



Tendances technologiques

Mégadonnées

Architecture d'entreprise, Direction générale du dirigeant principal
de la technologie

Version 0.1

Date : 2019-6-26



Shared Services
Canada

Services partagés
Canada

Canada

Table des matières

| | |
|--|-----------|
| Sommaire opérationnel | 3 |
| Sommaire technique | 3 |
| Utilisation au sein des entreprises..... | 5 |
| Utilité pour le gouvernement du Canada..... | 8 |
| Répercussions pour Services partagés Canada | 9 |
| Proposition de valeur | 9 |
| Difficultés..... | 10 |
| Éléments à considérer..... | 11 |
| Annexe A – Glossaire sur les méga données..... | 12 |

Sommaire opérationnel

Les mégadonnées représentent des données de grande variété, vélocité, véracité, valeur ou volume qui repoussent les limites des outils et des infrastructures traditionnelles et demandent des méthodes innovatrices économiques pour traiter ou extraire la valeur des données. Ce n'est pas une technologie comme telle, mais une combinaison de technologies nouvelles et traditionnelles qui fournit à l'organisation des informations qui pourront lui être utiles; c'est aussi la capacité de gérer un énorme volume de données disparates, à la vitesse nécessaire et dans le bon délai pour qu'il soit possible de faire des analyses et d'intervenir en temps réel.

Les mégadonnées sont étudiées en vue de mettre en relief, par exemple, des particularités cachées, des corrélations inconnues, des tendances du marché et des préférences de clients, ce qu'on appelle l'analyse des mégadonnées, pour que les organisations puissent prendre des décisions éclairées, une situation avantageuse si on veut augmenter la productivité, réduire les coûts, accélérer les résultats et améliorer les activités, entre autres.

Les mégadonnées sont d'utilisation très courante dans les milieux de la transformation, des sciences, de la santé, des banques et de la fabrication, aux côtés d'autres technologies et modes d'analyse de données.

Il est bon de savoir que les mégadonnées sont passées par toutes les étapes de la marche de l'engouement de Gartner. Gartner considère en effet que le concept a mûri et qu'il est maintenant très productif; les progrès réalisés dans ce domaine ont simplifié l'implantation de solutions de mégadonnées au sein d'une organisation. Loin de nous l'idée de banaliser la complexité du déploiement de mégadonnées : c'est seulement qu'elles sont arrivées au stade d'outil utilisable parmi tous ceux offerts. En fait, les mégadonnées ne sont pas une fin en soi; elles sont un moyen de mettre à profit d'autres technologies comme l'intelligence artificielle et l'apprentissage automatique, qui exigent de grandes quantités de données d'apprentissage pour se perfectionner.

Sommaire technique

Les mégadonnées ne sont qu'un concept et non une technologie en soi. C'est en fait un moyen de rendre accessibles des méthodes d'analyse impossibles avant l'avènement des superordinateurs et des microprocesseurs. L'exploitation des mégadonnées n'est limitée que par la capacité informatique et la compétence de l'équipe d'analyse de l'organisation. Des solutions puissantes sont à la portée de presque toutes les organisations maintenant que les outils abondent pour traiter la variété de données. Les mégadonnées sont aux yeux de Gartner un concept mature parce qu'elles sont facilement accessibles et que les solutions offertes sur le marché se sont hautement complexifiées. Les données en temps réel, ou en direct, peuvent maintenant être analysées et ainsi révéler de précieuses informations. Il en est survenu un revirement de

situation dans le traitement des données : par le passé, la direction mettait les données au profit de son intuition tandis que maintenant, ce sont les analystes qui fouillent dans les données et en font ressortir des tendances qui orientent les décisions stratégiques.

Si elles veulent exploiter les mégadonnées, les organisations doivent réorienter leurs façons de faire en fonction des données. Elles doivent en fait se doter d'un cadre qui régit le passage des données par les étapes préalables au traitement : l'extraction, la transformation et le chargement. Ces étapes n'ont rien de facultatif : les données bien nettoyées provenant de sources fiables ont plus de valeur, car il est plus facile d'en extraire des informations exploitables. Environ 80 % du temps et des efforts dévolus à un projet sont consacrés à la préparation des données parce que la plupart d'entre elles sont brutes, non structurées et non balisées.

Sur le plan strictement technique, pour que les données soient considérées comme des mégadonnées, leur volume doit être très important et leur nature, variée; elles seront stockées dans plusieurs bases de données et systèmes de fichiers distribués. Il faut pour ce faire faire appel au concept de parallélisme si on veut traiter les données simultanément en raison du volume de données, de leur disparité et de la complexité des tâches. De nombreuses solutions de mégadonnées se basent sur une architecture parallèle et distribuée appelée MapReduce, conçue par Google. MapReduce assure le traitement distribué et parallèle des requêtes effectuées sur de grands ensembles de données dans un environnement distribué. Il existe de nombreuses variantes de ce concept, et on s'affaire à l'heure actuelle à mettre au point des méthodes entièrement nouvelles de traitement des ensembles de données distribuées telles que le nuage de mégadonnées, où les données sont stockées et traitées entièrement dans un nuage.

Il faut savoir que les mégadonnées n'ont aucune utilité comme telles. C'est plutôt un concept dont dépendent d'autres technologies pour produire des informations qui seront, elles, utiles. L'exploration de données est le processus par lequel on tente de découvrir des tendances dans les mégadonnées, un processus faisant appel à des méthodes qui allient apprentissage autonome, statistiques et systèmes de base de données. L'analyse des données, si elle a recours à des méthodes similaires, est en revanche plutôt orientée sur l'information descriptive. À l'heure actuelle, on observe un changement marqué dans le traitement des mégadonnées, où la place importante de l'humain dans la manipulation des données fait place à l'intelligence artificielle pour le traitement des tâches courantes. Les mégadonnées ont en fait été un catalyseur dans l'évolution de l'intelligence artificielle parce que les algorithmes peuvent être mis à l'essai sur des données volumineuses et diversifiées. Comme les algorithmes d'apprentissage autonome de l'intelligence artificielle peuvent continuer de se perfectionner si on les alimente de nouvelles données, ils ne font que s'améliorer au fil du temps par voie de processus itératifs.

Les tâches d'analyse des mégadonnées se divisent en trois grandes catégories¹ :

- **Analyse descriptive** : Tâches consistant à résumer le comportement actuel et passé des données. L'objectif principal de l'analyse descriptive est de déterminer les causes du comportement des données et d'appliquer les connaissances ainsi acquises aux décisions futures. Plus de 80 % des analyses seraient descriptives en entreprise.
- **Analyse prédictive** : Tâches consistant à prédire ce qui est susceptible d'arriver sur la base des informations connues. Il est important de noter qu'aucune prévision n'est parfaite puisqu'elle est de nature probabiliste; on n'obtient que la probabilité d'occurrence par ce type d'analyse.
- **Analyse prescriptive** : Étape suivant l'analyse prédictive, où le modèle de traitement aboutit à un plan d'action. L'analyse prescriptive fait appel à un modèle prédictif composé de deux éléments : des données exploitables et un système qui surveille le résultat produit par l'intervention. Les utilisateurs peuvent ainsi voir les effets possibles de certaines modifications de paramètres et donc avoir une idée de ce qu'ils devraient faire pour obtenir les résultats souhaités.

Utilisation au sein des entreprises

Le marché des mégadonnées est considéré avoir atteint un stade de maturité : de nombreuses solutions commerciales existent pour presque tous les besoins perçus. Ces solutions peuvent être locales, autrement dit hébergées dans les serveurs d'une organisation; sur un nuage ou une combinaison des deux, une partie de l'information étant stockée sur le nuage et l'autre, localement. Selon une prévision de Statista (un portail allemand de statistiques en ligne), la taille du marché mondial des mégadonnées est actuellement de 49 milliards de dollars des États-Unis et devrait atteindre 103 milliards de dollars d'ici 2027ⁱⁱ. Étant donné que les solutions de mégadonnées jouissent d'une grande popularité et que de nombreux cas d'utilisation ont été étudiés et développés, le texte qui suit brosse un portrait plutôt vaste des utilisations possibles de ces mégadonnées.

Sciences

Les mégadonnées ont changé les modes de fonctionnement dans de nombreux domaines scientifiques. L'abondance des données et le coût peu élevé de traitement de l'information ont entraîné une modification de la façon dont on vérifie les hypothèses; ce n'est plus le scientifique qui formule des hypothèses et les vérifie, les données peuvent elles-mêmes suggérer des hypothèses et des relations entre les points de donnéesⁱⁱⁱ. Les scientifiques ont maintenant à leur disposition plus de moyens d'étudier la quantité de données générées par leurs expérimentations.

Voici quelques exemples d'utilisations scientifiques des mégadonnées :

- Stockage et traitement de grandes quantités de données expérimentales : Par exemple, le Grand collisionneur de hadrons du CERN génère quelque 22 pétaoctets de données chaque année qui sont analysées dans un réseau de plus de 150 centres de calcul dans le monde^v.
- Perfectionnement des méthodes de surveillance : La surveillance des données sur l'environnement, la géographie et les changements climatiques n'est plus limitée aux lieux d'échantillonnage. Grâce aux mégadonnées, il est dorénavant possible de combiner et d'analyser des ensembles de données mondiales, ce qui permet aux scientifiques de mieux comprendre l'évolution des conditions environnementales^v.
- Gestion des ressources en agriculture : Les organisations agricoles puisent leurs données de sources publiques, notamment des capteurs à distance et des ensembles de données accessibles à tous, pour optimiser l'utilisation des terres agricoles. Ainsi, les agriculteurs peuvent savoir quoi planter et même si leurs machines risquent de tomber en panne^{vi}.

Services de santé

Les données sur les patients sont recueillies et rendues anonymes afin de mieux comprendre l'effet des traitements et de certains médicaments pour des maladies particulières. Ces données, comme les meilleures pratiques, les revues médicales et les manuels scolaires, sont également intégrées aux systèmes d'aide à la décision clinique qui aident les médecins à recommander des plans de traitement. Certains de ces systèmes sont alimentés par l'intelligence artificielle, comme IBM Watson for Oncology^{vii}. Voici d'autres exemples d'utilisations de mégadonnées dans le domaine de la santé :

- activités de recherche et développement dans le domaine pharmaceutique;
- analyse du comportement des patients;
- analyse des demandes de remboursement et des coûts pour la détection de fraudes;
- analyse des alertes précoces;
- rentabilité des médicaments;
- technologies portatives pour des suivis santé.

Veille au marché et veille stratégique

Les mégadonnées rendent possibles, pour les gestionnaires et les spécialistes du marketing, la surveillance en temps réel de l'efficacité de leur travail. Voici des exemples de façons dont les organisations peuvent mieux adapter leurs offres et optimiser leur stratégie grâce aux mégadonnées :

- analyse du comportement des consommateurs;
- marketing local;
- optimisation des prix, tarification dynamique;
- analyse du panier d'épicerie;
- analyse des sentiments;
- taux d'attrition et de maintien;

- gestion des relations avec la clientèle, fidélisation des clients.

Banques/assurances/valeurs mobilières

Le secteur bancaire est l'un des premiers à avoir adopté les mégadonnées et dispose donc de solutions efficaces qui optimisent de beaucoup ses données. Les secteurs de l'assurance et des valeurs mobilières sont étroitement liés et bénéficient des mêmes avancées en matière de mégadonnées. Les exemples suivants illustrent des cas d'utilisations de mégadonnées dans ces secteurs :

- conseillers robotisés pour le commerce d'actions et les stratégies d'investissement;
- optimisation du service des guichets automatiques bancaires;
- évaluation du risque de crédit et cote de crédit;
- détection des fraudes et du blanchiment d'argent;
- catégorisation et segmentation de la clientèle pour optimiser l'offre de produits;
- négociation algorithmique, soit la prise en compte de données non structurées dans les médias sociaux et les fils d'actualités pour déclencher des transactions automatiques;
- coûts de rétablissement après sinistre.

Secteur manufacturier

Les mégadonnées ont également modifié les processus de fabrication et les ont rendus plus efficaces. À l'aide des données passées, les organisations peuvent optimiser et affiner leurs stratégies de gestion des opérations. En voici des exemples :

- optimisation de la chaîne d'approvisionnement en juste à temps;
- planification de la configuration de produits;
- modélisation de l'attribution des capacités et du réseau d'approvisionnement;
- assurance de la qualité automatisée;
- entretien anticipé;
- surveillance de sécurité selon les données transmises par capteur;
- établissement des valeurs marchandes et planification en fonction de la qualité du produit, de la saisonnalité, de la demande et d'autres facteurs liés à l'offre.

Communications, médias et divertissement

Tout comme les autres secteurs d'activité mentionnés, les milieux des médias et du divertissement ont été fortement secoués par l'arrivée de l'analyse des mégadonnées. Les plateformes de médias sociaux ont donné aux acteurs de ces milieux un moyen d'écouter ce que les consommateurs ont à dire, ce qui a eu une influence directe sur les activités de production et de distribution. En voici quelques exemples :

- algorithmes de recommandation de médias (Netflix, Spotify, YouTube);
- optimisation du contenu, en fonction des intérêts des consommateurs;
- identification de contenu (Shazam, YouTube);
- suivi des habitudes de vie pour déterminer les tendances de consommation futures dans le domaine de la mode.

Applications générales dans la sphère publique

Les États du monde entier tirent parti de leur relativement nouvelle capacité d'analyser de grandes quantités d'informations. C'est ce qui a donné le jour au nouveau concept d'analyse de données citoyennes. Il s'agit ici du recours aux données de citoyens pour surveiller l'efficacité des programmes publics et créer des initiatives adaptées. Les données peuvent également servir à faire ressortir des informations qui pourraient changer l'orientation des politiques publiques. Il serait alors possible de mesurer en temps réel l'efficacité des programmes et d'apporter des changements là où il le faut. Voici d'autres exemples d'utilisation de mégadonnées par des autorités publiques :

- gestion des crises;
- surveillance des foules;
- application préventive de la loi et efficacité des mesures d'application;
- optimisation de la circulation routière.

Utilité pour le gouvernement du Canada

À l'heure actuelle, de nombreux ministères du gouvernement du Canada utilisent ou comptent utiliser l'analyse de mégadonnées pour optimiser les données dont ils disposent. Aussi, certains offrent des services d'analyse sur demande. Le Conseil national de recherches du Canada dispose d'un centre de recherche sur les technologies numériques dont une équipe se consacre entièrement à l'analyse de données. Ses spécialistes des données et de l'apprentissage automatique peuvent être consultés sur demande. Ils fournissent des services pour organiser et analyser de grands ensembles de données afin de mettre au jour des tendances, donner des explications et faire des prédictions et ainsi de créer des informations utilisables^{viii}. De même, le Centre d'analyse de mégadonnées d'Innovation, Sciences et Développement économique Canada étudie par visualisation les données qu'il a recueillies sur le spectre sans fil du Canada^{ix}. Dans le même ordre d'idées, le gouvernement du Canada fournit directement des données de recherche sur les capacités de veille stratégique des entreprises sur le portail Canada.ca^x.

L'Agence du revenu du Canada envisage de nombreuses manières d'utiliser les données dont elle dispose actuellement pour trouver les contribuables soupçonnés de fraude. À l'aide de modèles d'apprentissage automatique, l'Agence a été en mesure de classer les contribuables selon la probabilité qu'ils produisent leur déclaration et paient leurs impôts^{xi}. Elle a essayé diverses méthodes de recouvrement pour déterminer celle qui fonctionne le mieux selon la catégorie de contribuable. L'Agence exploite également les données des médias sociaux des Canadiens qu'elle soupçonne d'évasion fiscale et combine ces renseignements pour effectuer des évaluations des risques^{xii}.

Agriculture et Agroalimentaire Canada donne un autre exemple d'utilisation de mégadonnées dans la fonction publique. À l'aide de données recueillies partout au Canada, le ministère a pu se créer des outils géomatiques publics pour l'étude de la surface terrestre. Les mégadonnées ont une fois de plus été utiles, cette fois à la création de cartes interactives qui montrent l'évolution des cultures et de l'utilisation des terres^{xiii}.

Répercussions pour Services partagés Canada

Proposition de valeur

Services partagés Canada (SPC) est déjà bien positionné au sein du gouvernement du Canada pour adopter les mégadonnées et réaliser tous les bienfaits potentiels qui y sont associés. Le marché étant parvenu à maturité, l'occasion est belle à saisir : en effet, l'utilisation de mégadonnées ne nécessite pas toujours des connaissances pointues. Comme l'utilité des mégadonnées n'est plus à prouver, il existe une grande diversité de solutions commerciales faciles d'accès qui répondent aux besoins des clients. Bien déployée, une solution de mégadonnées jette la lumière sur les possibles façons d'améliorer une organisation; SPC peut s'en servir pour affiner ses objectifs stratégiques et s'établir des objectifs clairs. Comme on le voit déjà au sein de SPC, l'analyse des données peut servir à mesurer et à comparer l'efficacité de programmes qui pourraient être déployés à grande échelle dans l'ensemble du gouvernement. Étant donné que SPC a pour mandat de chapeauter les centres de données du gouvernement du Canada, il est également en mesure de réaliser des stratégies de traitement des données et de créer de bout en bout une architecture de base de données grâce à laquelle les données demeureront exploitables des années durant. SPC pourrait également devenir l'Amazon de services Web du gouvernement du Canada et ainsi prêter l'espace informatique inutilisé des centres de données à la demande de ministères partenaires qui en ont besoin.

En tant que fournisseur de services au gouvernement du Canada, SPC doit mettre en priorité les services qui lui permettront d'améliorer le stockage, les réseaux, les communications, les capacités informatiques et la sécurité de son infrastructure et ainsi d'assurer la stabilité de l'environnement technologique du gouvernement. En supplément de son mandat central, SPC pourrait profiter des mégadonnées pour adapter ses interventions et ses offres aux besoins de ses partenaires et clients. SPC doit par contre songer au préalable à ses forces et aux moyens à sa disposition pour contrer les difficultés potentielles inhérentes à ces ajouts dans son offre.

Chaque jour, les diverses entités du gouvernement du Canada recueillent des données transactionnelles sur une variété de sujets. Depuis l'avènement de l'Internet des objets et l'utilisation des médias sociaux, de capteurs et de vidéos pour recueillir des données, ces dernières sont encore plus nombreuses au quotidien. Ces données pourraient être utilisées à l'infini sans que leur valeur s'en trouve diminuée. Si ces données sont bien

organisées, elles procureront un avantage indéniable pour les projets de mégadonnées du gouvernement du Canada. Dans le même ordre d'idées, le gouvernement du Canada a octroyé des sommes importantes à l'infrastructure numérique, une rampe de lancement potentielle des stratégies du gouvernement du Canada en matière de mégadonnées.

Difficultés

La mise en œuvre d'une solution de mégadonnées n'est pas sans accroc. En effet, avant que les outils ne soient adoptés, l'organisation doit se doter d'une méthode claire de prétraitement des données. Le prétraitement est nécessaire si on veut que les données aient une quelconque utilité. L'infrastructure pour héberger les mégadonnées demande un flot continu d'argent puisque la technologie évolue sans cesse et que les données générées sont de moins en moins structurées. Il faut aussi changer les mentalités dans l'organisation, de manière à ce que tous respectent et se transmettent les données, pour que ceux qui en ont besoin puissent en profiter.

Les questions de protection de la vie privée sont aussi importantes par rapport aux mégadonnées. Les données sensibles doivent être sauvegardées et chiffrées dans une infrastructure de sécurité informatique impénétrable pour en empêcher l'accès par des personnes malveillantes. Une institution peut facilement perdre la confiance des gens en cas d'atteinte à la protection des données et de fuites.

De plus, l'analyse de mégadonnées peut parfois aboutir à des informations de qualité discutable en raison de la nature des méthodes statistiques suivies. Deux points ressortent ici. Premièrement, l'analyse de mégadonnées porte généralement sur la corrélation et non sur la causalité. Une fois une corrélation mise au jour, on ne consacre souvent pas vraiment d'effort à essayer de comprendre le lien entre les variables. Sans une telle analyse, il est impossible d'évaluer le bien-fondé et l'efficacité financière des mesures prises en fonction de ces corrélations. Deuxièmement, les bases de données utilisées pour l'analyse peuvent ne pas être randomisées ou représentatives de la population à l'étude. Par exemple, les données sur les médias sociaux peuvent sous-représenter les points de vue de ceux qui ont une connaissance limitée de la technologie ou un accès limité à celle-ci, comme les aînés et les personnes à faible revenu. Un échantillon non représentatif limite les possibilités de généralisation des résultats et donc mine le processus décisionnel. On peut compenser ces lacunes méthodologiques par des techniques statistiques servant à déterminer la causalité et des échantillons de taille supérieure plus représentatifs.

L'utilisation des techniques d'apprentissage autonome issues de l'intelligence artificielle comporte aussi ses propres difficultés. Ces algorithmes pourraient demeurer trop obscurs pour qu'un observateur externe puisse comprendre comment ils aboutissent à leurs effets particuliers. Selon le type d'algorithme utilisé et le processus d'apprentissage qu'on lui a fait suivre, l'algorithme peut ne pas être en mesure de révéler exactement comment il a pris ses décisions. Il faut surveiller l'évolution des décisions prises par

l'intelligence artificielle reposant sur des mégadonnées pour éviter qu'elle n'intègre de mauvaises décisions.

Éléments à considérer

SPC devra tenir compte des conséquences de l'adhésion du Canada au réseau Digital 9. Ce réseau de collaboration entre pays dont le gouvernement est à la fine pointe en numérique vise collectivement la mise au point de systèmes informatiques publics libres et la réalisation d'initiatives de gouvernement ouvert dans le cadre desquelles les pays partenaires se partagent des données ouvertes^{xiv}. À titre de responsable des bases de données du gouvernement du Canada, SPC devra potentiellement respecter les principes proposés de développement numérique énoncés dans la charte de l'organisation. Parmi les pays du réseau, on observe une tendance à la démocratisation des données : autrement dit, les citoyens sont propriétaires de leurs propres données et le gouvernement agit en tant qu'intendant des données. L'Estonie est un exemple de gouvernement entièrement numérique : les citoyens ont accès à toutes les informations détenues par leur gouvernement (soins de santé, impôts, banques, etc.) grâce à la plateforme de données X-Road, qui relie les bases de données gouvernementales entre elles et extrait les informations demandées par un utilisateur dans un tableau de bord^{xv}. Le gouvernement du Canada et SPC devront aller dans cette direction s'ils ne veulent pas s'écarter du droit chemin en matière de technologies.

L'imminente directive et politique sur la gouvernance numérique force à augmenter le nombre de données interactives auxquelles les citoyens peuvent avoir accès. Le gouvernement du Canada devra envisager une architecture de données efficace et souple afin que les ministères puissent offrir des services numériques interopérables. Il faut aussi faire en sorte que le client n'ait plus à donner son information qu'une seule fois au lieu d'avoir à le faire régulièrement comme c'est le cas actuellement. Un autre point important, dans le contexte des données de gouvernement ouvert, est l'anonymisation de l'information. Si des renseignements sur un citoyen sont recueillis, compilés et utilisés dans le cadre de projets, ils doivent être protégés et rendus anonymes avant que les ensembles de données puissent être publiés. La directive et la politique sur la gouvernance numérique exigent également l'utilisation de logiciels libres chaque fois que c'est possible; il faudra donc en tenir compte lorsque nous adopterons de nouvelles technologies pour traiter les mégadonnées.

D'autres éléments importants sont le changement organisationnel et la gestion du changement. De nouvelles technologies et stratégies seront adoptées pour faciliter l'utilisation de solutions reposant sur des mégadonnées. La façon dont les données sont traitées devra également changer de fond en comble pour que les bonnes personnes aient accès aux bonnes données le moment venu.

Annexe A – Glossaire sur les mégadonnées

Volume : Qui dit mégadonnées dit volume élevé de données, élevé au point de rendre difficiles la gestion, le stockage et le traitement des données étant donné que ces dernières augmentent de façon exponentielle dans l'univers numérique.

Variété : Les mégadonnées sont des données insaisissables par les modes de collecte habituels (données structurées ou semi-structurées, comme les enquêtes, les ensembles de données administratives et les données opérationnelles/transactionnelles). Ce sont des données non structurées issues de diverses sources qui vont dans tous les sens : données de capteurs, texte, vidéo, audio, images, médias sociaux. Étant donné que la majorité des données de l'univers numérique ne sont pas structurées (85 % des données, contre 15 % de données structurées) et qu'elles sont en majorité constituées de bruit, il faut d'autres modes d'analyse technologique pour filtrer et classer les données numériques.

Vitesse : Les données non structurées sont souvent produites à grande vitesse, celles issues de capteurs, de compteurs intelligents ou même de réseaux sociaux (gazouillis) en étant de bons exemples. Cette vitesse sans précédent à laquelle des données sont créées ouvre la voie à l'analyse en temps réel, en particulier dans les domaines où des alertes doivent être envoyées à un système.

Véracité : Avant l'analyse des données, il faut en vérifier la véracité. Les mégadonnées viennent sous de nombreuses formes et proviennent de sources disparates; par conséquent, il faut pouvoir retracer le chemin emprunté pour donner les résultats de l'analyse et justifier ces résultats.

Valeur : Les mégadonnées bousculent la manière dont on decode et met à profit les données, loin des statistiques traditionnelles. L'analyse de mégadonnées exige de nouveaux procédés et de nouvelles façons de faire des prévisions pour dévoiler les tendances révélatrices de précieuses informations.

ⁱ Analytics: Predictive, Descriptive & Prescriptive. Big Data LDN. 17 octobre 2016. <https://bigdataldn.com/analytics-predictive-descriptive-prescriptive/>.

ⁱⁱ Forecast of Big Data market size based on revenue, from 2011 to 2027. Statista. <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>.

ⁱⁱⁱ How Big Data is Changing Science. Tom Chivers. Mosaic. 2 octobre 2018. <https://phys.org/news/2018-10-big-science.html>.

^{iv} Cern: Where the Big Bang meets Big Data. Nick Heath. Tech Republic. 22 mai 2012. <https://www.techrepublic.com/blog/european-technology/cern-where-the-big-bang-meets-big-data/>.

^v Big Data: Explaining its uses in Environmental Sciences. Matthew Mason. <https://www.environmentalscience.org/data-science-big-data>.

^{vi} Big Data: Explaining its uses in Environmental Sciences. Matthew Mason. <https://www.environmentalscience.org/data-science-big-data>.

^{vii} Ce que Watson for Oncology peut faire pour votre organisation. IBM. <https://www.ibm.com/ca-fr/marketplace/clinical-decision-support-oncology>.

^{viii} Centre d'analyse des données. Conseil national de recherches du Canada. <https://nrc.canada.ca/fr/recherche-developpement/installations-cnrc/centre-danalyse-donnees>.

^{ix} Centre de recherches sur les communications Canada. Sciences et innovation. http://www.crc.gc.ca/eic/site/069.nsf/fra/h_00045.html.

^x Recherche et renseignements d'affaires. Gouvernement du Canada. <https://www.canada.ca/fr/services/entreprises/recherche.html>.

^{xi} Sommaire de l'évaluation des facteurs relatifs à la vie privée – exploration de données. <https://www.canada.ca/fr/agence-revenu/services/a-propos-agence-revenu-canada-arc/protection-vos-renseignements-personnels/evaluation-facteurs-relatifs-a-vie-privee/sommaire-evaluation-facteurs-relatifs-a-vie-privee-exploration-donnees.html>.

^{xii} Canada Revenue Agency Monitoring Facebook, Twitter Posts of Some Canadians. Elizabeth Thompson. CBC News. 19 janvier 2017. <https://www.cbc.ca/news/politics/taxes-cra-facebook-big-data-1.3941416>.

^{xiii} La géomatique : Il y a une carte pour ça! Agriculture et Agroalimentaire Canada. <http://www.agr.gc.ca/fra/nouvelles/realisations-scientifiques-en-agriculture/la-geomatique-il-y-a-une-carte-pour-ca/?id=1492700093284>.

^{xiv} Digital 5 Charter. Décembre 2014. <https://www.ict.govt.nz/assets/Uploads/D5Charter-signed-accessible.pdf>.

^{xv} Estonia, the Digital Republic. Nathan Heller. 18 décembre 2017. <https://www.newyorker.com/magazine/2017/12/18/estonia-the-digital-republic>.