

Statistique Canada, Environnement de fichiers couplés, une vue d'ensemble

(Novembre 2023)



Statistique
Canada

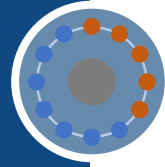
Statistics
Canada

Canada





Présentation Aperçu



**Environnements de fichiers couplés à statistique Canada :
(EFC-E, ECDS et ECDO)**



**Comment combler les lacunes de données en utilisant les
Environnements de données couples de statistique Canada**



Processus d'accès aux données couplées



Environnements de fichiers couplés à Statistique Canada



Qu'est -ce qu'un Environnement de fichiers couplés

- **Définition:**

L'environnement de couplage de données fournit de l'information sur les entreprises, les particuliers et les questions économiques en facilitant la création d'ensembles de données combinant des données d'enquêtes et des données administratives.

Le couplage de données prend les renseignements personnels sur les candidats d'un ensemble de données et les relie à d'autres ensembles de données sur les personnes ou les entreprises correspondantes afin de permettre l'analyse, la comparaison et l'évaluation intersectionnelle statistique anonymisée.

- **Avantages:**

Comblent les lacunes en matière de données ; Faible coût ; Faible charge de réponse ; Couverture et la mise-à-jour continue.

- **Cadre juridique :**

Mandat ; Approbations; Données administratives et confidentialité

- **Objective:**

Répondre aux besoins analytiques des décideurs et des chercheurs à l'aide de microdonnées



Environnement de fichiers couplés à statistique Canada

- **Environnement de fichiers couplables entreprises (EFC-E)**

Objective : Fournir des microdonnées, des informations démographiques et des caractéristiques sur les entreprises, l'entrepreneuriat et la propriété **au niveau de l'entreprise**.

Lien page web: <https://www.statcan.gc.ca/fr/aperçu/statcan/efc>

- **Environnement de couplage de données sociales (ECDS)**

Objective : Permettre le couplage **au niveau de l'individu** entre de multiples ensembles de données dans le domaine social afin de faciliter la recherche statistique sociale et économique pancanadienne.

Lien page web: <https://www.statcan.gc.ca/fr/ecds/aperçu>

- **Environnement de couplage de données ouvertes (ECDO)**

Objective: Fournir au public des microdonnées ouvertes sur une variété de services, commodités et d'infrastructures **au niveau de l'emplacement et de la géographie**.

Lien page web: <https://www.statcan.gc.ca/fr/ecdo>



Avantages des fichiers couplés

- Comble les lacunes en matière de données
- Réduit le fardeau de réponse
- Réduit les coûts de couplage d'enregistrements
- Cadre de travail mise-à-jour en permanence, avec de nouvelles données et sources
- Permet de couvrir l'ensemble de la population des entreprises et des individus sur de longues périodes



Comment combler les lacunes en matière de données à l'aide d'environnements de fichiers couplés de Statistique Canada



Il y a une demande croissante pour de l'information nécessitant à la fois des caractéristiques d'entreprises, individuelles et géographiques

□ Comment réconcilier des informations provenant de plusieurs sources de données:

- Requièrè un identifiant commun (Clé)
- Requièrè un environnement cohèrant (Une mètrique commune -Enterprise; Individu; situation géographique...)

□ Développement de nouveaux indicateurs:

- Caractèristiques de la propriété ; Qualité de vie ; équitè; diversité et indicateurs d'inclusion; nouvelles taxonomies et plus...





Environnement de Fichiers couplables entreprises (EFC-E)



Environnement de fichiers couplables - Entreprises (EFC-E)

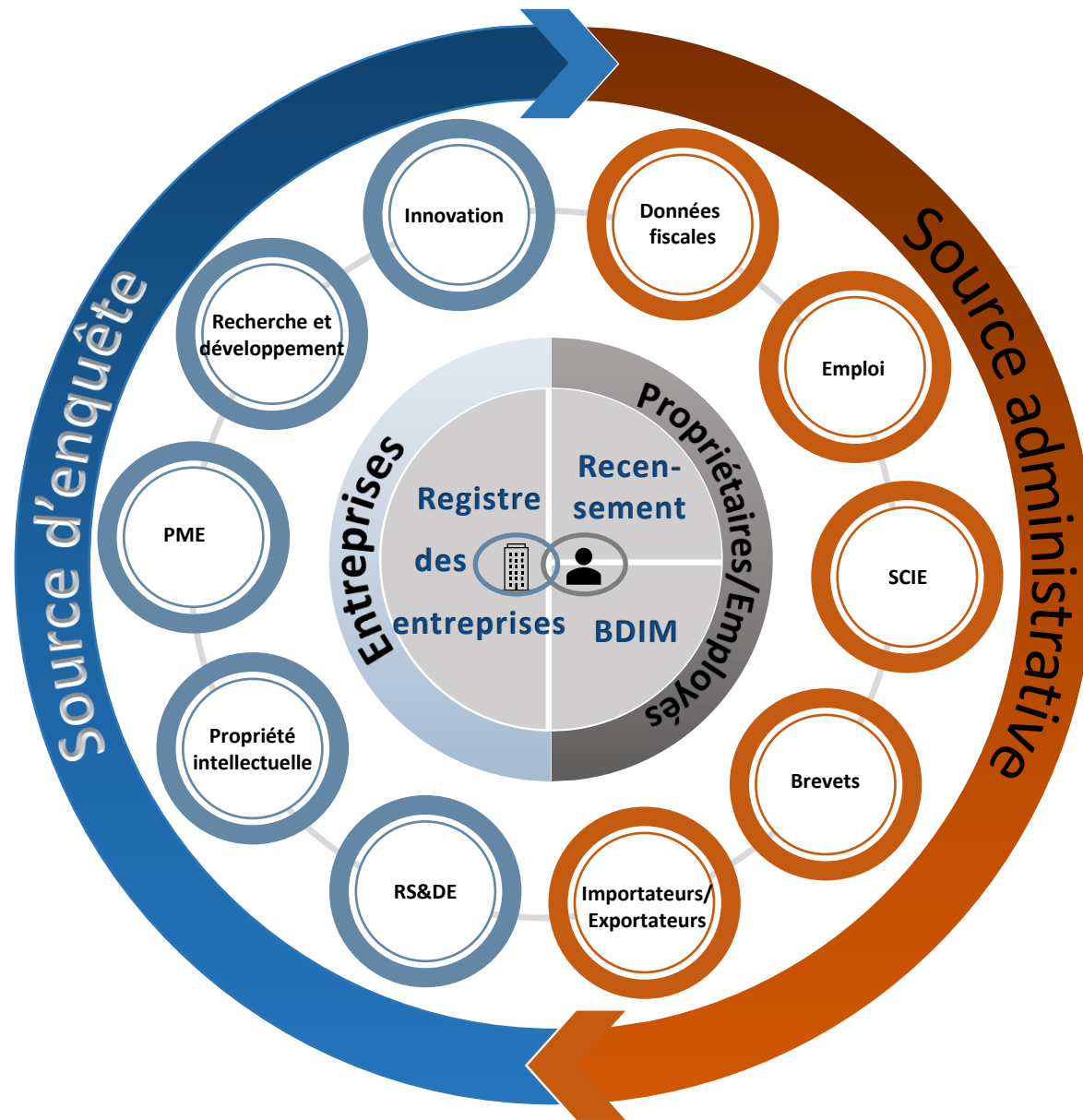
Couvre toutes les entreprises canadiennes

Plus de 30 sources de données

Facilite la création de fichiers de données administratives et d'enquêtes liées afin d'informer sur les enjeux économiques des entreprises

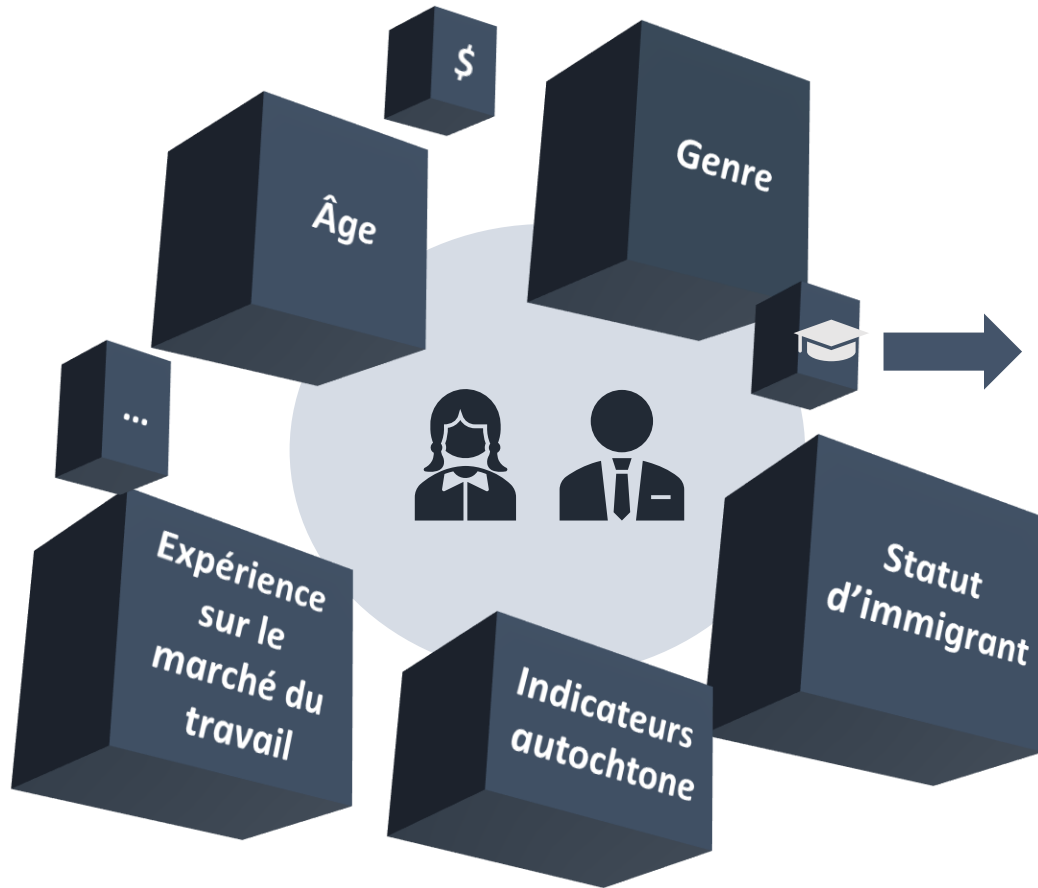
Opportunités analytiques :

- Évaluer la performance des entreprises
- Évaluation d'impact sur les programmes
- Tableaux personnalisés
- Indicateurs sur la diversité et le genre



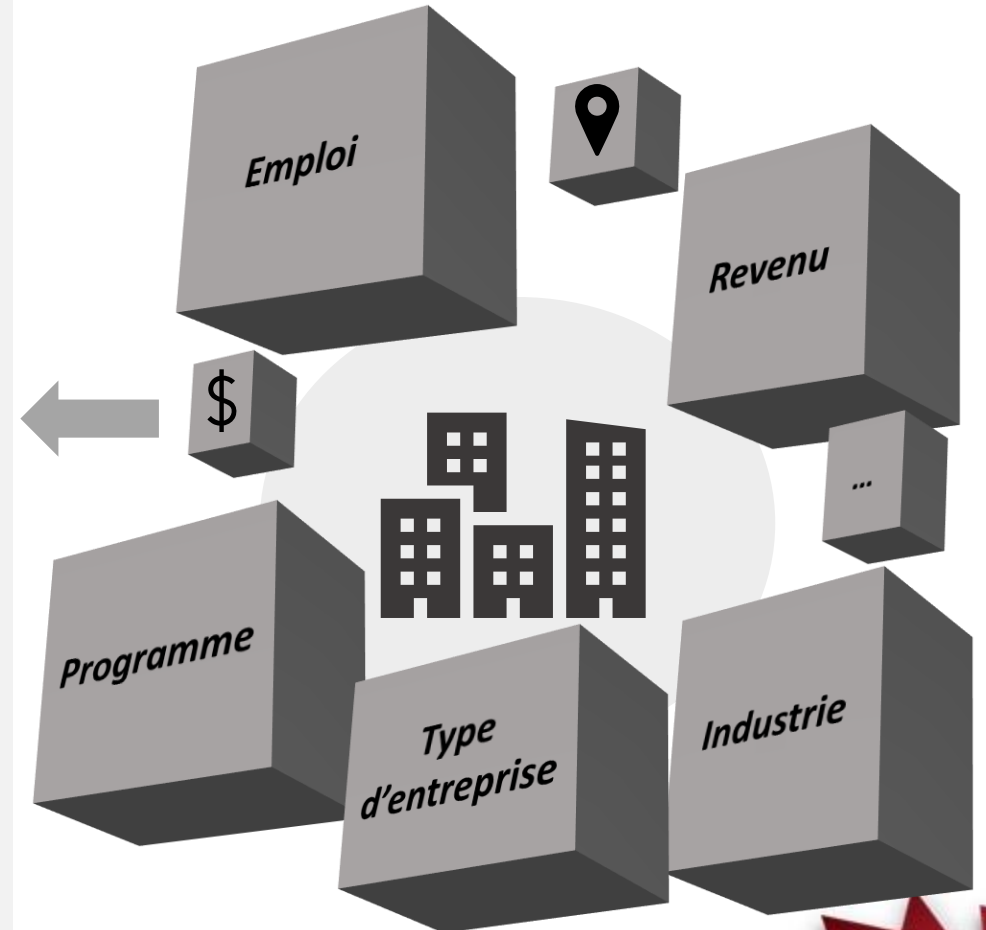
Quelles données sont disponibles pour les perspectives analytiques et la prise de décision?

Caractéristiques des individus



Comprendre non seulement l'impact des programmes sur les entreprises, mais aussi analyser les caractéristiques individuelles des propriétaires et de la main-d'œuvre

Caractéristiques des entreprises



EFC-E : Nouveaux couplages pour répondre aux lacunes de données

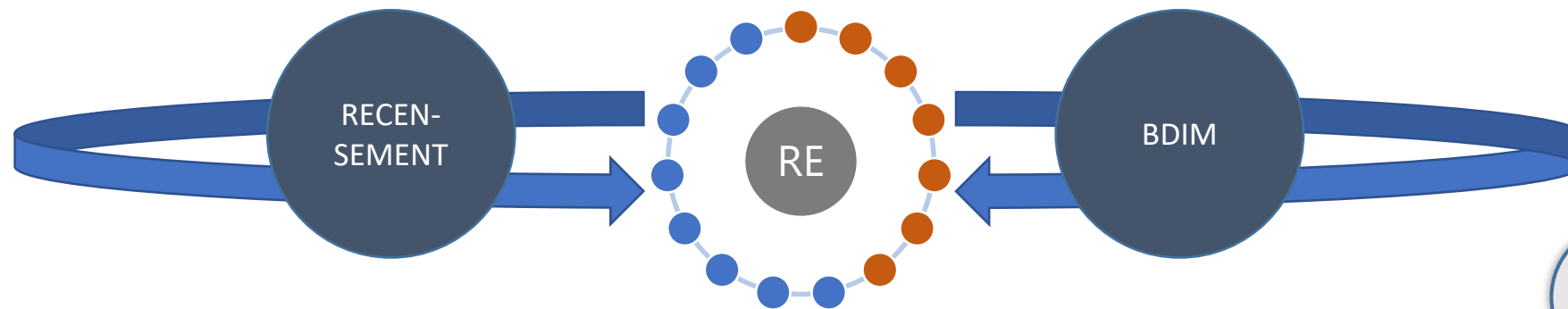
T2
Annexe 50

TCOP

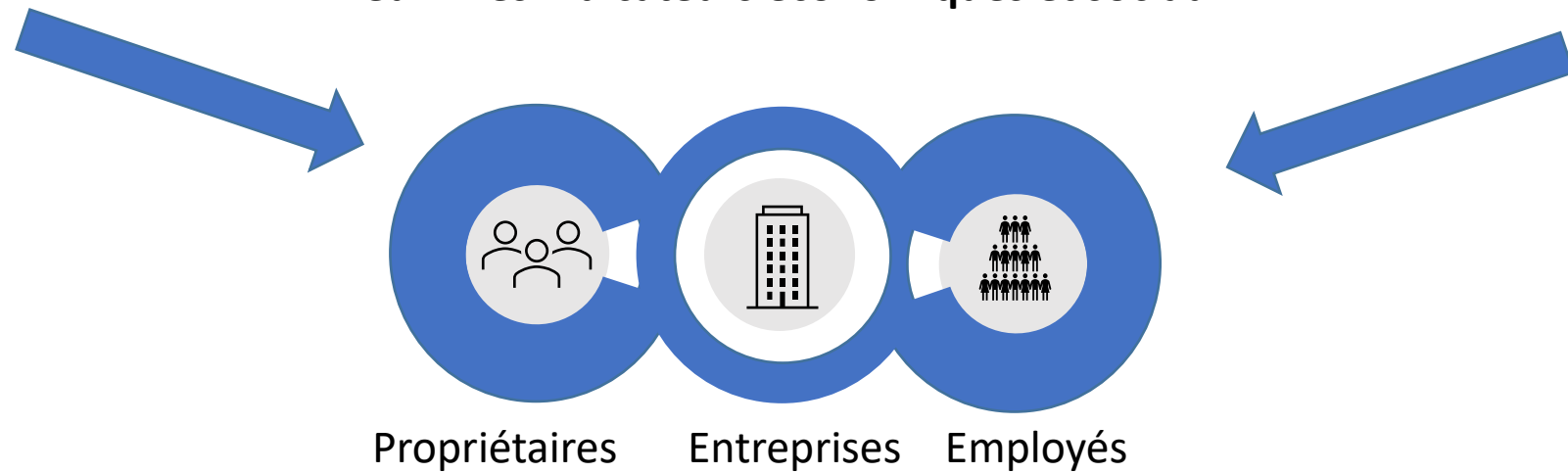
T1FD

BDA

Recensement & BDIM



Réunir les indicateurs économiques et sociaux



T4

BDA

Recensement & BDIM

Annexe 50

Liste des actionnaires détenant au moins 10 % des actions pour toutes les entreprises privées incorporées

TCOP

Fichiers de NEs et coordonnées sur les propriétaires (inclut les entreprises non-incorporées)

T1FD

Propriétaires d'entreprises non-incorporées avec revenu T1

BDA

Données administratives sur les individus (contribuables)

T4

Données administratives sur la rémunération et les avantages imposables versés aux employés



Produits EFC_E – Utilisation de EFC_E pour l'analyse du genre et de la diversité Autres résultats relâchés

- Analyse des propriétaires noirs pour les entreprises qui reçoivent du soutien à la croissance et à l'innovation en entreprise du fédéral (SCIE), 2020
- Un portrait des minorités de langue officielle (MLO) qui détiennent des entreprises recevant du soutien à la croissance et l'innovation du fédéral (SCIE), 2020

SOUTIEN FÉDÉRAL À LA CROISSANCE ET À L'INNOVATION POUR LES ENTREPRISES APPARTENANT AUX MINORITÉS DE LANGUE OFFICIELLE, 2020

En 2020, on comptait au Canada **344 225** entreprises appartenant à des minorités de langue officielle¹ (MLO), ce qui représente **6,3 %** de toutes les entreprises^{2,3} au pays. Parmi celles-ci, **1 775** (0,5 %) ont reçu **437 595 990 \$** (12,7 %) de l'ensemble du soutien fédéral à la croissance et à l'innovation en entreprise.

Comparativement à celles n'ayant pas obtenu de soutien, les entreprises des MLO qui ont obtenu du soutien⁴ étaient :

- 10,1 fois** plus susceptibles d'être des entreprises de taille moyenne (100 à 499 employés)
- 33,1 fois** plus susceptibles d'être de grandes entreprises (500 employés ou plus)
- 4,8 fois** plus susceptibles de déclarer des revenus d'au moins 2 millions de dollars
- 1,9 fois** moins susceptibles d'appartenir à des femmes

Près de la moitié (**48,7 %**) des entreprises appartenant à des MLO qui ont obtenu du soutien étaient concentrées dans le secteur des services professionnels, scientifiques et techniques et dans celui de la fabrication, comparativement à **14,9 %** des entreprises appartenant à des MLO qui n'ont pas obtenu de soutien.

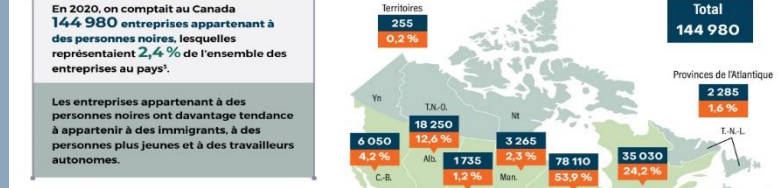
Comparativement aux entreprises appartenant aux MLO qui n'ont pas obtenu de soutien, celles en ayant obtenu étaient :

- 10,5 fois** plus nombreuses dans la fabrication
- 8,6 fois** plus nombreuses dans l'extraction minière, l'exploitation en carrière, et l'extraction de pétrole et de gaz
- 3,2 fois** plus nombreuses dans l'industrie de l'information et l'industrie culturelle
- 22,6 fois** plus nombreuses dans les services immobiliers et les services de location et de bail

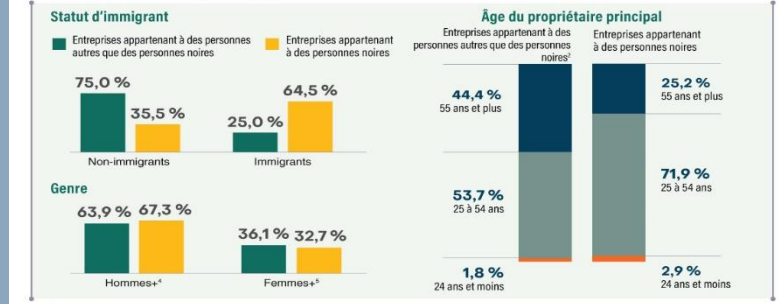
Notes :
1. Dans cette infographie, les entreprises appartenant à des minorités de langue officielle (MLO) désignent les entreprises appartenant à des personnes dont l'anglais est la première langue officielle parlée (PLOP) au Québec, et celles appartenant à des personnes dont le français est la PLOP ailleurs au Canada. Les entreprises dont la PLOP du propriétaire est l'anglais et le français ont été réparties également entre les catégories « anglais seulement » et « français seulement ».
2. Les propriétaires d'entreprise comprennent les propriétaires principaux d'entreprises nouvelles, d'entreprises actives ou d'entreprises inactives, mais existantes, qui exercent des activités au Canada et qui sont couplés à un enregistrement du Recensement de 2021, y compris ceux ayant un revenu provenant d'un travail autonome.
3. Le propriétaire principal d'une entreprise désigne la personne la plus susceptible d'en être le propriétaire (ou le gestionnaire) parmi tous les propriétaires et titulaires de charge connus. Si une entreprise compte plus d'un propriétaire, le plus grand part du revenu de l'entreprise détermine qui en est le propriétaire principal.
4. Le propriétaire principal d'une entreprise ayant obtenu du soutien est le bénéficiaire ultime de soutien fédéral à la croissance et à l'innovation en entreprise.

ENTREPRISES APPARTENANT À DES PERSONNES NOIRES au Canada, 2020

→ Répartition des entreprises appartenant à des personnes noires¹ par région²



→ Pourcentage d'entreprises appartenant à des personnes noires et d'entreprises appartenant à des personnes autres que des personnes noires, selon les caractéristiques du propriétaire



→ Pourcentage d'entreprises appartenant à des personnes noires et d'entreprises appartenant à des personnes autres que des personnes noires, selon les caractéristiques de l'entreprise



Notes :
1. Les propriétaires d'entreprise comprennent les propriétaires principaux d'entreprises actives qui exercent des activités au Canada, y compris ceux ayant un revenu provenant d'un travail autonome. Le propriétaire principal d'une entreprise est la personne la plus susceptible d'en être le propriétaire (ou le gestionnaire) parmi tous les propriétaires et titulaires de charge connus. Ce propriétaire principal est la personne qui obtient la plus grande part de la propriété.
2. Si de multiples propriétaires existent et si c'est le cas, le propriétaire principal est celui avec le plus grand revenu.
3. Le total des pourcentages ne correspond pas toujours à 100 % en raison de l'arrondissement.
4. Les propriétaires qui n'ont pu être couplés au recensement ont été ignorés.
5. Cette catégorie comprend les hommes et peut également inclure certaines personnes non binaires.
6. Cette catégorie comprend les femmes et peut également inclure certaines personnes non binaires.
Sources : Statistique Canada, Environnement de fichiers couplables – Entreprises, 2020, Soutien de la croissance et de l'innovation en entreprise, 2020, Recensement de la population, 2021, 2006, 2016 et 2021, enquête nationale auprès des entreprises, 2011, et Base de données longitudinales sur l'immigration, 2020.

Environnement de couplage de données sociales (ECDS)



Environnement de couplage de données sociales (ECDS)

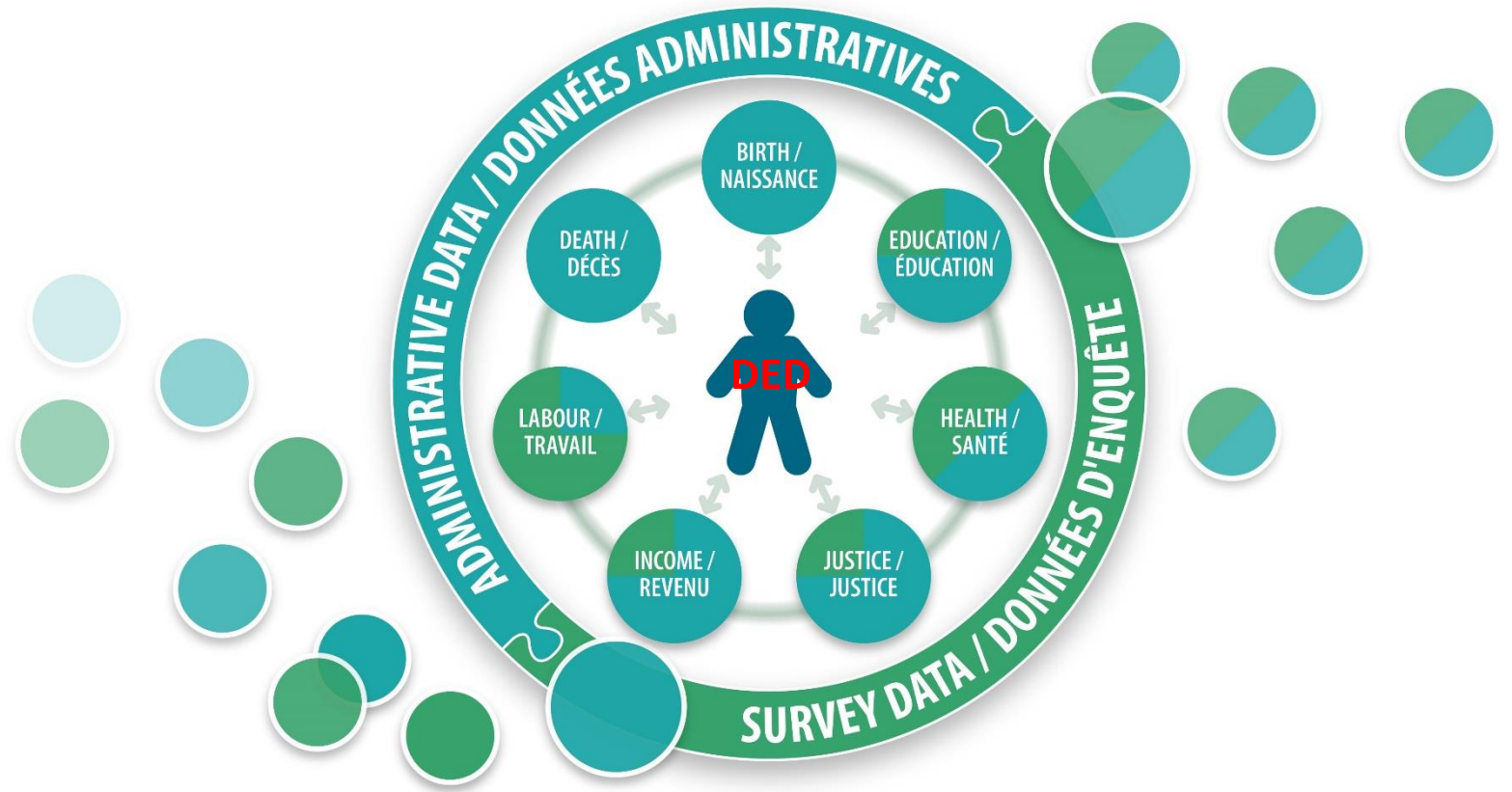
Couverture: base de données historiques de Canadiens et Canadiennes construite en combinant des fichiers de données administratives (dépôt d'enregistrements dérivés DED)

Plus de 160 sources de données déjà appariées

Facilite la création d'ensembles de données couplées à des fins d'analyse sociale.

Opportunités analytiques:

- Utilise des données administratives et d'enquêtes existantes pour répondre à d'importantes questions de recherche et éclairer les politiques socio-économiques.
- Permet de générer de nouvelles informations sans collecte de données supplémentaires



Potentiel de l'ECDS

Réduit les coûts d'un projet de recherche.

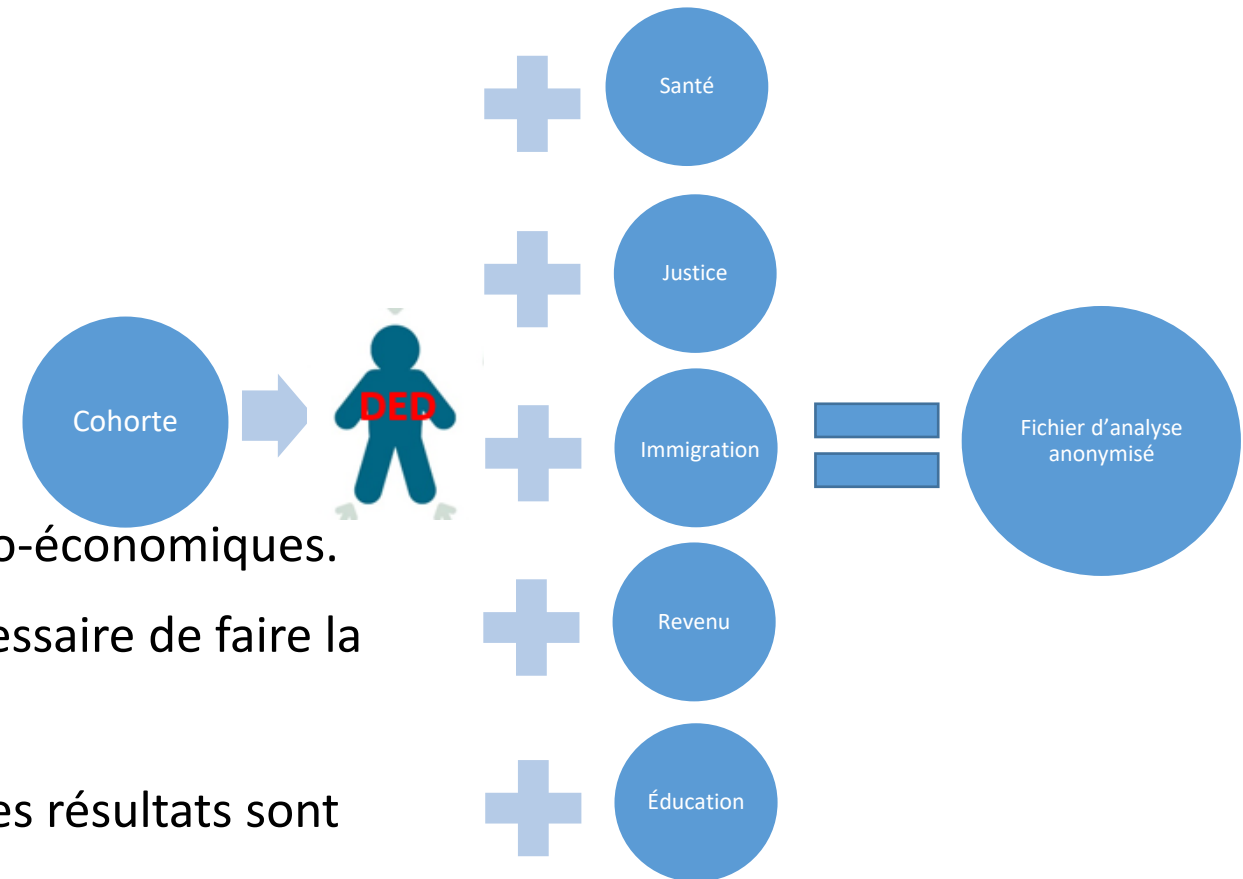
Approche d'appariement de données standardisée

Réduit le fardeau de réponse.

Utilise des données administratives et d'enquêtes existantes pour répondre à d'importantes questions de recherche et éclairer les politiques socio-économiques.

Génère de nouvelles informations sans qu'il soit nécessaire de faire la collecte de données supplémentaires.

Les fichiers de l'ECDS sont couplés une seule fois et les résultats sont réutilisés pour créer les fichiers d'analyse.



L'ECDS dans les données à Statistique Canada

Base de données longitudinales sur les immigrants (BDIM)

Plateforme longitudinale entre l'éducation et le marché du travail (PLEMT)

Cohortes santé et environnement du recensement canadien (CSERCan)

Base de données relationnelle sur la justice pénale (BRJP)

Recensement de la population

Enquête sur la sécurité financière (ESF)

Enquête sur la santé dans les collectivités canadiennes (ESCC)

Enquête sociale générale (ESG)

Enquête sociale canadienne (ESC)

[DRD linkage status - Social Data Linkage Environment \(SDLE\) - Environnement de couplage de données sociales \(ECDS\) - Statcan Confluence](#)

Projets d'appariement en collaboration avec l'ECDS

Résultats observés au niveau de la justice

- Profils de contacts répétés avec les différents systèmes de justice pénale provinciaux (ON, SASK, etc.)
- Projet de mise en liberté sous caution et de détention provisoire en Ontario

Résultats observés au niveau de la santé de la population immigrante

- Projet canadien d'appariement avec les traitements contre le cancer
- Appariement de l'Enquête sur la santé dans les collectivités canadiennes et la base de données longitudinales sur les immigrants.

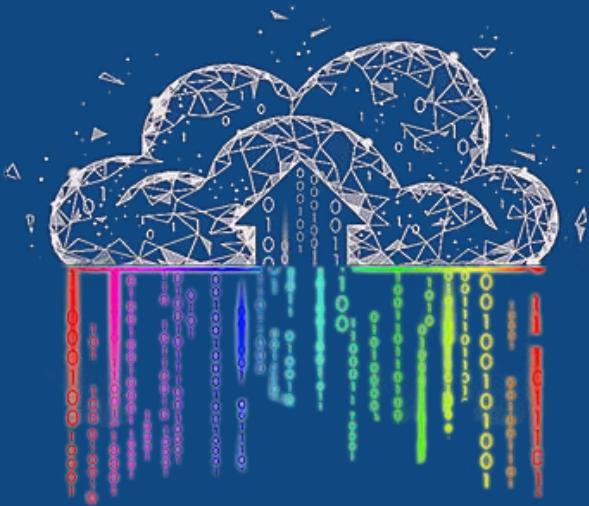
Résultats observés pour des populations spécifiques

- Étude sur les opioïdes en Colombie-Britannique pour la municipalité de Surrey
- Étude sur le cancer et la mortalité dans les forces armées canadiennes

Résultats observés au niveau de l'éducation

- Programme canadien de prêts aux étudiants et prêts à l'apprentissage





Environnement de couplage de données ouvertes (ECDO)



Environnement de couplage de données ouvertes (ECDO)

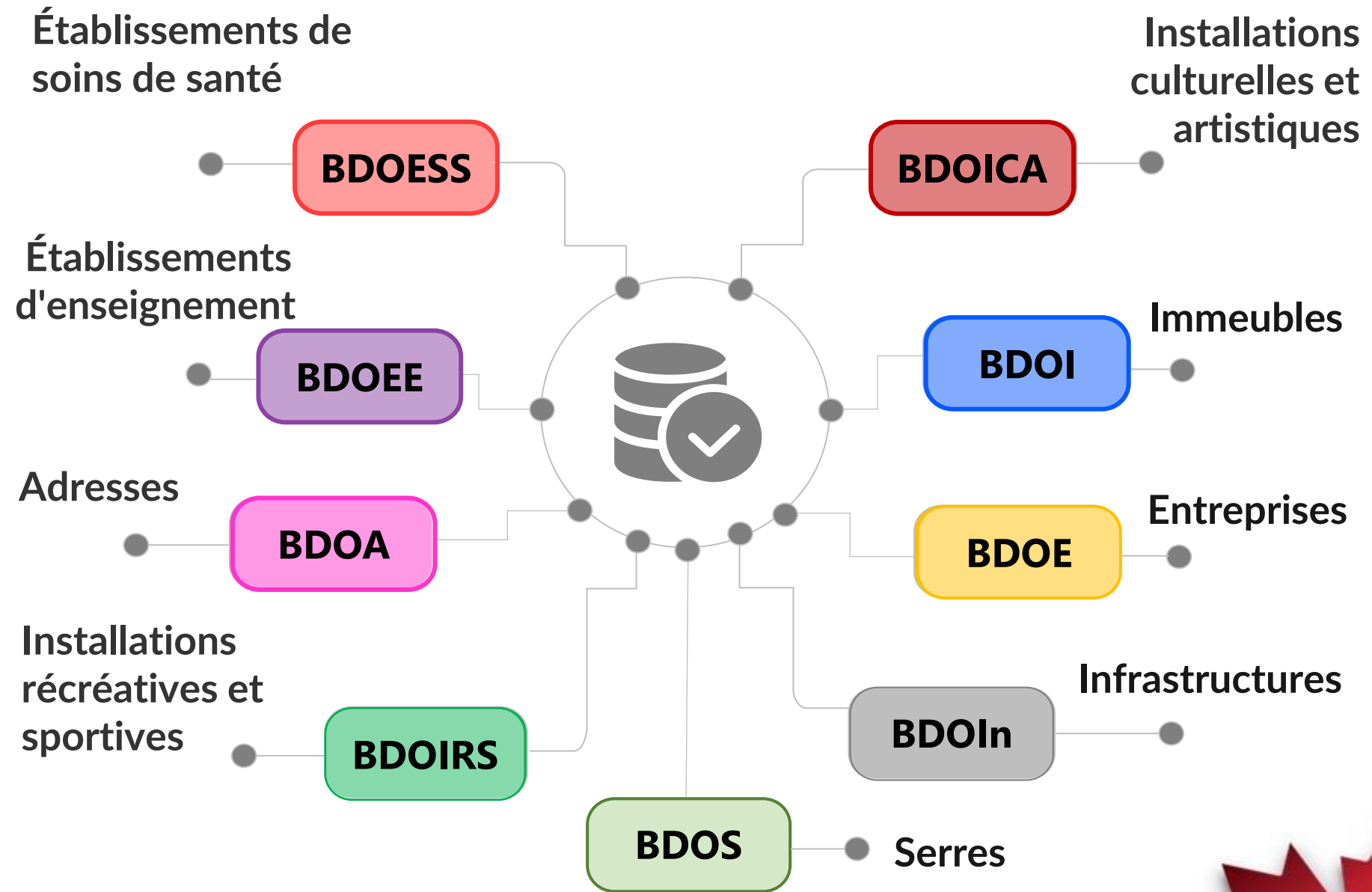
La couverture varie en fonction du sujet

Des centaines de sources de données ouvertes

Facilite la recherche ouverte sur les installations publiques et privées à travers le Canada en harmonisant les données provenant de tous les niveaux de gouvernement et d'autres entités en contact avec le public.

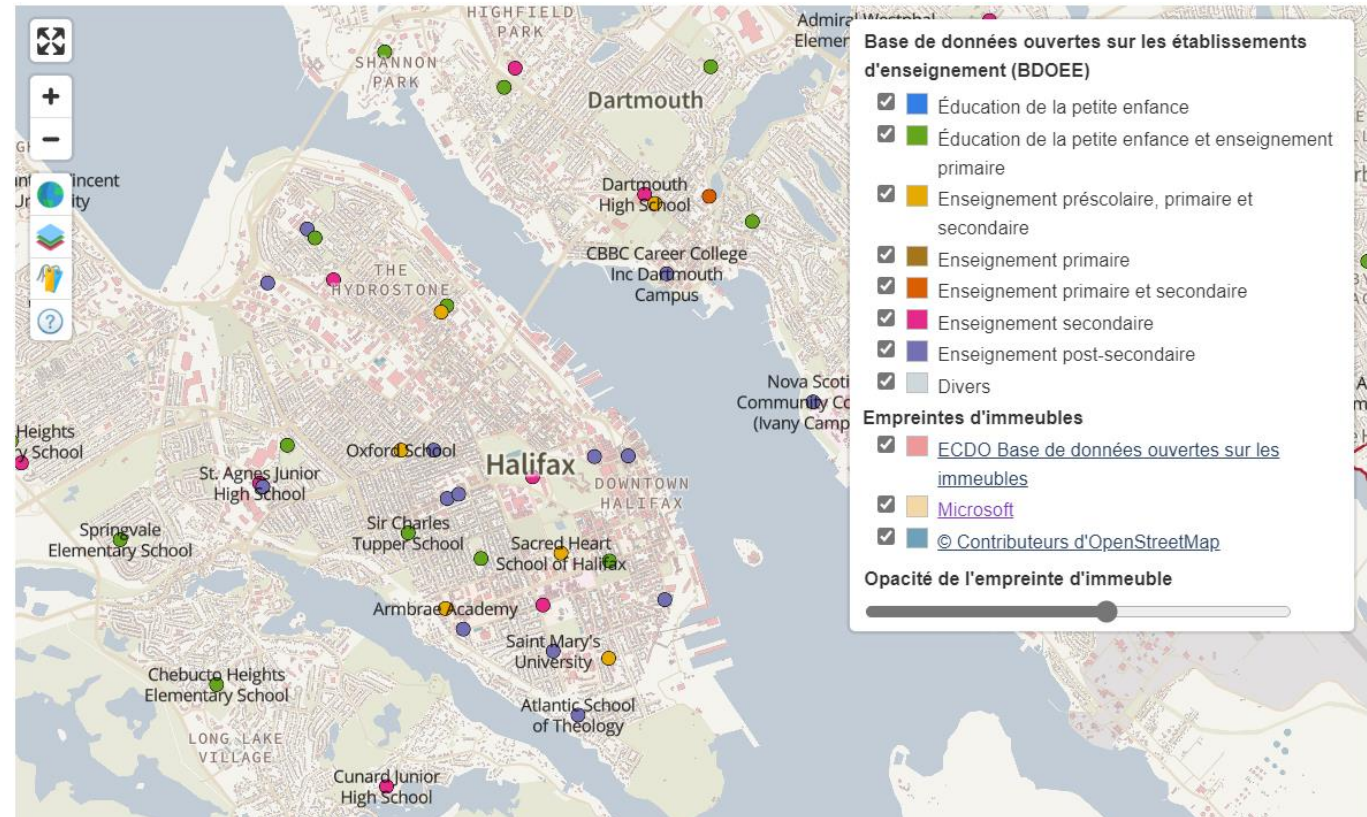
Possibilités d'analyse :

- Licence du gouvernement ouvert permet un partage et une collaboration aisé
- Contribue à des produits vedettes tels que les Mesures spatiales de l'accès



Données ouvertes et plus

- Code source ouvert
 - Traitement de données avec Python
 - Métadonnées transparentes
- Collaborations ouvertes
 - OpenAddresses
 - Relations communautaires
- Un visualiseur ouvert



[Visualiseur de l'environnement de couplage de données ouvertes](#)

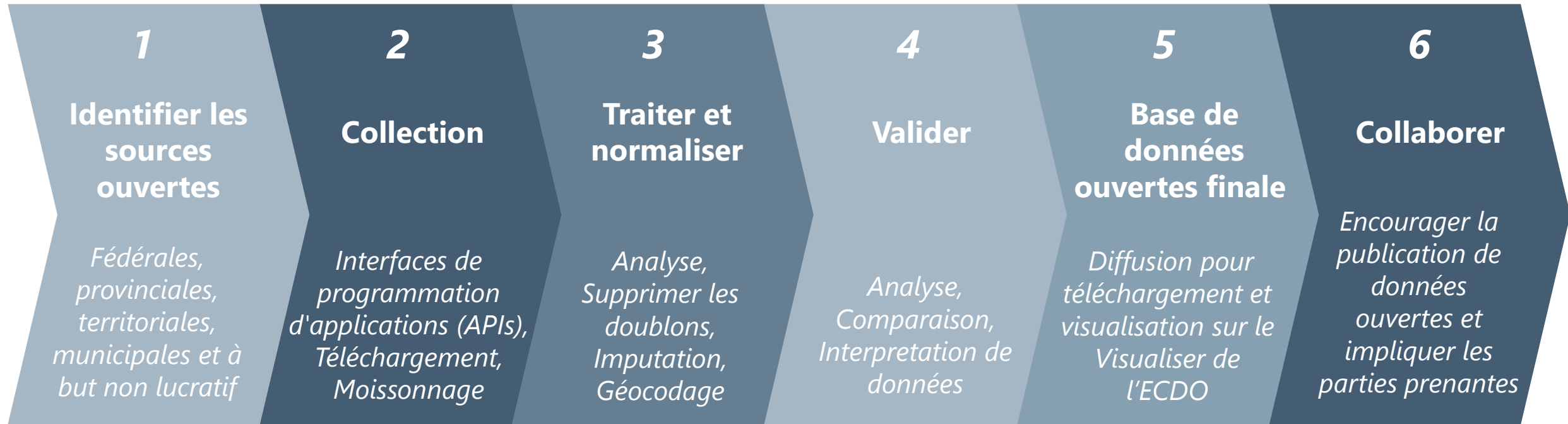
Un exemple - La base de données ouvertes sur les infrastructures

- une collection de données ouvertes contenant l'emplacement d'une sélection d'infrastructures à travers le Canada
- Mise à disposition sous la [Licence du gouvernement ouvert – Canada](#)
- Contenant :
 - Ponts et tunnels
 - Eau potable
 - Eaux pluviales et eaux usées
 - Déchets solides
 - Pistes piétonnes et cyclables



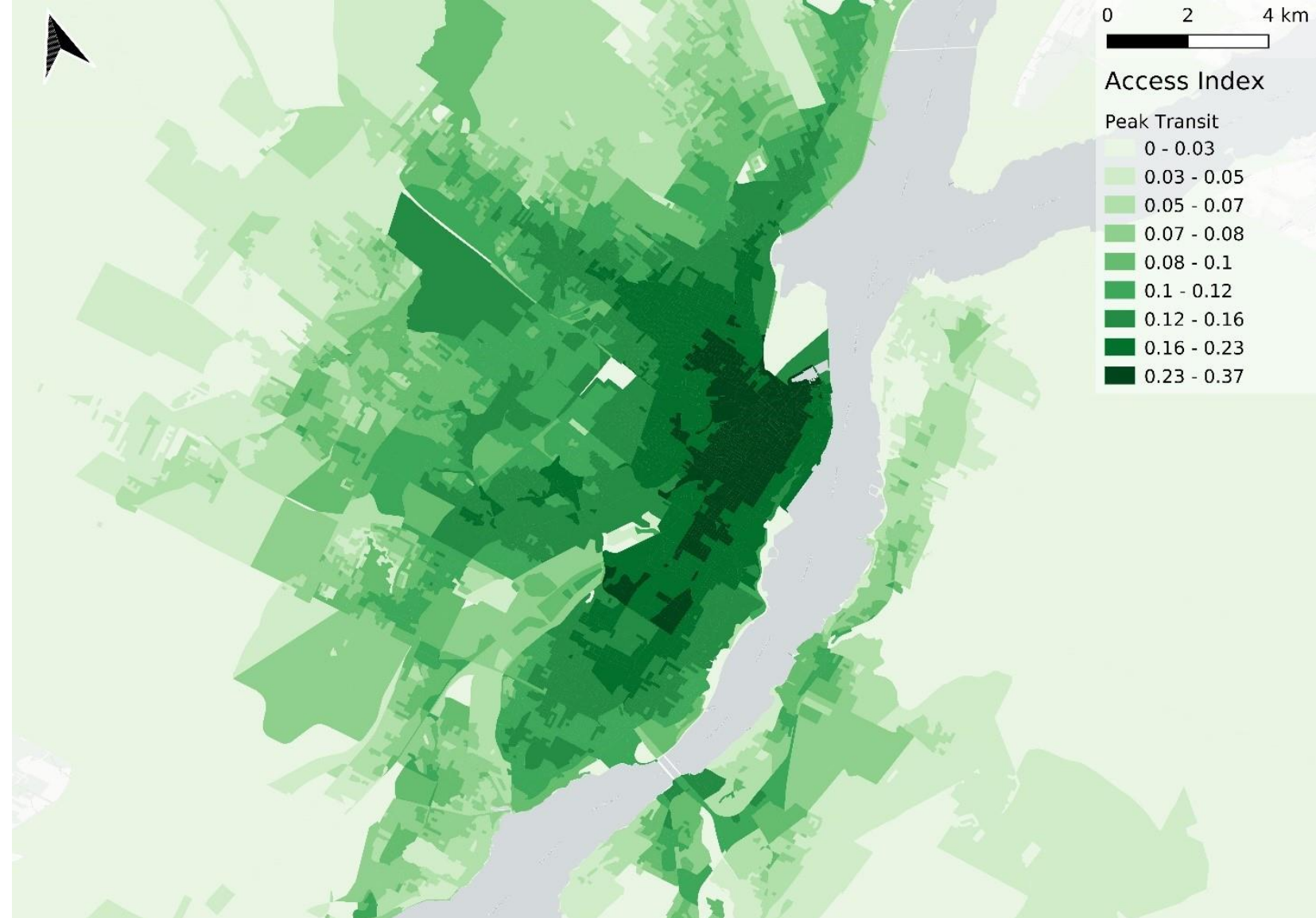
[Visualiseur de l'environnement de couplage de données ouvertes](#)

Processus de l'ECDO



Comment l'ECDO est-il utilisé ?

- Chaque produit de l'ECDO peut être utilisé pour contribuer à l'analyse d'un sujet donné
- Par exemple, les [Mesures spatiales de l'accès](#) explorent l'accès à sept commodités et services à travers le Canada par quatre modes de transport actif et public
- La Base de données ouvertes sur les établissements d'enseignement a été utilisée pour calculer l'accès à l'enseignement postsecondaire par les transports en commun



Accès à l'enseignement post-secondaire, heures de pointe des transports en commun, ville de Québec



Développements et améliorations itératifs de l'ECDO

- Elargissement de la matière :
- Ajout de nouvelles variables:
- Novembre 2023 : Publication de la Base de données ouvertes sur les entreprises
- Printemps 2024 : Mise à jour de la Base de données ouvertes sur les établissements d'enseignement
 - Identification de nouvelles écoles des minorités de langues officielles (EMLO)
 - Ajout d' écoles d'immersion française
- Mai 2023 : Publication de la Base de données ouvertes sur les infrastructures
- Printemps 2024 : Base de données ouvertes sur les infrastructures V 2.0
- Février 2023, Publication de la Base de données ouvertes sur les serres

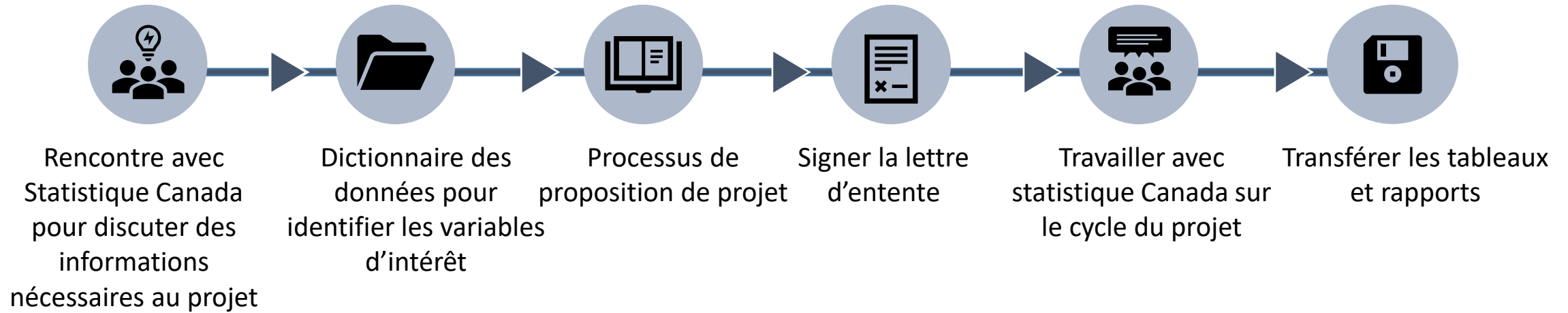


Processus d'accès au couplage de données

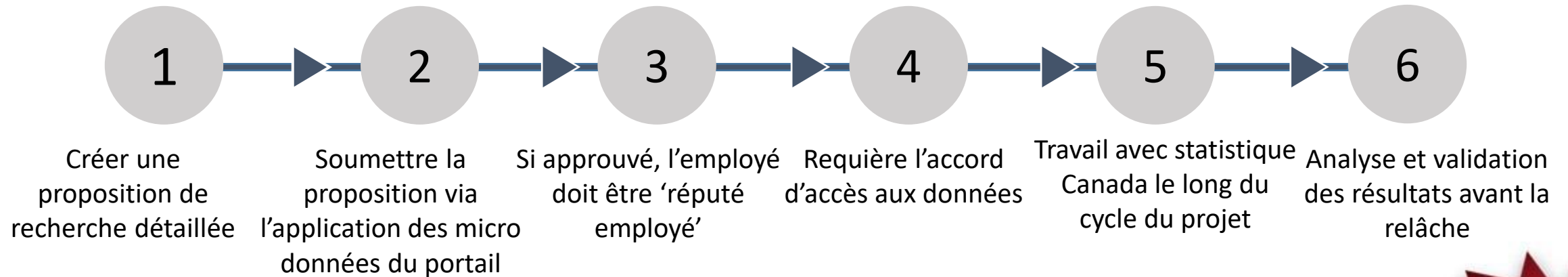


Processus pour initier un projet de couplage

Coûts recouvrables



Division d'accès aux données (DAD)



Nous vous remercions! Restez branchés.

Vous avez des questions?

Contactez-nous :

Julio Rosa, Chief

Environnement de Fichiers couplables entreprises (EFC-E)

613-796-4371 julio.rosa@statcan.gc.ca

Chantal Poirier, Chief

Environnement de couplage de données sociales (ECDS)

613-854-8237 chantal.poirier@statcan.gc.ca

Bjenk Ellefsen, Chief

Environnement de couplage de données ouvertes (ECDO)

(343) 573-2370 Bjenk.Ellefsen@statcan.gc.ca



ANNEXE 1

Sources de données administratives (EFC-E)

https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=6000

- Registre des entreprises (RE) - 1996 à 2022 (SDDS 1105)
- Programme d'analyse longitudinale de l'emploi (PALE) - 2000 à 2019 (SDDS 8013)
- Index général des renseignements financiers (IGRF - T1) - Entreprises non constituées en sociétés - 2005 à 2021
- Index général des renseignements financiers (IGRF-T2) - Entreprises constituées en sociétés - 2000 à 2021
- T4 (État de la rémunération Payée) fichier supplémentaire - 1997 à 2021
- FPP (Fichier Principal Personnel) du propriétaire principal de l'entreprise (avec des variables du recensement 2006 à 2021 [SDDS 3901] et de la base de données longitudinales sur l'immigration 2000 à 2021 [SSDS 5057]) - 2007 à 2014
- BDA (Base de données T1 améliorée) du propriétaire principal de l'entreprise (avec des variables du recensement 2006 à 2021 [SDDS 3901] et de la base de données longitudinales sur l'immigration 2000 à 2021 [SSDS 5057]) - 2012 à 2021
- Compte de retenues sur la paye (PD7) - 2001 à 2021
- Registre des exportateurs - 2010 à 2021 (SDDS 2201)
- Registre des importateurs - 2012 à 2021 (SDDS 2201)
- Brevets (Office de la propriété intellectuelle du Canada) -2001 à 2006
- United States Patent Office (USPTO) - Entreprises canadiennes seulement - 2000 à 2011
- Soutien à la croissance et à l'innovation en entreprise - 2007-2008 à 2020-2021 (SDDS 5304)
- Programme des données fiscales agricoles (PDFA) - 2016-2019 (SDDS 3447)
- Organismes de bienfaisance - 2003-2020
- Base de données de recherche sur les brevets canadiens (BDRBC) - 2001-2015
- Base de données sur la diversité et les compétences (BDDC) - 2001-2019 (SDDS 5228)



ANNEXE 2

Sources de données d'enquête(EFC-E)

https://www23.statcan.gc.ca/imdb/p2SV_f.pl?Function=getSurvey&SDDS=6000

- Enquête annuelle Recherche et développement dans l'industrie canadienne (RDIC) - 2000 à 2020 (SDDS 4201)
- Balance des paiements : investissements directs canadiens à l'étranger (IDCE) - 2000 à 2013 (SDDS 1537)
- Balance des paiements : investissements directs étrangers au Canada (IDEC) - 2000 à 2013 (SDDS 1537)
- Échanges de services commerciaux (ESC) - 2000 à 2020 (SDDS 1536)
- Enquête sur l'innovation et les stratégies d'entreprise (EISE) - 2009, 2012, 2017, 2019 (SDDS 5171)
- Enquête sur l'innovation (INNO) - 2003, 2005 (SDDS 4218)
- Enquête sur le commerce électronique et la technologie (ECET) - 2000 à 2007 (SDDS 4432)
- Enquête sur les technologies de pointe (ETP) - 2007, 2014 (SDDS 4223)
- Enquête sur la commercialisation de l'innovation (ECI) - 2007 (SDDS 5140)
- Enquête sur le financement et la croissance des petites et moyennes entreprises (EFCPME) - 2004, 2007, 2011, 2014, 2017, 2020 (SDDS 2941)
- Enquête sur la gestion de la propriété intellectuelle (EGPI) - 2010 (SDDS 5183)
- Enquête sur la technologie numérique et l'utilisation d'Internet (ETNUI) - 2012, 2013, 2019, 2021 (SDDS 4225)
- Enquête sur le coût de la mise en conformité à la réglementation (SDD 5093) - 2011, 2016
- Demande pour les dépenses de recherche scientifique et développement expérimental (RS&DE) - T661 - 2000 à 2021
- Enquête sur la connaissance et l'utilisation de la propriété intellectuelle (CUPI) - 2019 (SDDS 5291)



Types of Linkages

Deterministic or exact match:

- Candidate record pairs are compared using a set of comparison functions that only require exact similarity. Usually, this type of straightforward match is based on **unique identifiers** (Business number; SIN number...)

Probabilistic match:

- This approach calculates match weights based on the similarity (Names; Addresses; Postal codes...).

