, les normes et les processus liés aux données de l'organisation et de leur finalité.

Principes intermédiaires

1.1.7 Se tenir au courant des grandes tendances en matière de données et de leurs répercussions possibles sur le gouvernement du Canada.

1.1.8 Utiliser les services, les ressources et le soutien disponibles par l'intermédiaire des centres d'expertise (Bureau du dirigeant ou de la dirigeante principale des données, division des normes de données/gouvernance, etc.) et d'autres sources fiables.

1.1.9 Comprendre des concepts complexes relatifs aux données, aux statistiques et à l'analyse.

1.1.10 Connaître les données qui existent au sein de l'organisation et leur utilisation pour informer ou soutenir la prise de décision.

1.1.11 Travailler dans le cadre d'un partenariat pluridisciplinaire avec des partenaires essentiels tels que les équipes responsables de la gestion de l'information, du numérique, des technologies de l'information, de la [protection des renseignements personnels et de la sécurité](#), et des affaires juridiques, ainsi que les experts en la matière et les utilisatrices et utilisateurs finaux.

Principes avancés

1.1.12. Habilitier ses collègues dans leur apprentissage et leur utilisation des principaux termes relatifs aux données quand ils parlent de données et travaillent avec celles-ci.

1.1.13 Montrer l'exemple et contribuer à la création d'un effectif qui connaît bien les données et d'une culture axée sur les données au sein de son organisation et dans l'ensemble du gouvernement.

1.1.14 Cerner et éliminer les obstacles afin de permettre à d'autres de travailler le plus efficacement possible avec des données (p. ex. en assurant l'accès à la formation, à l'infrastructure et aux logiciels).

1.1.15 Investir dans le talent, les systèmes et l'infrastructure qui permettent d'utiliser les données et d'y accéder en temps opportun pour la prise de décisions.

1.2 Éthique et confidentialité relativement aux données

Comprendre les considérations éthiques relatives à la collecte, à l'utilisation, à l'interprétation et le partage des données.

Principes fondamentaux

1.2.1 Connaître la signification de l'éthique des données, de la gouvernance, du consentement, des biais et de la discrimination, de l'inclusion, de l'équité et de la responsabilité.

1.2.2 Comprendre et respecter les principes et les normes clés en matière d'éthique, de confidentialité, de droit et de sécurité, y compris, mais sans s'y limiter, les éléments suivants :

- [Loi sur la protection des renseignements personnels](#)
- [Loi sur la statistique](#)
- [Politique sur la sécurité du gouvernement](#)
- [Politique sur la protection de la vie privée](#)
- [Évaluations des facteurs relatifs à la vie privée](#)
- [Niveaux de sécurité](#)
- [Analyse comparative entre les sexes plus](#)
- [Énoncé de politique des trois conseils : éthique de la recherche avec des êtres humains](#)
- [Politique modèle sur l'intégrité scientifique](#)
- [Directive sur la prise de décision automatisée](#)
- Données désagrégées

Principes intermédiaires

1.2.3 Savoir comment protéger et partager les données confidentielles.

1.2.4 Comprendre les implications éthiques de l'utilisation de l'IA.

1.2.5 Déterminer les indicateurs de biais et s'assurer que les politiques, programmes ou services ne renforcent pas les préjugés involontaires.

1.2.6 S'assurer que les données sont utilisées conformément à leur usage prévu en consultant les propriétaires ou les intendants et intendantes des données.

1.2.7 Appliquer des processus et des procédures pour garantir une approche éthique de la recherche et des données tout au long du cycle de vie de celles-ci.

1.2.8 Déterminer les questions éthiques, les implications en matière de confidentialité et de sécurité et les obstacles à l'accessibilité.

Principes avancés

1.2.9 Comprendre les concepts avancés tels que la nécessité, la proportionnalité, la sensibilité et l'explicabilité.

1.2.10 Évaluer les données pour déterminer la présence de biais, leur représentativité, leur exactitude et leur validité. Déterminer les étapes pour remédier aux problèmes et les mettre en œuvre au besoin.

1.2.11 Tester des idées et des solutions potentielles à l'aide d'un large éventail de données diverses afin de remettre en question les hypothèses et de garantir des solutions inclusives en raison de leur conception.

1.2.12 Concevoir et mettre en œuvre des politiques, des processus et des procédures pour assurer des approches éthiques de la recherche et de l'utilisation des données.

1.3 Prise de décisions fondées sur des données probantes

La prise de décision éclairée par les données probantes est le processus qui consiste à distiller et à diffuser les meilleures données probantes disponibles issues de la recherche, de la pratique et de l'expérience. Il s'agit de l'utilisation de données probantes pour éclairer et améliorer les politiques, les programmes, les opérations et la prestation de services à la population canadienne.

Principes fondamentaux

1.3.1 Privilégier l'utilisation des connaissances et des informations recueillies par l'intermédiaire de données plutôt que les témoignages anecdotiques simples.

1.3.2 Contribuer à répondre aux questions liées aux données opérationnelles et à les résoudre.

1.3.3. Consulter les autorités compétentes (par exemple, les experts et les expertes en la matière, les dirigeants et dirigeantes communautaires) pour déterminer ce qui est considéré comme des données de haute qualité dans un contexte particulier.

Principes intermédiaires

1.3.4 Repérer les données pour éclairer la prise de décision et soutenir l'évaluation de leur pertinence. Consigner quand et où les données requises ne sont pas saisies, collectées ou manquantes.

1.3.5 Utiliser les données pour comprendre les besoins des utilisateurs et utilisatrices et pour concevoir et développer des produits, des programmes et des services qui répondent à ces besoins (voir [Normes relatives au numérique du gouvernement du Canada : Guide de lecture](#) pour plus de détails).

1.3.6 Utiliser les données et les analyses pour évaluer le bien-fondé et l'incidence des solutions ou des décisions avant leur mise en œuvre.

Principes avancés

1.3.7 Communiquer les complexités et les nuances des résultats d'analyse à des auditoires non spécialisés afin de promouvoir la compréhension et la valeur ajoutée opérationnelle.

1.3.8 Développer et mettre en œuvre des politiques ou des initiatives visant à développer le talent, les systèmes ou l'infrastructure qui permettent l'utilisation et l'accès en temps opportun aux données pour la prise de décisions.

1.3.9 Évaluer la pertinence des données pour éclairer la prise de décision en intégrant des perspectives éthiques, méthodologiques et spécifiques au sujet ou au contexte.

2. Gouvernance, collecte et intendance des données

2.1 Gouvernance, intendance et normes relatives aux données

Gérer et mettre en œuvre les politiques, procédures et normes, et adhérer à celles-ci, en vue de favoriser la disponibilité, la facilité d'utilisation, l'intégrité, la sécurité et l'accessibilité des données au sein de l'organisation et dans l'ensemble du gouvernement.

Principes fondamentaux

2.1.1 Connaître la gouvernance et la souveraineté des données, notamment la [souveraineté des données des Premières Nations](#), les normes, les directives, les processus, y compris les normes d'accessibilité.

2.1.2 Déterminer et signaler de manière proactive les problèmes et conflits en matière de gouvernance, d'intendance et de normes relatives données.

Principes intermédiaires

2.1.3 Appliquer les politiques, procédures et normes clés pour la collecte, l'accès et la gestion des données.

2.1.4 Collaborer et négocier pour :

- assurer une compréhension commune des données ;
- gérer l'accès aux données ;
- déterminer les implications et les obstacles liés à la confidentialité, à la sécurité ou à l'accessibilité des données.

Principes avancés

2.1.5 Assurer une bonne gestion et une supervision des données tout en limitant la duplication des données et en respectant l'objectif prévu pour celles-ci.

2.1.6 Choisir des méthodes, des processus ou des outils qui réduisent la duplication des données, et utiliser les données là où elles se trouvent, dans la mesure du possible.

2.1.7 Surveiller les questions et changements relatifs aux données pour s'assurer que celles-ci sont activement gérées, accessibles et améliorées afin de maintenir des données fiables et utilisables sur le plan organisationnel.

2.1.8 Évaluer les priorités organisationnelles pour l'utilisation opérationnelle et stratégique des données.

2.1.9 Évaluer les besoins en matière de données pour les projets, les programmes, les rapports ministériels et les politiques afin de s'assurer que les données sont gérées en fonction de leur valeur et de leur utilisation.

2.1.10 Évaluer les exigences relatives à la conservation et à l'élimination des données.

2.1.11 Adapter ou mettre en œuvre la gouvernance des données dans le contexte opérationnel de l'organisation.

2.1.12 Développer des règles, des principes de gouvernance et des lignes directrices avant la collecte, la création, l'utilisation ou le partage des données.

2.1.13 Développer des politiques, une gouvernance ou des normes qui équilibrent et garantissent la sécurité, l'interopérabilité et l'accès aux données.

2.2 Collecte des données

Utiliser des méthodes, des processus, des outils, des plateformes et des logiciels pour recueillir des données.

Principes fondamentaux

2.2.1 Comprendre le rôle de la collecte de données et connaître les méthodes et outils courants de collecte de données.

2.2.2 Déterminer et utiliser les données existantes avant de collecter de nouvelles données.

2.2.3 Adhérer aux politiques, aux lois, aux processus et aux normes lors de la collecte de données.

- [Loi sur la protection des renseignements personnels](#)
- [Loi sur la statistique](#)
- Lois particulières aux ministères et agences en matière de collecte de données
- [Politique sur la sécurité du gouvernement](#)
- [Politique sur la protection de la vie privée](#)
- [Évaluations des facteurs relatifs à la vie privée](#)
- [Niveaux de sécurité](#)
- [Analyse comparative entre les sexes plus](#)
- [Énoncé de politique des trois conseils : Éthique de la recherche avec des êtres humains](#)
- [Modèle sur de politique l'intégrité scientifique](#)

Principes intermédiaires

2.2.4. Avoir une connaissance approfondie des méthodes et des outils de collecte de données, et de la manière de les appliquer pour garantir la qualité, la cohérence et le respect des délais.

2.2.5. Contribuer à l'atténuation ou à la résolution des problèmes liés à la collecte de données (par exemple, problèmes d'échantillonnage, données manquantes, erreurs et incohérences, etc.).

2.2.6 Concevoir et mettre en œuvre des méthodes de collecte de données qui s'accordent directement la ou les questions énoncées.

2.2.7 Concevoir et mettre en œuvre des méthodes de collecte de données automatisées, le cas échéant.

2.2.8 Évaluer ou valider les méthodes et les outils de collecte de données.

2.3 Qualité et valeur des données, et confiance à l'égard de celles-ci

Avoir confiance en nos données est une question de qualité. Nettoyer, traiter et transformer les données contribue à garantir l'accès à des données et des informations précises, fiables et de grande valeur. Établir des partenariats solides favorise une responsabilité partagée en matière d'intégrité des données.

Principes fondamentaux

2.3.1 Connaître le cadre de la [qualité des données](#) et les différentes dimensions à travers lesquelles on peut évaluer la qualité (par exemple, définitions des métadonnées, normes, interprétabilité, cohérence, pertinence, accessibilité, respect des délais).

2.3.2 Comprendre l'incidence négative d'une mauvaise gestion des données sur les opérations et les décisions de l'organisation.

Principes intermédiaires

2.3.3 Examiner les données pour en assurer la validité, l'exactitude et l'exhaustivité.

2.3.4 S'aligner sur les normes en matière de données pour garantir et améliorer la qualité de celles-ci.

2.3.5 Communiquer avec les parties prenantes et les autorités reconnues en matière de données pour établir des relations qui renforcent la confiance au sein d'une organisation, avec les parties prenantes et avec la population canadienne.

2.3.6 Déterminer et analyser les valeurs aberrantes et les anomalies dans les données et utiliser des méthodes de résolution des problèmes (par exemple, l'analyse des causes profondes) pour améliorer la qualité des données.

Principes avancés

2.3.7 Assurer et évaluer la qualité et l'interopérabilité des données afin de maximiser la valeur et la confiance.

2.3.8 Utiliser ou développer des systèmes de gestion des données de référence (SGDR) pour assurer la cohérence et l'intégrité des données et de leur utilisation.

2.3.9 Sensibiliser les employés et employées à l'existence de banques de données utiles à l'intégration et à la réutilisation.

2.3.10 Développer ou coordonner une formation supplémentaire à la saisie des données lorsque la qualité ou la validité des données saisies pose problème.

2.3.11 Assurer la traçabilité des données et effectuer la piste de vérification pour garantir l'utilisation appropriée des données.

2.3.12 Établir des partenariats pluridisciplinaires dans l'ensemble de l'organisation pour favoriser une responsabilité commune à l'égard de la gouvernance, des normes et de l'intendance des données, et pour renforcer la confiance dans toute l'organisation.

2.3.13 Évaluer la pertinence d'un ensemble de données (c'est-à-dire qu'il vaut mieux parfois ne pas avoir de données que d'avoir de mauvaises données). Consigner et partager ses raisons de manière transparente.

2.4 Accès, sécurité et interopérabilité

Garantir l'accès, la sécurité et l'accessibilité des données organisationnelles.

Principes fondamentaux

2.4.1 Connaître les politiques, directives et règlements relatifs à la sécurité et au stockage des données.

2.4.2 Comprendre les coûts et les risques liés à la perte de données, et prendre des mesures pour réduire la perte de données et d'informations.

2.4.3 Comprendre les risques associés à la duplication des données et prendre des mesures pour réduire la duplication des données et des informations.

2.4.4 Assurer la confidentialité et la sécurité des données en utilisant les niveaux appropriés de permissions et d'accès.

Principes intermédiaires

2.4.5 Soutenir l'utilisation de données en libre-service et l'accès aux données dès la conception.

2.4.6 Connaître les banques de données, les données ouvertes et les données accessibles qui sont disponibles pour la (ré) utilisation et l'intégration.

2.4.7 Déterminer et hiérarchiser la publication d'ensembles de données de grande valeur sur le [portail du gouvernement ouvert](#) en vue de leur réutilisation et de leur interopérabilité.

2.4.8 Protéger les renseignements personnels en utilisant des méthodes appropriées de dépersonnalisation, d'anonymisation, de masquage ou d'autres méthodes similaires.

2.4.9 Publier les données dans la mesure du possible dans des formats lisibles par machine.

Principes avancés

2.4.10 Comprendre une variété de méthodes de conversion pour soutenir l'interopérabilité.

2.4.11 Développer la provenance et les cartographies pour soutenir la préservation des données.

2.4.12 Utiliser, créer ou gérer les métadonnées, l'indexation des bases de données ou l'analyse et la synthèse des informations pour améliorer l'accès, la recherche et l'utilisation des informations.

2.4.13 S'efforcer de déterminer et de supprimer les obstacles au partage et à la communication des données, et soutenir les données ouvertes et de la transparence au sein du gouvernement travaillant en mode [ouvert par défaut](#).

3. Analyse et évaluation

3.1 Questions et définition des difficultés

Comprendre quand les données peuvent être utilisées pour éclairer ou soutenir la prise de décision, ainsi que le processus d'interprétation des données permettant de préciser des problèmes et des questions de recherche.

Principes fondamentaux

3.1.1 Savoir quand et comment les données peuvent être utilisées pour éclairer une décision.

3.1.2 Utiliser la pensée critique et poser des questions pour définir les données nécessaires, et la façon dont elles seront recueillies et utilisées.

3.1.3 Considérer l'importance et l'utilisation des données dès le départ et non pas après coup.

Principes intermédiaires

3.1.4 Formuler des questions autour des besoins organisationnels et des utilisateurs et utilisatrices auxquels il est possible de répondre à l'aide de données.

3.1.5 Déterminer les données préexistantes qui peuvent être utilisées pour soutenir ou éclairer la prise de décision.

3.1.6 Explorer les données et utiliser les cadres pertinents pour déterminer les problèmes, soutenir la recherche ou générer des questions relatives à des situations pratiques.

Principes avancés

3.1.7 Travailler au sein d'équipes multidisciplinaires en tant que spécialiste pour éclairer divers domaines à l'aide d'approches axées sur les données.

3.1.8 Appliquer sa compréhension des besoins, des mandats et des orientations de l'organisation et des besoins des utilisateurs et utilisatrices afin de générer des exigences opérationnelles associées aux données et d'améliorer la définition des problèmes au début de nouveaux projets, politiques, programmes, prestation de services et autres.

3.2 Analyses et science des données

Reconnaître les tendances et les modèles relatifs aux données afin de définir les relations et de générer des informations. Accéder aux données, les manipuler, les consulter et les analyser en utilisant une variété de méthodes, d'outils et de processus.

Principes fondamentaux

3.2.1. Connaître les fonctionnalités de base de logiciels communs, effectuer des calculs simples et créer des graphiques et des tableaux.

3.2.2 Savoir analyser des données pour répondre à des questions organisationnelles simples.

3.2.3 Considérer l'utilisation d'outils de sources ouvertes et librement disponibles par rapport à ceux offerts sur le marché.

3.2.4 Connaître la [Directive sur la prise de décision automatisée](#) et [l'outil d'évaluation de l'incidence algorithmique](#).

Principes intermédiaires

3.2.5 Utiliser des méthodes analytiques de base et des outils connexes, qualitatifs ou quantitatifs, pour générer des idées.

3.2.6 Prendre des questions énoncées et développer un plan d'analyse comprenant l'évaluation des données et des méthodes pertinentes.

3.2.7 Comprendre les diverses méthodes statistiques et analyses, et les outils connexes, et savoir quand leur utilisation est appropriée.

3.2.8 Diffuser ouvertement les données et les résultats d'analyse afin de soutenir d'autres initiatives gouvernementales, conformément aux [normes relatives au numérique](#).

3.2.9 Reconnaître et explorer les modèles, les relations et les tendances au sein des données et entre les sources de données afin de générer des idées en utilisant une variété de méthodes.

3.2.10 Fixer des objectifs clairs avant d'entreprendre des activités d'analyse afin de garantir des analyses à valeur ajoutée et consulter toutes les parties prenantes pour s'assurer que la collecte de données et les plans d'analyse correspondent à cette valeur.

3.2.11 Effectuer des analyses en utilisant des données et des méthodes pertinentes pour une consommation par un large éventail de publics.

3.2.12 Évaluer les résultats des analyses et les comparer à d'autres résultats.

3.2.13 Échanger ouvertement les données et les résultats d'analyse pour soutenir le travail des autres équipes.

3.2.14 Accéder à des données provenant de différentes sources et les manipuler, par exemple en utilisant des fichiers plats ou du langage SQL (Structured Query Language).

3.2.15 Appliquer des méthodes et des outils analytiques, qualitatifs ou quantitatifs, communs aux données.

3.2.16 Appliquer la [directive sur la prise de décision automatisée](#), y compris la réalisation d'une [évaluation de l'incidence algorithmique](#), afin de déterminer, d'évaluer et d'atténuer les risques associés au déploiement d'un système de prise de décision automatisée.

Principes avancés

3.2.17 Utiliser un large éventail de méthodes analytiques avancées et les outils connexes, qualitatifs ou quantitatifs, pour générer des idées.

3.2.18 Relier entre elles des données provenant de sources multiples afin d'en accroître l'utilité et la valeur.

3.2.19 Effectuer des analyses exploratoires pour soutenir le ciblage des problèmes, déterminer les zones de risque et plus encore.

3.2.20 Construire et valider des modèles statistiques à partir de données.

3.2.21 Utiliser des solutions de modélisation appropriées, précises, valides et efficaces pour une variété de types de données complexes.

3.2.22 Utiliser des outils et techniques avancés pour effectuer l'exploration des données, tels que le forage de données, le moissonnage du Web ou l'apprentissage automatique.

3.2.23 Évaluer et sélectionner les outils en fonction de leurs avantages, du cas d'utilisation, de la facilité d'accès et de l'inclusivité, et prendre en compte les options de sources ouvertes lorsqu'elles sont disponibles.

3.2.24 Prendre en compte la conception numérique et de la plateforme ainsi que les considérations à l'échelle de l'organisation et du gouvernement lors de la sélection des outils.

3.2.25 Évaluer les données en fonction de leur utilité, de leur qualité, de leur exactitude, de leur pertinence et de leur partialité potentielle.

3.2.26 Être en mesure d'appliquer l'automatisation pour la collecte des données.

Pour plus de détails sur les compétences en science des données, veuillez consulter les [profils de compétences](#) élaborés par le [Réseau de la science des données pour la fonction publique fédérale](#).

3.3 Narration et visualisation

Traduire les données dans un format accessible pour aider les autres à voir et à comprendre les tendances, les valeurs aberrantes et les profils dans les données.

Principes fondamentaux

3.3.1 Présenter l'information à l'aide d'éléments visuels, de présentations accessibles ou d'histoires pour :

- aider les autres à comprendre un sujet ;
- éclairer une discussion ;
- rendre compte d'un progrès ;
- soutenir la prise de décision ou la résolution de problèmes.

3.3.2 Intégrer l'accessibilité et l'inclusion dans les éléments visuels et le contenu en suivant les pratiques exemplaires, telles que les [Normes relatives au numérique du gouvernement du Canada : Directives](#).

3.3.3 Créer des tableaux et des représentations graphiques de données qui sont exacts et informatifs.

3.3.4 Inclure des références, des étiquettes et des citations correctes et pertinentes.

Principes intermédiaires

3.3.5 Présenter des informations globales, en racontant des histoires complètes plutôt qu'en présentant des résultats sélectifs ou incomplets.

3.3.6 S'assurer que les présentations de données sont directement liées aux questions initiales ou à la ligne de pensée.

Principes avancés

3.3.7 Évaluer les besoins du public, sa familiarité avec les données et sa compréhension du sujet.

3.3.8 Évaluer les récits et les visualisations pour en vérifier l'exactitude et les fausses représentations.

3.3.9 Communiquer les pratiques exemplaires et les détails sur les outils de visualisation des données aux équipes chargées des données et à d'autres personnes afin d'éviter les erreurs courantes et de rendre les visualisations de données plus efficaces.

3.3.10 Fournir une capacité d'exploration dans les rapports et les informations récapitulatives pour permettre une investigation plus approfondie.

3.3.11 Développer des tableaux de bord, des infographies et des visualisations interactives à l'aide de différents logiciels, y compris des outils courants d'informatique décisionnelle (par exemple, PowerBI, Tableau) et/ou des bibliothèques spécialisées (par exemple, D3.js, seaborn, plotly).

3.4 Évaluation des résultats

Évaluer de manière critique l'efficacité des politiques, des programmes, des services et des décisions en utilisant la recherche, l'évaluation, et les techniques analytiques.

Principes fondamentaux

3.4.1 Comprendre les principes de l'évaluation et la manière dont les données sont utilisées, interprétées et appliquées dans le cadre du suivi, de l'amélioration ou de la démonstration des résultats d'un programme ou d'un projet.

3.4.2 Connaître les méthodes et techniques analytiques de base pour mesurer et suivre la mise en œuvre ou le rendement d'un projet ou d'un programme.

Principes intermédiaires

3.4.3 Utiliser les données dans le but de mesurer les résultats relatifs aux politiques, à la mise en œuvre des programmes, à la législation, à la réglementation, à la prise de décision et plus encore.

3.4.4 Communiquer régulièrement avec les utilisateurs et utilisatrices, pour recueillir des commentaires et des données tout au long de la mise en œuvre d'idées ou de solutions afin de tenir compte de leur fonctionnement ou de leur rendement, et utiliser les commentaires en vue d'apporter des corrections, si nécessaire (voir [Normes relatives au numérique du gouvernement du Canada : Guide de lecture](#) pour plus de détails).

3.4.5 Comparer les résultats avec d'autres recherches et conclusions.

3.4.6 Conserver et tenir à jour les données et informations originales utilisées dans la prise de décision pour analyser les résultats par rapport à l'intention.

3.4.7 Recueillir des données de suivi pour évaluer l'efficacité des décisions ou des solutions.

3.4.8 Déterminer les principales conclusions à tirer des tableaux, des graphiques et des diagrammes afin de les intégrer à d'autres informations dans les processus décisionnels futurs.

3.4.9 Analyser les données pour comprendre l'incidence que les solutions déterminées peuvent avoir sur divers groupes de personnes, politiques, programmes, services, processus, etc.

3.4.10 Évaluer et mesurer les résultats de la prise de décision en utilisant des données.

3.4.11 Utiliser les commentaires des utilisatrices et utilisateurs finaux pour établir des exigences évolutives afin d'ajuster les politiques, les programmes, les services et autres pour répondre aux besoins des utilisateurs et utilisatrices.

3.4.12 Obtenir l'accord des parties prenantes sur les critères d'évaluation avant la mise en œuvre.

3.4.13 Mettre en œuvre les changements apportés aux politiques, aux programmes, à la prestation de services et à d'autres fonctions gouvernementales sur la base du retour d'information continu et de la recherche sur les utilisateurs et utilisatrices.

3.4.14 Surveiller les résultats des décisions automatisées afin de se prémunir contre des résultats non intentionnels.

4. Systèmes de données et architecture

4.1 Architecture des données intégrées

Comprendre les principes des systèmes de données et concevoir des solutions pour maximiser la facilité d'utilisation, l'intégrité, la sécurité, l'évolutivité et l'accessibilité des données et des informations.

Principes fondamentaux

4.1.1 Comprendre les outils de gestion des données, notamment les catalogues de données, les lacs de données et les entrepôts de données.

4.1.2 Connaître le [Conseil d'examen de l'architecture intégrée du gouvernement du Canada](#) ainsi que son rôle, ses mandats et sa fonction au sein du gouvernement.

4.1.3 Comprendre le rôle que joue le lac de données intégrées dans la recherche et l'exploration des données, y compris l'utilisation de l'intelligence artificielle (IA).

Principes intermédiaires

4.1.4 Prendre en compte les besoins architecturaux de l'organisation, de l'informatique et des données lors de l'élaboration de solutions organisationnelles.

4.1.5 Tenir compte du cycle de vie complet du système lors du développement ou du déploiement de l'architecture organisationnelle.

4.1.6 Évaluer les exigences du système (telles que le stockage, la conservation et le calendrier) pour s'assurer que l'architecture organisationnelle est conçue pour être utilisable, interopérable et apporter de la valeur aux utilisateurs et utilisatrices.

Principes avancés

4.1.7 Développer et assurer la surveillance des systèmes d'IA, ce qui inclut, sans s'y limiter, la santé, les biais et la dérive des données.

4.1.8 Concevoir, mettre en œuvre ou surveiller les systèmes qui permettent l'accès à la source, en utilisant les SGDR appropriés, ce qui permet de réduire la duplication de l'information.

4.1.9 Anticiper et préciser les points problématiques dans le développement des systèmes et des solutions.

4.1.10 Mettre au point et démontrer des solutions pour les clients et clients dans leurs propres environnements pour permettre un développement ultérieur.

4.1.11 Construire des logiciels d'automatisation pour faire fonctionner les systèmes nécessaires au stockage des données, à la gestion des données, aux codes de science des données, à la formation distribuée, au dépôt de modèles, au dépôt de fonctionnalités, à la livraison continue, au service de modèles et à la surveillance.

4.1.12 Exploiter les systèmes d'IA de production et s'assurer qu'ils sont disponibles, évolutifs et performants.

4.1.13 Construire et transmettre l'expertise technique nécessaire pour analyser et recommander des solutions organisationnelles pour rendre l'IA ou les modèles analytiques avancés opérationnels.

4.1.14 Définir l'architecture et les processus de déploiement requis pour l'IA, de l'ingestion des données à la production et l'entretien.

4.1.15 Fournir des conseils techniques à la direction et aux autres scientifiques des données en ce qui concerne l'opérationnalisation des modèles.

Pour plus de détails sur les compétences en science des données, veuillez consulter les [profils de compétences](#) élaborés par le [Réseau de la science des données pour la fonction publique fédérale](#).

4.2 Systèmes de données

Utiliser des outils, des logiciels, des plateformes et des processus pour collecter, organiser, stocker, gérer et protéger les données.

Principes fondamentaux

4.2.1. Connaître les approches, outils et techniques de base pour assurer une bonne gestion des données.

Principes intermédiaires

4.2.2 S'assurer que la conception des systèmes est conforme aux besoins de l'organisation.

4.2.3 Garantir la sécurité des données et des informations grâce à des niveaux appropriés de permissions et d'accès.

4.2.4 Développer, mettre en œuvre et tester les flux de travail du système, et effectuer le dépannage de ceux-ci.

4.2.5 Développer des canalisations de données de bout en bout sur la base d'une compréhension approfondie des outils disponibles et de l'écosystème de gestion des données, du cycle de vie des données et des problèmes organisationnels, afin de garantir que les solutions analytiques sont efficaces, prévisibles et durables.

4.2.6 Être en mesure :

- d'utiliser les pratiques exemplaires de codage pour générer un travail reproductible et vérifiable ;
- d'extraire, de transformer et de charger des données à l'aide des outils et pratiques appropriés ;
- d'utiliser les outils et techniques d'analyse des données, y compris les logiciels de sources ouvertes.

4.2.7 Tenir compte de toutes les interdépendances sur le plan organisationnel qui nécessitent des processus, des normes ou la préservation et le stockage des données pour maximiser la valeur organisationnelle.

4.2.8. Extraire les données des sources de données en utilisant les outils et pratiques appropriés.

4.2.9 Transformer les données en utilisant une variété d'outils.

4.2.10 Connaître les fonctions d'extraction, de transformation et de chargement et leur utilisation appropriée.

Principes avancés

4.2.11 Évaluer, construire ou mettre en œuvre des structures de données telles que des lacs de données, des magasins de données, des bases de données relationnelles ou des bases de données en mémoire.

4.2.12 Concevoir ou mettre en œuvre des systèmes et des processus de données qui utilisent les données à la source plutôt que de les répliquer.

4.2.13 Concevoir, mettre en œuvre ou évaluer les outils, méthodes et processus de collecte et de transformation des données.

4.2.14 Évaluer la valeur du stockage dans le nuage par rapport au stockage sur site.

4.2.15 Utiliser le nuage et les logiciels en tant que service (SaaS) tels que les progiciels de gestion intégrés (PGI), les applications de gestion des services ou l'architecture distribuée.

4.2.16 Tenir à jour et soutenir la disponibilité et la longévité des données grâce à des systèmes de sauvegarde et des plans de récupération des données appropriés.

4.2.17 Développer et recommander des indicateurs de performance pour mesurer l'efficacité, l'efficacité et l'incidence des solutions de gestion de l'information.

4.2.18 Déterminer les problèmes avec les parties prenantes, comprendre où la science des données peut ajouter de la valeur en soutenant la prise de décision stratégique et opérationnelle pour créer un effet, et concevoir des solutions et des mesures de science des données pour ces problèmes.

Annexe A : Définitions des termes clés

Algorithme

Ensemble d'instructions qui vise la résolution d'un problème ou la réalisation d'un calcul.

Analyse des causes profondes

Une gamme d'approches, de techniques et d'outils permettant de déterminer systématiquement les mécanismes de causalité qui sous-tendent les racines potentielles d'un problème.

Analytique

La découverte, l'interprétation et la communication de modèles significatifs dans les données. Il s'agit de transformer les données en information utile à la prise de meilleures décisions. L'analytique repose souvent sur l'application de statistiques, d'algorithmes et de logiciels pour rechercher des tendances potentielles, analyser les effets de certaines décisions ou évaluer le rendement d'une politique, d'un programme, or d'un service donné.

Base de données

Une collection importante et organisée de données. Elle permet d'accéder aux données, de les manipuler et de les mettre à jour facilement.

Dérive des données

Variations des données de production par rapport aux données qui ont été utilisées pour tester et valider le modèle avant de le déployer dans l'environnement de production. La dérive des données entraîne une dégradation de la performance du modèle.

Données

Les données sont des faits, des chiffres, des observations ou des enregistrements qui peuvent se présenter sous la forme d'image, de son, de texte ou de mesure physique (p. ex. distance, poids, longueur d'onde).

Données brutes

Une collection de chiffres ou de caractères qui n'ont pas encore été nettoyés ou traités.

Données non structurées

Données qui n'ont pas de modèle de données préétabli ou qui ne sont pas organisées d'une manière préétablie. Elles ne sont donc pas stockées dans des champs fixes d'une base de données. Par exemple, le texte contenu dans un rapport.

Données ouvertes

Données structurées qui sont lisibles par machine, librement partagées, ainsi qu'utilisées et constituées sans restrictions.

Données qualitatives

Données décrivant les attributs ou les propriétés qu'un objet possède. Elles représentent des variables subjectives qui ne sont parfois pas faciles à mesurer (p. ex., le goût, la couleur des yeux).

Données quantitatives

Données exprimant une quantité, un nombre ou une gamme. Habituellement, des unités de mesure sont associées aux données (p. ex., des mètres pour la taille, des kilogrammes pour le poids ou des degrés Celsius pour la température).

Données structurées

Données qui ont été organisées en champs fixes au sein d'une base de données ou d'un fichier, de sorte qu'elles sont facilement consultables et peuvent être facilement analysées. Par exemple, les données d'une feuille de calcul Excel.

Ensemble de données

Une collection de données. Il correspond souvent au contenu d'une seule base de données ou d'un seul fichier de données.

Entrepôt de données

Un système qui regroupe des données provenant de plusieurs sources dans un magasin de données unique, central et cohérent pour prendre en charge l'exploration de données, l'intelligence artificielle et l'apprentissage automatique.

Extraction, transfert, chargement

Un processus d'intégration de données qui combine des données provenant de plusieurs sources de données dans un magasin de données unique et cohérent qui est chargé dans un entrepôt de données ou un autre système cible.

Gestion des données de référence

Une approche visant à maintenir des références précises, cohérentes et normalisées aux données dans toute l'organisation.

Gestion des données maîtres

Les technologies, les outils et les processus permettant de garantir la coordination des données maîtres dans l'ensemble de l'organisation.

Information

Données converties dans un contexte significatif et utile. Connaissances saisies dans n'importe quel format, comme des faits, des événements, des choses, des processus ou des idées, qui peuvent être structurés ou non — y compris des concepts qui, dans un certain contexte, ont une signification particulière.

Intelligence artificielle

Toute technologie de l'information qui exécute des tâches pour lesquelles il faut habituellement faire appel à l'intelligence biologique, comme comprendre le langage parlé, apprendre des comportements ou résoudre des problèmes.

Interopérabilité

La capacité de systèmes, dispositifs, applications ou produits différents à se connecter et à communiquer ensemble de manière coordonnée, sans effort de la part de l'utilisateur final.

Lac de données

Un système de stockage ou un référentiel qui contient de grandes quantités de données brutes.

Lac de données de l'organisation

Un lac de données à l'échelle d'une organisation, visant à stocker et à partager l'information.

Littératie en matière de données

La capacité de comprendre, de créer et de communiquer des données sous forme d'information significative.

Magasin de données

Un sous-ensemble d'un entrepôt de données qui est axé sur un secteur d'activité, un ministère ou un domaine en particulier.

Modèle

Une représentation mathématique d'un processus du monde réel. Un modèle prédictif prévoit un résultat futur en fonction de comportements passés.

Nettoyage des données

Le processus consistant à préparer des données pour l'analyse en supprimant, corrigeant ou documentant les données qui sont incomplètes, inexactes, dupliquées ou mal formatées.

Renseignements personnels

Renseignements enregistrés sur une personne identifiable, autres que les coordonnées professionnelles. Il s'agit entre autres de tout renseignement qui peut être relié à une personne ou l'identifier par référence ou association.

Requête en langage structuré

Un langage de programmation servant à gérer les bases de données relationnelles et à effectuer des opérations sur leurs données.

Science des données

Domaine multidisciplinaire qui combine la méthode scientifique, la programmation informatique, les statistiques et les affaires pour extraire des informations significatives des données et éclairer la prise de décision. Bien que la science des données ait des points communs avec les statistiques, ces deux termes ne doivent pas être utilisés de manière interchangeable, car ils diffèrent dans leurs processus et leurs applications.

Statistique

Dans le contexte de l'analyse des données, les statistiques désignent un type d'information obtenu par des opérations mathématiques sur les données. Les statistiques désignent également la science du développement et de l'étude des méthodes de collecte, d'analyse, d'interprétation et de présentation des données.

Tableau de bord

Outil de gestion de l'information utilisé pour suivre et représenter visuellement les indicateurs de rendement clé, les mesures et les points de données clés.

Annexe B : Ressources supplémentaires

[Bibliothèque et Archives Canada](#)

[CODATA Research Data Management Terminology](#)

[Communauté des données du GC](#)

[Communauté de pratique sur les données intégrées du GC](#)

[Consortia Advancing Standards in Research Administration Information Research Data Management Glossary](#)

[Gestion des données de recherche](#)

[Répertoire des ressources de l'écosystème des données](#)