

## Exploration des données ouvertes et des projets ouverts : La BDOI – Une base de données ouverte expérimentale sur les immeubles

Alessandro Alasia, Jean Le Moullec et Marina Smailes<sup>1</sup>

### Résumé

Cette initiative exploratoire vise à accroître l'utilisation et l'harmonisation des données ouvertes sur les immeubles provenant de sources gouvernementales en vue de contribuer à la mise sur pied d'une base de données complète, exhaustive et ouverte sur les immeubles au Canada. Le travail exploratoire a mené à la création d'une première version de la Base de données ouvertes sur les immeubles (BDOI), un répertoire centralisé et harmonisé des données sur les immeubles publiées en vertu de la Licence du gouvernement ouvert du Canada. La version actuelle (version 1.0) de la BDOI contient environ 4,3 millions d'empreintes d'immeubles et 10 variables. L'une des principales caractéristiques de ce travail est qu'il exécuté au moyen d'une approche fondée sur un projet ouvert, qui consiste à privilégier d'autres options (données ouvertes et source ouverte) lorsqu'on met en œuvre des projets, dans le but de faciliter la collaboration et d'améliorer la transparence et l'efficacité.

Mots-clés : Données ouvertes; microdonnées des municipalités; empreintes d'immeubles; projets ouverts.

### 1. Introduction

#### 1.1 Exploration des données ouvertes et des projets ouverts

Au cours des deux dernières années, le Laboratoire d'exploration et d'intégration des données (LEID) du Centre des projets spéciaux sur les entreprises de Statistique Canada a entrepris plusieurs projets ayant pour thème les données ouvertes et les sources ouvertes. Le lancement d'un projet pilote consistant à évaluer l'approche participative comme instrument de collecte de données pour les statistiques officielles a été la première étape des travaux amorcés en 2016. Il était lié aux empreintes d'immeubles et à leurs attributs (leurs adresses). Pendant le projet pilote, ainsi que lors des projets subséquents, les nombreux avantages et possibilités des microdonnées ouvertes et des logiciels ouverts ont été mis en évidence. On a également découvert que l'utilisation combinée de ces solutions gratuites et accessibles, plutôt que de logiciels fermés<sup>2</sup>, facilite grandement le partage de données et la collaboration non seulement avec les autres ministères, mais également avec le grand public.

Les leçons tirées pendant et après le projet pilote ont jeté les bases qui ont mené à l'élaboration du concept de *projet ouvert*. Le concept est lié à l'idée selon laquelle lorsqu'on favorise, dans la mesure du possible, des solutions ouvertes pour la mise en œuvre de projets, de nombreux avantages directs et indirects peuvent en découler, notamment une plus grande transparence et efficacité dans la réalisation des projets.

Les travaux présentés ici portent sur la création d'une base de données ouverte sur les immeubles, qui constitue la première initiative de l'équipe du LEID consistant à mettre en œuvre des travaux conformément au concept de *projet ouvert*. On décrit également dans les pages qui suivent ce qu'est la base de données ouverte sur les immeubles et comment les données sont compilées. Le contenu présenté ici se trouve également dans le document sur les métadonnées qui accompagne la BDOI lorsqu'on la télécharge, bien qu'il ne soit pas tout à fait identique.

#### 1.2 L'importance croissante des données ouvertes pour les organismes statistiques nationaux

---

<sup>1</sup> Alessandro Alasia, Statistique Canada, 100 promenade du pré Tunney, Ottawa (Ontario), Canada, K1A 0T6; Jean Le Moullec, Statistique Canada, 100 promenade du pré Tunney, Ottawa (Ontario), Canada, K1A 0T6; Marina Smailes Statistics Canada, Statistique Canada, 100 promenade du pré Tunney, Ottawa (Ontario), Canada, K1A 0T6

<sup>2</sup> Par exemple, les logiciels privés ou encore, les données dont l'utilisation est assortie de restrictions.

La transformation numérique influence tous les domaines de notre vie sociale et de notre économie, faisant du flux de données et d'information un aspect essentiel de pratiquement toutes les activités commerciales. La transformation a une incidence sur les organismes nationaux de statistiques (ONS). À titre d'organismes publics ayant pour mandat principal la collecte, la production et la gestion de données, plusieurs ONS ont amorcé ces dernières années un processus de modernisation, influencé en grande partie par les occasions et la nécessité de réagir aux nouvelles réalités dans un contexte d'abondance de données.

Des acteurs locaux, comme des organismes publics infranationaux ou des organisations privées, produisent de plus en plus de données numériques, et il s'agit d'un aspect particulièrement pertinent pour les statistiques territoriales. Les entreprises du secteur privé et les gouvernements nationaux ne sont pas les seuls à prendre le virage numérique; les administrations locales en font tout autant, de sorte qu'un volume de données sans cesse croissant est généré par les gouvernements municipaux et provinciaux. Les données administratives et provenant de capteurs produites par les villes alimentent toujours plus les systèmes statistiques nationaux et elles sont de plus en plus susceptibles de servir à la production de statistiques territoriales. Ces entités sont devenues des producteurs actifs de statistiques locales en raison de la numérisation croissante des dossiers administratifs qu'elles compilent, gèrent et diffusent. Le développement des portails de données ouvertes locaux et municipaux n'est qu'un indice témoignant de cette tendance.

Le site Web du Gouvernement ouvert du Canada définit les données ouvertes comme étant des données structurées, lisibles par machine, qui peuvent être librement partagées, utilisées et mises à profit par quiconque, sans restriction<sup>3</sup>. Ces paramètres définis, les données ouvertes deviennent une source d'informations qui peuvent être utilisées de façon créative et améliorée, et qui sont accessibles librement à l'ensemble des organismes publics et privés. Les initiatives liées aux données ouvertes des gouvernements fédéral, provinciaux et municipaux contribuent à favoriser la distribution et l'uniformisation des données dans ce vaste écosystème.

Les données ouvertes et les outils ouverts peuvent être utilisés pour accélérer le partage des connaissances. L'utilisation de données ouvertes de sources non officielles est de plus en plus répandue et elle est facilitée par la diffusion des données sur des plateformes en ligne. Même si l'utilisation d'outils « ouverts » sera toujours limitée par des contraintes relatives à la protection de la vie privée, un grand éventail de questions de politique peut être abordé dans un espace public : les indicateurs des infrastructures et des installations publiques, des lieux récréatifs, des déplacements à pied, et des indicateurs similaires de la qualité de vie en milieu régional et urbain. La disponibilité sans cesse croissante de microdonnées ouvertes géoréférencées offre aux ONS des possibilités uniques pour l'établissement de statistiques infranationales.

### **1.3 Survol de la base de données ouverte sur les immeubles**

Pour explorer l'utilisation des données ouvertes en vue d'établir les statistiques officielles et de soutenir la recherche géospatiale dans divers domaines, on a entrepris des travaux pour créer une base de données sur les empreintes d'immeubles qui soit accessible, harmonisée et axée sur les données ouvertes ayant été publiées par plusieurs palliers de gouvernement au Canada<sup>4</sup>. Dans le présent document, on décrit en détail le processus de collecte, de compilation et d'uniformisation des divers ensembles de données sur les empreintes d'immeubles ayant servi à la création de la Base de données ouverte sur les immeubles (BDOI), disponible en vertu de la Licence du gouvernement ouvert – Canada<sup>5</sup>.

La version actuelle (version 1.0) de la BDOI contient environ 4,3 millions d'empreintes d'immeubles. On prévoit mettre à jour périodiquement la base de données à mesure que de nouveaux ensembles de données ouverts sur les immeubles sont rendus disponibles par des sources gouvernementales. Cet ensemble de données vectorielles est fourni sous forme de fichier « shapefile » compressé à l'échelle provinciale ou territoriale.

---

<sup>3</sup> Voir : Gouvernement du Canada, [Données ouvertes](#).

<sup>4</sup> Inclut des gouvernements municipaux, régionaux et provinciaux.

<sup>5</sup> Voir : Gouvernement du Canada, [Licence du gouvernement ouvert – Canada](#).

L'ensemble de données est accessible sur le site Web de Statistique Canada, à l'adresse suivante : <https://www.statcan.gc.ca/fra/donnees-ouvertes-immeubles/index>

## 2. Sources des données

De nombreuses sources de données ont été utilisées pour créer la BDOI. Les fournisseurs de données, qui comprennent les niveaux de gouvernement municipal, régional et provincial, sont indiqués au tableau 2-1. Les sources de données sont attribuées à leur fournisseur respectif, conformément aux exigences de la licence, à l'annexe A du document sur les métadonnées<sup>6</sup>. Le cas échéant, la version de la licence est également indiquée dans cette annexe. Pour en savoir plus sur les licences individuelles, le lecteur devrait consulter directement les informations fournies dans les portails de données ouvertes des fournisseurs de données en question.

**Tableau 2-1**

**Liste des fournisseurs de données compris dans la version actuelle de la Base de données ouverte sur les immeubles**

Alberta	Colombie-Britannique		Nouveau-Brunswick	Territoires du Nord-Ouest
1. Banff 2. Canmore 3. Chestermere 4. Cochrane 5. Edmonton 6. Grande Prairie 7. Lethbridge 8. Strathcona County	1. Chilliwack 2. Kamloops 3. Kelowna 4. Nanaimo 5. New Westminster 6. North Vancouver	7. Prince George 8. Saanich 9. Squamish 10. Surrey 11. Vancouver 12. Victoria 13. White Rock	1. Fredericton 2. Moncton 3. Saint John	1. Yellowknife
Nouvelle-Écosse	Ontario		Québec	Saskatchewan
1. Cape Breton 2. Halifax 3. Nova Scotia	1. Barrie 2. Brampton 3. Brantford 4. Burlington 5. Caledon 6. Cambridge 7. Région de Durham 8. Guelph 9. Hamilton 10. Comté de Huron 11. Kingston 12. Kitchener 13. Muskoka	14. Newmarket 15. Niagara Falls 16. Région de Niagara 17. Comté de Norfolk 18. Oakville 19. Ottawa 20. St. Catharines 21. Toronto 22. Waterloo 23. Région de Waterloo 24. Welland 25. Région de York	1. Longueuil 2. Montréal 3. Québec 4. Repentigny 5. Rimouski 6. Shawinigan 7. Sherbrooke	1. Regina

## 3. Période de collecte des données

Les données ont été recueillies à partir des portails de données ouvertes entre janvier et août 2018. La date de téléchargement des ensembles de données de chaque municipalité utilisés dans la BDOI est indiquée à l'annexe A du document sur les métadonnées<sup>7</sup>. Il convient de faire remarquer au lecteur que la date du téléchargement ne correspond

<sup>6</sup> Voir : Statistique Canada, [La BDOI – Une base de données ouverte sur les immeubles expérimentale](#).

<sup>7</sup> *Ibid.*

pas à la période de référence des données. Idéalement, la période de référence serait clairement indiquée; toutefois, comme les données composant les ensembles de données proviennent de différents fournisseurs de données, on peut difficilement connaître la période de référence réelle.

La période au cours de laquelle les données sous-jacentes sur les immeubles ont été recueillies ou mises à jour par la municipalité n'est pas toujours indiquée dans les portails sur les données ouvertes. Si le lecteur a besoin de renseignements précis sur la période de référence des données, il devrait communiquer avec le fournisseur de données approprié, indiqué à l'annexe A du document sur les métadonnées<sup>8</sup>.

#### 4. Population cible

Le but de la BDOI est de constituer un répertoire centralisé et harmonisé des empreintes d'immeubles dont les données proviennent de sources ouvertes de gouvernements d'un bout à l'autre du Canada. En compilant cette base de données, on a privilégié les immeubles où des activités importantes (p. ex., des activités économiques) sont susceptibles de se produire. Les empreintes inférieures à 10 m<sup>2</sup> ont donc été supprimées, car on les juge trop petites pour intéresser la plupart des utilisateurs. À mesure que le travail sur la BDOI expérimentale avance, les définitions et les seuils évolueront. Il est important de rappeler au lecteur qu'il est toujours possible d'obtenir directement les données non modifiées dans les portails de données ouvertes des divers fournisseurs de données.

#### 5. Méthodologie de compilation

Chaque ensemble de données sur les immeubles a été téléchargé du portail de données ouvertes respectif du gouvernement les ayant publiés. Les fournisseurs des 61 ensembles de données ont chacun joint un ensemble de variables différent pour les empreintes d'immeubles. Seules les variables communes à toutes les sources de données ont été incluses dans la base de données finale. Avant de les fusionner, tous les ensembles de données ont été uniformisés et nettoyés pour assurer l'harmonisation des mesures et des unités. Dans les rares cas où des ensembles de données municipaux se chevauchaient sur le plan géographique, une zone tampon de cinq mètres a été établie pour l'un des ensembles de données, et tous les immeubles des autres ensembles de données qui empiétaient sur cette zone tampon ont été supprimés.

#### 6. Dictionnaire de données

Le dictionnaire de données ci-contre décrit les variables contenues dans la BDOI exploratoire.

Variable - Object ID	
Nom	OBJECTID (alias FID)
Format	Code d'identification de l'objet (propre à ArcGIS)
Source	Générée à l'interne lors du traitement des données
Description	Identification unique d'objet ou de champ générée automatiquement lors du traitement des données

  

Variable - Forme	
Nom	Forme
Format	Géométrie (propre à ArcGIS)
Source	Générée automatiquement à partir des propriétés géométriques des ensembles de données polygonales
Description	Décrit le type de données vectorielles dans l'ensemble de données

  

Variable - Longitude	
Nom	Longitude
Format	Double

<sup>8</sup> *Ibid.*

Source	Générée d'après le centre du polygone de l'immeuble
Description	Longitude, en degrés décimaux, du centre du polygone de l'immeuble

Variable - Latitude	
Nom	Latitude
Format	Double
Source	Générée d'après le centre du polygone de l'immeuble
Description	Latitude, en degrés décimaux, du centre du polygone de l'immeuble

Variable – Identificateur unique de la subdivision de recensement	
Nom	SDRIDU
Format	Nombre entier
Source	Limites de la subdivision canadienne de recensement, 2016 (Produit GeoSuite de Statistique Canada)
Description	Numéro d'identification de la subdivision de recensement où l'immeuble est situé

Variable - Nom de la subdivision de recensement	
Nom	CSDNAME
Format	Chaîne
Source	Limites de la subdivision canadienne de recensement, 2016 (Produit GeoSuite de Statistique Canada)
Description	Nom de la subdivision de recensement

Variable – Fournisseur de données	
Nom	Data_prov
Format	Texte (chaîne)
Source	Créée à partir de la source de l'ensemble de données ayant servi d'intrant
Description	Nom de la municipalité, région ou province/territoire ayant fourni l'ensemble de données

Variable – Identificateur unique de l'immeuble	
Nom	Build_ID
Format	Texte (chaîne)
Source	Générée en concaténant l'identificateur unique de la subdivision de recensement avec un code séquentiel à 6 chiffres unique à chaque immeuble de cette subdivision de recensement
Description	Identificateur unique de chaque immeuble, d'après la subdivision de recensement où il est situé

Variable – Périmètre	
Nom	Shape_Length
Format	Flottant
Source	Calculée à l'aide des attributs géométriques du polygone de l'immeuble
Description	Périmètre de l'immeuble en mètres (nom spécifique à ArcGIS)

Variable - Superficie	
Nom	Shape_Area
Format	Flottant
Source	Calculée à l'aide des attributs géométriques du polygone de l'immeuble
Description	Superficie de l'immeuble en mètres carrés (nom spécifique à ArcGIS)

## 7. Cliché d'enregistrement

Un exemple de cliché d'enregistrement en format tableau est présenté ci-contre (tableau 7-1). Les champs marqués d'un astérisque sont ceux qui sont indexés automatiquement par le logiciel GIS utilisé lors du traitement.

Les sources de données ouvertes augmenteront au Canada, et nous espérons que les prochaines mises à jour enrichiront la BDOI en y ajoutant plus de variables et en élargissant la couverture géographique.

**Tableau 7-1****Exemple de cliché d'enregistrement de la BDOI**

OBJECTID*	Forme*	Longitude	Latitude	SDRIDU	SDRNOM	Fourn_donn	Imm_ID	Shape_Length	Shape_Area
1	Polygone	-115.561757	51.18907	4815035	Banff	Banff	48150350000001	16.560241	16.963528
2	Polygone	-115.569331	51.171372	4815035	Banff	Banff	48150350000002	87.531972	330.625531
3	Polygone	-115.569616	51.178173	4815035	Banff	Banff	48150350000003	104.044015	573.938947

## 8. Exactitude des données

Toutes les empreintes d'immeubles de la BDOI sont tirées de sources gouvernementales de données ouvertes. En général, les ensembles de données obtenus des divers portails de données ouvertes ont été laissés tels quels, à l'exception d'un traitement d'uniformisation des sources afin de constituer une seule base de données. On prévient le lecteur que les méthodologies utilisées ainsi que la quantité de renseignements disponibles sur ces méthodologies varient d'un fournisseur de données à l'autre. Pour en savoir plus sur les méthodologies d'un fournisseur de données, veuillez cliquer sur le lien menant à son site Web qui se trouve à l'annexe A du document sur les métadonnées<sup>9</sup>.

Durant la phase du traitement des ensembles de données en vue de créer la BDOI, plusieurs étapes ont été suivies pour accroître l'exactitude des données et uniformiser les données de sortie. Quelques immeubles dont la géométrie a été jugée improbable ont été supprimés. Pour assurer la cohérence des mesures et des unités (p. ex. les mesures de la superficie et du périmètre de l'immeuble), on a systématiquement retiré certaines variables des ensembles de données sous-jacents et on les a recalculées en se fondant sur une projection normalisée.

Il convient de noter que lorsque des ensembles de données d'entrée se chevauchent sur le plan géographique, les résultats dépendent du choix de la largeur de la zone tampon entre les deux ensembles (cinq mètres) et des segments d'immeubles chevauchants qui seront supprimés. Le choix des segments à conserver repose sur plusieurs facteurs heuristiques; la préférence a été accordée aux ensembles de données qui semblaient avoir des polygones plus détaillés ou compter un plus grand nombre d'immeubles.

La BDOI expérimentale ne contient que les données ouvertes gouvernementales qui étaient disponibles au moment de sa compilation, et ne devrait donc pas être interprétée comme étant un répertoire exhaustif et objectif de tous les immeubles du Canada. À l'heure actuelle, la BDOI ne couvre pas tout le Canada; la plupart des données proviennent des grandes villes. Les données manquantes s'expliquent par le fait qu'il y a des régions pour lesquelles on n'a pas pu trouver de données ouvertes gouvernementales sur les immeubles. Lorsque plus de données ouvertes seront accessibles et ajoutées à la BDOI, on s'attend à ce que la couverture soit meilleure.

## 9. Conclusion

Les microdonnées ouvertes des gouvernements municipaux, provinciaux et fédéral offrent d'immenses possibilités pour l'intégration des données, et Statistique Canada a un rôle à jouer dans cette intégration. La BDOI est un exemple, peut-être le premier, de la capacité des ONS de travailler dans un contexte de microdonnées ouvertes.

Durant la réalisation de travaux visant à améliorer la couverture de la BDOI, d'autres domaines statistiques sont examinés pour créer des bases de microdonnées ouvertes mieux harmonisées (p. ex., les adresses d'immeubles, les répertoires d'entreprises, les infrastructures publiques, ainsi que d'autres). Ces bases de données, ainsi que la possibilité de les relier, aideront les ONS à participer à la création et à la mise à jour d'un ensemble d'infrastructures de données de base qui est essentiel pour soutenir l'économie canadienne et la société dans un monde toujours plus numérisé et axé sur les données.

---

<sup>9</sup> *Ibid.*

Les nouvelles possibilités s'accompagnent de nouveaux défis, et on ne comprend pas encore tout à fait la plupart des défis que doivent relever les ONS qui travaillent dans un milieu de microdonnées ouvertes. Par exemple, l'utilisation de microdonnées ouvertes géoréférencées est limitée à des statistiques à caractère non personnel et non confidentiel, mais la définition exacte des données qui entrent dans ces catégories peut varier d'une administration à l'autre et elle est influencée en partie par la perception du public. Des questions concernant notamment la propriété, l'ouverture et le partage des données feront inévitablement partie de la discussion, qui pourrait avoir des implications uniques dans un système de gouvernance à plusieurs niveaux qui reconnaît habituellement les données ouvertes produites au niveau infranational.

## **Bibliographie**

Statistique Canada, (2019), *La BDOI – Une base de données ouverte sur les immeubles expérimentale, Document de métadonnées : concepts, méthodologie et qualité des données.*