

Statistique Canada et les données ouvertes

Bill Joyce¹

Résumé

La présente communication vise à fournir un bref aperçu de l'engagement de Statistique Canada à l'égard des données ouvertes. Nous aborderons tout d'abord la façon dont les principes des données ouvertes sont intégrés dans les pratiques courantes de diffusion de Statistique Canada. Puis, nous discuterons de la participation de l'organisme à l'initiative de données ouvertes de l'ensemble du gouvernement. Cette participation comporte deux volets : Statistique Canada est le principal fournisseur de données au portail de données ouvertes du gouvernement du Canada, mais il joue aussi un rôle important en arrière-plan, en tant que fournisseur de services responsable du développement et de la mise à jour du portail de données ouvertes (qui fait maintenant partie du portail plus large Gouvernement ouvert).

Mots Clés : Données ouvertes; Statistique Canada, Gouvernement du Canada.

1. Introduction

La présente communication vise à fournir un bref aperçu de l'engagement de Statistique Canada à l'égard des données ouvertes. Nous aborderons tout d'abord la façon dont les principes des données ouvertes sont intégrés dans les pratiques courantes de diffusion de Statistique Canada. Puis, nous discuterons de la participation de l'organisme à l'initiative de données ouvertes de l'ensemble du gouvernement. Cette participation comporte deux volets : Statistique Canada est le principal fournisseur de données au portail de données ouvertes du gouvernement du Canada, mais il joue aussi un rôle important en arrière-plan, en tant que fournisseur de services responsable du développement et de la mise à jour du portail de données ouvertes (qui fait maintenant partie du portail plus large Gouvernement ouvert).

2. Adoption des principes des données ouvertes dans nos activités de diffusion normales permanentes

À Statistique Canada, au cours des dernières années, on est passé graduellement et continuellement des produits imprimés aux produits en ligne et des produits payants aux produits gratuits. En 2012, Statistique Canada a fait un grand pas dans cette direction, lorsque tous les produits de données courants en ligne sont devenus gratuits. Parallèlement, nous avons supprimé tous les droits de licence et de redevances et, avec l'adoption d'une licence ouverte, nous avons supprimé les obstacles qui existaient auparavant quant à la redistribution de nos données.

Une part importante des données qui se trouvent dans notre site Web (données figurant dans [CANSIM](#), par exemple) sont depuis toujours téléchargeables et lisibles par machine, mais nous progresserons davantage dans cette direction, au cours des mois à venir, au moment où nombre de nos « publications de données » traditionnelles (en HTML et PDF) seront remplacées par des tableaux reposant sur des bases de données qui seront téléchargeables. Parallèlement, nous offrirons une IPA gratuite pour tous les utilisateurs, afin qu'ils puissent recueillir directement des données de nos bases de données agrégées publiques, au moment où cela leur convient, 24 heures par jour, 7 jours par semaine.

¹Bill Joyce, Statistique Canada, 100 promenade Tunney's Pasture, Ottawa, ON, K1A 0T6I (bill.joyce@canada.ca)

Nous sommes en mesure de dire que nous respectons les principes des données ouvertes dans nos activités normales permanentes de diffusion des données.

La définition de données ouvertes que nous utilisons est la suivante : les données ouvertes sont gratuites, leur redistribution est libre et elles sont lisibles par machine. Cette définition est conforme à celle utilisée partout dans le monde, mais certaines organisations vont plus loin et incluent le concept de « données couplées ». Dans l'évaluation à cinq étoiles des données ouvertes de Sir Tim Berners Lee, les 4^e et 5^e étoiles se rapportent aux données couplées. Les données formatées pour le Web sémantique représentent une autre façon de désigner ce concept, et des formats comme RDF et SPARQL sont utilisés.

Nous ne produisons pas de données formatées pour le Web sémantique. Les utilisateurs des données ne souhaitent pas obtenir des données dans ces formats, et certains pourraient prétendre que le concept est académique et théorique dans une certaine mesure. Par ailleurs, nous devons nous assurer que nous ne nous rendons pas compte, à un moment donné, que la demande pour alimenter le Web sémantique est réelle et que nous avons « manqué le bateau ». La position de Statistique Canada est que nous surveillerons le développement des données couplées, mais que nous n'investirons pas dans ce format pour le moment.

Nos fichiers de microdonnées, et même nos fichiers de microdonnées à grande diffusion, ne sont pas directement disponibles en ligne, et ils ne sont pas considérés non plus techniquement comme des données ouvertes. Cela ne veut pas dire que nous ne voulons pas augmenter l'accessibilité des microdonnées. Nous disposons d'une gamme d'options pour nous assurer que des microdonnées sont mises à la disposition des chercheurs qui en ont besoin, grâce à l'utilisation de l'accès à distance, de centres de données de recherche sécuritaires, etc. Nos fichiers de microdonnées à grande diffusion, toutefois, comportent des restrictions strictes en matière de licence, y compris, par exemple, l'interdiction de coupler ou de fusionner des fichiers. C'est pourquoi une véritable licence ouverte ne s'applique pas, et nous ne considérons pas ces données comme étant une composante de notre répertoire de données ouvertes.

Des totalisations personnalisées sont fournies à nos utilisateurs contre recouvrement des coûts. Ce produit continue d'être fourni avec une licence ouverte, et aucun droit de licence ou redevance n'est appliqué, mais le coût est fondé sur le temps qu'il faut à nos employés pour produire les travaux personnalisés. Nous croyons toujours pouvoir dire que nous respectons les principes des données ouvertes, étant donné que nous tentons toujours de fournir le plus grand nombre de données courantes gratuitement.

3. Statistique Canada comme principal fournisseur de données au portail de données ouvertes à l'échelle du gouvernement

Depuis le tout premier projet pilote d'un portail gouvernemental de données ouvertes au Canada, Statistique Canada est le principal fournisseur de données.

La saisie d'écran suivante, tirée de ouvert.canada.ca, montre que 72% des fichiers de données non géospatiales sont fournis par Statistique Canada.

Portail du gouvernement ouvert

Au moment où nous nous apprêtons à lancer cette recherche intégrée améliorée du Portail du gouvernement ouvert, il se pourrait que vous remarquiez que le nombre d'enregistrements sur l'information ouverte a été réduit. Veuillez prendre note qu'il vous est possible d'accéder aux publications qui n'y figurent plus en visitant les sites suivants : [Publications du gouvernement du Canada](#) et [Bibliothèque et Archives Canada](#). Si vous avez des commentaires ou des questions, [contactez-nous!](#)

Recherche de dossiers

Proposez un jeu de données

Recherche...

8 891 dossiers trouvés

Trier par date de modification

Non spatiales

Régime de pensions du Canada (RPC) - Nombre de nouvelles prestations, selon l...

Cette base de données fournit l'information sur le Régime de pensions du Canada (RPC) - Nombre de nouvelles prestations, selon le lieu de résidence et le type de prestation. Elle fait partie du Bulletin statistique sur le Régime de pensions du Canada et la Sécurité de la vieillesse - une...

Organisation: Emploi et Développement social Canada

Formats des données: CSV

Sécurité de la vieillesse (SV) - Nombre de nouvelles prestations, selon la pr...

Cette base de données fournit l'information sur le nombre de nouvelles prestations, selon la province et le type pour la Sécurité de la vieillesse (SV). Elle fait partie du Bulletin statistique sur le Régime de pensions du Canada et

Filtres de recherche [Effacer tout](#)

Type de portail

Données ouvertes (8891)

[Effacer tout](#)

Type de collection

Géospatiales (112700)

Non spatiales (8891)

Cartes Ouvertes (200)

Publication (54)

[Effacer tout](#)

Organisation

Statistique Canada (6418)

Agriculture et Agroalimentaire Canada (1603)

Ministère des Finances Canada (131)

Bibliothèque et Archives Canada (104)

Emploi et Développement

En tant qu'organisme statistique, nos données font toujours l'objet d'un contrôle de confidentialité et sont conçues pour être publiées. Ainsi, nous avons un grand avantage par rapport aux autres ministères fédéraux qui doivent, pour la première fois, adopter une approche « d'ouverture par défaut » en ce qui a trait à leurs fichiers de données.

Dans le portail de données ouvertes fédéral (qui fait maintenant partie du portail plus large Gouvernement ouvert), on ne conserve pas de double de nos ensembles de données. On y trouve plutôt un catalogue ou registre central, et chaque enregistrement comprend une indication vers nos serveurs où les fichiers de données téléchargeables originaux sont conservés.

En ce qui a trait aux formats, la plupart des données de Statistique Canada sont disponibles en CSV et XML (SDMX-ML).

4. Statistique Canada comme fournisseur de services pour ouvert.canada.ca

Le Secrétariat du Conseil du Trésor du Canada est le ministère du gouvernement fédéral responsable du portail Gouvernement ouvert (qui a vu le jour comme portail de données ouvertes). Dans sa recherche d'un dirigeant pour la mise en œuvre technique du portail, il a considéré que Statistique Canada avait l'expérience et l'expertise appropriées requises.

Par conséquent, outre qu'il fournit des données, Statistique Canada fournit aussi les professionnels des sciences de l'informatique pour le développement et la mise à jour du portail du gouvernement fédéral et coordonne la fourniture de l'infrastructure des TI requise pour appuyer le portail.

Un aperçu rapide d'ouvert.canada.ca fournira une idée des types de services offerts. D'un point de vue technique, nous utilisons le système de gestion du contenu Drupal, qui appuie la publication de contenu statique, de blogues, de commentaires et d'évaluation, etc. Le catalogue de données principal utilise une technologie appelée [CKAN](#).

5. Conclusion

Il est clair que le concept de données ouvertes correspond bien à la mission et au mandat des organismes statistiques nationaux. Statistique Canada s'engage à offrir aux Canadiens un libre accès aux renseignements importants dont ils ont besoin pour assumer efficacement leur rôle de citoyens et de décideurs dans un monde en rapide évolution.