



Statistics
Canada

Statistique
Canada

Accès au microdonnées d'entreprise de Statistique Canada

www.statcan.gc.ca



CANADA 150

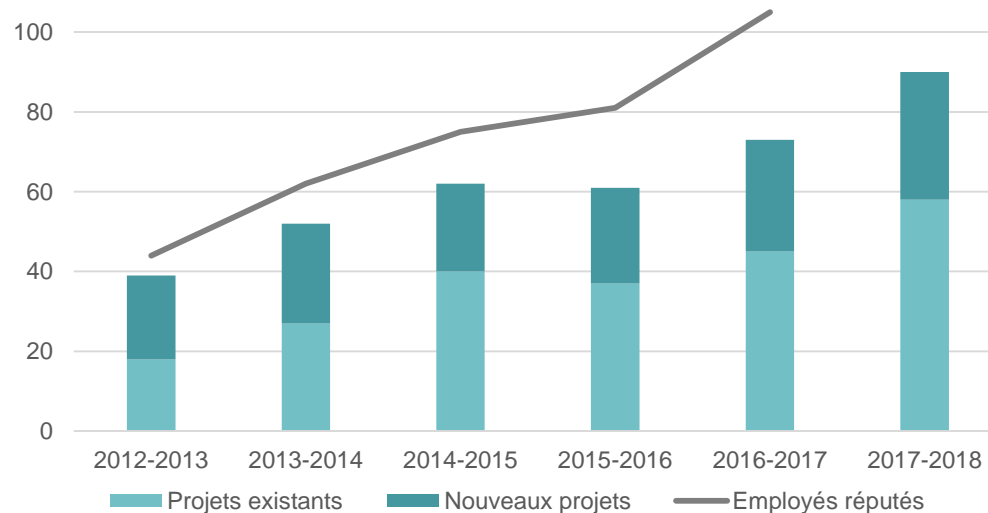
L'histoire du Canada
racontée en chiffre

Lydia Couture
Gestionnaire de programme

26 février 2018

Survol du CDRE

- Établi en 2011 au bureau central de Statistique Canada afin d'offrir un accès sécurisé au microdonnées d'entreprises pour des projets de recherche analytiques approuvés
- Institutions desservies:
 - Environ 30 institutions académiques canadiennes et étrangères
 - 16 agences et ministères fédéraux, gouvernements provinciaux et organismes sans but lucratif





Améliorer l'accès aux microdonnées d'entreprises

- 1) Projet pilote
- 2) Données synthétiques
 - Atelier du CDRE 2016 : intro aux données synthétiques sophistiquées et leur méthodologie - Lars Vilhuber
- 3) Collaboration avec le Partenariat Productivité
- 4) Améliorer le processus de demande d'accès

Projet pilote

- Statistique Canada annoncera un appel de proposition pour utiliser les microdonnées d'entreprises synthétiques dans les Centres de données de recherche (CDRs)
- Des projets analytiques approuvés se concentrant sur l'analyse multivariée seront menés dans les CDRs en 2018-2019
 - Les chercheurs développeront leurs modèles en utilisant les données synthétiques dans les CDRs
 - Les résultats sur les données synthétiques ne seront pas publiés
 - Les programmes seront envoyés au CDRE par les analystes des CDRs pour être roulés sur les données réelles
 - Les résultats finaux seront publiés par courriel par le CDRE



Projets de données synthétiques

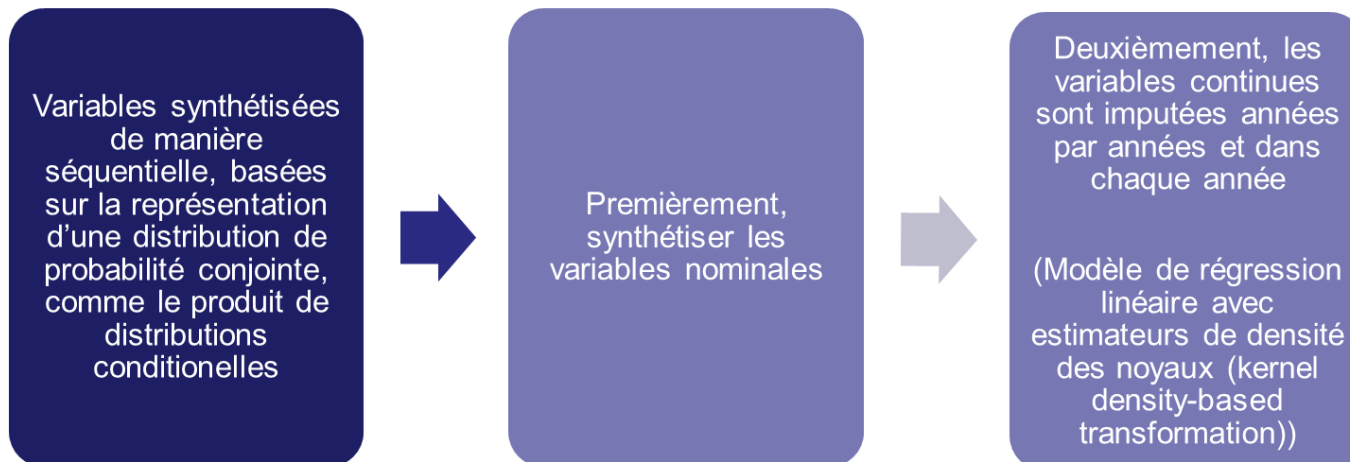
- Développement de données synthétiques plus sophistiquées pour faciliter l'accès à l'extérieur du bureau central de
 - Base de données synthétiques d'entreprises longitudinale
 - Fichier à divulgation limitée de l'Enquête sur le financement et la croissance des petites et moyennes entreprises (EFCPME) couplée aux données fiscales de l'ARC

Base de données synthétiques d'entreprises longitudinale

- Version complètement synthétique du fichier du Programme d'analyse longitudinale de l'emploi (PALE) (2001 to 2015)
- Le PALE inclut les entreprises avec salariés
 - Feuille T4 sur la rémunération payée
 - Registre des entreprises
 - Enquête sur l'emploi, la rémunération, et les heures de travail (EERH)
 - Variables : entrées, sorties, paie, unité moyenne de main d'oeuvre (paie divisée par la moyenne annuelle de rémunération de l'EERH), et industrie
 - Utilisé pour étudier la dynamique des entreprises et de l'emploi (par exemple entrée et sortie d'entreprises, création et destruction d'emplois, survie et croissance d'entreprises, firmes à haute croissance, transitions)

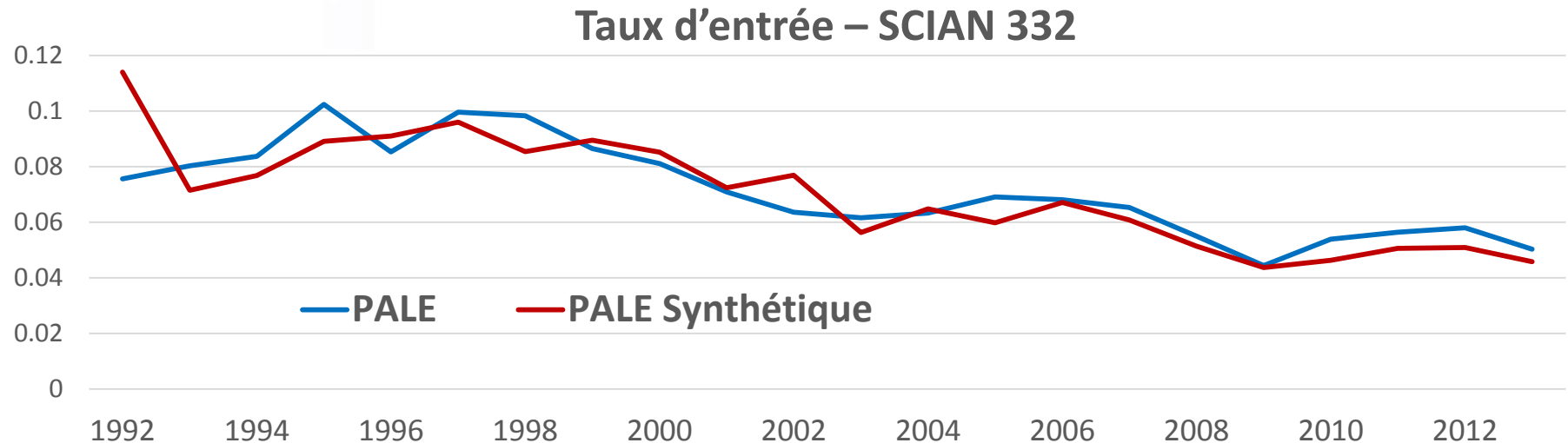
Méthodologie

- Suivant l'approche développée au U.S. Census Bureau
 - Colaboration avec la direction de la méthodologie, Cornell University, Partenariat Productivité
- Toutes les variables sont modélisées et ensuite simulées par industrie
 - Aucun risque de ré-identification directe, toutes les observations étant simulées
 - Les données synthétiques peuvent être utilisées pour produire des agrégats détaillés semblables aux originaux, incluant ceux qui sont considérés confidentiels
- Création de données synthétiques en remplaçant les données réelles avec des sélections d'un modèle ajusté aux données d'origine.
- Stratégie se rapprochant de l'imputation multiple





Validité analytique des données synthétiques



- Magnitude et tendance similaire pour les taux d'entrées, de sorties, l'emploi et la paie au cours de la période



Validité analytique des données synthétiques

- Teste la validité en utilisant une analyse de régression

$$emp_i = \alpha + \beta emp_{i,t-1} + \delta pay_i + \vartheta age_i + \theta Naics_i + \epsilon_i$$

Variable	PALE	Erreur Standard	PALE Synthétique	Erreur Standard
Constante	-9.08	0.05	-6.83	0.06
Emp_t-1	0.12	0.00	0.30	0.01
Paie	0.86	0.00	0.65	0.01
Age 4-7	-0.04	0.01	-0.05	0.01
Age 8 -10	-0.03	0.01	-0.04	0.01
Age 10-22	-0.02	0.01	-0.05	0.01
Age 23 et plus	-0.02	0.01	-0.04	0.01

Confidentialité

- Pas de risque direct de ré-identification
- Ventilation par industrie limitée pour le fichier synthétique
- L'utilisation du fichier principal pour obtenir les résultats finaux permettra d'utiliser des niveaux d'industrie et de géographie plus détaillés, ainsi que des variables du fichier T2-PALE



Partenariat Productivité

- Équipe d'experts des secteurs académique, privé, et public
- Réduire les obstacles à l'accès aux données d'entreprises pour les chercheurs académiques et les étudiants en offrant du financement pour des projets de recherche sur la productivité
- Chercheur de liaison de Statistique Canada : Beryl Li

<https://productivitypartnership.ca>



Partenariat Productivité (2)

- Partenaire important pour le développement de données synthétiques
 - Post-doc: M. Jahangir Alam
- Collaboration et appui pour mieux définir les étapes requises pour les demandes d'accès au CDRE



Indicateurs

Durée moyenne des demandes d'accès au CDRE

- Projets de ministères fédéraux: 3 mois
 - 2 mois, si l'on exclut les projets nécessitant un couplage
- Projets académiques: 6 mois
- Différence: principalement due aux attestations de sécurité et aux revues par les pairs

Site internet – révisions à venir en 2018

- Étapes requises pour les demandes d'accès au CDRE bien définies
- Révisions des exigences pour les propositions de recherche
- Information sur le processus d'attestation de sécurité
- Section sur les conflits d'intérêt

Prochaines étapes

- Développer des données synthétiques additionnelles :
 - Enquêtes d'entreprises : SFGSME, etc.
 - Données fiscales
- Élargir l'utilisation des données synthétiques de microdonnées d'entreprises et leur accès à l'extérieur du CDRE et du QG de StatCan à court terme
- À venir : Webinar-Séminaire sur le pilote



Thank you! Merci !

Courriel du CDRE :
statcan.cder-cdre.statcan@canada.ca