



# Données de mobilité et loi 25

## Les enjeux pour la planification des transports

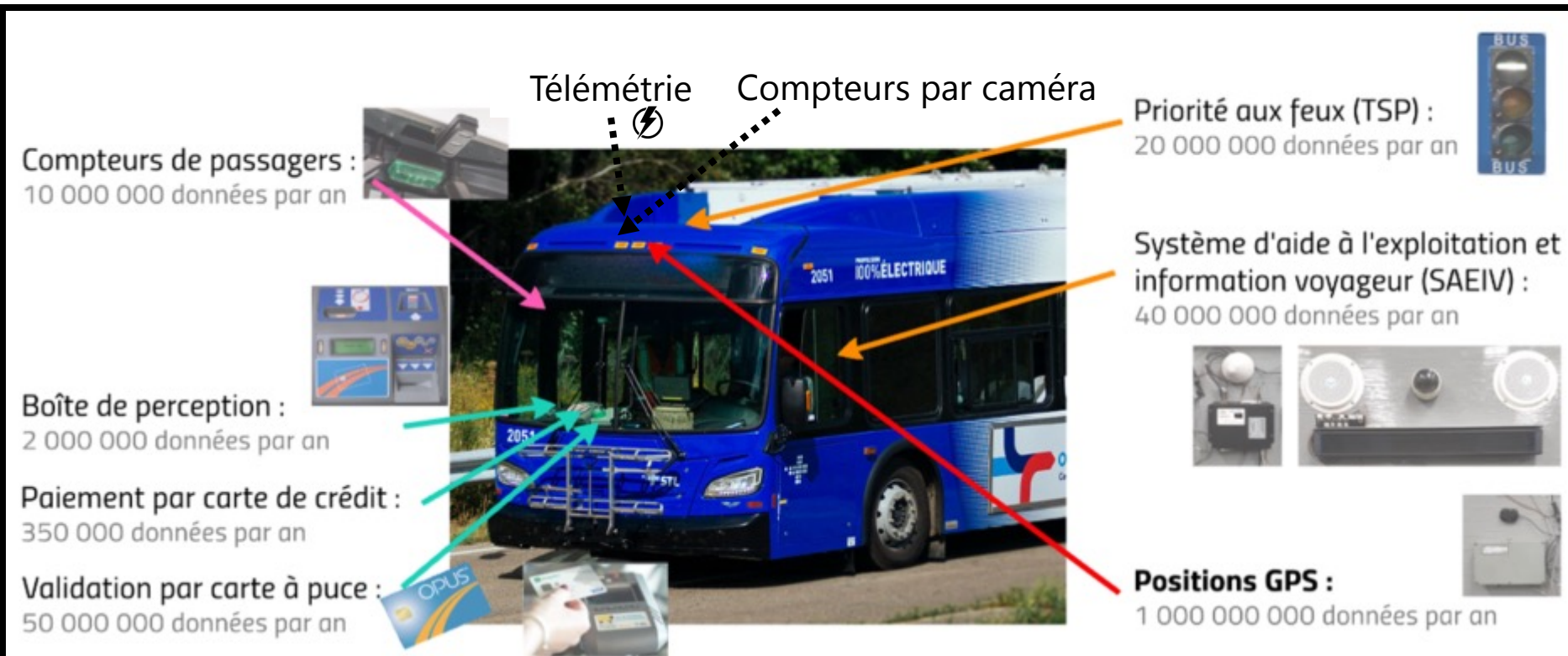
Vincent Dionne, ing., M.Sc.A.  
Chef, Développement et données massives

Symposium « Anonymisation des données »  
29 avril 2024 – Université de Montréal



# Aux origines

## Les diamants bruts



### Autres sources

- Sondages et enquêtes
  - Recensements
- Fournisseurs de données de mobilité (app et cellulaires )
  - Données ouvertes

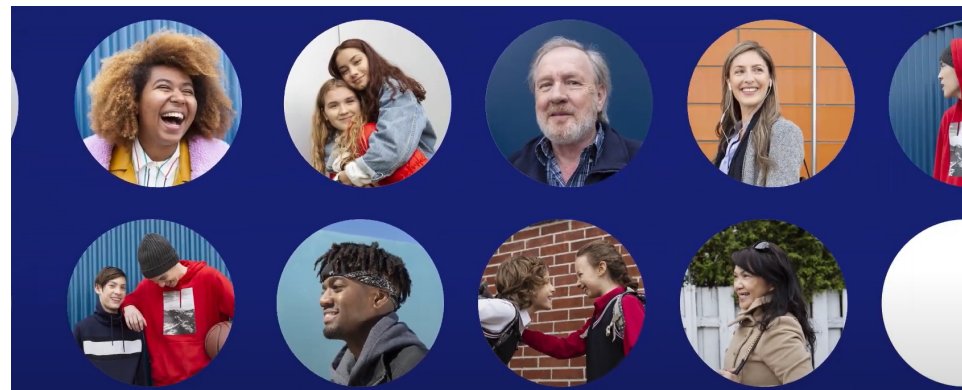
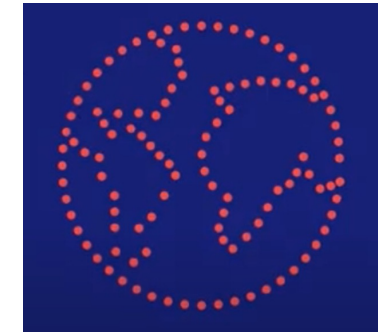
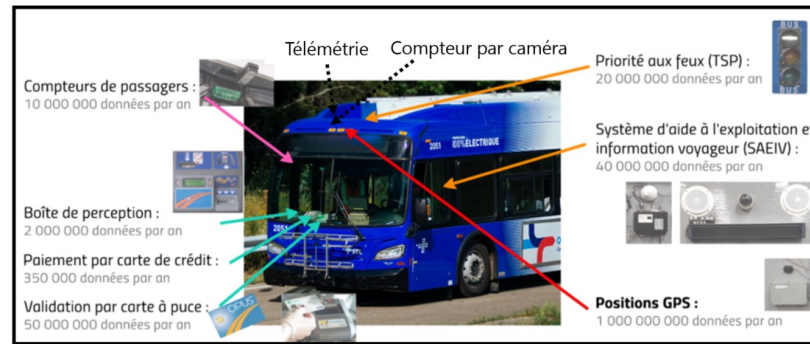
# Des données

## Pour la cause



### hors du

## Exploiter et valoriser cette ressource pour le bien commun



- Optimiser l'offre de service
- Améliorer l'expérience client



# Le diamant cubique



**Comment tirer un maximum de valeur des données de mobilité, tout en s'assurant de masquer les facettes qui touchent la vie privée?**



# Nos données - Loi 25

## 1<sup>er</sup> réflexe

La planification ou l'opération des transports publics\* ne requiert pas de données personnelles des clients.

- ✘ Nom
- ✘ Adresse
- ✘ Dossier médical ou autre
- ✘ Numéro de carte de crédit

Si requis, l'anonymisation est simple, il suffit de retirer/détruire cette information.





# On a seulement besoin de...

Bref, l'approche totalement désagrégée supporte un réseau multimodal qui permet le suivi d'un cheminement de l'origine à la destination tout en préservant l'ensemble des caractéristiques du ménage et de la personne rattachée.

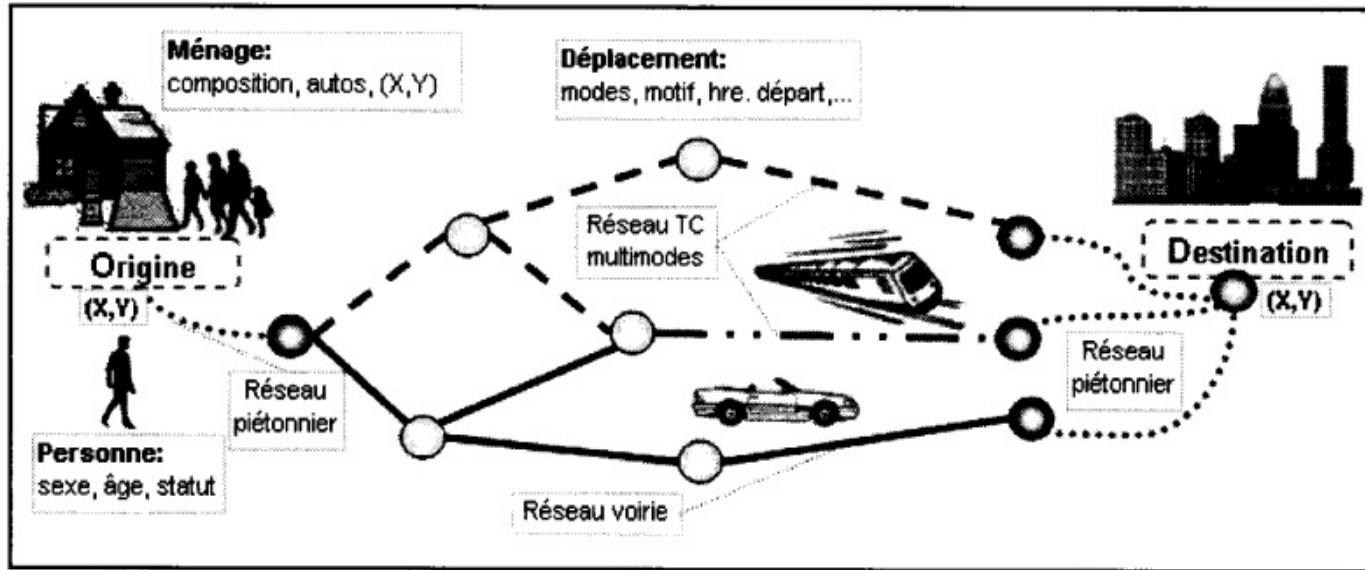


Figure 2.3: Itinéraire totalement désagrégé

## Valeur inestimable pour:

- Planification
- Caractérisation de la demande
- Modélisation
- Mesures d'impact des projets
- Diffusion d'information / marketing

## Plusieurs sources de données:

Enquête OD, Cartes à puce  
GPS...





# Renseignements personnels?

Un renseignement est personnel lorsqu'il concerne une personne physique et qu'il en permet, directement ou indirectement, l'identification. Voici ce qui le caractérise :

- il doit faire connaître quelque chose à quelqu'un;
- il doit avoir un rapport avec une personne physique;
- il doit être susceptible de distinguer cette personne par rapport à une autre ou de reconnaître sa nature

## Renseignements relatifs au travail

Dossier disciplinaire, motifs d'absence, dates de vacances, salaire, évaluation du rendement, heures d'entrée et de sortie liées au lieu de travail, etc.

## Renseignements d'identification

Adresse, numéro de téléphone, sexe, âge, numéro d'assurance sociale, numéro d'assurance maladie, identifiant numérique, etc.

## Renseignements relatifs à la situation sociale ou familiale

Documents qui attestent l'état civil, le fait qu'une personne ait ou non des enfants ou qu'elle reçoive des prestations d'aide sociale ou de chômage, etc



# Dépersonnalisé ou anonymisé?

Un renseignement est dépersonnalisé quand il empêche d'identifier **directement** la personne concernée.

Un renseignement est anonymisé quand il est **raisonnable** de prévoir dans les circonstances qu'en tout temps et de façon **irréversible**, il ne permet plus d'identifier **directement ou indirectement** une personne.







# Beaucoup plus de questions... (EFVP)

## Les questions:

Questions	Sous-questions
Quoi?	<ul style="list-style-type: none"> <li>✓ Quels <b>types de renseignements</b> personnels seront collectés, communiqués, utilisés ou conservés dans le cadre de ce projet?</li> <li>✓ Quelle est la nature de ces renseignements (ex. sont-ils <b>sensibles</b>)?</li> </ul>
Pourquoi?	<ul style="list-style-type: none"> <li>✓ Pourquoi souhaitez-vous recueillir, utiliser, communiquer ou conserver des renseignements personnels?</li> <li>✓ Quelle est la <b>finalité</b> de l'utilisation de ces renseignements dans le cadre de votre projet?</li> <li>✓ En quoi l'accès à ces renseignements est-il nécessaire à l'exercice des fonctions des personnes qui y auront accès?</li> </ul>
Combien?	<ul style="list-style-type: none"> <li>✓ Quelle <b>quantité</b> de renseignements personnels sera impliquée dans votre projet?</li> <li>✓ Combien de personnes seront concernées par votre projet (nombre absolu ou proportion)?</li> <li>✓ Quel est le volume ou l'étendue des renseignements personnels concernés?</li> <li>✓ Quelle est la durée envisagée du projet?</li> <li>✓ Quelle est l'extension géographique projetée?</li> </ul>
Qui?	<ul style="list-style-type: none"> <li>✓ Quelles catégories de personnes auront accès à ces renseignements dans l'organisation ou à l'extérieur (tiers)?</li> </ul>
Comment?	<ul style="list-style-type: none"> <li>✓ Comment ou par quels moyens les renseignements personnels seront-ils collectés, utilisés, communiqués ou conservés au sein (ou à l'extérieur) de l'organisation?</li> <li>✓ Comment l'organisation disposera-t-elle de ces renseignements une fois que la finalité justifiant leur collecte (ou leur communication ou leur utilisation) sera atteinte?</li> <li>✓ Quelle sera la méthode de destruction (ou d'anonymisation) utilisée?</li> </ul>
Où?	<ul style="list-style-type: none"> <li>✓ Où ces renseignements seront-ils <b>répartis</b> et conservés au sein (ou à l'extérieur) de l'organisation?</li> <li>✓ Sur quel(s) type(s) de <b>support</b> et dans quelles conditions seront-ils conservés?</li> </ul>

## LA question:

En a-t-on vraiment besoin? (Utilité)

Peut-on atteindre notre objectif à l'aide de données anonymisées? ou dépersonnalisées?

Trouver l'équilibre

Assurer notre mission ⇔ Niveau d'anonymisation (ou de dépersonnalisation)

# Enquêtes OD ... depuis 1970

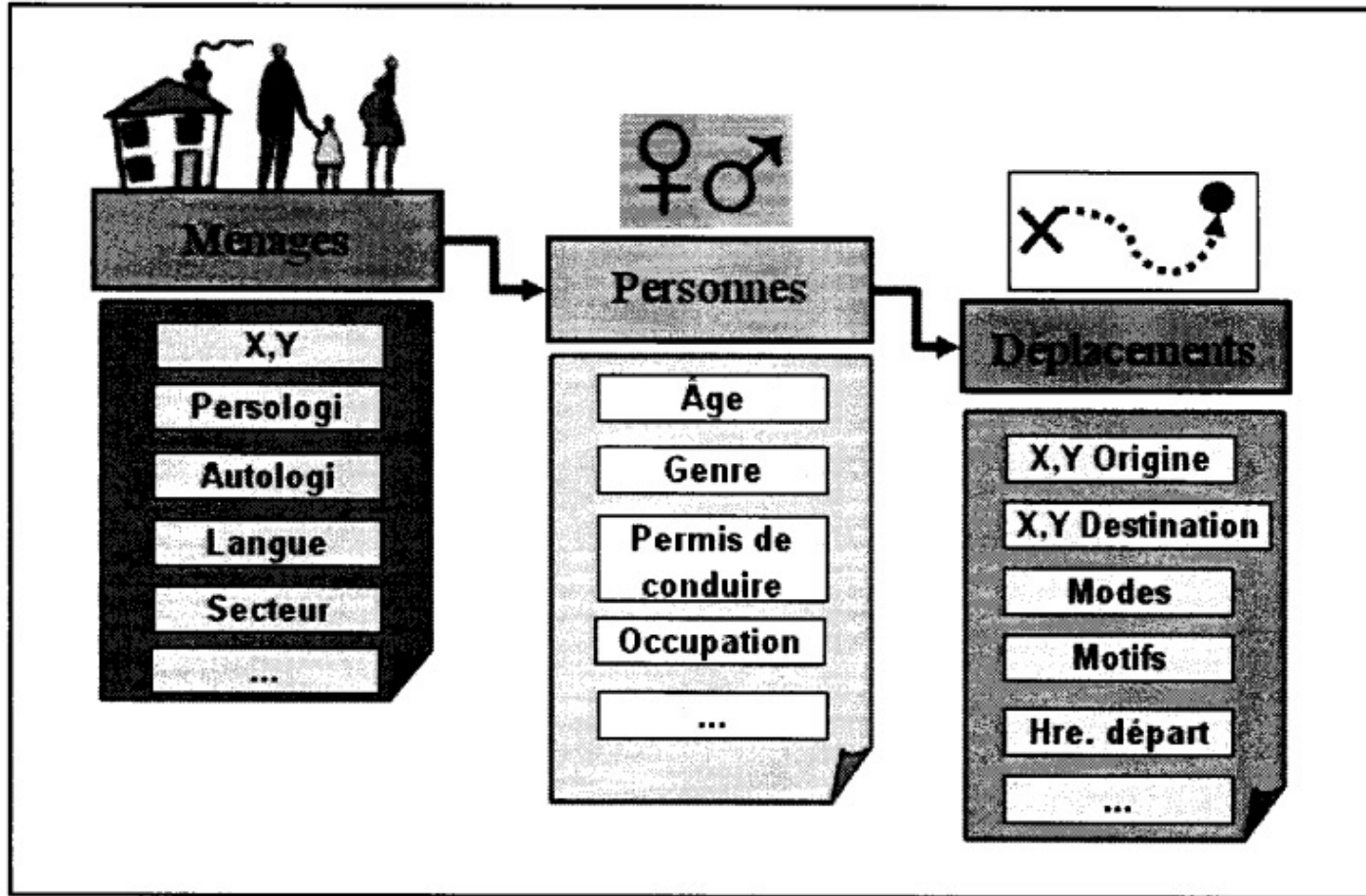


Figure 2.4: Variables de l'enquête Origine-Destination

On peut identifier une personne indirectement via le X Y du domicile.

En a-t-on vraiment besoin?

→ La réponse simple, non.

Possible d'anonymiser ?

→ Oui!

Déplacer aléatoirement du XY?

→ À quel niveau?

Ça dépend...

→ Plusieurs niveaux? Avec droit d'accès différents ?



# Validations de cartes à puce

Un diamant brut pour l'aide à la planification. « L'enquête ultime »

- Embarquement à une heure précise
- Séquence d'embarquements

Une donnée dépersonnalisée et anonyme.

- Estimation du lieu de domicile
- Estimation de l'arrêt de descente
- Estimation du lieu d'activité
- Estimation du motif
- Estimation du segment de clientèle

**La fascination de la descente**

(2006) Moreney et al. Contributions des données de cartes à puces à la compréhension des rythmes d'activités en transport

Les données de cartes à puces ont la capacité de révéler différents aspects de régularité et de variabilité des comportements des usagers du service de transport en commun, mesures impossibles à effectuer avec les données classiques d'enquêtes Origine-Destination

(2007) Trepanier et al. Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System

"The system ties transaction data to detailed transit operational data which are essential for operational planning"

(2008) Chapleau et al. THE ULTIMATE SURVEY FOR TRANSIT PLANNING: COMPLETE INFORMATION WITH SMART CARD DATA AND GIS

"thus provides an extra dimension to analyse activity pattern and travel behaviour"

(2010) Trepanier. L'exploitation des données de cartes à puce à des fins de planification des transports collectifs urbains

La carte à puce, de par la versatilité et la puissance du médium, constitue une des plus intéressantes avancées technologiques

Munizaga, Santiago  
Wang, Oyster, London  
Nassir, Minneapolis

(2014) Robinson et al., Methods for pre-processing smartcard data to improve its quality.

there are many secondary benefits of smartcard systems including better understanding of travel patterns and behaviour of travellers

"allowing multi-day analysis on variability of transit use"

(2015) He et Trépanier. ESTIMATING THE DESTINATION OF UNLINKED TRIPS IN PUBLIC TRANSPORTATION SMART CARD FARE COLLECTION SYSTEMS



# La population synthétique

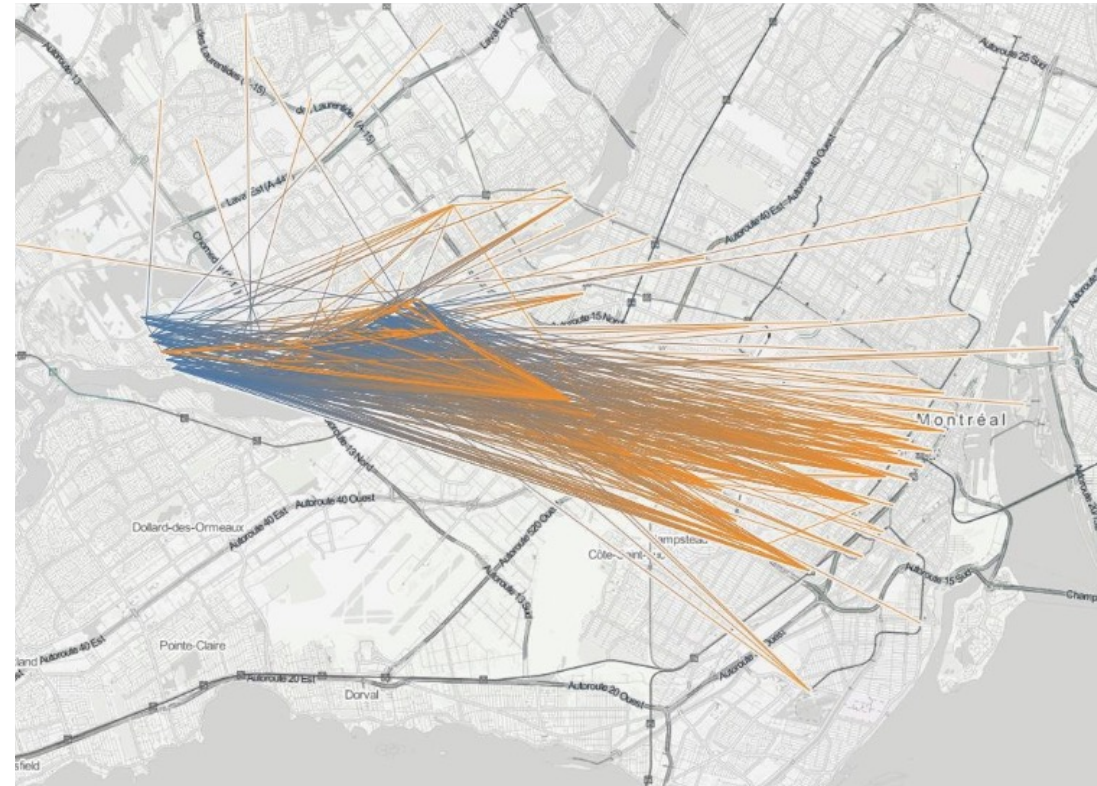
D'un enregistrement réel...

- Identifier un arrêt de montée et un arrêt de descente
- + Synthétiser des activités et des personnes ménages qui reproduit statistiquement des observations provenant d'autres sources de données. (recensement, emploi, étude , etc).

**On s'approche du concept de la population synthétique!**

Plusieurs recherches récentes ou en cours.

- Prometteur, mais haut niveau de complexité
- \$ pour des données en continu



**Origines et destinations des utilisateurs de la ligne 144, le 27 mars 2019**

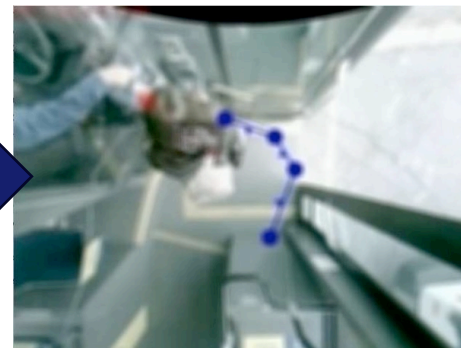
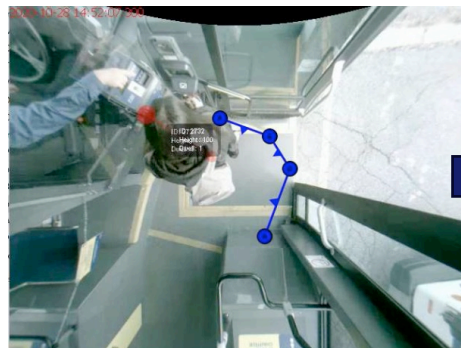
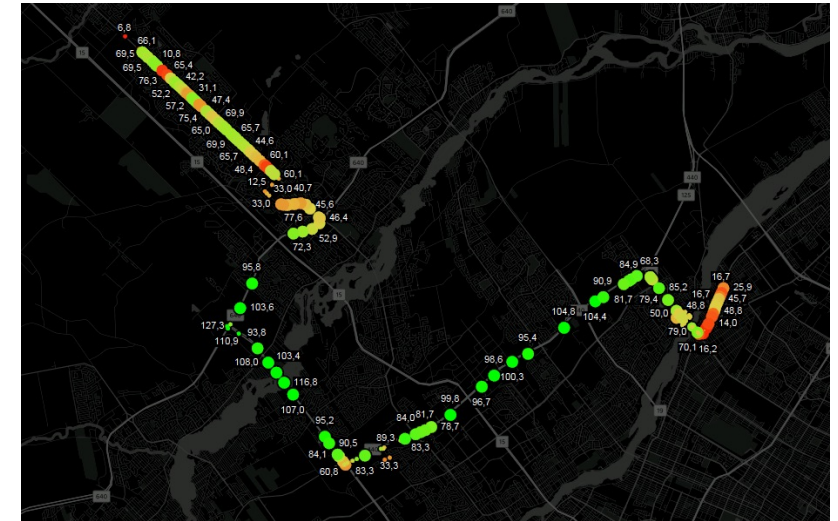


# Les autres données ...

- Trace GPS des véhicules, notamment au transport adapté.
- Des applications mobiles et de la géolocalisation
- Des enquêtes (le cas de Montréal Trajet)
- Technologies ... compteurs de passagers par caméra

Beaucoup mieux outillé avec l'EFVP

- Justification
- Accès restreint
- Dépersonnalisation
- Anonymisation de base





# Quelques réflexions...

Justifier, Justifier, Justifier =>Utilité (valeur ajoutée)

Risque, niveau de sensibilité, ampleur VS utilité VS degré d'anonymisation

Complexité de la technique d'anonymisation (Coût)

Et le potentiel d'utilité futur! Les besoins peuvent évoluer!

→ En attendant... on sécurise!

→ Et on anonymise, mais on conserve une clé d'encryptage sécurisée?





# En conclusion

*Nous avons besoin de lois qui permettent aux Canadiens de profiter des avantages de la technologie, dans l'intérêt public, tout en préservant leur droit à la vie privée. Le moment est venu de montrer aux Canadiens qu'ils peuvent avoir les deux à la fois.*

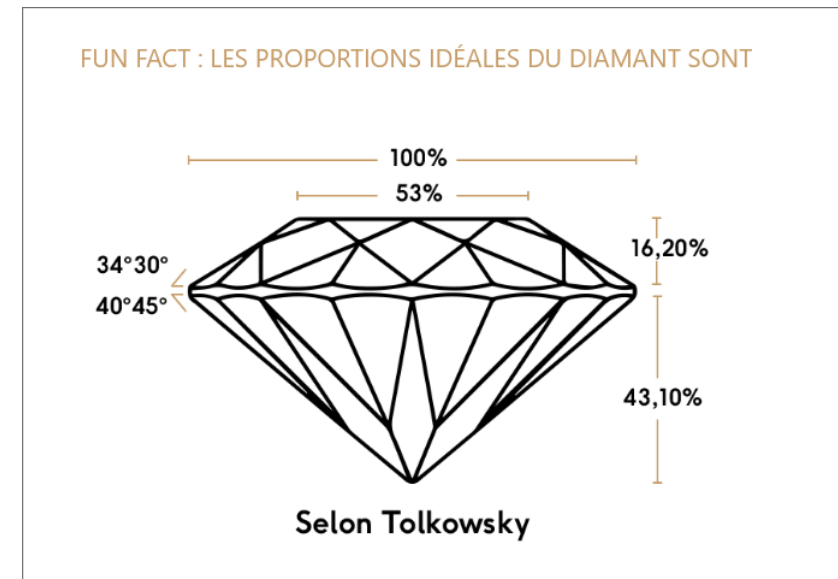
*Données et vie privée : une question de valeur. Daniel Therrien . Commissaire à la protection de la vie privée au Canada – 8 sept. 2020*

Être exemplaire en matière de protection des renseignements personnels.

Les données représentent un actif stratégique qui nous permettent de rester compétitif et efficient => Réaliser notre mission!

+ Contribuer à des projets de recherche

→ Équilibre pour conserver toute la richesse des données sans les dénaturer



# C'est prévu?!



## Utilisation des renseignements dépersonnalisés par un organisme public

L'article 65.1 de la Loi sur l'accès est modifié (...)

Cette exception autorise un organisme public à **utiliser un renseignement personnel** lorsque celui-ci est nécessaire à des fins d'étude, de recherche ou de production de statistiques, dans la mesure où ce dernier est **dépersonnalisé**.

Cette nouvelle exception vise un équilibre entre la protection des renseignements personnels et leur utilisation dans un objectif d'innovation

Un organisme public peut utiliser des renseignements dépersonnalisés à des fins d'étude, de recherche ou de production de statistiques dans le but :

- d'en **évaluer la performance** ou de mesurer la qualité des produits et des services offerts, **dans une perspective d'amélioration continue**;
- de **servir l'intérêt public**, par exemple en menant une recherche sur un sujet précis en lien avec sa mission afin d'élaborer un nouveau programme ou service.



# Questions?

**Merci!**

- La protection de la vie privé ...
- c'est **notre** fougère !

