

Partenariat de découverte des données du recensement du Canada (PDDRC)
<https://cddp-pddr.ca/>

RAPPORT DE CONSULTATION DU PDDRC #2

Diffuser les données du recensement : L'apprentissage par l'expérience

**ATELIER VIRTUEL
16 FÉVRIER 2022**

Zack Taylor, Western University
Susan Mowers, University d'Ottawa
Julien Doris, Université d'Ottawa
Shannon Leitch, University d'Ottawa
Leanne Trimble, University of Toronto

27 mai 2022

English version available:
Disseminating Census Data: Learning from Experience

Contenu

Aperçu des leçons tirées	3
Audience, convivialité, découverte et accès	3
Limites à l'innovation	3
Confidentialité des données	3
Introduction	4
Vue d'ensemble	4
Enquête auprès des panélistes	5
Thème 1 : Audience, convivialité, découverte et accès	8
Définir l'audience	8
Compromis accès-coûts-convivialité	10
Maximiser la découverte et l'accès (pour la bonne audience)	11
Leçons tirées	11
Thème 2 : Durabilité à long terme des projets	12
Viabilité financière	12
Viabilité technologique	13
Planifier l'inévitable : stratégies d'archivage et de préservation.	13
Intendance durable des données	13
Leçons tirées	14
Thème 3 : Frontières de l'innovation	14
Les progrès dans les outils d'analyse et de visualisation en ligne	14
De nouveaux outils d'accès aux données	15
La création de produits à valeur ajoutée	15
La numérisation	15
Le couplage de microdonnées	15
Leçons tirées	16
Thème 4: Confidentialité des données	16
Changer les normes et les exigences en matière de protection de la vie privée	16
Vers plus de flexibilité ? Le compromis entre l'accès et la protection de la vie privée	17
Les techniques de la divulgation des données	17
Leçons tirées	17
Appendice A : Panélistes et projets	18
Panel 1 – Défis et innovations	19
Panel 2 – Diffuser les microdonnées : Perspectives internationales	23
Panel 3 – Diffuser les microdonnées : Perspectives canadiennes	25
Appendice B : Vue des paramètres du portail de découverte du recensement.	30

Aperçu des leçons tirées

Audience, convivialité, découverte et accès

1. Déterminer une audience au projet de diffusion du recensement est une décision primordiale intervenant dès le départ. D'importantes décisions sur la conception du canal de diffusion ne peuvent être prises sans clarté sur l'identité et sur les besoins des groupes d'utilisateurs. Pour rencontrer les besoins des différents groupes d'utilisateurs et dans un contexte de raréfaction des ressources, le projet doit soigneusement formuler un compromis entre l'accès, la convivialité et le coût. Cela nécessite la prise en considération des cas d'utilisation les plus probables et la durabilité à long terme des projets.
2. Les données et les portails de diffusion spécialement conçus par des chercheurs universitaires, à l'aide de subventions limitées dans le temps, sont vulnérables aux pertes de financement, au roulement du personnel et aux changements dans les priorités institutionnelles (Paprica, 2021)¹.
3. Les projets doivent inclure une stratégie de préservation et d'archivage dès le départ, y compris l'identification d'une partie-prenante responsable de la gestion des données et des obligations légales associées.

Limites à l'innovation

4. L'accélération de la puissance de calcul et du stockage ont permis de débloquer une série d'innovations en matière de traitement et de diffusion des données.
5. À mesure que le rythme de l'innovation augmentera, le double défi de la viabilité budgétaire et de la viabilité technologique ne fera qu'augmenter.
6. Le changement technologique signifie que l'analyse des microdonnées deviendra de plus en plus réalisable pour les chercheurs du milieu universitaire, du gouvernement et du secteur privé, ce qui posera des défis pour la sécurité et la confidentialité des données.

Confidentialité des données

7. Les normes sociales et les exigences légales au regard de la confidentialité des données sont en mutation, particulièrement dans le sens d'une protection qui tend à se renforcer.
8. On ne sait pas quel impact à long terme cela aura sur l'utilisation des données du recensement et sur d'autres données pour la recherche.
9. Différents compromis risque-récompense concernant la tension entre l'accès et la vie privée sont possibles.

¹ Alison Paprica. (14 mars 2021). *Risks for academic research projects, an empirical study of perceived negative risks and possible responses*. (Rapport scientifique no. arXiv:2103.08048). Cornell University. <https://doi.org/10.48550/arXiv.2103.08048>

Introduction

Vue d'ensemble

Des recensements ou des dénombrements de la population ont été conduits sur le territoire que l'on connaît de nos jours sous le nom Canada depuis 1665-66 et qui portait alors le nom de Nouvelle-France. Le recensement au Canada est l'ensemble le plus utile de données économiques, sociales et culturelles. Il constitue un outil de recherche essentiel pour la formalisation de nouveaux savoirs et pour disposer d'une meilleure compréhension des conditions de vie d'une population sur un territoire donné, au passé comme au présent. Le Partenariat de découverte des données du recensement du Canada est un projet de deux ans qui entend améliorer l'accès et les conditions de recherche des sources de données qualitatives, quantitatives et spatiales existantes pour les recensements de la population au Canada. Les livrables clés du projet permettront de penser et de défendre une preuve conceptuelle de portail de découverte bilingue des données du recensement ainsi que des recommandations tournées vers l'avenir et sur lesquelles un tel projet pourrait être construit.

À ces fins, le projet a conduit une série de consultations de parties prenantes. En collaboration avec le **Réseau pour les tendances économiques et sociales** de l'Université *Western Ontario*, le projet a tenu la seconde de ses consultations, le 16 février 2022, sur le sujet ***Diffuser les données du recensement : l'apprentissage par l'expérience***. L'atelier sur invitation était axé sur les défis associés à la diffusion des données du recensement contemporain et historique pour les chercheurs universitaires et pour le public. L'objectif était de mieux comprendre comment les créateurs et les gestionnaires des données du recensement, à l'intérieur comme à l'extérieur du gouvernement, abordent les dilemmes associés à leur diffusion ainsi que de créer une communauté internationale de pratique sur ces questions.

Comme notre objectif était d'obtenir un large éventail de points de vue, nous avons invité des présentateurs d'organismes gouvernementaux de statistique et de projets universitaires portant sur les données du recensement au Canada, en Irlande, au Royaume-Uni et aux États-Unis. Comme nous mettons l'accent sur le libre accès dans l'intérêt du public, nous n'avons pas souhaité convier à cet atelier des fournisseurs de données à but lucratif.

L'atelier s'est déroulé autour de trois axes thématiques animés par des chercheurs du projet. Les présentations des 14 intervenants ont été suivies de questions de membres invités provenant de la communauté statistique gouvernementale.

Panel 1: Enjeux et innovations

- Jonathan Schroeder, *IPUMS–National Historical GIS*, États-Unis
- Humphrey Southall & Paula Aucott, *A Vision of Britain Through Time Project*, Royaume-Uni.
- Glen Hohlmann et David Price, Statistique Canada.
- Elaine Castonguay, Statistique Canada.

Panel 2 : Diffuser les microdonnées, perspectives internationales

- Deirdre Lynch, *Central Statistics Office*, Irlande.
- Callum Foster, *Office for National Statistics*, Royaume-Uni.
- Lara Cleveland, *IPUMS*, États-Unis.

Panel 3 : Diffuser les microdonnées, perspectives canadiennes

- Alexandra Clemence, Bibliothèque et Archives Canada.
- Sandra Sawchuk, *Mount St. Vincent University*.
- Susan Mowers, Université d'Ottawa.
- Kris Inwood, *Les populations canadiennes*.
- Michael Haan, *Western University*.

Ce rapport de consultation contient un résumé thématique des présentations de l'atelier et tire des leçons pour les futures initiatives de diffusion du recensement. Les présentations sont hébergées sur le site Internet du projet.

Enquête auprès des panélistes

Nous avons interrogé les panélistes sur les origines, les objectifs et les modalités de financement de leur projet en amont de l'atelier. Cela nous a apporté des réponses utiles, dont certaines sont résumées dans les tableaux 1 à 3. Les réponses complètes au sondage et les renseignements biographiques des panélistes et de leur projet sont reproduits en annexes.

Tableau 1: Origine du projet

	Projet de recherche académique	Initiative gouvernementale
<i>IPUMS–NHGIS</i> (États-Unis)	✓	
<i>A Vision of Britain Through Time</i> (Royaume-Uni)	✓	
Diffusion du recensement (StatCan)		✓
Géographie du recensement (StatCan)		✓
<i>IPUMS– International</i>	✓	
<i>Central Statistics Office</i> (CSO-Irlande)	✓	
<i>Office for National Statistics</i> (ONS-Royaume-uni)		✓
Bibliothèque et archives Canada (BAC)		✓
Infrastructure de recherche sur le Canada au XXème siècle (IRCS)	✓	
Partenariat de découverte des données du recensement du Canada (PDDRC)	✓	
Les populations canadiennes (LPC)	✓	
Centres de données de recherche (CDR-Canada)	✓	

Bien que l'origine déclarée des projets (tableau 1) ne soit pas surprenante – dans l'ensemble, les projets gouvernementaux ont été lancés par les gouvernements et les projets universitaires ont été lancés par des chercheurs universitaires – nous avons constaté des variations importantes dans les objectifs des projets (tableau 2). Bien que pratiquement tous les projets aient donné la priorité à un accès viable pour les utilisateurs, ils ont fait des compromis différents entre la préservation des formats originaux et le traitement ou le reformatage des données originales pour faciliter les couplages thématiques ou géographiques au fil du temps. Plusieurs projets gouvernementaux et universitaires accordaient expressément de l'importance à l'innovation, tandis que d'autres ne le faisaient pas. Les organismes gouvernementaux en particulier, ont donné la priorité aux rapports réguliers. Cela n'est pas surprenant compte tenu de leur structuration organisationnelle et de leurs responsabilités.

Tableau 2 : Objectifs du projet

	Accès viable	Libre accès	Préservation des formats originaux	Croisement temporels	Harmonisation inter-temporelle	Innovation	Rapports réguliers
IPUMS–NHGIS	✓	✓		✓	✓	✓	
A Vision of Britain	✓			✓	✓		
Diffusion du recensement (StatCan)	✓	✓	✓				✓
Géographie du recensement (StatCan)	✓	✓		✓	✓		
IPUMS–Int'l	✓			✓	✓	✓	
CSO-Irlande			✓		✓		✓
ONS-RU	✓	✓	✓	✓	✓	✓	✓
BAC	✓		✓				
IRCS	✓	✓				✓	
PDDRC	✓	✓			✓		
LPC	✓			✓	✓	✓	
Centres de données de recherche (CDR-Canada)	✓		✓	✓	✓	✓	✓

Les répondants ont également indiqué une gamme de sources de financement pour leurs projets (tableau 3). Les services gouvernementaux de collecte et de diffusion de statistiques dépendent du financement des programmes gouvernementaux, tandis que les projets universitaires sont généralement lancés avec le soutien des gouvernements, souvent par le biais de subventions concurrentielles de soutien à la recherche universitaire, et parfois aussi avec le soutien interne des universités et le soutien externe des fondations et des organismes de bienfaisance. Au fur et à mesure que les fonds de démarrage sont épuisés, les projets universitaires se tournent vers d'autres sources de financement, y compris vers des fondations et des organismes de bienfaisance, et dans certains cas, vers des licences et des publicités. Comme nous l'avons relevé plus en détail ci-dessous, l'enquête a révélé d'importantes préoccupations quant à la viabilité financière des projets universitaires et quant à l'entretien continu de l'infrastructure d'hébergement suite à l'achèvement du projet.

Tableau 3 : Sources de financement du projet

	Fonds de programmes gouvernementaux	Subventions des gouvernements incluant l'aide à la recherche académique	Fondations ou oeuvres de bienfaisance à but non-lucratif	Universités	Recouvrement des coûts à partir du paiement des utilisateurs, des licences, des publicités Internet
IPUMS–NHGIS		Principalement	✓	✓	Dons
A Vision of Britain		Pour le démarrage	Pour le démarrage		Actuellement
Diffusion du recensement (StatCan)	✓				
Géographie du recensement (StatCan)	✓				
IPUMS–International		Principalement	✓		
CSO-Irlande	✓				
ONS-RU	✓				
BAC	✓				
IRCS		✓			
PDDRC		✓			
LPC		✓			
CDR-Canada	✓				

Thème 1 : Audience, convivialité, découverte et accès

Le premier thème qui a émergé des discussions de l'atelier porte sur la relations entre trois facteurs :

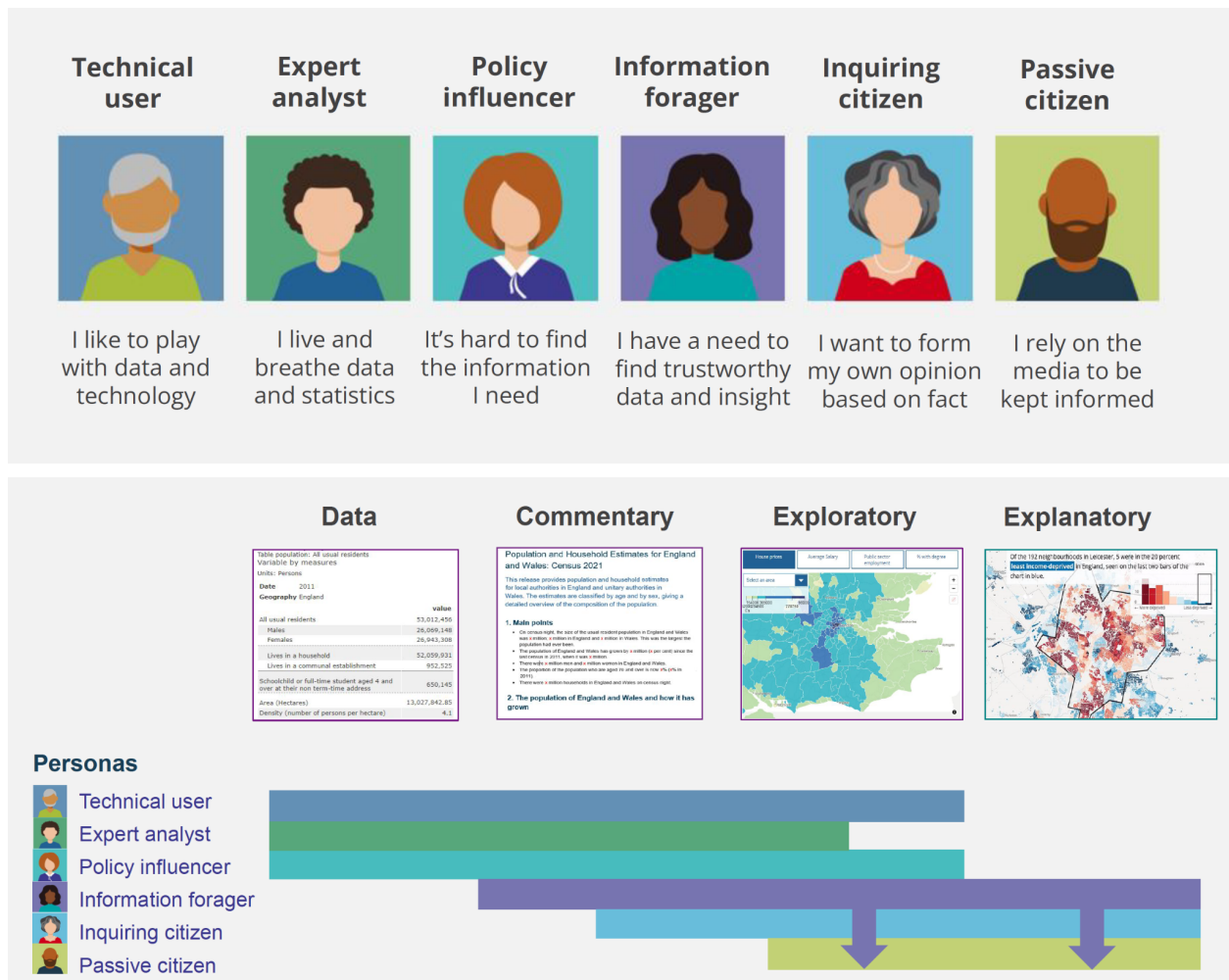
- *La convivialité* (l'appropriation de la complexité de l'outil par l'audience visée),
- *La découverte* (la capacité pour les usagers de trouver ce qu'ils recherchent ou de trouver des canaux de diffusion en premier lieu) et,
- *L'accès* (la disponibilité des données pour les utilisateurs).

Ces facteurs dépendent du public visé par le projet de diffusion. Un projet conçu pour des utilisateurs spécialisés ou expérimentés incarnera des choix différents de ceux conçus pour le grand public ou pour les utilisateurs professionnels occasionnels, tels que les journalistes.

Definir l'audience

Dans sa présentation, Callum Foster du Bureau des statistiques nationales du Royaume-Uni a présenté une classification systématique des publics des projets. Il a expliqué dans quelle mesure les stratégies de diffusion et de conception d'outils en ligne par le Bureau des statistiques nationales (l'ONS) étaient dictées par les besoins de groupes d'utilisateurs distincts. La figure 1 présente les personnes utilisatrices, des plus expérimentées aux moins expérimentées, et identifie les produits qui leur conviennent. Alors que les utilisateurs les plus expérimentés (utilisateurs techniques et analystes experts) veulent explorer et interagir directement avec les données, les citoyens curieux et les citoyens passifs sont intéressés à recevoir des informations immédiates et de brèves informations explicatives, y compris des visuels, pour trouver des données. Les influenceurs politiques et les glaneurs d'informations se situent quelque part entre ces deux groupes.

Figure 1: Des produits différents pour différents types d'utilisateurs



Source: Foster, diapositives 2 à 4.²

² Callum Foster. (16 février 2022). *Thinking about Census users and products, and ONS's Census 2021 dissemination and innovation* [communication]. Disseminating Census data: Learning from

Les autres panélistes étaient également conscients du public cible de leur projet, généralement des utilisateurs expérimentés : des chercheurs et des analystes du gouvernement et des universités, ainsi que du secteur à but non lucratif et des entreprises. Par exemple, les infrastructures de recherche NHGIS et IPUMS s'adressent aux utilisateurs expérimentés qui souhaitent manipuler directement les données, de sorte que le résultat propose un téléchargement de données organisées et personnalisées plutôt que des outils d'analyse et de visualisation interactifs, avec accès à des données longitudinales et transversales. (Jonathan Schroeder a comparé NHGIS à l'application Internet à but lucratif *Social Explorer* avec laquelle il partage des données. *Social Explorer* inclut des outils de cartographie et de visualisation sur la base de l'interrogation de l'utilisateur.) De même, les extraits du projet de *l'infrastructure de recherche sur le Canada au XXème siècle* (IRCS) et du projet sur *Les populations canadiennes* sont des produits de données destinés à être manipulés et analysés par les utilisateurs eux-mêmes.

En contraste avec certains autres, le projet *A Vision of Britain Through Time* a été expressément conçu pour être utilisé par le grand public, sans nécessiter de compétences techniques particulières. En effet, Humphrey Southall a précisé qu'ils ont organisé des groupes de discussion sur les besoins des utilisateurs et des essais d'utilisation au début de leur projet. Leur public est constitué de gens ordinaires intéressés par des lieux particuliers; d'où l'accent mis sur l'accès aux données longitudinales plutôt que transversales. De plus, la mission publique des fonds du recensement de Bibliothèque et Archives Canada (BAC) vise à répondre principalement aux demandes d'utilisateurs tels que les généalogistes, les historiens et autres professionnels.

Les agences statistiques gouvernementales doivent se tourner vers un large éventail de publics potentiels. Elles diffusent de l'information sur les lieux spécifiques aux utilisateurs non qualifiés au moyen de portails en ligne, de l'analyse et de commentaires sous la forme de rapports, ainsi que d'ensembles de données complets pour être analysés par les utilisateurs, quelquefois dans des contextes sécurisés, comme pour les microdonnées du réseau des Centres de données de recherche du Canada ou des installations équivalentes dans les autres pays.

Compromis accès-coûts-convivialité

La typologie des personnes et des produits d'audience de Foster illustre les compromis axés sur le public, entre le fait de rendre les données accessibles à un large public et la complexité du canal de diffusion, ou encore entre l'accès et le coût. Rendre des tableaux de données complets et disponibles au téléchargement (ou interrogeables par API) est relativement peu coûteux mais nécessite que l'utilisateur soit qualifié. Les outils d'exploration, de visualisation et d'analyse basés sur Internet sont coûteux à construire et à

expérience, virtuel workshop (Diffuser les données du recensement : L'apprentissage par l'expérience, atelier virtuel).

<https://cddp-pddr.ca/en/committees/user-needs/disseminating-census-data-event>

maintenir, mais peuvent être conçus pour permettre une utilisation par un plus large éventail de personnes. En général, plus l'ensemble des choix d'interaction disponibles pour l'utilisateur est limité ou plus les données mises à disposition sont organisées au lieu d'être exhaustives, plus il est facile pour les utilisateurs de découvrir et de récupérer des données. Dans le même temps, ces contraintes peuvent rendre les données moins utiles aux utilisateurs avertis.

Maximiser la découverte et l'accès (pour la bonne audience)

Les données ne sont utilisables que si elles peuvent être découvertes par leurs publics cibles. En tant que fournisseurs officiels, les agences gouvernementales peuvent tenir la découverte pour acquise. Glen Hohmann a fait remarquer que Statistique Canada est en train de transférer ses données du recensement destinées au public vers le nouveau référentiel centralisé de données du gouvernement et des entreprises du gouvernement du Canada. Cela peut faciliter la découvrabilité et l'accès à divers groupes d'utilisateurs.

Les projets destinés aux utilisateurs spécialisés et aux experts peuvent séduire directement les groupes d'utilisateurs. Jonathan Schroeder et Lara Cleveland ont noté que l'équipe qui administre les projets NHGIS et IPUMS-International passe beaucoup de temps à les promouvoir lors de réunions et d'événements universitaires et professionnels. En revanche, le projet IRCS s'est concentré presque entièrement sur la numérisation et la validation des données, avec une attention limitée accordée à la découvrabilité continue.

Humphrey Southall a noté un élément important de la découverte lorsque le public visé est constitué de gens ordinaires. Il convient alors de s'assurer que les données relatives à des lieux et des concepts particuliers apparaissent dans les recherches sur Internet, même si les pages sont activement générées par des requêtes provenant des bases de données. Le portail *The Vision of Britain* a été conçu pour être exposé aux moteurs de recherche des pages Internet relatives à des endroits particuliers.

Enfin et en guise de transition vers le deuxième thème, il y a la question du libre accès. Tous les projets décrits diffusent leurs données aux utilisateurs sans frais, à moins qu'ils ne soient limités par la législation sur la protection de la vie privée ou à des conditions particulières de licence. Néanmoins, la pression pour recouvrer les coûts par le biais de licences ou de frais d'utilisation augmente à mesure que les autres sources de financement se font plus rares.

Leçons tirées

1. Déterminer une audience au projet de diffusion du recensement est une décision primordiale intervenant dès le départ. D'importantes décisions sur la conception du canal de diffusion ne peuvent être prises sans clarté sur l'identité et sur les besoins des groupes d'utilisateurs.

2. Pour rencontrer les besoins des différents groupes d'utilisateurs et dans un contexte de raréfaction des ressources, le projet doit soigneusement formuler un compromis entre l'accessibilité, la convivialité et le coût. Ce compromis est fortement requis dans la plupart des cas.

Thème 2 : Durabilité à long terme des projets

Un deuxième thème qui est ressorti des présentations et des discussions était celui de la durabilité à long terme des projets. Au moins trois dimensions sont à relever :

- *La viabilité financière* (d'où l'argent viendra-t-il?),
- *La viabilité technologique* (qui maintiendra et mettra à jour le matériel et les logiciels?), et
- *La viabilité des données* (qui surveillera dans le temps long les droits et les exigences en matière de licences ainsi que les évolutions des régimes de confidentialité des données?).

Viabilité financière

La plupart des projets s'inquiètent des ressources. Les projets universitaires en cours sont très préoccupés par la viabilité financière. En effet, le modèle de recherche universitaire standard consiste à financer des projets déterminés et spécifiques avec des subventions limitées dans le temps. À moins que les universités, les fondations ou les organismes de bienfaisance n'assument la responsabilité du financement des projets sur une base ouverte, les projets doivent se tourner vers des stratégies de monétisation, y compris vers la facturation de produits de données et de travaux d'analyse souhaitables, sapant ainsi le libre accès.

Tout en bénéficiant du financement des programmes, les organismes gouvernementaux de statistiques font face à leurs propres pressions fiscales et politiques. Le recensement coûte cher et il y a toujours des pressions pour réduire les coûts. Comme on l'a vu aux États-Unis et au Canada il y a dix ans, il y a aussi des pressions politiques de la part de certains milieux pour rendre la collecte des données gouvernementales moins intrusive, en rendant les sondages volontaires et en éliminant les questions jugées de nature trop personnelle. Les organismes ont réagi à ces pressions du mieux qu'ils le pouvaient en mettant l'accent sur la continuité et en s'appuyant de plus en plus sur les données administratives liées.

En fin de compte, les projets non gouvernementaux et universitaires ont des objectifs différents de ceux des programmes gouvernementaux. Les agences statistiques gouvernementales sont à la fois tournées vers le présent et vers l'avenir. Leur mandat est de diffuser les données actuelles aux décideurs et aux parties prenantes du secteur privé en temps opportun. Elles ont peu d'intérêt à élargir, à maintenir et à diffuser des ensembles de données historiques en numérisant des documents imprimés, en stockant des données originelles dans des formats de fichiers modernes ou en attachant des identificateurs

géographiques et conceptuels modernes aux données historiques. Il revient *de facto* aux projets académiques d'effectuer ces tâches mais avec des ressources beaucoup plus limitées.

Viabilité technologique

Les panélistes associés à des projets universitaires ont souligné la difficulté à suivre les changements technologiques. L'un des aspects consiste à s'assurer que le matériel et les logiciels – parfois prêts à l'emploi, souvent écrits sur mesure et qui appuient le canal de diffusion – soient maintenus et mis à jour. À cela s'ajoute le problème de la mise à jour des fichiers de données visant à refléter l'évolution des formats et des normes de métadonnées. Ces tâches nécessitent une expertise spécialisée et sont coûteuses à effectuer, de telle sorte que la viabilité technique et la viabilité financière sont indissociables.

Planifier l'inévitable : stratégies d'archivage et de préservation.

La discussion s'est ensuite tournée vers la question de savoir quoi faire si un projet en cours est contraint de s'arrêter subitement, en raison d'une perte de ressources financières ou en raison d'un roulement dans la composition de la direction et des expertises du projet. Ce problème a été mis de l'avant dans le cadre du projet *The Vision of Britain*. Humphrey Southall a reconnu que le logiciel et l'infrastructure de la gestion de base des données qui pilotent le site Internet sont défaillants et qu'il n'y a pas de ressources disponibles pour le mettre à jour ou pour en moderniser l'interface. Comme plan de secours, il met en œuvre une stratégie d'archivage des tableaux longitudinaux à partir du Service des données du Royaume-Uni. Bien que cela implique l'abandon de leur modèle novateur de données relationnelles et semble compromettre la découverte et l'accès par leur public principal – les gens ordinaires - cela permet toutefois que les données soient stockées en lieu sûr par un intendant de confiance et soient bien dotées en ressources.

Cette mise en garde suggère l'importance d'une stratégie d'archivage et de préservation à long terme pour tout projet de création ou de re-formatage de données. Les obstacles à l'archivage et à la préservation permanente peuvent diminuer avec l'émergence de référentiels de données bien établis et avec un financement de base qui soit permanent, tels que pour *Scholars GeoPortal*, pour les mises en œuvre du *Harvard's Dataverse*, pour le Service des données du Royaume-Uni ou encore pour l'ICPSR. Certaines de ces infrastructures sont neutres quant au format, c'est-à-dire qu'il est possible de contribuer et de récupérer les données hébergées dans une variété de formats, résolvant ainsi partiellement le problème de l'obsolescence du format des fichiers. Cependant, étant donné que ces installations stockent généralement des tableaux ou des fichiers de limites spatiales sans outil de sélection et de filtrage (par emplacement ou par variables), elles sont généralement et essentiellement utiles que pour les spécialistes.

Intendance durable des données

Une stratégie d'archivage et de préservation à long terme peut garantir que les données ne soient pas « perdues » si leurs créateurs se retirent ou disparaissent de leur gestion. Un autre problème est le maintien de la conformité aux règles d'accès dans le temps et aux règles de licence négociées avec les fournisseurs privés. Par exemple, il a été noté que les fichiers de microdonnées de 1931, 1941 et 1951 de l'IRCS (qui sont actuellement couverts par la règle de confidentialité des 92 ans du gouvernement du Canada), entreront dans le domaine public de concert avec l'expiration de la période de 92 ans (par exemple 2023 pour les microdonnées de l'IRCS pour 1931). Mais il n'est pas clair à qui incombe le travail de veiller à ce que les fichiers sous embargo soient ensuite transférés dans des dépôts ouverts. De plus, Kris Inwood a décrit comment les fichiers de microdonnées (d'échantillonnage à 100 %) du projet *Les populations canadiennes* ont été produits en partenariat avec un fournisseur privé dont les conditions de licence excluent certaines utilisations. Encore une fois, il n'est pas clair de qui sera responsable des fichiers de données sur le long terme.

Leçons tirées

3. Les données et les portails de diffusion spécialement conçus par des chercheurs universitaires, à l'aide de subventions limitées dans le temps, sont vulnérables aux pertes de financement, au roulement du personnel et aux changements dans les priorités institutionnelles.
4. Les projets doivent inclure une stratégie d'archivage et de préservation dès le départ, y compris l'identification d'une partie responsable pour gérer les données et les obligations légales associées.

Thème 3 : Frontières de l'innovation

L'un des principaux objectifs de l'atelier était d'identifier les frontières de l'innovation. En synthétisant les discussions et les présentations, nous avons identifié cinq développements importants :

- Les progrès dans les outils d'analyse et de visualisation en ligne.
- De nouveaux outils d'accès aux données (APIs).
- La création de produits à valeur ajoutée.
- La numérisation des documents papier.
- Le couplage de microdonnées.

Ces enjeux ont été facilités par l'augmentation de la puissance de calcul et l'émergence d'un stockage en nuage, rapide et peu coûteux.

Les progrès dans les outils d'analyse et de visualisation en ligne

La démonstration faite par Callum Foster a permis de mettre en évidence que le prototype d'application Internet du Bureau des statistiques nationales du Royaume Uni est capable de compiler des microdonnées tout en appliquant des critères de contrôle à la divulgation en temps réel. Cette démonstration a été un des points d'orgue de l'atelier. Alimentés par le stockage de bases de données dans le *Cloud* et d'algorithmes optimisés, ces outils brouillent la frontière entre les analyses effectuées sur des ordinateurs locaux rapides et la commodité des applications Internet. Ils brouillent également la frontière entre les données agrégées et les microdonnées. Ces outils émergents peuvent présenter des opportunités alternatives aux totalisations spéciales et personnalisées qui sont coûteuses pour créer des tableaux croisés multidimensionnels.

De nouveaux outils d'accès aux données

Il a également été question de faciliter la récupération de données personnalisées par des spécialistes grâce à l'utilisation d'interfaces de programmation d'applications (API) – il s'agit essentiellement de bibliothèques de codes pour les logiciels couramment utilisés en science et en analyse des données et qui peuvent interroger par la programmation, des bases de données distantes sans récupérer manuellement des données via un interface utilisateur graphique basé sur Internet. Le NHGIS et l'IPUMS-International développent des API pour l'extraction directe de données dans des logiciels de statistiques modernes.

La création de produits à valeur ajoutée

Une troisième innovation est l'élaboration de systèmes de codage normalisés pour les variables et les unités géographiques, ainsi que l'élaboration d'outils de correspondance et de transition qui relient les unités géographiques à travers le temps. IPUMS-NHGIS a été un chef de file dans ce domaine en créant des tableaux qui transcrivent les données historiques dans le format des unités contemporaines. IPUMS-International a également consacré un travail considérable à l'élaboration d'un dictionnaire de données complet et d'outils de métadonnées permettant une comparaison transnationale.

La numérisation

Plusieurs projets universitaires canadiens (IRCS et LPC) ainsi que Bibliothèque et Archives Canada se sont engagés dans la numérisation de données agrégées historiques et des microdonnées. Cela a lentement mais sûrement permis de combler les lacunes dans la disponibilité des données numérisées. Le projet *The Vision of Britain* et le NHGIS ont effectué des travaux similaires, respectivement au Royaume-Uni et aux États-Unis. Le Bureau des statistiques nationales du Royaume-Uni et le Bureau central irlandais des statistiques ont également participé à la numérisation des données agrégées historiques et des microdonnées. Kris Inwood a noté qu'à mesure que les microdonnées historiques complètes des dénombrements de la population seront disponibles, il deviendra possible de croiser des variables non publiées dans les volumes imprimés originaux.

Le couplage de microdonnées

À mesure que la puissance de calcul augmente, il est de plus en plus possible de relier les données administratives et les données du recensement au niveau de l'individu, du ménage, de la propriété, ou encore de relier les données du recensement à d'autres ensembles de données d'enquête (par exemple ceux relatifs à la santé, à l'immigration, etc.). Michael Haan a souligné que l'utilisation des ensembles de données à accès restreint sont liés à Statistique Canada et concernent les résultats en matière de santé et d'établissement des immigrants et des réfugiés. Deirdre Lynch a expliqué comment les données du recensement irlandais ont été liées aux données sur les propriétés et les services publics pour être utilisées dans le cadre de la planification à long terme de l'utilisation des terres et des infrastructures. Glenn Hohlmann a également mentionné comment les données sur le revenu des particuliers et des ménages, diffusées dans les produits du recensement canadien, sont maintenant tirées des données des déclarants. À mesure que le couplage de microdonnées devient plus réalisable, les gestionnaires en données et les chercheurs sont confrontés à des questions importantes concernant la protection de la vie privée et l'utilisation éthique des données qui est le sujet du quatrième et dernier thème.

Leçons tirées

5. L'accélération de la puissance de calcul et du stockage a permis de débloquer une série d'innovations en matière de traitement et de diffusion des données.
6. À mesure que le rythme de l'innovation augmentera, le double défi de la viabilité budgétaire et de la viabilité technologique ne fera qu'augmenter.
7. Le changement technologique signifie que l'analyse des microdonnées deviendra plus réalisable pour les chercheurs du milieu universitaire, du gouvernement et du secteur privé, ce qui posera des défis pour la sécurité et pour la confidentialité des données.

Thème 4: Confidentialité des données

Le dernier thème est la confidentialité des données. Les participants ont relevé trois problèmes :

- *L'évolution des normes et des exigences en matière de protection de la vie privée.*
- *Une plus grande flexibilité dans l'accès aux microdonnées.*
- *De nouvelles techniques de contrôle de la divulgation appliquées aux données agrégées.*

Changer les normes et les exigences en matière de protection de la vie privée

Les participants ont reconnu que nous vivons dans une époque où les attitudes à l'égard de la protection de la vie privée évoluent. D'une part, les entreprises de « big tech » collectent et monétisent continuellement des données selon les mouvements et les habitudes de consommation des gens ordinaires. D'autre part, les gens se méfient de plus en plus des efforts de collecte de données des gouvernements. Du Règlement général sur la protection des données (RGPD) aux nouvelles règles du Conseil du Trésor du Canada, les cadres juridiques et réglementaires ne cessent d'évoluer en exigeant une autorisation proactive des utilisateurs et la divulgation de la façon dont les données seront utilisées. Michael Haan s'est demandé si nous n'étions pas trop absolutistes à vouloir éviter l'identification des individus, suggérant que nous devrions réfléchir aux risques pratiques et probables de le faire dans la recherche universitaire.

Vers plus de flexibilité ? Le compromis entre l'accès et la protection de la vie privée

L'utilisation d'installations sécurisées pour accéder aux microdonnées est bien établie mais il existe une demande croissante pour une plus grande flexibilité dans la façon dont on peut y accéder. Michael Haan a fait remarquer que le réseau des Centres de données de recherche du Canada héberge environ 500 fichiers de microdonnées et qu'il y a des discussions constantes sur la façon dont ils peuvent être utilisés plus largement sans compromettre la vie privée. Les expériences menées pendant les restrictions liées à la pandémie, sur l'accès à distance à des ensembles de données hébergés de manière centralisée au Canada et en Irlande indiquent différentes options. Le modèle du « client léger » de données adopté par Statistique Canada peut permettre un accès à distance plus important avec une plus grande puissance de traitement.

Les techniques de la divulgation des données

Callum Foster a fait remarquer qu'une variété de techniques sont utilisées pour assurer la protection de la vie privée dans le contrôle de la divulgation. L'anonymisation des microdonnées est une pratique de longue date, tout comme la suppression des comptes qui tombent en dessous des seuils établis. De manière plus récente cependant, sont adoptés l'échange de dossiers individuels entre les ménages et l'application de la « perturbation des données » aux fichiers de microdonnées.

Leçons tirées

8. Les normes sociales et les exigences légales au regard de la confidentialité des données sont en mutation, particulièrement dans le sens d'une protection qui tend à se renforcer.
9. On ne sait pas quel impact à long terme cela aura sur l'utilisation des données du recensement et sur d'autres données pour la recherche.
10. Différents compromis entre les risques et les incitatifs pour régler la tension entre l'accès aux données et la protection de la vie privée sont possibles.

Appendice A : Panélistes et projets

L'information suivante a été obtenue des panélistes en amont de l'atelier.

Panel 1 – Défis et innovations

Panéliste

Jonathan Schroeder

Chercheur scientifique

IPUMS

University of Minnesota

Courriel : jps@umn.edu

Bio: Jonathan Schroeder est géographe au « Center for Data Integration » de l'IPUMS à l'Université du Minnesota. Il est notamment chef de projet et chercheur principal pour le *National Historical Geographic Information System* (NHGIS) de l'IPUMS. Il dirige l'intégration des nouvelles données sommaires du recensement américain et le développement de séries chronologiques, de fichiers SIG et d'autres ressources. Il contribue également au projet IPUMS USA en dirigeant le développement de variables géographiques intégrées pour les microdonnées du recensement américain.

Description du projet: IPUMS fournit des données du recensement et d'enquêtes du monde entier, intégrées dans le temps et l'espace. L'IPUMS se compose de plusieurs projets et sites Internet qui fournissent différents types de données. Jonathan Schroeder gère l'un de ces projets : le *National Historical Geographic Information System* (NHGIS) de l'IPUMS, qui donne accès à des tableaux sommaires et à des séries chronologiques de données sur la population, le logement, l'agriculture et l'économie des États-Unis, ainsi qu'à des fichiers de limites géographiques compatibles avec les SIG, pour les années allant de 1790 à aujourd'hui et pour tous les niveaux de la géographie du recensement des États-Unis.

Provenance du projet : Sous la forme d'un projet de recherche académique

But du projet : Rendre les données accessibles au public pour une utilisation générale

Sources de financement : Le financement principal provient de subventions des gouvernements, y compris le soutien à la recherche universitaire. « Notre institut a reçu des contributions correspondant à tous les autres types mentionnés. Nous ne demandons aucun paiement aux utilisateurs, mais acceptons les dons ».

Conditions de financement : Sujet à des validations fréquentes.

Préoccupations concernant le financement : « Oui. La majeure partie de notre financement provient de subventions à durée limitée, pour lesquelles nous devons régulièrement présenter une nouvelle demande. Nous avons économisé suffisamment au chapitre des bénéfices indirects des subventions pour couvrir les intervalles occasionnels non financés, mais nous ne pouvons pas le faire indéfiniment ».

À qui s'adresse le projet : Le public en général, des instances gouvernementales, la communauté des chercheurs universitaires, des organismes à but non lucratif ou de bienfaisance, des entreprises. Les chercheurs universitaires et les étudiants constituent le principal public auquel s'intéressent nos organismes de financement. Nous leur accordons donc la priorité dans notre travail, mais nous accueillons et soutenons les utilisateurs de toutes sortes.

Objectifs du projet : Un accès durable aux données et aux matériaux pour notre public; relier les données dans le temps (par exemple pour plusieurs cycles de recensement); assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement); innover en matière de diffusion des données (par exemple sur le format d'accès aux données, la convivialité, ou la liaison des données); l'accès libre aux données. Dans le cadre de nos projets de données internationaux, nous accordons également la priorité à la mise en relation et à la cohérence des données entre les pays.

Co-panéliste

Humphrey Southall

Professeur en géographie historique
School of the Environment, Geography and Geoscience
University of Portsmouth
Courriel: humphrey.southall@port.ac.uk

Bio: Humphrey Southall est professeur de géographie historique à l'Université de Portsmouth, au Royaume-Uni. Il est titulaire d'un diplôme de premier cycle et d'un diplôme de troisième cycle de l'Université Cambridge et a également étudié à l'Université Harvard. Il dirige les travaux du *Great Britain Historical GIS Project* (GIS Geographical Information System) depuis 1994 et est membre du *Census Advisory Group* de l'*Office of National Statistics* depuis 1999.

Co-panéliste

Paula Aucott

Chercheuse associée
University of Portsmouth
Courriel : paula.aucott@port.ac.uk

Bio: Paula est « *Senior Research Associate* » à la *School of the Environment, Geography and Geosciences* de l'Université de Portsmouth. Elle travaille sur le projet de SIG historique de la Grande-Bretagne (*Great Britain Historical GIS project*) depuis 2000, en mettant l'accent sur le contenu. Ses recherches portent sur les SIG historiques, les répertoires géographiques et la cartographie historique de l'utilisation des terres.

Description du projet : (Fourni par Humphrey Southall). « Depuis 1994, je dirige le projet SIG historique de la Grande-Bretagne (*Great Britain Historical GIS Project*), une entité auto-définie et financée par de nombreuses subventions distinctes des conseils de recherche gouvernementaux et des organismes de recherche caritatifs, chacune ayant son propre titre de projet. Jusqu'en 2001, l'activité principale consistait à cartographier les unités statistiques changeantes de la Grande-Bretagne depuis 1801, y compris les quelque 16 000 paroisses, et à informatiser ou à obtenir d'une autre manière des versions numériques des statistiques historiques associées. Le recensement de la population est notre source statistique la plus importante, mais nous couvrons également les statistiques de l'état civil, le chômage et la Loi sur les pauvres (*Poor Law*), les statistiques agricoles, etc. En 2001, nous avons obtenu une nouvelle subvention importante de la Loterie nationale britannique, qui nous a permis de créer le site Internet « *A Vision of Britain through Time* » (www.visionofbritain.org.uk) comme ressource pour le grand public. Plutôt que d'être un autre système de téléchargement, l'architecture de données unique de VoB se concentre sur le rassemblement de données pour la localité sélectionnée par l'utilisateur à partir d'enquêtes géographiques nombreuses et diverses

de la Grande-Bretagne. En particulier, les données du recensement sont présentées principalement sous forme de séries chronologiques locales, à la fois pour les unités de déclaration historiques originales et, par le biais d'une redistribution basée sur les SIG (« superposition vectorielle »), pour les districts des autorités locales modernes. Le système a été considérablement étendu en 2009. Je suis maintenant semi-retraîtée, mais nous avons toujours un membre du personnel à temps plein, financé principalement par les revenus des licences de données ».

Provenance du projet : Sous la forme d'un projet de recherche académique.

But du projet : Rendre les données accessibles au public pour une utilisation générale.

Règles concernant la diffusion : « Non. Nous travaillons exclusivement avec des données agrégées et des données de limites associées ».

Sources de financement : Financement initial par le gouvernement, les associations caritatives et la loterie nationale. Aujourd'hui, il s'agit principalement de licences de données et de publicités sur le site Internet.

Conditions de financement : « Comme indiqué plus haut, nous dépendons désormais des revenus tirés de l'octroi de licences pour les données des limites, des publicités Google et de certains travaux de conseil ».

Préoccupations concernant le financement : « Oui ! Une partie de nos revenus de licence est à long terme, mais pas assez, donc nous dépendons de ventes ponctuelles ».

À qui s'adresse le projet : « Le public en général, des instances gouvernementales, la communauté des chercheurs universitaires. Notre principale activité est la gestion d'un site Internet de données ouvertes et il existe certainement quelques exemples d'utilisation par les entreprises et les organismes non-gouvernementaux ou sans but lucratif. Nous avons récemment fourni des données numériques sur les limites des unités du recensement historiques à la division des recensements de l'*Office of National Statistics* et nous avons redécoupé les données du recensement historiques pour une équipe interministérielle travaillant sur le Livre blanc sur le nivellement (« Levelling-up » White Paper), de sorte que le « Gouvernement » soit clairement inclus ».

Objectifs du projet : « Un accès durable aux données et aux matériaux pour notre public; relier les données dans le temps (par exemple pour plusieurs cycles de recensement); assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement) et la récupération partielle des coûts grâce à la génération de revenus ».

Co-panéliste

Glen Hohlmann

Directeur adjoint

Division des opérations du recensement

Statistique Canada

Courriel : glen.hohlmann@statcan.gc.ca

Bio: Glen Hohlmann est le Directeur adjoint de la Division des opérations du recensement de Statistique Canada, responsable du sous-projet de Diffusion du recensement. Il est titulaire d'une maîtrise en géographie de l'Université Carleton. Au cours de ses 20 ans de carrière à Statistique Canada, M. Hohlmann a occupé un certain nombre de postes dans les Divisions du recensement et de la géographie. Pour le recensement de 2021, il dirige pour la deuxième fois le sous-projet de diffusion du recensement. Il se concentre actuellement sur la diffusion des données du recensement de 2021 et s'assure que les produits du recensement sont opportuns, pertinents et accessibles afin de mieux servir

les Canadiens.

Co-panéliste

David Price

Gestionnaire

Tableaux de données du recensement

Statistique Canada

Courriel : david.price2@statcan.gc.ca

Description du projet : L'objectif global du Recensement de la population de 2021 est de faire en sorte que les Canadiens aient accès à des renseignements statistiques à jour, pertinents et de qualité sur l'économie et la société canadiennes qui sont en pleine évolution, pour alimenter les débats de fond, la recherche et la prise de décisions sur des questions d'ordre social et économique. Grâce au recensement de la population, Statistique Canada produit des renseignements sur les changements qui touchent la population canadienne et l'évolution des caractéristiques démographiques, ethnoculturelles et socio-économiques. Ces renseignements posent les fondements à la prise des décisions dans les secteurs public et privé, ainsi que pour la recherche et l'analyse dans des domaines d'intérêt pour les Canadiens. L'objectif du sous-projet de diffusion est de produire et de mettre les données pertinentes entre les mains des utilisateurs en temps voulu et de s'assurer qu'ils disposent des outils pour les utiliser et les comprendre. En accord avec les objectifs généraux du programme du recensement, le sous-projet de diffusion du recensement vise à promouvoir l'utilisation des données autant que possible, en les rendant accessibles et pertinentes pour les besoins d'information d'un large éventail d'utilisateurs.

Provenance du projet : Sous la forme d'une initiative gouvernementale.

But du projet : Rendre les données accessibles au public pour une utilisation générale.

Règles concernant la diffusion : Non, tous les utilisateurs ont un accès identique.

Sources de financement : Fonds gouvernementaux.

Conditions de financement : Sujet à des validations fréquentes.

Public du projet : Le public en général, des instances gouvernementales, la communauté des chercheurs universitaires, des organismes à but non lucratif ou de bienfaisance, des entreprises, tous les groupes d'utilisateurs.

Objectifs du projet : Fournir l'accès aux données du recensement à tous les Canadiens.

Panéliste

Elaine Castonguay

Directrice adjointe

Centre de géomatique statistique

Statistique Canada

Courriel : elaine.castonguay@statcan.gc.ca

Bio: Je suis la directrice adjointe au Centre de géomatique statistique de Statistique Canada. Je suis responsable du sous-projet du Registre des adresses et de la géographie (RAGéo) et de la tâche de diffusion des données géographiques du recensement (Géo Diffusion). Combinés, les sous projets RAGée et GéoDiffusion supportent toutes les activités du recensement en fournissant les bases de données, des adresses, les régions géographiques et le réseau routier pour supporter toutes les

activités de collecte, de traitement de données, la certification, la pondération, l'évaluation et la diffusion.

Description du projet : (cela apparaît dans les documents de l'atelier).

Provenance du projet : Sous la forme d'une initiative gouvernementale.

But du projet : Rendre les données accessibles au public pour une utilisation générale.

Règles concernant la diffusion : Comme je suis responsable des données géographiques - spatiales et non spatiales, il n'y a pas de règles de confidentialité, peu importe les utilisateurs.

Sources de financement : Fonds gouvernementaux.

Conditions de financement : Financement stable et continu.

Préoccupations concernant le financement : Non, le développement de concepts géographiques pour le recensement, la mise à jour de ces concepts et la diffusion sont requis pour la diffusion des données démographiques et statistiques du recensement

Public du projet : Le public en général, des instances gouvernementales, la communauté de la recherche académique, des organismes à but non lucratif ou de bienfaisance, Des entreprises

Objectifs du projet : Fournir l'accès aux données

Panel 2 – Diffuser les microdonnées : Perspectives internationales

Panéliste

Deirdre Lynch

Statisticienne

Diffusion du recensement

Bureau central des statistiques

Courriel : deirdre.lynch@csso.ie

Bio: Je m'appelle Deirdre Lynch et je suis statisticienne au *Central Statistics Office* d'Irlande depuis 4 ans. J'ai une formation en économie et en géographie, et j'aime beaucoup la démographie. J'aime travailler sur les données du recensement car je m'intéresse toujours aux questions auxquelles elles peuvent répondre. Je travaille au sein d'une équipe formidable de 7 personnes dans la section de diffusion du recensement et nous répondons constamment aux demandes de données de recensement du public. Nous entreprenons notre prochain recensement le 3 avril 2022, alors nous attendons cette période avec impatience. Mon travail consiste principalement à établir des tableaux, à répondre aux questions parlementaires, à créer des fichiers généraux de recherche (« Research Master Files » ou RMF), à approuver les résultats, à planifier le prochain recensement, à visualiser les données et bien plus encore.

Description du projet : Titre de la présentation - Une perspective irlandaise : Fichiers de microdonnées de recherche du recensement. Je présenterai un aperçu général de la manière dont l'Irlande a créé et continuera de créer des RMF (*Research Microdata Files*) spécifiques au recensement. Dans cette présentation, je donnerai un aperçu général des contrôles d'accès, du portail d'application et de l'approche basée sur les règles utilisées pour générer les RMF et les approbations de sortie. J'indiquerai ensuite où l'on peut trouver des informations complémentaires sur le site Internet du CSO.

Provenance du projet : Sous la forme d'un projet de recherche académique.

But du projet : Utiliser les données pour résoudre une question de recherche ou un problème public particulier.

Règles concernant la diffusion : Oui, seuls les chercheurs universitaires et les organisations à but non lucratif (ou les entreprises mandatées pour effectuer ces recherches) peuvent avoir accès à nos RMF.

Sources de financement : Il n'y a aucun coût associé à l'accès à notre RMF du recensement et il est créé en vertu du mandat des statisticiens chargés de la diffusion du recensement. Il s'agit donc de fonds gouvernementaux mais il n'y a pas de subvention disponible. La mise en place de la nouvelle unité de coordination de la recherche, chargée d'héberger et de contrôler les accès / applications, etc. a toutefois entraîné des coûts importants qui ont eux aussi été financés par le gouvernement.

Conditions de financement : Il s'agit d'un rôle permanent puisqu'il est inclus dans le *Central Statistics Office* et que celui-ci est permanent.

Préoccupations concernant le financement : Je pense que cette question n'est pas applicable ici étant donné que le financement est sous la responsabilité du statisticien du CSO qui est une profession permanente de la fonction publique.

Public du projet : La communauté des chercheurs universitaires

Objectifs du projet : Préservation des formats originaux, assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement), la préparation de rapports réguliers.

Panéliste

Callum Foster

Directeur des opérations, des produits et de la diffusion

2021 Census

Bureau des statistiques nationales

Courriel : callum.foster@ons.gov.uk

Bio: J'ai participé aux trois derniers recensements pour l'Angleterre et le Pays de Galles, en travaillant à chaque fois sur la partie des résultats et de la diffusion des opérations. Entre 2002 et 2012, j'ai également été responsable de la gestion d'un site Internet et d'un service de « *Neighbourhood Statistics* » qui rassemblait des données sur les petites zones provenant de différentes sources gouvernementales et les rendait disponibles via un seul service axé sur la compréhension des zones de privation sociale. Je travaille actuellement sur les préparatifs de la publication des résultats du recensement de 2021, notamment sur des moyens novateurs pour offrir plus de flexibilité aux utilisateurs.

Description du projet : Gestionnaire responsable de la production et de la diffusion des données de recensement à l'*Office for National Statistics*. Cela inclut la planification et la préparation des systèmes de sortie et de diffusion pour le recensement de 2021, ainsi qu'un service de soutien continu à l'utilisation des données du recensement de 2011 et des précédents. Nous sommes également en train de numériser une grande partie de notre contenu historique dans le but d'améliorer l'accès aux données du recensement au fil du temps.

Provenance du projet : Sous la forme d'une initiative gouvernementale

But du projet : Rendre les données accessibles au public pour une utilisation générale et aussi mettre d'autres ensembles de données à la disposition de publics spécifiques dans des cadres plus sûrs et mieux gérés. Il s'agit également de concevoir et de développer des produits et des systèmes pour le recensement actuel, tout en fournissant un soutien aux clients et un accès continu aux données des recensements précédents.

Règles concernant la diffusion : Oui, dans la mesure où nous permettons à certains publics d'utilisateurs d'avoir un accès plus large aux données selon certaines conditions.

Sources de financement : Fonds gouvernementaux

Conditions de financement : Sujet à des validations fréquentes.

Préoccupations concernant le financement : Pas vraiment, mais j'aimerais toujours pouvoir disposer de plus de fonds pour pouvoir en faire plus.

Public du projet : Le public en général, des instances gouvernementales, la communauté des chercheurs universitaires, des organismes à but non lucratif ou de bienfaisance, des entreprises, écoliers, des groupes communautaires spécifiques et d'autres groupes.

Objectifs du projet : Un accès durable aux données et aux matériaux pour notre public; préservation des formats originaux; relier les données dans le temps (par exemple pour plusieurs cycles de recensement); assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement); innover en matière de diffusion des données (par exemple sur le format d'accès aux données, la convivialité, ou la liaison des données), l'accès libre aux données, la préparation de rapports réguliers, l'implication de partenaires spécifiques.

Panéliste

Lara Cleveland

Gestionnaire de projet, IPUMS International

Chercheuse scientifique, ISRDI

Courriel : cleveland@umn.edu

Bio: Je dirige le projet Données de recensement et d'enquêtes internationales (*International census and survey data*) de l'IPUMS. En plus de superviser l'intégration des données et le développement des métadonnées pour le projet, je suis également responsable de la planification stratégique et des relations avec les partenaires. J'ai commencé à travailler pour l'IPUMS (*Integrated Public Use Microdata Series*) en 2005 en codant et en harmonisant des variables professionnelles historiques tout en terminant mes études supérieures. Je suis passée à IPUMS-International en 2008 et je suis titulaire d'un doctorat en sociologie de l'Université du Minnesota. Mes intérêts de recherche comprennent les données et les méthodes, l'harmonisation et le développement des variables géographiques, les professions et le travail, et les pratiques de normalisation mondiales. J'ai fait partie des équipes qui ont mené les travaux sur l'estimation de la variance, la documentation des échantillons, l'harmonisation géographique, les indices d'actifs des ménages, l'enclave de données de recherche à accès sécurisé et l'élaboration de matériels didactiques pour l'IPUMS.

Panel 3 – Diffuser les microdonnées : Perspectives canadiennes

Panéliste

Alexandra Clemence

Cheffe, Orientation et Généalogie

Bibliothèque et Archives Canada

Courriel : alexandra.clemence@bac-lac.gc.ca

Bio: Actuellement cheffe de l'orientation et de la généalogie à Bibliothèque et Archives Canada, j'ai auparavant travaillé comme bibliothécaire de référence au gouvernement et dans des établissements universitaires.

Description du projet : Bibliothèque et Archives Canada possède une vaste collection de documents de recensement canadiens de 1640 à 1926, et pour Terre-Neuve de 1671 à

1945. L'équipe de généalogie de BAC aide à mettre en contexte et à diffuser les données du recensement au public, et planifie les futures publications de documents.

Provenance du projet : Sous la forme d'une initiative gouvernementale.

But du projet : Rendre les données accessibles au public pour une utilisation générale

Règles concernant la diffusion : De façon générale, tous les utilisateurs ont le même accès en vertu de la législation sur la vie privée, cependant les chercheurs associés à des ministères peuvent avoir accès à des fichiers contenant des informations personnelles dans le cadre de certaines tâches ou projets.

Sources de financement : Fonds gouvernementaux.

Conditions de financement : Financement stable et continu.

Préoccupations concernant le financement : Puisque la dissémination des données du recensement est obligatoire, le financement à long terme n'est pas une préoccupation, cependant le degré d'indexation et les méthodes de dissémination peuvent être affectées.

Public du projet : Le public en général, la communauté des chercheurs universitaires.

Objectifs du projet : Un accès durable aux données et aux documents pour notre public; la préservation des formats originaux.

Co-panéliste

Sandra Sawchuk

Bibliothécaire

Mount Saint Vincent University

Courriel : sandra.sawchuk@msvu.ca

Bio: Sandra Sawchuk est la bibliothécaire responsable des données et de l'expérience des utilisateurs aux Bibliothèque et archives de l'Université Mount Saint Vincent à Halifax, en Nouvelle-Écosse. Elle occupe actuellement le poste de coordonnatrice de la formation pour la région de l'Atlantique dans le cadre de l'Initiative de démocratisation des données (IDD). Elle a une formation universitaire en humanités numériques, et ses intérêts de recherche comprennent le sauvetage et la réutilisation des données. Elle a récemment coécrit un article sur la reproductibilité computationnelle et participe actuellement à une subvention de partenariat de deux ans du CRSH visant à améliorer l'accès aux données historiques du recensement du Canada.

Co-panéliste

Susan Mowers

Bibliothécaire de recherche (Données)

Bibliothèque

Université d'Ottawa

Courriel smowers@uottawa.ca

Bio: Susan Mowers est bibliothécaire de données à l'Université d'Ottawa. Par le biais de ses activités liées aux services de données et à l'enseignement, Susan promeut la littératie, la découverte, le référencement, l'accès et l'utilisation des données. Susan a co-écrit un guide sur le référencement des microdonnées publiques canadiennes, a présidé la communauté de données de l'OCUL et participe à une subvention de partenariat de deux ans du CRSH (PDDRC). Elle poursuit des travaux sur le recensement historique.

Description du projet : Les panélistes présenteront la portée et les moyens d'accès aux précieuses données et documents du CCRI pour les recensements décennaux de 1911

à 1951. On espère que le projet de PDDRC (Partenariat de découverte des données du recensement du Canada), en couvrant les données et les documents de recensement de l'infrastructure de recherche sur le Canada au 20^e siècle (IRCS), améliorera la connaissance et la facilité de repérage de cette information. Le projet IRCS est une initiative pancanadienne, multidisciplinaire et multi-institutionnelle dont l'objectif est de développer des bases de données à partir des listes nominatives des recensements canadiens tenus entre 1911 et 1951.

Provenance du projet : Sous la forme d'un projet de recherche académique.

But du projet : Rendre les données accessibles au public pour une utilisation générale (CDDPP).

Règles concernant la diffusion : La capacité du PDDRC d'améliorer l'accessibilité à certaines parties des données historiques du recensement canadien est limitée en vertu de la période de non-divulgaration de 92 ans de la Loi sur la statistique. En raison de la valeur des données historiques du recensement et des documents connexes, nous voulons rendre accessible la plus grande partie possible des données de l'IRCS.

Sources de financement : Fonds gouvernementaux.

Conditions de financement : Subvention unique.

Préoccupations concernant le financement : Tenir à jour et entretenir le portail de découverte.

Public du projet : Le public en général, la communauté des chercheurs universitaires.

Objectifs du projet : Un accès durable aux données et aux matériaux pour notre public, assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement), l'accès libre aux données.

Panéliste

Kris Inwood

Professeur

Économie et Histoire

University of Guelph

Courriel : kinwood@uoguelph.ca

Bio: Kris Inwood est membre du corps professoral de l'Université de Guelph depuis 1986. Il est associé conjointement au Département d'histoire et au Département d'économie. Ses recherches actuelles portent sur la mobilité sociale et sur les différences ethniques en matière de niveau de vie et d'incarcération sur la base des données administratives et du recensement historique.

Description du projet : Le projet *The Canadian Peoples* (TCP) / Les Populations canadiennes (LPC) prépare les données du recensement à un dénombrement complet de 1852 à 1921 pour qu'elles puissent être utilisées à des fins de recherche.

Provenance du projet : Sous la forme d'un projet de recherche académique.

But du projet : Utiliser les données pour résoudre une question de recherche ou un problème public particulier.

Règles concernant la diffusion : Oui.

Sources de financement : Subventions gouvernementales, incluant les subventions à la recherche.

Conditions de financement : Fondation canadienne pour l'innovation.

Préoccupations concernant le financement : La maintenance et la mise à jour des données posent problème.

Public du projet : La communauté de la recherche académique.

Objectifs du projet : Un accès durable aux données et aux matériaux pour notre public, relier les données dans le temps (par exemple pour plusieurs cycles de recensement); assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement); innover en matière de diffusion des données (par exemple sur le format d'accès aux données, la convivialité, ou la liaison des données).

Panéliste

Michael Haan

Professeur associé

Sociologie

Western University

Courriel : mhaan2@uwo.ca

Bio: Michael Haan (doctorat de l'Université de Toronto en 2006) est Professeur agrégé et directeur académique du CDR de l'Université Western. De 2005 à 2010, il a été titulaire d'une charge professorale à l'Université de l'Alberta, où il a également été récipiendaire de la bourse de recherche Winspear-Archer en immigration et politique sociale. De 2010 à 2015, il a été titulaire de la chaire de recherche du Canada en politique démographique et sociale à l'Université du Nouveau-Brunswick. Il est également chercheur associé au *Prentice Institute for Global Population and Labour* de l'Université de Lethbridge et au *McGill Centre for Population Dynamics*. Ses recherches portent sur la démographie, l'établissement des immigrants, l'intégration au marché du travail et le développement des données. Le professeur Haan est fréquemment sollicité par les gouvernements provinciaux et fédéral pour des conseils stratégiques dans les domaines de l'immigration, des services d'établissement, du marché du travail canadien et du vieillissement de la population. Il est actuellement chercheur ou cochercheur dans le cadre de projets de recherche de plus de dix millions de dollars axés sur l'établissement des immigrants, le développement de communautés d'accueil et l'identification des facteurs qui prédisent la rétention des nouveaux arrivants. Depuis l'obtention de son doctorat en 2006, il a déjà publié plus de 75 articles et rapports sur ces questions.

Description du projet : Les centres de données de recherche (CDR) sont situés sur des campus universitaires partout au Canada et leur personnel est composé d'employés de Statistique Canada. Ces centres sont accessibles aux chercheurs et chercheuses accrédités affiliés à l'organisation d'accueil. La pandémie de COVID-19 a clairement souligné la nécessité pour Statistique Canada de passer de l'infrastructure physique des CDR vers de nouveaux modes d'accès aux données.

Provenance du projet : Sous la forme d'un projet de recherche académique.

But du projet : Utiliser les données pour résoudre une question de recherche ou un problème public particulier.

Règles concernant la diffusion : Non.

Sources de financement : Fonds gouvernementaux.

Conditions de financement : Sujet à des validations fréquentes.

Préoccupations concernant le financement : Non.

Public du projet : Le public en général, des instances gouvernementales, la communauté des chercheurs universitaires, des organismes à but non lucratif ou de bienfaisance, des entreprises.

Objectifs du projet : Un accès durable aux données et aux matériaux pour notre public; préservation des formats originaux; relier les données dans le temps (par exemple pour

plusieurs cycles de recensement); assurer la cohérence des données dans le temps (pour plusieurs cycles de recensement); innover en matière de diffusion des données (par exemple sur le format d'accès aux données, la convivialité, ou la liaison des données); l'accès libre aux données, la récupération partielle des coûts grâce à la génération de revenus.

Appendice B : Vue des paramètres du portail de découverte du recensement.

Cette annexe résume les commentaires des panélistes concernant les fonctionnalités de leur plateforme de données du recensement.

Only in English

For aggregate census data

National Historical Geographic Information System (NHGIS)

<https://www.nhgis.org/>

Jonathan Schroeder

Targeted at users who want to download data.

Easy data retrieval steps:

- 1 Landing page is literally a Get data button that takes users to 4 filters: Years, Geographic levels, Datasets and Topic.
- 2 Immediately after this page, users are taken to dataset results by Source data, Time series tables, and GIS files.
- 3 Time series tables are built on harmonized census variables and value codes, and geography. These table lists include metadata on the type of geographic integration, years, geographic levels, short descriptive table names.

In development,

- API support, R support
- Looking at new tabulations from historical microdata

Email: gbhgis@port.ac.uk

Mention of new interface of the **US Census Bureau**

<https://data.census.gov/cedsci> - Targeted at all users. Some difficulties, e.g., finding basic tables.

Vision of Britain

<https://www.visionofbritain.org.uk/>

Humphrey Southall

- Targeted at the general public, and not researchers or others wanting to map data
- Designed to find basic information immediately

Access:

1-a - *Main means of access: their hierarchical geographic structure is exposed to Google. Most of their users come to their site from a web search - they land on the "Place page"*

- 1-b - Their website landing page: search by place (via name or map).
- 2 Place page
 - immediate place information, e.g., summarized gazetteer information
 - Sidebar (the spine). Option: "Units and statistics"
- 3 Units and statistics links page. Units = administrative units. 2 tabs: Units named after this place (by themes), Units covering this place
- 4. Select unit (geographic level)
- 5 Place through time page, See tabs: Historical statistics, Key Findings ...
- 6 Key findings page, Views: Chart, Table, Definitions and source information.
 - Project funding supported focus group testing of a demo site. Major funding was received in 2001 to build the website.
 - Delivers very popular local area census time series. Users prefer these to cross section data. Site complexity is hidden from users.
 - Geographic structure is exposed to google, landing users on its Place pages. About 300 time series based on selected published tables
 - All data is in one database @19,498,650 rows (statistical units). They took complex administrative geographic units and grouped them into "places" primarily based on name.

Census of Population (Statistics Canada)

<https://www12.statcan.gc.ca/census-recensement/index-eng.cfm>

Glen Hohlmann, David Price

- Targeted at all users
- Includes, but not limited to,
 - [Census Program Data Viewer](#) (interactive map, 2016, 2021)
 - by year: Profiles & Data tables
- (A) Formats:
 - 1) Beyond 20/20. Why? (a) Usability: Ease of use / Supports data use for a wide audience, (b) Feedback received from Census 2021 consultations to not get rid of B2020 e.g., for the Data Files (i.e., multidimensional tables), (c) The 2021 Census Data tables will contain many billions of cells this year
 - 2) CSV and SDMX (all tables). Why? a) Neutral format / versatile (CSV), b) Preservation formats
- (B) Usage. The Census Profiles and Data Tables are well-used. The Census Profiles receive approximately 5 million views a year.

Geography (Statistics Canada)

<https://www.statcan.gc.ca/en/geography>

Elaine Castonguay

- (A) [Geography page](#)
 - Includes intercensal products
- (B) [Census Geography page \(2021 Census\)](#)
 - Spatial information (GIS) products
 - Maps
 - Attribute information products
 - Reference documents & more

- (C) [Standard geographic classification \(SGC\)](#)
 - Standard geography, harmonized over time, includes Reference Maps. The SGC (2021) is in its 11th edition.
 - The main goal is to support the Census

For census microdata - International
(includes mention of aggregate census data)

Central Statistics Office (Ireland)

<https://data.cso.ie/#>

Deirdre Lynch

CSO open data

A) PxStat (aggregate census data tables)

- Targeted at all users from [CSO web site](#)
 1. Browse
 2. People and Society
 3. various Censuses of Ireland

Notes:

- CSO has moved away from 2 and 3 dimensional tables with geography
- Every year about 1,200 tables
- [PxStat](#) is a Data Dissemination Management System for publishing Statistics in Open Data formats

B) Special tabulations by small geographies created for public organizations

1. [Census 2016 Small Area Population Statistics](#)
 2. [SAPMAP](#)
- Targeted at power users and users interested in their area
 - Popular map interface

C) Anonymized Microdata Files (AMF) via [Irish Social Science Data Archive](#) and via [IPUMS-International](#)

Office for National Statistics (UK)

[Census 2021](#)

Callum Foster

Demonstrated two innovative tools for providing flexible aggregate data tables from underlying census microdata. The tools address the needs of two different user groups, and make more data available than the standard, or ready-made outputs (which however, still provide blanket geographic coverage (all areas), are good for their level of comparability and for users who may not want to build their own tables.

INNOVATIVE TOOLS:

A) First tool: flexible tables

- Targeted at a more expert statistician user and uses Contabula to which the ONS contributed
 1. Choose your population (e.g., "People with households, England and Wales")

2. Choose your geographic level, e.g., "Lower layer super output area". When you click on the area, "Your table" displays the number of cells / areas, and the name of the dataset (People with households, England and Wales)
 3. Select specific (by name) areas, all areas (whole population), or all areas within a larger area.
 4. Choose your variables (and breakdown detail: number of categories)
 5. (Disclosure control process starts, and *interface displays* "Your table" details including the disclosure success rate as the number of areas that pass out of total number of areas, and an infographic)
 6. In relation to your success rate, you can adjust the level of chosen detail and add or take away variables.
 7. You can view or download your data
 8. Several download options include, (a) changing your dataset, (b) geographic level (e.g., Middle Layer Super Output Level or Local Authority),(c) geographic area, and (d) variables. Seamlessly calling up the relevant selection screen with your success rate information. Filter table.
 9. Supports different output formats
- (B) Second tool is a potential interface that sits on top of the first tool.
- Population Explorer is an exploratory tool
 - As an exploratory tool it brings data to life, and quickly provides insights from the data in a more accessible form
 - It is targeted at a more policy influencer type user who needs data to support their everyday work, but doesn't use data everyday.
1. Select a topic, e.g., Arrival in UK, length of time, and passports held
 2. Shows percent of total population in England and Wales, median age, age profile, a choropleth map, and top 5 areas with highest and lowest percentages
 3. Shows key indicators for General health, Marital status, Social grade, Economic activity, Distance to work, Mode of travel to work
 4. Can scroll over the map to identify geographical differences (% of population)
- (C) Third tool - Flex-It
- Targeted at the public
 - ONS to introduce Flex-It tools for various ready-made standard Census 2021 tables so that, e.g., users can swap variables.

IPUMS-International

<https://international.ipums.org/international/>

Lara Cleveland

- Narrower audience than other IPUMS projects: Not to give data to for profit entities

Database user tools and standards

1. Access via ...

- Different modes of dissemination for academic users
 - gives microdata because academic users want the microdata on their computers
- Metadata is essential for responsible use (rely on good sample design documentation)
- Increasing interest shown by the policy sector

2. Facilitating and managing open dissemination via ...
 - Harmonization / standardization (e.g., variables) is extremely helpful and efficient
 - Use registration to alert users to changes/errors, make connections
 - Web automation for fulfillment of requests
 - Standardized linking fields - especially geography
 - Three codebook options:
 - General level of codes (IPUMs harmonized codes),
 - Detailed level of codes (IPUMS harmonized codes + the comparable country codes),
 - Fully harmonized codes (only the cross-censuses *common* harmonized codes)

For census microdata - Canada

Library and Archives Canada

17 Census databases, 1825 - 1926

<https://www.bac-lac.gc.ca/eng/census/Pages/census.aspx#b>

Alexandra Clemence

Census metadata includes:

- column headings
- instructions to enumerators and other resources
- section about relevant issues & problems..

Database searching

- Each database currently has its own search tool. Typical search tool:

1. Search screen

- Searching by: keywords, names, ages, provinces, names and numeric identifiers of Districts and Sub-Districts, page numbers.

- Includes search help

2. Click on the entry

3. Digitized image appears (click on the image to see the full image for the census return)

4. Click the item number for archival reference and to suggest a correction

CCRI (Canadian Century Research Infrastructure)

Covering various web sites

Susan Mowers and Sandra Sawchuk

Demonstration of [CCRI Guide \(preliminary version\)](#) covering CCRI resources found on various platforms and where to find them.

1. The five CCRI Canadian Census microdata files (“databases”), (1911, 1921, 1931, 1941, 1951) form the core of CCRI’s data
2. The five microdata files feature harmonized census variables and harmonized geography.
3. The guide also covers metadata, spatial data, reference maps, and introductory selected published Census tables.

4. The CCRI archive "web gateways"

Where available in French or bilingually, or in English, these include:

- [University of Alberta](#)
- [University of Alberta Dataverse](#)
- [l'Université du Québec à Trois Rivières](#)
- [University of Toronto](#)
- [University of Ottawa](#)
- Alternate platforms: data repositories housing CCRI resources

5. Further CCRI partner university websites

Where available in French or bilingually, or in English, these include:

- [The CCRI project lead website at the University of Ottawa](#)
*Note, the above website was unavailable at the time of writing.
Please note instead [the 2019 website capture](#) via the Internet Archive Wayback Machine.*
- [CCRI York University Centre website](#)
- [CCRI-ICRS 1911 Sample of the 1911 Canadian Census website \(Université du Québec à Trois-Rivières\)](#)