



uOttawa

Faculté des sciences sociales  
Faculty of Social Sciences

# Applications de l'analyse des mégadonnées dans le secteur public: possibilités et défis

Eric Champagne ([echampagne@uottawa.ca](mailto:echampagne@uottawa.ca))

Francis Gaudreault ([fgaud009@uottawa.ca](mailto:fgaud009@uottawa.ca))

*Avec l'assistance de Ula Abu Rashed*

Centre d'études en gouvernance

Université d'Ottawa



# Objectifs

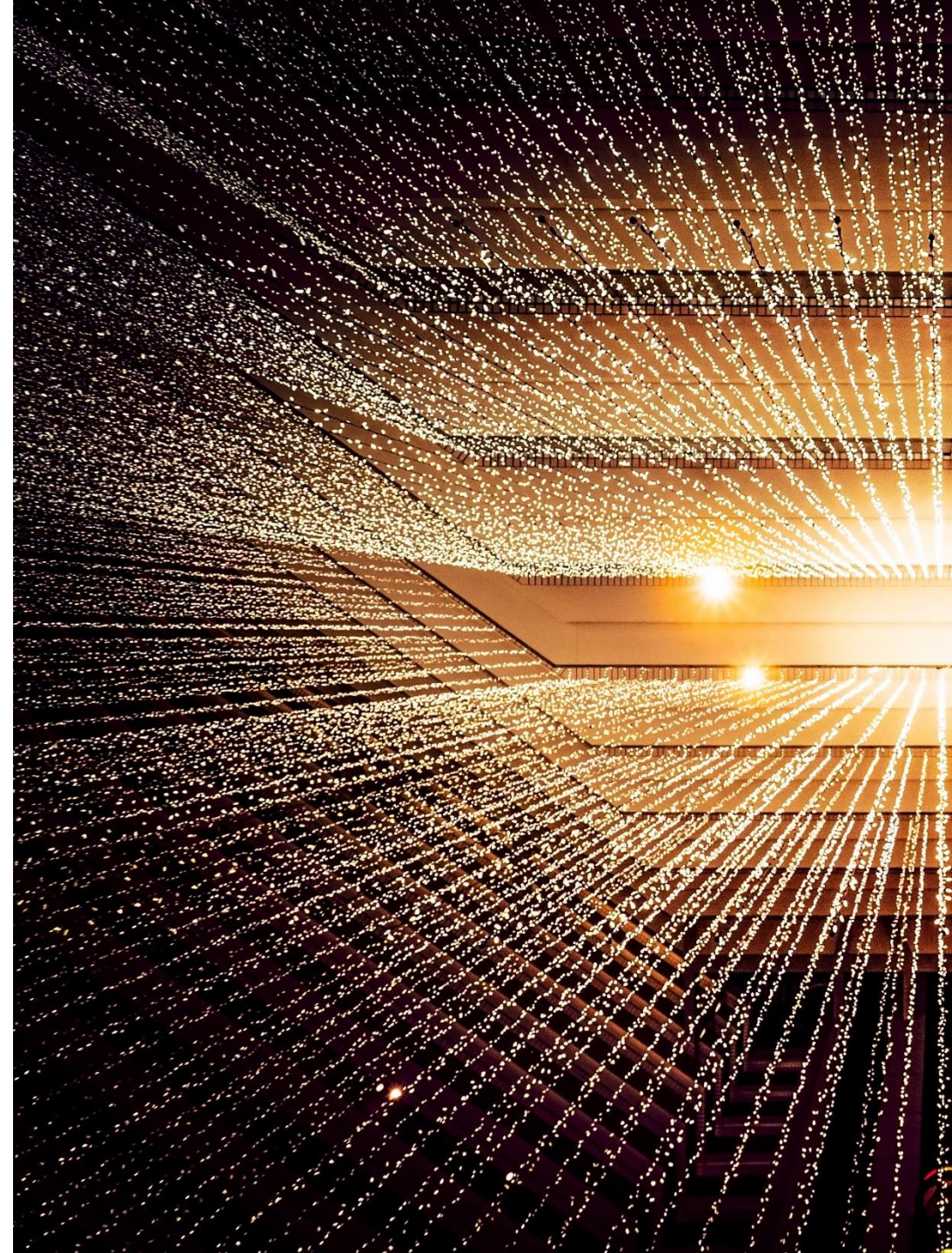
A photograph of a young child with red hair aiming a bow at a target. An adult wearing a red hat is looking on. The target has a yellow flag on top and a bullseye in the center.

- 1. Contexte et définition des mégadonnées pour le secteur public**
- 2. Les applications dans le secteur public**
- 3. Les défis organisationnels et les enjeux éthiques dans le secteur public**

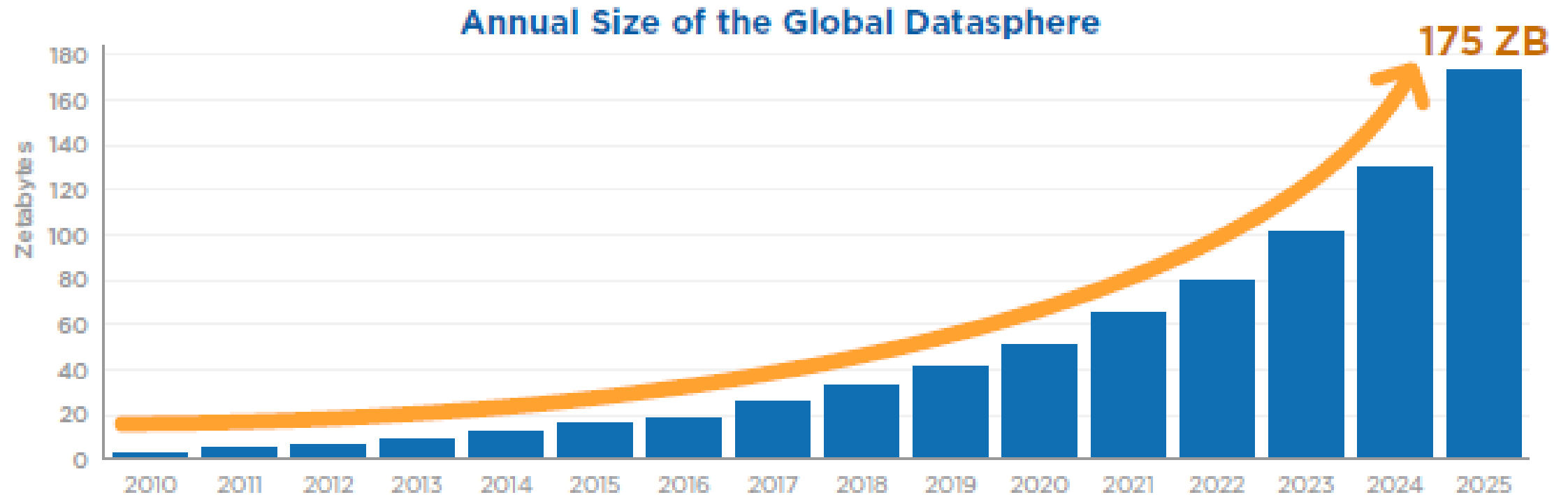


# Révolution numérique

- bouleversement profond de la société et de l'économie provoqué par l'essor informatique (appareils mobiles) et l'Internet combinés
- Accroissement des échanges numériques (utilisation d'internet, courriels, réseaux sociaux, messagerie, applications, transactions, publications de documents, photos, vidéos, technologies de machine à machine, objets connectés à l'Internet, G5...)
- Crée une explosion quantitative de données numériques



# Croissance exponentielle de création de données numériques



Source: Data Age 2025, sponsored by Seagate with data from IDC Global DataSphere, Nov 2018

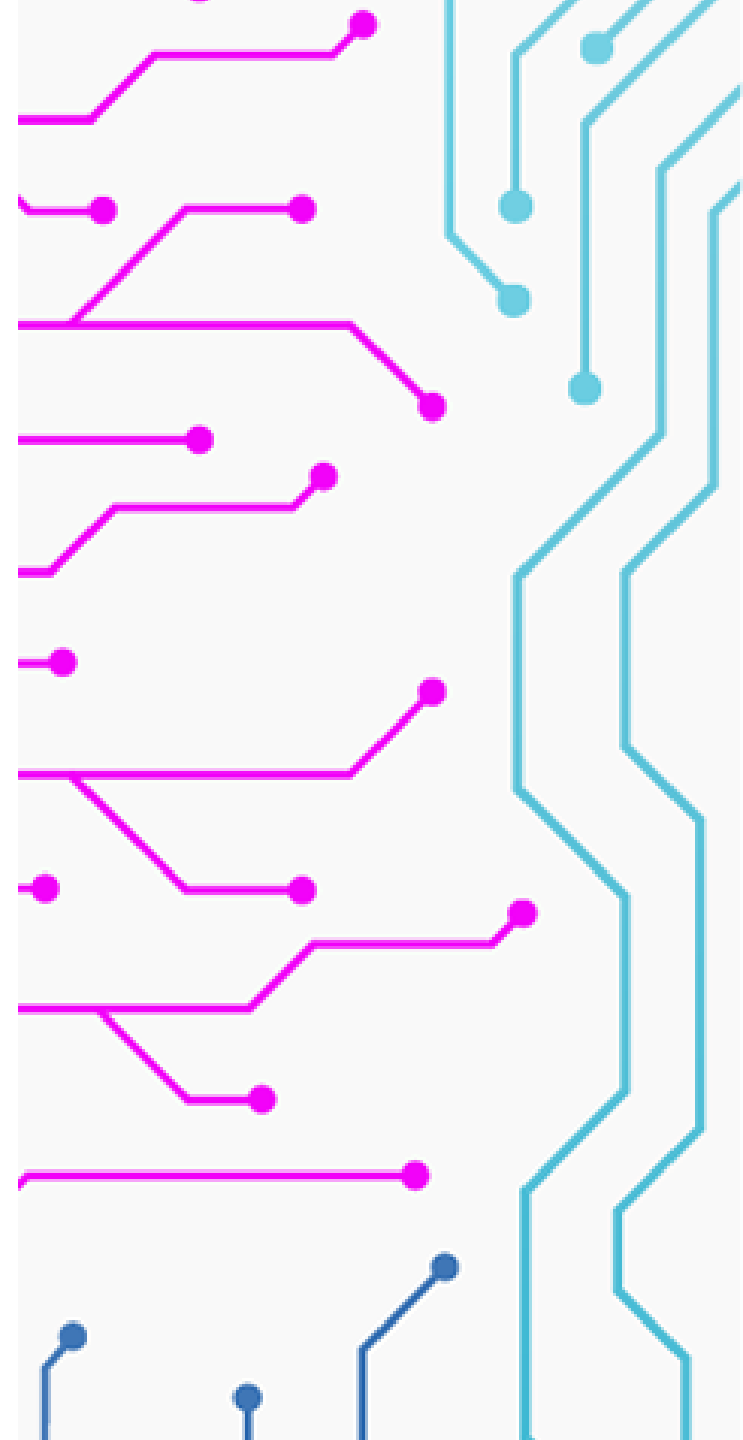
Source: IDC (2018). The Digitization of the World: From Edge to Core, p. 6.

<https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>

# Feuille de route de la Stratégie de données pour la fonction publique fédérale

- En 2018, le Gouvernement du Canada a annoncé la feuille de route de la stratégie des données pour la fonction publique fédérale.
- La stratégie:
  - vise à favoriser les nouvelles opportunités d'utilisation des mégadonnées et de l'IA dans les initiatives et projets gouvernementaux;
  - soutenir le développement et la mise en oeuvre de plateformes de mégadonnées et de gouvernement ouvert;
  - supporter l'effort de modernisation à travers la mise en place d'une fonction publique centrée sur le citoyen, la transparence, la collaboration et le numérique.

Source: <https://www.canada.ca/fr/conseil-prive/organisation/greffier/publications/strategie-donnees.html>



# Brainstorming



**Quelles initiatives numériques  
ont été prises dans vos  
organisations respectives?**

Prenez une minute pour y réfléchir

# Mégadonnées - définitions

- Désigne l'énorme quantité de données numériques collectées à partir d'une multitude de sources (sites internet, médias sociaux, photos et vidéos numériques, transactions, signaux GPS, téléphones portables, etc.)
- Elles sont trop volumineuses, brutes ou non structurées pour être analysées à l'aide de techniques statistiques classiques ou des systèmes de bases de données conventionnels (Excel; SPSS).





# Les trois V

## Volume

- Grand nombre de données et d'informations à traiter

## Variété

- Grande diversité d'informations provenant de multiples sources, non-structurées, non-organisées, accès libre, etc.

## Vélocité

- Génération, collecte et partage de données rapides.

# Mégadonnées versus données traditionnelles



=IF(B9<E5,B9*C4)					
	A	B	C	D	E
1					
2					
3	<b>Tax Band</b>	<b>Income Salary</b>	<b>Income Tax Level</b>		
4	1	< \$29,100	15%		
5	2	= > \$29,100	25%		29100
6	3	= > \$72,000	28%		72000
7					
8	<b>Employee</b>	<b>Income</b>	<b>Tax Due</b>		
9	Jenny	\$26,000	=IF(B9<E5,B9*C4)		
10	Ralf	\$45,350			
11	Kurt	\$102,000			
12					
13					
14					



# Analyse des mégadonnées

- L'analyse donne vie aux mégadonnées et crée la nouvelle valeur dans les grands ensembles de données
- Nouvelles techniques d'analyse capables d'intégrer le volume, la variété et la vélocité des données
- Nouveaux outils/logiciels d'analyse différents pour la collecte et l'analyse

# Nouveaux logiciels

- Stockage et traitement de grand volumes de données
- Intégration de multiples bases de données – non-organisées
- Souvent associés à l'infonuagique et à l'intelligence artificielle



mongoDB



cloudera®



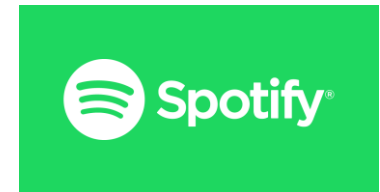
# Usages courants

- Consommation
- Affaires
- Recherche



# Information personnalisée et consommation

- Demande d'information personnalisée (reconnait vos besoins selon votre localisation, personnalisation, etc.)
- Anticipe vos besoins et apprend vos préférences (restaurants)
- Recommandations selon votre usage et profil d'utilisation (musique, films, livres)
- Besoin de transport, services ou hébergement (évaluation de la qualité des services)
- Collaboration de quartier



# Révolution des affaires

- Publicité ciblées – selon vos recherches précédentes
- Marketing prédictif – selon les informations démographiques et le cycle de vie (graduation; mariage; retraite; enfants; salaire; etc.)
- Détection de la fraude – selon vos habitudes le système peut anticiper la fraude et vous aviser ou aviser votre compagnie de crédit



# Application en recherche scientifique et universitaire

- Astronomie
- Médecine
- Génétique
- Pharmacologie
- Sciences sociales (psychologie, sociologie, sciences politiques)
- Linguistique (Google books)





# Brainstorming



## **Quels bénéfices des mégadonnées pour le secteur public?**

Prenez une minute ou deux pour y réfléchir

# Les bénéfices et les applications dans le secteur public



# Les bénéfices potentiels pour le secteur public



- Améliore le **bien-être des citoyens** (contrôle de la propagation de la maladie et recherche scientifique)
- Améliore la **prestation de services** (transport et services de santé)
- Assure la **sécurité nationale** (services secrets et le blanchiment d'argent)
- **Prévision** des **catastrophes** naturelles

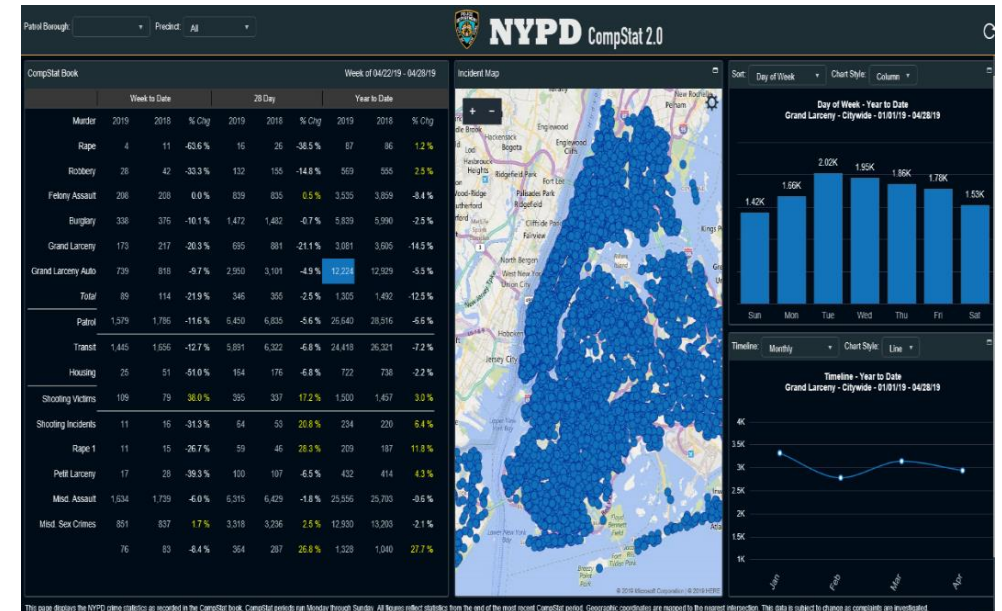
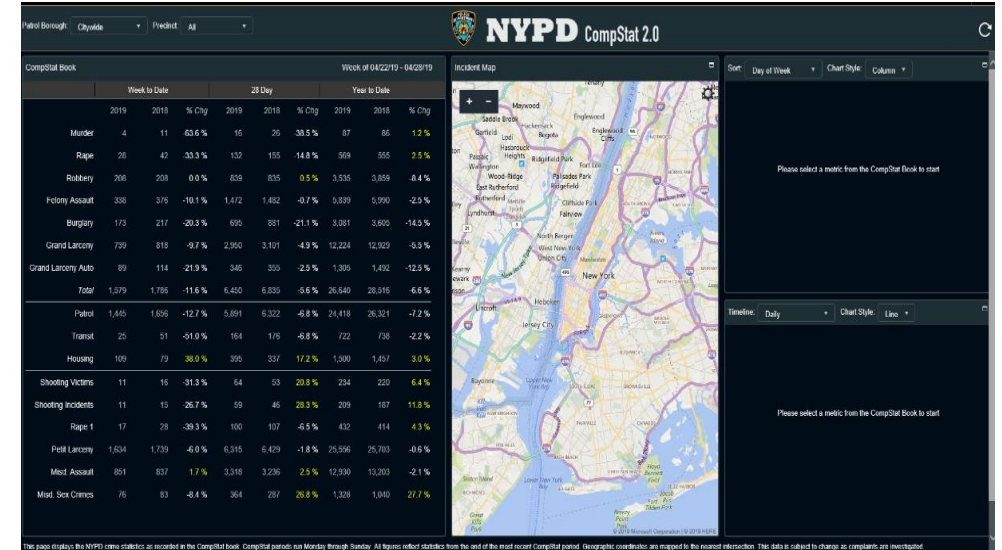
# Les bénéfices ou applications pour le secteur public (2)

- **Moteur d'innovation:** Fournit de **nouvelles idées et prévisions** qui améliore la prise de décision des décideurs publics;
- Permet **d'évaluer les besoins** des communautés, **planifier** les services et mesurer leur performance;
- **Élaboration de politiques fondées sur des faits** (soutient les rapports publics);
- Améliore **l'efficacité et l'efficience** des services publics (meilleure mesurabilité, meilleure productivité);



# COMPSTAT– Gouv. Local -NY, U.S.A.

- Le service de Police the New York a créé en 1995 le COMPSTAT (Computerized statistics). Maintenant aussi présent à Halifax et Vancouver;
- COMPSTAT est un système de gestion du rendement basé sur les mégadonnées;
- Système de mesure de la productivité et de responsabilisation des employés face aux niveau de crime et perception de sécurité;
- Le système d'informations géographiques permet de cartographier les crimes et les plaintes selon les zones.



# COMPSTAT– Fonctionnement

- **Rapport et rencontre hebdomadaire :**
  - Analyse statistiques de plaintes, arrestations, temps supplémentaire généré;
  - Permet aux gestionnaires/exécutifs de discerner les tendances et anomalies;
  - Permet de faire des comparaisons entre les gestionnaires (+ ou – de crime);
- **Rapport de profil du gestionnaires:**
  - Permet aux cadres d'analyser les performances des employés;
  - Fournit informations sur les employés, expérience, formation, évaluations.



# COMPSTAT– Bénéfices


- Meilleure rédition de compte et transparence (gouv. ouvert);
- Appuyer l'effort de réduction des crimes;
- Améliorer la sécurité des citoyens grâce à l'anticipation et la prédiction des crimes;
- Améliore l'efficacité et la mesurabilité de la performance chez les employés.

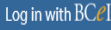


# Le Logiciel ICM \*Integrated Case Management – BC

- La province de la Colombie Britannique a mis sur pied une transformation numérique dans le but de permettre aux ministères des services sociaux de centraliser les services sociaux.
- Le logiciel ICM permet de:
  - Numériser les services sociaux;
  - Partager les données entre trois ministères de service sociaux (Min. du Dev. Social et de l'Innovation; Min. de l'Enfance et de la famille; Min. de la technologie, de l'Innovation et des Services Citoyens);
  - Fournir de meilleurs services aux citoyens.
- Par exemple: le portail de service ICM permet fournisseurs de service de garde d'accéder aux options de libre-service liées services sociaux.



 [Log in to icm.ext.gov.bc.ca](https://icm.ext.gov.bc.ca)

Log in with  BCeID

User ID  
Use a Business BCeID

Password

[Continue](#)

[Forgot your user ID or password?](#)

No account?  
[Register for a BCeID](#)

Or log in with:  
[Log in with IDIR](#)

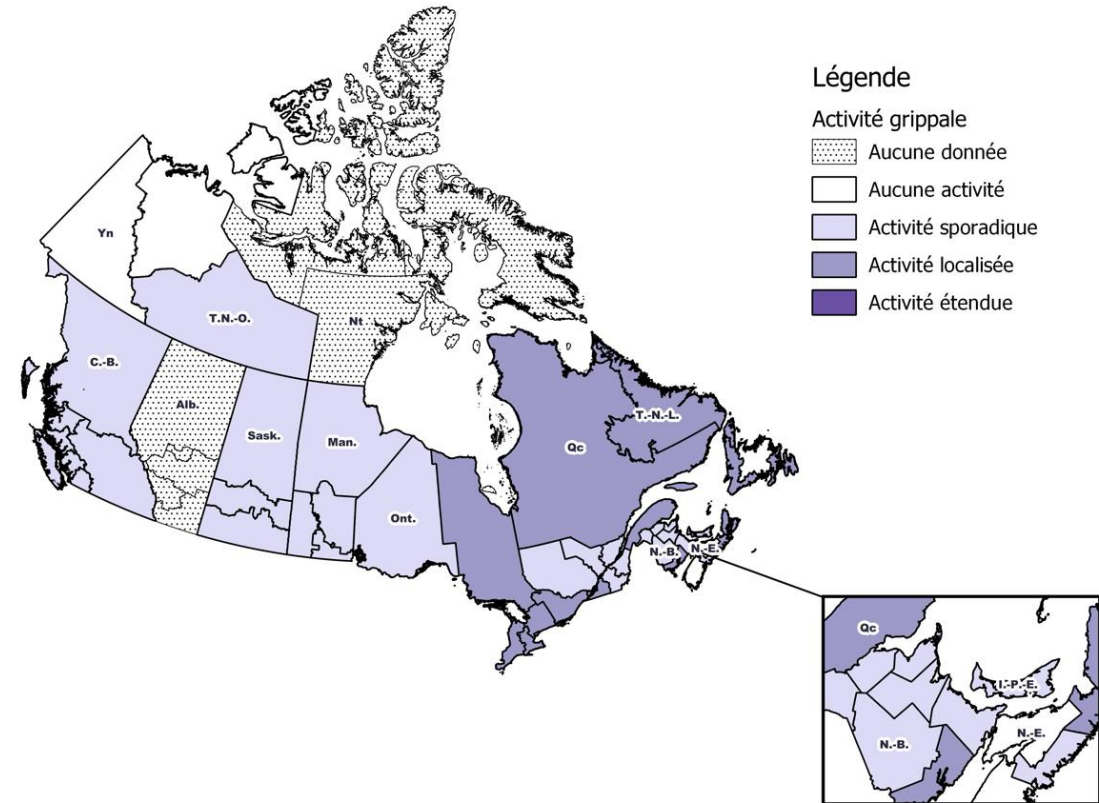
Need help?  
[Contact the BCeID Help Desk](#)



# Santé publique

- Génétique médicale – meilleurs traitements
- Prédiction et surveillance - propagation de grippe ou Ébola
- Cybersanté (accès à des diagnostic et soins à distance)

Santé Canada, Surveillance de l'influenza :  
Du 28 avril au 4 mai 2019



<https://www.canada.ca/fr/sante-publique/services/publications/maladies-et-affections/surveillance-influenza/2018-2019/semaine18-28-avril-4-mai-2019.html>

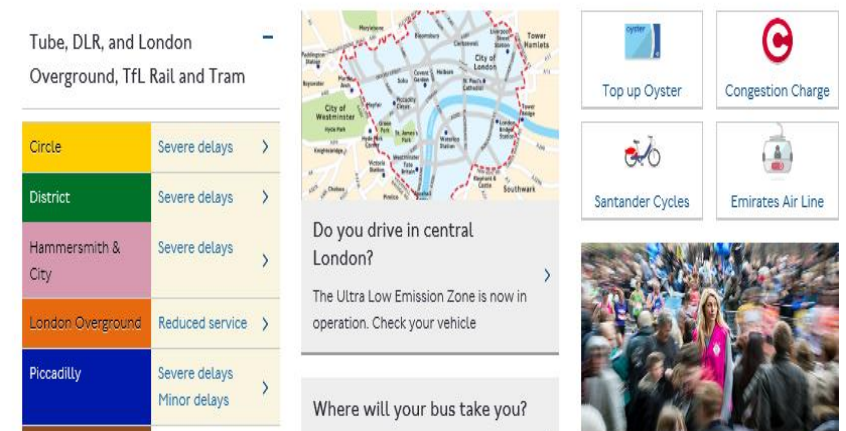
# Villes intelligentes

- Application des technologies de l'information et de l'internet des objets aux fonctions urbaines:
  - Réseaux électriques intelligents
  - Bâtiments intelligents (économie d'énergie)
  - Trottoirs laboratoires (Sidewalk Toronto)
  - Gestion des déchets et de eau (conomie et environnement)
  - Sécurité urbaine (senseurs, cameras, biométrie)
- *Cela est rendu possible par l'analyse de mégadonnées et l'intelligence artificielle*



# Mobilité intelligente

- L'usage des technologies de l'information et du numérique dans le domaine des transports:
  - Améliore la qualité des services de transport en commun (GPS sur les autobus)
  - Améliore l'offre multimodale (l'utilisation de plusieurs moyens de transports)
  - Infrastructure pour les automobiles électriques et autonomes
  - Information et gestion du trafic (trajets personnalisés)



# Brainstorming



**Quels sont les défis organisationnels et les risques éthiques associés à l'analyse des mégadonnées?**

Prenez une minute ou deux pour y réfléchir

# Les défis et les enjeux éthiques





# La gouvernance des mégadonnées

- La gouvernance des données: gérer formellement les données dans une organisation et en garantir la valeur;
- Réalisée par une combinaison de **personnes et de processus**.
- La **technologie** est utilisée pour simplifier et automatiser certains aspects du processus.
- Bonne gouvernance: les personnes et les processus restent clés (imp. **Formation**)

# Les défis liés à la mise en oeuvre dans le secteur public

- Surmonter le statu quo et la **résistance au changement** (aversion au risque);
- **Manque d'intégration** entre les divisions et les ministères; (rend combinaison des données difficile)
- Le **coût élevé** de l'analyse et le stockage des mégadonnées (informatique privée vs public);



# Les enjeux éthiques: vie privée et protection des informations

- Le **respect de la vie privée**: utilisation des informations personnelles (statut social, culturel, religieux ou professionnel etc.);
- Le cas de **Cambridge analytica**: utilisation des données pour influencer l'opinion, le vote.
- **Protection des informations** (sécurité en ligne: coût élevé et confiance du public)





# Les enjeux éthiques: la question de la confiance envers le gouvernement

- Très faible **confiance de la population** envers le gouvernement (Confiance envers l'IA: 17%; peur de partager ses informations) (BCG, 2018);
- Plus confiance dans le privée: Test d'ADN (ancestry; 23 and me)
- **Solutions:**
- Besoin de **construire la légitimité** de l'analyse des mégadonnées: inspirer la confiance et assurer la compréhension;
- Être **centré sur la perspective des citoyens** et non seulement le gouvernement (écouter la société civile);



# Les enjeux éthiques: La question de la confiance envers le gouvernement

- Utiliser mégadonnées pour plus de **transparence**: Gouv. Ouvert; mieux expliquer les dépenses du gouvernement; permettre les examens externes;
- **Éduquer la population et les fonctionnaires** sur l'IA et l'analyse des mégadonnées;
- Besoin d'une **approche mesurée** (quick wins) pour inspirer la confiance et assurer le succès (approche agile).



# Conclusion

- Mégadonnées dans le secteur public sont inévitables
- Il faut s'assurer de bien faire les choses (enjeux éthiques)
- Participation citoyenne et transparence (confiance)
- Formation des fonctionnaires



# REFERENCES

- Aggarwal, A. (2016). Opportunities and Challenges of Big Data in Public Sector. In Anil Aggarwal (Ed.), *Managing big data integration in the public sector* (pp. 289-301). Hershey, PA: IGI Global.
- Manzoor, A. (2016). Emerging Role of Big Data in Public Sector. In Anil Aggarwal (Ed.), *Managing big data integration in the public sector* (pp. 268-288). Hershey, PA: IGI Global.
- Huffman, T., Olesen, M., Green, M., Leckie, D., Liu, J., & Shang, J. (2018). Agricultural Data Analytics for Environmental Monitoring in Canada. In Feras Batarseh & Ruixin Yang (Eds.), *Federal Data Science: Transforming Government and Agricultural Policy Using Artificial Intelligence* (pp. 55–79). Academic Press.
- Debenham, D. (2016). Big Data Analytics, Big Financial Institutions, and Big Money Fraud Litigation. *Banking & Finance Law Review*, 32(1), 103-143.
- Barbero, M., Coutuer, J., Jackers, R., Moueddene, K., Renders, E., Stevens, W., ... Versteede, D. (2016). Big data analytics for policy making. Retrieved from European Commission DG INFORMATICS (DG DIGIT) website: [https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg\\_digit\\_study\\_big\\_data\\_analytics\\_for\\_policy\\_making.pdf](https://joinup.ec.europa.eu/sites/default/files/document/2016-07/dg_digit_study_big_data_analytics_for_policy_making.pdf)
- Kim, G., Trimi, S., & Chung, J. (2014). Big-Data Applications in the Government Sector. *Association for Computing Machinery. Communications of the ACM*, 57(3), 78.
- Tomar, L., Guicheney, W., Kyarisiima, H., & Zimani, T. (2016). Big Data in the Public Sector: Selected Applications and Lessons Learned [IDB-DP-483]. Retrieved from Institutions for Development Sector website: <https://publications.iadb.org/en/big-data-public-sector-selected-applications-and-lessons-learned>
- Troester, M. (2013). *Big Data Meets Big Data Analytics Three Key Technologies for Extracting Real-Time Business Value from the Big Data That Threatens to Overwhelm Traditional Computing Architectures*. Retrieved from <https://eric.univ-lyon2.fr/~ricco/cours/slides/sources/big-data-meets-big-data-analytics-105777.pdf>

# REFERENCES

- Quintero, D., de Souza Casali, D., Correia Lima, M., Gabor Szabo, I., Olejniczak, M., Rodrigues de Mello, T., & Carlos dos Santos, N. (2015). *IBM Software Defined Infrastructure for Big Data Analytics Workloads*. IBM Redbooks.
- Wilder-James, E. (2014). Defining Big Data. Forbes. Retrieved from <https://www.forbes.com/sites/eddumbill/2014/05/07/defining-big-data/#29ff1fcf5667>
- Azyan, L. (2017). How Government Can Use Big Data to Solve Public Problems. Retrieved from LGEO Research website: <http://lgeoresearch.com/government-big-data/>
- Province of British Columbia. (2019). Welcome to the ICM Service Provider Portal. Retrieved from [https://www2.gov.bc.ca/assets/gov/family-and-social-supports/child-care/icm\\_service\\_provider\\_portal\\_guide.pdf](https://www2.gov.bc.ca/assets/gov/family-and-social-supports/child-care/icm_service_provider_portal_guide.pdf)
- FireBrary. (2015). About the FireBrary. Retrieved from <http://www.firebrary.com/about-the-firebrary>
- Marr, B. (2016). *Amazing Big Data At NASA: Real Time Analytics 150 Million Miles From Earth*. Retrieved from <https://www.forbes.com/sites/bernardmarr/2016/04/14/amazing-big-data-at-nasa-real-time-analytics-150-million-miles-from-earth/#380c716f5cc4>
- Marr, B. (2019a). How Amsterdam Fire Department uses Big Data in practice. Retrieved from Bernard Marr & Co. website: <https://bernardmarr.com/default.asp?contentID=1082>
- Marr, B. (2019b). Transport for London: Big Data and public transport. Retrieved from Bernard Marr & Co. website: <https://bernardmarr.com/default.asp?contentID=693>
- Marr, B. (2019c). How NASA uses Big Data in practice. Retrieved from Bernard Marr & Co. website: <https://bernardmarr.com/default.asp?contentID=703>
- Al Nuaimi, E., Al Neyadi, H., Mohamed, N., & Al-Jaroodi, J. (2015). Applications of big data to smart cities. *Journal of Internet Services and Applications*, 6(1), 1-15.

# REFERENCES

- Fernando Almeida. (2018). Big Data: Concept, Potentialities and Vulnerabilities. *Emerging Science Journal*, 2(1), Emerging Science Journal, 01 March 2018, Vol.2(1).
- Garimella, N., Gong, W., & Zheng, Y. (2015). *Deploying a big data solution using IBM Spectrum Scale Best practices to make smart decisions for optimal performance and scalability*. Retrieved from IBM website: [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=2ahUKEwjJ14vV2YviAhXKT98KHfphCdkQFjACegQIAxAC&url=https%3A%2F%2Fwww.ibm.com%2Fdeveloperworks%2Fcommunity%2Fwikis%2Fform%2Fanonymous%2Fapi%2Fwiki%2Ffa32927c-e904-49cc-a4cc-870bcc8e307c%2Fpage%2Fab5c2792-feef-4a3a-a21b-d22c6f5d728a%2Fattachment%2F80d5c300-7b39-4d6e-9596-84934fcc4638%2Fmedia%2FDeploying a big data solution using IBM Spectrum Scale v1.7.5.pdf&usg=AOvVaw21wPpGMKwkcKSiFGBzHGlu](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&ved=2ahUKEwjJ14vV2YviAhXKT98KHfphCdkQFjACegQIAxAC&url=https%3A%2F%2Fwww.ibm.com%2Fdeveloperworks%2Fcommunity%2Fwikis%2Fform%2Fanonymous%2Fapi%2Fwiki%2Ffa32927c-e904-49cc-a4cc-870bcc8e307c%2Fpage%2Fab5c2792-feef-4a3a-a21b-d22c6f5d728a%2Fattachment%2F80d5c300-7b39-4d6e-9596-84934fcc4638%2Fmedia%2FDeploying%20a%20big%20data%20solution%20using%20IBM%20Spectrum%20Scale%20v1.7.5.pdf&usg=AOvVaw21wPpGMKwkcKSiFGBzHGlu)
- Johnson, N. B. (2016). *How you can use Data Analytics to change government*. Retrieved from GovLoop website: <https://www.govloop.com/wp-content/uploads/2016/01/DataAnalyticsGuide.pdf>
- SAS. (2019). Analytics what it is and why it matters. Retrieved from [https://www.sas.com/en\\_ca/insights/analytics/what-is-analytics.html](https://www.sas.com/en_ca/insights/analytics/what-is-analytics.html)
- Jin, X., Wah, B., Cheng, X., & Wang, Y. (2015). Significance and Challenges of Big Data Research. *Big Data Research*, 2(2), 59-64.
- de Fremery, R. (2018). Big Data and Government: How the Public Sector Leverages Data Insights. Retrieved from Hortonworks website: <https://fr.hortonworks.com/article/big-data-and-government-how-the-public-sector-leverages-data-insights/>
- Transport for London. (2019). Status updates. Retrieved from TfL website: <https://tfl.gov.uk/tube-dlr-overground/status/>
- Soderstrom, T. (2017). Exploring Space Through Streaming Analytics. Retrieved from NASA website: <https://www.nasa.gov/content/exploring-space-through-streaming-analytics>

# REFERENCES

- Davies, R. (2016). Big data and data analytics: The potential for innovation and growth Retrieved from European Parliamentary Research Service website:  
[http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS\\_BRI\(2016\)589801\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/BRIE/2016/589801/EPRS_BRI(2016)589801_EN.pdf)
- Shahat, A., Elragal, A., & Bergvall-Kåreborn, B. (2017). *Big Data Analytics and Smart Cities: A Loose or Tight Couple?* In International Conference on Connected Smart Cities, Libson, 20-22 July 2017.
- Neirotti, P., De Marco, A., Cagliano, A. C., Mangano, G., & Scorrano, F. (2014). Current trends in Smart City initiatives: Some stylised facts. *Cities*:38, 25-36.
- Su K, Li J, Fu H. (2011). Smart city and the applications. In International Conference on Electronics, Communications and Control (ICECC), Ningbo: China, 9-11 September 2011. Retrieved from the IEES website:  
<https://ieeexplore.ieee.org/abstract/document/6066743>
- Oracle. (2019). PeopleSoft Enterprise Payroll Interface. Retrieved from Oracle website:  
<http://www.oracle.com/us/products/applications/peoplesoft-enterprise/human-capital-management/053274.html>
- Agriculture and Agri-Food Canada. (2009). Government of Canada Human Resource Management System (Oracle PeopleSoft Enterprise application), version 8.9. Retrieved from Agriculture and Agri-Food Canada website:  
<http://www.agr.gc.ca/eng/about-us/offices-and-locations/access-to-information-and-privacy-services/privacy-impact-assessments/government-of-canada-human-resource-management-system-oracle-peoplesoft-enterprise-application-version-89/?id=1258395335107>
- Burrington, I. (2018). Policing Is an Information Business. *Urban Omnibus*. Retrieved from  
<https://urbanomnibus.net/2018/06/policing-is-an-information-business/>

# REFERENCES

- NYDP. (2019). NYPD CompStat 2.0. Retrieved from the website: <https://compstat.nypdonline.org/2e5c3f4b-85c1-4635-83c6-22b27fe7c75c/view/89>
- Arora, A., Benay, A. & Mendelsohn, M. (2019). Report to the Clerk of the Privy Council: A Data Strategy Roadmap for the Federal Public Service. Retrieved from the Privy Council Office website: <https://www.canada.ca/en/privy-council/corporate/clerk/publications/data-strategy.html>