Principes fondamentaux et impact des mégadonnées

Une initiative de l'AIT numérique

Présentée par le Conseil des technologies de l'information et des communications

Bienvenue à l'AIT numérique du CTIC - Principes fondamentaux et impact des mégadonnées

L'AIT numérique du CTIC est un programme novateur d'apprentissage intégré au travail qui aide les étudiants et les étudiantes à acquérir des connaissances et des compétences pour faire progresser leur apprentissage et acquérir une expérience de travail enrichissante.

APERÇU DU COURS

Bienvenue au cours Principes fondamentaux et impact des mégadonnées. Ce cours du CTIC est une excellente occasion d'apprentissage pour toutes nos participantes et tous nos participants qui pourront se familiariser avec l'une des industries les plus passionnantes aujourd'hui. Ce cours constitue également une source de connaissances sur les technologies des mégadonnées et leur impact actuel sur les entreprises du monde entier, et plus particulièrement au Canada.

À PROPOS DE CE COURS

Comme vous l'avez peut-être remarqué (d'après le titre du cours), il y a deux blocs d'apprentissage principaux : principes fondamentaux des mégadonnées et impact juridique et éthique au Canada. Chacun d'eux contient trois modules pour un total de six modules.

Titre	Learning goals	Learning goals
1	Valeur des mégadonnées et applications potentielles	Apprendre les technologies des mégadonnées, les principes fondamentaux et leur relation avec d'autres technologies comme l'intelligence artificielle.
2	Travailler avec les données	Analyser les architectures des mégadonnées, les méthodologies d'intégration des données et les entreprises fournissant des données en tant que service.
3	Equipes de données et rôles	Décrire les équipes de données, les rôles typiques, les activités et les outils.
4	Modèles d'entreprise et analyse stratégique	Exploiter les outils existants pour analyser les modèles d'entreprises et les stratégies des entreprises et déterminer la valeur des projets de mégadonnées.
5	Obligations légales	Comparer la réglementation sur les données et l'intelligence artificielle dans le monde.
6	Impacts éthiques	Explorer les approches, les lignes directrices et les principes de l'éthique des données.
	Étude de cas	Systèmes d'identité auto-souveraine

Nous pensons que ces modules vont vous permettre d'avoir une vision actualisée de l'industrie des mégadonnées et de leurs applications de pointe.

Nous vous recommandons également de lire ce rapport du CTIC sur les mégadonnées et l'économie de l'intelligence artificielle : *Aperçu des talents numériques pour 2025* https://www.ictc-ctic.ca/wp-content/uploads/2021/08/apercu-des-talents-numeriques-pour-2025.pdf et les informations actualisées fournies dans le rapport CTIC sur les perspectives 2025 pour les talents numériques : *Construire la future main-d'œuvre canadienne dans le domaine de l'intelligence artificielle dans le meilleur des mondes (postpandémie)* https://www.ictc-ctic.ca/wp-content/uploads/2021/03/ICTC_Report_Building_FRE.pdf ainsi que *L'ère de demain - La main-d'œuvre amplifiée par l'intelligence artificielle du Canada* https://www.ictc-ctic.ca/wp-content/uploads/2020/02/canadas-ai-workforce-FINAL-FRENCH-2.24.20.pdf .

Il explore l'immense opportunité des mégadonnées pour les entreprises, les chercheuses et les chercheurs d'emploi sur le marché canadien. Nous nous référerons à des sections spécifiques de ce document tout au long du cours.

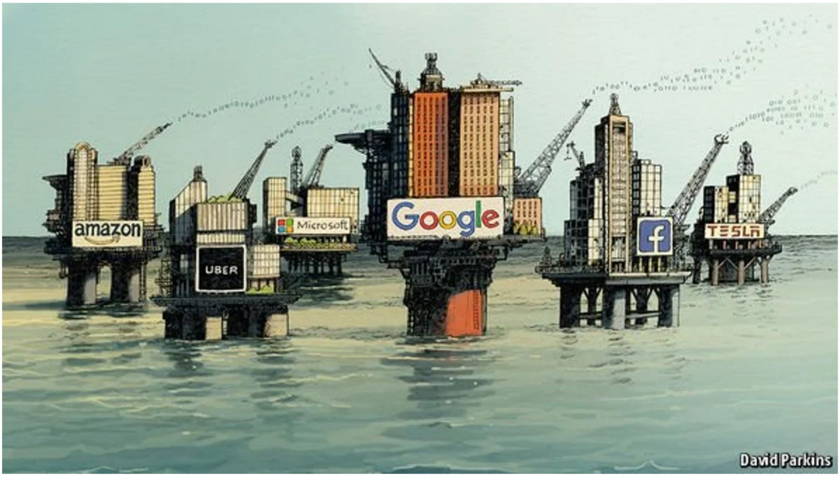
Principes fondamentaux et impact des mégadonnées

Une initiative de l'AIT numérique

Présentée par le Conseil des technologies de l'information et des communications

1.1 Introduction aux mégadonnées et exemples dans diverses industries à travers le monde

Les mégadonnées ou *big data* sont souvent considérées comme un mot à la mode, mais la réalité est que cette technologie (ou cet ensemble de technologies) est déjà en train de transformer les industries et les organisations partout dans le monde. Les données sont considérées comme le « nouveau pétrole » depuis des années, en raison de leur énorme potentiel, mais il est nécessaire de les raffiner pour en extraire la valeur réelle. Les entreprises investissent des millions de dollars pour créer des équipes des données et pour développer de nouvelles applications guidées par les données.



PARKINS, David. *Data is the new oil* [illustration], The Economist, licence inconnue. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (en anglais)

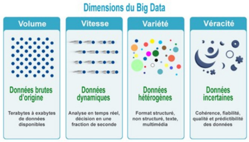
[Description du média : peinture de sept plateformes pétrolières sur l'eau, chacune portant le logo d'une entreprise technologique. Les logos comprennent Amazon, Uber, Microsoft, Google, Facebook et Tesla]

Mais qu'entendons-nous par mégadonnées? Existe-t-il aussi ce qu'on appelle les petites données? Si oui, comment fait-on la différence? S'agit-il uniquement de la taille?

Techniquement, les mégadonnées ne sont pas seulement « méga ». Cela signifie que nous ne considérons pas seulement le volume (quantité ou taille des données), mais nous considérons aussi d'autres dimensions telles que la variété, la vitesse, la véracité, etc. Remarque : il existe une panoplie de versions avec trois, quatre, cinq, sept, et même dix « V ». N'hésitez pas à faire des recherches sur les autres dimensions, mais ces quatre dimensions sont les plus pertinentes en tant que caractéristiques des données :

- Volume (généralement massif)
- Variété (sources diverses et formats divers)
- Vitesse (rapidité de traitement)
- Véracité (données crédibles et réelles)

Lecture optionelle pour des détails additionnels https://blog.avanci.fr/le-big-data-et-les-4-v-de-la-donnee


Dimensions du Big Data
Volume
Données brutes d'origine
Terabytes à exabytes de données disponibles
Vitesse
Données dynamiques
Analyse en temps réel, décision en une fraction de seconde
Variété
Données hétérogènes
Format structuré, non structuré, texte, multimédia
Véracité
Données incertaines
Cohérence, fiabilité, qualité et prédictibilité des données

IBM et al. ***Les « 4 V » ou principaux attributs des mégadonnées*** **[infographie], licence inconnue.** http://www.astrosurf.com/luxorion/big-data-mining.htm

[Description du média : infographie intitulée « Les quatre V des mégadonnées », à savoir le volume, la vitesse, la variété et la véracité]

Il existe des applications où le volume n'est pas si pertinent, notamment les systèmes de l'IdO (Internet des objets) dans lesquels nous collectons de petits morceaux de données à des fins de suivi ou d'analyse... et c'est tout à fait bien. Ces petits morceaux peuvent être utilisés de façon spécifique ou ils peuvent être intégrés ensemble et devenir des données à haut volume. Quelque chose de similaire se produit pendant le processus de formation de certains modèles d'apprentissage machine, soit parce qu'ils ne nécessitent pas un volume très élevé, soit parce que l'ensemble de données de formation est basé sur de plus petits échantillons de données.

Comme vous le voyez, il existe différentes façons de définir les mégadonnées et la définition dépendra de la source. Voyons ce que disent certaines autorités en la matière comme l'Université de Nantes :

I GOOGLE SEARCH
PASSPORT
PASSPORT SCAN
SHOPPING HISTORY
BARCODE SCAN
SUBSCRIBE

Qu'est-ce que le Big Data? **Université de Nantes, [vidéo YouTube], CC BY 3.0.**

https://www.youtube.com/watch?v=eVSfJhssXUA

[Description du média : animation des sous-titres]

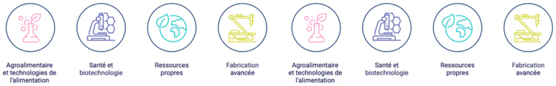
Maintenant que nous avons compris ce que nous appelons mégadonnées, leurs dimensions et leur valeur stratégique pour les organisations du monde entier, regardons quelques cas concrets. Quels sont les exemples typiques d'applications des mégadonnées? En voici quelques-unes :

- Télécommunications : suivi en temps réel des pannes de réseau au niveau national
- Banque : détection des fraudes sur les cartes de crédit
- Assurance : données provenant des capteurs des voitures pour des primes d'assurance personnalisées en fonction du comportement de conduite
- Chaîne d'approvisionnement : activités internationales de suivi de bout en bout et alertes proactives
- Commerce de détail : gestion centralisée du stock des magasins au niveau international
- Internet : moteur de recherche basé sur d'énormes points de données

ALTMANN, Gerd. ***Industry 4.0*** **[image], licence de Pixabay.** https://pixabay.com/photos/industry-industry-4-web-network-3087393/

[Description du média : homme tenant une tablette, portant un costume gris. À droite de la tablette, des icônes de forme hexagonale associées à l'industrie technologique s'emboîtent les unes dans les autres]

Voici d'autres bons exemples d'industries stratégiques qui utilisent les mégadonnées.

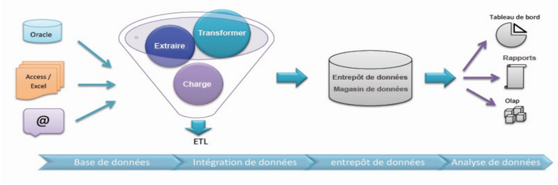


Principes fondamentaux et impact des mégadonnées

Une initiative de l'AIT numérique

Présentée par le Conseil des technologies de l'information et des communications

2.0 Travailler avec les données



SPCTEAMWORK. ***Schéma expliquant l'intégration de base de données*** **[image], Wikimedia Commons, CC BY 4.0.**
https://commons.wikimedia.org/wiki/File:Sh%C3%A9mas_d%27int%C3%A9gration_de_donn%C3%A9es_1.jpg

[Description du média : homme hispanique aux cheveux bruns courts et à la barbe, portant un chandail gris avec un col noir. L'arrière-plan est de couleur jaune uni]

Transcription

Bienvenue au module 2. On est ici après le module 1, où vous avez vu les [principes] fondamentaux déjà de la partie mégadonnées, on a parlé des tendances de l'industrie, la différence par rapport à l'intelligence artificielle, plusieurs choses comme ça. On va aller plus en détail, maintenant, on va commencer à parler de O.K. une fois qu'on a les données de l'entreprise, comment est-ce qu'on va les garder? C'est quoi les architectures des données qu'on va utiliser pour les garder et les utiliser plus tard en fonction des objectifs d'affaires? Il y a différentes façons de le faire. On va parler des méthodes de collecte des données, de comment est-ce qu'on va aller collecter les données. Et on va ramener des notions aussi très intéressantes comme celle des données ouvertes, de quoi on parle quand on parle d'open data? Les gouvernements, les industries et les entreprises qui vont mettre des données à disposition des gens afin d'innover ou d'étudier, de faire des projets académiques. Donc on va parler de tout ça. Comme on a fait déjà pour le module 1, on va avoir une petite activité à la fin du module. Vous allez avoir du matériel un peu interactif, multimédia, donc profitez-en bien et on se voit à la fin du module 2. Merci beaucoup.

APPRENTISSAGE INTÉGRÉ AIT NUMÉRIQUE AU TRAVAIL

Principes fondamentaux et impact des mégadonnées

Une initiative de l'AIT numérique

Présentée par le Conseil des technologies de l'information et des communications

4.1 Identification des opportunités guidées par les données

La première étape pour trouver des opportunités guidées par les données au sein des organisations est de réfléchir à l'adéquation entre les objectifs potentiels du projet et les objectifs existants de l'entreprise. Cette étape est cruciale car aucune entreprise ne lancera un nouveau projet de données s'il n'est pas censé apporter un avantage concurrentiel pour l'organisation. Aucune dirigeante ni aucun dirigeant ne prendra le risque de mettre en œuvre un projet guidé par les données s'il n'est pas aligné sur ses indicateurs de rendement clés (IDC).

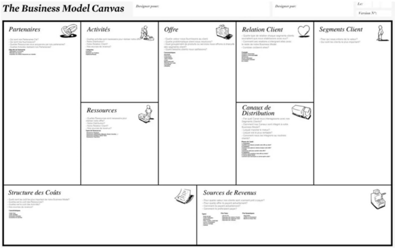
HARTAWAN, Nugroho Dwi. ***Questions idea*** **[image], licence de Pixabay.**

https://pixabay.com/vectors/question-questions-man-head-2519654

[Description du média : illustration de deux têtes se faisant face, l'une avec un point d'interrogation dans la tête, l'autre avec une ampoule dans la tête. Les têtes sont noires sur un fond vert]

Pour ce faire, et même avant de commencer à énumérer et examiner les cas potentiels guidés par les données, il convient de prendre en compte deux aspects principaux :

Le modèle économique de l'entreprise - Il s'agit de la manière dont les entreprises gagnent de l'argent et de l'analyse de leurs principales activités. Il existe un modèle très utile et universel pour cette analyse, le déjà classique *Business Model Canvas*, créé par Alexander Osterwalder en 2005. Vous n'avez pas besoin de l'utiliser systématiquement, mais il peut s'agir d'un bon document de travail pour organiser les informations et garder à l'esprit le modèle économique de l'entreprise lors de la discussion autour de nouveaux projets guidés par les données.

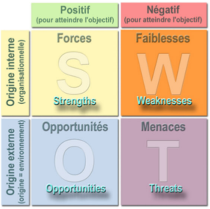


Business Model Canvas **[cadre], basé sur l'œuvre d'Alexander Osterwalder, Agir & entrependre, CC BY 3.0.** https://www.agiretentreprendre.fr/business-model-canvas
[Description du média : capture d'écran du cadre du *Business Model Canvas* du site Web ci-dessus]

Comment l'utiliser?

Une fois que vous avez complété le *Business Model Canvas* (BMC) avec les informations de l'entreprise, vous pouvez extraire des éléments tels que des activités clés et des canaux de distribution spécifiques pour imaginer des domaines potentiels pour les cas d'utilisation des mégadonnées. Par exemple, une visibilité partagée d'un outil pour les parties prenantes de la chaîne d'approvisionnement, des tableaux de bord opérationnels pour soutenir la direction dans la prise de décisions, etc. Vous pouvez également explorer la structure des coûts et les sources de revenus pour imaginer la valeur monétaire des projets potentiels.

- **L'analyse stratégique interne et externe de l'entreprise -** À cette fin, il existe également un cadre standard qui aide à identifier les différents aspects stratégiques, la matrice FFPM.



Wikipedia, basé sur l'œuvre d'Albert Humphrey et de l'Institut de recherche de Stanford **[cadre], CC BY 2.5.** https://fr.wikipedia.org/wiki/SWOT_(m%C3%A9thode_d%27analyse)#/media/Fichier:SWOT_grapheFL.jpg

Principes fondamentaux et impact des mégadonnées

Une initiative de l'AIT numérique

Présentée par le Conseil des technologies de l'information et des communications

4.4 Calcul du rendement du capital investi des projets de mégadonnées

Nous devons définir les objectifs du projet dès le départ pour décider de la « définition du succès ».

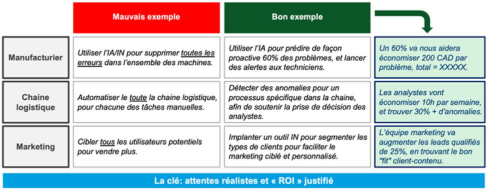
WITT, Malachi. *Rencontre entreprise* [image], licence de Pixabay.

https://pixabay.com/fr/photos/rencontre-entreprise-architecte-2284501/

[Description du média : vue d'ensemble de trois personnes (deux hommes, une femme) assises autour d'un bureau en bois, travaillant sur des plans détaillés avec une calculatrice et un ordinateur portable]

a) Les hypothèses de valeur

Pour cela, nous devons choisir des projets basés sur des attentes réalistes et mesurables. Voici quelques bons et mauvais exemples ainsi que des hypothèses de valeur (par exemple, économies, augmentation des revenus, etc.).



	Mauvais exemple	Bon exemple	
Manufacturier	*Utiliser l'IA/IN pour supprimer toutes les erreurs dans l'ensemble des machines.*	*Utiliser l'IA pour prédire de façon proactive 60% des problèmes, et lancer des alertes aux techniciens.*	*Un 60% va nous aidera économiser 200 CAD par problème, total = XXXXX.*
Chaine logistique	*Automatiser le toute la chaine logistique, pour chacune des tâches manuelles.*	*Détecter des anomalies pour un processus spécifique dans la chaine, afin de soutenir la prise de décision des analystes.*	*Les analystes vont économiser 10h par semaine, et trouver 30% + d'anomalies.*
Marketing	*Cibler tous les utilisateurs potentiels pour vendre plus.*	*Implanter un outil IN pour segmenter les types de clients pour faciliter le marketing ciblé et personnalisé.*	*L'équipe marketing va augmenter les leads qualifiés de 25%, en trouvant le bon "fit" client-contenu.*

La clé: attentes réalistes et « ROI » justifié

GONZALEZ SANCHEZ, Adrian. *Exemples de projets de mégadonnées et hypothèse de valeur* [image]. https://drive.google.com/file/d/1ViAknC-kCQu5xGmC6AWR7CMlEITeL5WO/view?usp=sharing

[Description du média une matrice de décision carrée de quatre par quatre. Le texte complet de la description se trouve dans l'hyperlien ci-dessus]

b) Rendement du capital investi (RCI)

Une fois que nous avons une estimation claire de la valeur potentielle du projet, nous pouvons calculer le rendement du capital investi (RCI). Investopedia utilise le terme ROI comme équivalent de RCI. Il définit le ROI comme suit :

> *« Une mesure de performance utilisée pour évaluer l'efficacité ou la rentabilité d'un investissement ou pour comparer l'efficacité d'un certain nombre d'investissements différents. Le ROI tente de mesurer directement le montant du retour sur un investissement particulier, par rapport au coût de l'investissement. Pour calculer le ROI, le bénéfice (ou rendement) d'un investissement est divisé par le coût de l'investissement. Le résultat est exprimé en pourcentage ou en ratio ».*

$$ROI = \frac{Bénéfice}{(Capitaux\ propres + capitaux\ empruntés)} X\ 100$$

Le bénéfice est la différence entre la valeur du projet (section 4.4.a) et le coût total estimé (y compris les ressources humaines, les licences de logiciel, le matériel, etc.).

Financé par le gouvernement du Canada | Canada

Centre d'expertise national à but non lucratif pour l'économie numérique, nous sommes la source de confiance pour les conseils stratégiques fondés sur des données probantes, de recherches prospectives et de programmes de renforcement des capacités créatives.