

LA RÉUTILISATION DES DONNÉES OUVERTES AU QUÉBEC : DU DÉVELOPPEMENT ÉCONOMIQUE À LA TRANSPARENCE GOUVERNEMENTALE

[Christian Boudreau](#)

I.I.S.A. | « [Revue Internationale des Sciences Administratives](#) »

2021/4 Vol. 87 | pages 771 à 786

ISSN 0303-965X

ISBN 9782802771265

DOI 10.3917/risa.874.0771

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-internationale-des-sciences-administratives-2021-4-page-771.htm>

Distribution électronique Cairn.info pour I.I.S.A..

© I.I.S.A.. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.



La réutilisation des données ouvertes au Québec : du développement économique à la transparence gouvernementale

*Christian Boudreau*¹

Résumé

Basé sur l'histoire des données ouvertes au Québec, le présent article traite de la réutilisation de ces données par différents acteurs de la société en vue d'obtenir les retombées économiques, administratives et démocratiques souhaitées. À partir de l'analyse des mesures gouvernementales et des pratiques de la communauté en matière de réutilisation des données, l'étude montre que les retombées des données ouvertes semblent peu concluantes en termes de croissance économique. En revanche, leurs retombées paraissent prometteuses du point de vue de la transparence gouvernementale en permettant à différents acteurs de la société civile de surveiller l'intégrité et la performance des activités de l'État. À l'ère des données numériques et des réseaux, l'État doit être vu non seulement comme une plateforme propice à l'innovation, mais aussi comme un terrain d'observation faisant l'objet d'une surveillance étroite par divers acteurs motivés par des visées politiques et sociales.

Remarques à l'intention des praticiens

Bien que les retombées économiques des données ouvertes ne se sont pas montrées concluantes jusqu'ici, du moins au Québec, les administrations publiques ne doivent pour autant cesser d'investir dans l'ouverture de leurs données. Les résultats de l'étude tendent à montrer que les retombées des données ouvertes en termes de transparence sont, quant à elles, suffisamment porteuses pour poursuivre la libération des données gouvernementales, ne serait-ce que pour appuyer les activités d'évaluation et de planification des programmes et des services publics.

Mots-clés : Données ouvertes, Réutilisation, Transparence, Développement économique, Surveillance sociale.

¹ **Christian Boudreau**, École nationale d'administration publique, Québec (Canada) Courriel: christian.boudreau@enap.ca.

Introduction

Depuis la directive sur le gouvernement ouvert adoptée par l'administration d'Obama en 2009, les données ouvertes ont connu un important essor à travers le monde. Par données ouvertes, nous entendons des données numériques, granulaires et gratuites que les administrations publiques (ci-après administrations) rendent accessible sur le Web dans des formats et des conditions qui assurent leur libre réutilisation. En libérant leurs données, les administrations souhaitent stimuler l'innovation et l'économie tout en améliorant leur offre de services par le développement d'applications à coût presque nul. Elles entendent aussi favoriser la transparence de leurs actions et une saine gestion de leurs ressources par une meilleure diffusion de l'information et une participation citoyenne plus éclairée. Les données ouvertes sont donc vues comme une ressource publique exploitable par tous et à laquelle il convient de donner une deuxième vie en raison de leurs retombées économiques, administratives et démocratiques.

Si les portails de données ouvertes constituent un rouage important dans la libération des données gouvernementales, il ne suffit pas d'en créer pour qu'il y ait des retombées immédiates en termes de croissance économique, d'amélioration des services et de participation citoyenne (Zuiderwijk et al., 2014). Les données ouvertes ont peu de valeur en elles-mêmes; c'est leur réutilisation par divers intermédiaires qui en crée (Janssen et coll., 2012). La réutilisation des données ouvertes doit être prise ici au sens large, en commençant par leur téléchargement jusqu'au développement d'applications en passant par le traitement, l'interprétation et la diffusion de ces données.

En s'appuyant sur l'histoire des données ouvertes au Québec, le présent article porte sur la réutilisation de ces données par différents acteurs de la société. À partir de l'analyse des interventions gouvernementales et des pratiques d'intermédiaires en matière de réutilisation des données, l'étude montre que les retombées des données ouvertes semblent jusqu'ici peu concluantes en termes de croissance économique. Au Québec, comme au Canada et dans bien d'autres pays dits industrialisés, les retombées économiques ont pourtant occupé une place prépondérante dans la rhétorique gouvernementale visant à faire la promotion des données ouvertes, particulièrement au début des années 2010, au point de comparer les administrations à des plateformes de données dont la réutilisation contribuerait à la croissance économique et à l'amélioration des services publics par le développement d'applications mobiles (O'Reilly, 2010). En revanche, les retombées des données ouvertes en termes de transparence gouvernementale paraissent plus prometteuses du fait qu'elles permettraient à différents acteurs de la société de surveiller l'intégrité et d'évaluer la performance des activités des administrations. À l'ère des données numériques, il convient donc de voir les administrations non seulement comme des plateformes propices à l'innovation, mais aussi comme des terrains d'observation pouvant faire l'objet d'une surveillance par des acteurs motivés par des visées politiques et sociales plutôt que par des intérêts économiques.

L'article est structuré de la façon suivante. Dans un premier temps, nous abordons la contribution essentielle d'intermédiaires dans la démocratisation des

données ouvertes à partir d'écrits scientifiques. Nous présentons ensuite notre démarche méthodologique qui s'appuie sur une reconstitution de l'histoire des données ouvertes au Québec. Par la suite, nous nous attardons, d'une part, aux mesures déployées par les administrations québécoises, en particulier le gouvernement provincial et les municipalités, pour favoriser la réutilisation des données ouvertes et, d'autre part, aux pratiques de réutilisation des données ouvertes par la communauté d'intermédiaires québécoise. Enfin, nous revenons sur l'importance du rôle des intermédiaires dans la transparence des administrations en contexte numérique.

La nécessaire réutilisation des données ouvertes

Les administrations jouent un rôle essentiel dans la réutilisation des données ouvertes. Elles doivent d'abord s'assurer que les données qu'elles libèrent sont de qualité suffisante (ex. : à jour, exactes et complètes) pour que les intermédiaires soient en mesure de les réutiliser. Il doit aussi déployer des mesures pour faciliter le repérage, l'exploitation et la visualisation des données ouvertes (Susha et al., 2014). Ces mesures prennent habituellement la forme d'un catalogue regroupant des jeux de données à l'intérieur d'un portail ainsi que de normes et de licences communes d'utilisation des données (ex. : *Creative Commons*). Elles peuvent aussi consister en des interfaces de programmation d'applications (API) qui permettent d'interroger les données et d'automatiser leur téléchargement. Quant aux outils de visualisation, ils contribuent à la démocratisation des données ouvertes en rendant celles-ci accessibles à un plus large public, notamment en les superposant à des cartes géographiques facilement utilisables (Janssen et coll., 2012). Les administrations ont aussi intérêt à faire la promotion des données ouvertes auprès de leur communauté d'intermédiaires ou à consulter celle-ci dans la priorisation des données à libérer (Susha et al., 2014). Enfin, les administrations peuvent organiser des activités de création, en particulier des hackathons, pour mettre à profit les capacités d'idéation et d'innovation de la communauté. Les hackathons consistent en des concours où s'affrontent des équipes de développeurs et d'utilisateurs dans la conception de prototypes à partir de l'exploitation de données. Des prix sont décernés au terme de ces concours.

Dans leur rôle de producteur de données ouvertes, les administrations s'adressent surtout à des intermédiaires (Susha et al., 2014). Alors que ceux-ci perçoivent ces données comme une matière première à exploiter, les citoyens ordinaires y voient une longue série de chiffres insérés dans d'innombrables fichiers dont ils ne savent que faire. Par leurs réutilisations des données ouvertes, les intermédiaires contribuent à transformer celles-ci en information intelligible pour les citoyens qui, le plus souvent, n'ont ni les compétences techniques, ni les connaissances statistiques pour les exploiter et les analyser (Susha et coll. 2014; Janssen et coll., 2012). En diminuant les barrières d'utilisation des données ouvertes, les intermédiaires établiraient le pont entre les producteurs de données, les administrations, et leurs utilisateurs finaux, les citoyens.

Les intermédiaires s'apparentent à un groupe hétérogène aux contributions complémentaires, d'où la notion d'écosystème ou de communauté

d'intermédiaires (Harrison et coll., 2102). On y retrouve des développeurs indépendants, des jeunes pousses (start-up), des activistes numériques citoyens (ci-après hactivistes), des journalistes de données, des chercheurs, des étudiants, des organismes sans but lucratif (OSBL) et bien d'autres acteurs. Ces intermédiaires peuvent contribuer de diverses façons et à différents moments au cycle de réutilisation des données. Ils peuvent demander la libération de données ou en libérer eux-mêmes à partir d'opérations d'extraction de données sur des sites Web (Web scraping), comme ils peuvent évaluer la qualité des données libérées, les nettoyer, les convertir et les amalgamer dans des formats exploitables. Ils peuvent aussi créer des interfaces de visualisation ou autres applications qui exploitent les données ouvertes, comme ils peuvent analyser ces données pour en apprendre sur le fonctionnement des administrations et sur leurs décisions puis d'en informer la population et les autorités concernées. Quant aux motivations des intermédiaires à réutiliser les données ouvertes, elles peuvent s'inscrire dans des visées commerciales (ex : développer des applications mobiles), politiques ou démocratiques (ex : surveiller l'intégrité des administrations et en informer la population), évaluatives ou scientifiques (ex : analyser l'efficacité des programmes et la qualité des services) et même ludiques (ex : réutiliser des données pour le plaisir de la chose). Les intermédiaires peuvent agir comme demandeur, producteur, évaluateur, auditeur, développeur et communicateur dans cette réutilisation des données (Gonzalez-Zapata et coll., 2015). Ils constitueraient un passage obligé dans la concrétisation des retombées des données ouvertes, notamment en démocratisant l'accessibilité aux données.

L'apport des intermédiaires dans la réutilisation des données ouvertes ne se limite donc pas au développement économique et à l'amélioration des services publics par le développement d'applications Web. Les données ouvertes peuvent aussi permettre aux intermédiaires d'agir comme chien de garde dans un souci de transparence et d'imputabilité gouvernementale. Dans cette ère de la donnée ouverte et du *big data*, les administrations ne sont plus seulement des agents qui exercent une surveillance bureaucratique à l'égard des conduites de ses administrés, mais aussi des agents que divers acteurs surveillent à partir de données qu'elles libèrent (Boudreau et Caron, 2016). S'ajoute à la traditionnelle surveillance bureaucratique une surveillance sociale qui permet à la société civile et aux autres acteurs non gouvernementaux « de suivre les agissements des organisations et d'influencer les pouvoirs en place en alertant l'opinion publique » (Boudreau, 2006 : 63; voir aussi Castells, 2001). Des acteurs de la société seraient d'autant plus enclins à surveiller les administrations que les (in)activités gouvernementales s'accompagnent d'irrégularités administratives ou de problèmes sociaux.

La surveillance des administrations par le numérique, tout comme le développement d'applications, n'est pas une activité à la portée du citoyen ordinaire. Elle interpelle des acteurs de la société civile qui, en plus d'être politiquement engagés, savent repérer et exploiter les données, que ce soit pour développer des applications civiques ou pour traiter les données de manière à détecter des problèmes et à alerter les autorités et la population (Harrison et coll., 2012). Ainsi, la libération des données gouvernementales ne conduit pas automatiquement

à une participation citoyenne éclairée et accrue. Insuffisante en elle-même, la libération requiert aussi l'intervention d'intermédiaires intéressés et techniquement compétents qui, par diverses pratiques de réutilisation, contribuent à la démocratisation des données numériques et, du coup, à la transparence gouvernementale. Cette nécessaire intervention des intermédiaires dans la transparence et surveillance sociale des administrations par le numérique n'est pas propre au Québec. Worthy (2015) l'a aussi constatée dans diverses activités de surveillance sociale des marchés publics au Royaume-Uni à la suite de la libération des données sur l'octroi des contrats gouvernementaux.

Si la surveillance des administrations par la société civile est devenue un trait de la société en réseau (Castells, 2001) ou de la société du risque (Beck, 1990), la capacité de cette surveillance sociale demeure tributaire de la qualité de l'information que les administrations rendent disponibles ainsi que de la pertinence de cette information en termes d'imputabilité gouvernementale. Or, la qualité et la pertinence des données ouvertes laissent parfois à désirer en raison d'erreurs de saisie et de la réticence des administrations à publier des données jugées politiquement sensibles, limitant la capacité des acteurs de la société civile à surveiller étroitement les administrations (Boudreau et Caron, 2016; voir aussi Goeta et al., 2014).

Méthodologie

Afin de repérer et de situer dans le temps les mesures gouvernementales et les pratiques d'intermédiaires en matière de réutilisation des données, nous avons reconstitué et analysé l'histoire des données ouvertes au Québec, dont les premières manifestations remontent à 2010. Pour ce faire, nous avons procédé au repérage et à l'analyse d'articles de journaux des principaux quotidiens québécois pour tisser la trame de cette histoire à partir de la base de données *Eureka*. L'utilisation des mots clés « données ouvertes » et « open data » nous a permis de repérer environ 1300 articles dont le contenu a été classé selon la chronologie des événements et codé en fonction des administrations concernées, à savoir les villes québécoises² et le gouvernement du Québec, et les mesures que ces administrations ont été déployées pour favoriser la réutilisation des données.

Nous avons étendu la recherche documentaire aux informations publiques produites par les administrations québécoises (ex : politiques, portails de données, fil Twitter, guides, communiqués, etc.) et par leur communauté d'intermédiaires (ex : rapports, communiqués, mémoires, etc.) afin de cerner les mesures gouvernementales et les pratiques d'intermédiaires relatives aux données ouvertes.

Nous avons aussi mené une vingtaine d'entrevues (de mai 2017 à janvier 2018) pour préciser ou valider les événements marquants de l'histoire des données ouvertes au Québec ainsi que les mesures gouvernementales et les pratiques d'intermédiaires en la matière. (Voir tableau I) Les répondants ont été sélectionnés à partir d'une stratégie d'échantillonnage dite boule de neige qui consiste à demander aux personnes rencontrées en entrevue de suggérer

² En particulier Montréal, Québec, Sherbrooke et Gatineau.

d'autres personnes à interviewer en lien avec l'histoire des données ouvertes au Québec. Toutes les personnes dont le nom a été évoqué par plus de trois répondants ont été rencontrées en entrevue. Cette technique nous a permis d'atteindre une saturation empirique des catégories au sens « où la collecte de données [supplémentaires] apparaît répétitive ou stérile eu égard au phénomène concerné » (Paillé et Muchielli, 2008, p. 272).

Tableau I : Nombre des personnes rencontrées en entrevue selon leur profil

Représentants de l'administration publique	<ul style="list-style-type: none"> • 5 gestionnaires de municipalité québécoise • 4 gestionnaires et 1 professionnel d'organisme ou ministère provincial • 2 gestionnaires et 2 professionnels de l'équipe du gouvernement ouvert du Secrétariat du Conseil du Trésor (SCT)³
Représentants de la communauté d'intermédiaires	<ul style="list-style-type: none"> • 3 journalistes de données • 2 entrepreneurs de jeune pousse • 2 hactivistes • 1 chercheur

Nous avons aussi analysé l'ensemble des échanges d'un groupe de discussions, nommé *Montréal Ouvert*, sur la plateforme *Google Groupes*, pour mettre en lumière l'éventail des pratiques de réutilisation d'une importante communauté d'intermédiaires. Particulièrement actif entre 2012 et 2016, ce groupe de discussions a donné lieu à de nombreux échanges faisant état des pratiques courantes de réutilisation des données ouvertes au Québec⁴.

Une attention particulière a été portée à la réutilisation des données ouvertes provenant du Système électronique d'appel d'offres (ci-après SEAO) afin d'illustrer les mesures et les pratiques en matière de réutilisation à partir d'un cas réel. Contenant des renseignements sur tous les contrats de plus de 25 000 \$ octroyés par les organismes publics du Québec, y compris les municipalités, ce jeu de données a été mis en exergue dans l'étude en raison de sa popularité auprès de la communauté d'intermédiaires.

L'ensemble des données issues des entretiens et des documents ont été codés à l'aide d'un logiciel d'analyse qualitative, *MaxQda*. Une première vague de codage a permis d'ordonner chronologiquement les événements qui ont marqué l'histoire des données ouvertes au Québec et d'y repérer les mesures ou pratiques déployées par les administrations et les intermédiaires dans ce domaine. Une deuxième vague de codage a permis de regrouper les résultats autour des mesures gouvernementales et des stratégies de réutilisation en matière de données ouvertes tout en faisant ressortir l'importance des retombées potentielles de ces données en termes de transparence gouvernementale.

³ Le SCT exerce une responsabilité provinciale dans le déploiement des données ouvertes au Québec.

⁴ Environ 1000 fils de discussions y ont été lancés. L'ensemble de ces fils de discussion a été codé.

Résultats

Des mesures gouvernementales visant la réutilisation des données ouvertes

Création d'un portail intégré. Montréal est la première ville au Québec à créer, en octobre 2011, un portail de données ouvertes. D'autres villes québécoises lui emboîtent le pas, dont Québec (février 2012), Gatineau (janvier 2013) et Sherbrooke (novembre 2013). Quant au gouvernement du Québec, il convie, en mai 2012, les principaux ministères et organismes provinciaux à libérer au moins trois jeux de données et à les déposer dans le portail provincial de données ouvertes, lancé en juin 2012. Au départ, les données ouvertes étaient de qualité inégale: « *Les premières vagues de données ont été faites à partir de ce qui était disponible et prêt à être diffusé facilement [...] On a eu un peu n'importe quoi* ». (Professionnel du SCT) Plusieurs données étaient publiées dans des formats propriétaires (ex: PDF et Excel) « compliqués à utiliser, si tu n'avais pas la suite Office ». (Professionnel d'organisme provincial) Dans le cas du portail provincial, « aucune consigne n'avait été donnée quant à la saisie [...] Certains mettaient jour, mois, année alors que d'autres mettaient année, mois, jour ». (Représentant provincial) La granularité des données laissait aussi, parfois, perplexe. On pouvait y retrouver « *des tableaux finaux, donc des données agrégées. Mais, on n'avait pas les données brutes qui ont permis de constituer ces tableaux* ». (Gestionnaire du SCT)

Afin de favoriser une réutilisation des données ouvertes, le gouvernement du Québec et les quatre municipalités québécoises susmentionnées⁵ adoptaient, en février 2014, une licence commune, la *Creative Commons 4.0*. En plus d'accorder les mêmes droits de réutilisation d'un portail à l'autre, cette licence est l'une des moins restrictives, protégeant les droits d'auteur par une simple mention de la source. Le gouvernement du Québec et cinq municipalités⁶ lancaient, en avril 2016, le portail *Données Québec* qui intègre les données du précédent portail provincial et des 5 portails municipaux. Depuis, d'autres municipalités québécoises ont intégré leurs données ouvertes à *Données Québec*, dont Blainville, Longueuil, Repentigny, Rimouski et Shawinigan. Le principal avantage de ce portail réside dans la normalisation de ses données. « *Avec les villes, on s'est donné des standards simples. Par exemple, quand tu insères une date, elle est dans le format année, mois et jour [...] On s'est aussi donné des standards pour les fichiers* ». (Gestionnaire du SCT) La création d'un portail intégré et l'adoption d'une licence commune procurent aux développeurs un accès simplifié à un plus large marché. « *Quand quelqu'un développe une application, si ça marche pour une ville, ça marche pour les autres villes qui diffusent sur Données Québec [...] Tu as donc un plus grand intérêt à développer* ». (Gestionnaire de municipalité)

Outre les tentatives de normalisation des données ouvertes, l'équipe responsable de *Données Québec* adopte une approche qui privilégie la qualité sur la quantité. Ainsi, plus de 150 jeux de données sont retirés de *Données Québec*, en 2016, en raison de lacunes dans la mise à jour ou la granularité de ces jeux.

⁵ Gatineau, Montréal, Québec et Sherbrooke.

⁶ La ville de Laval s'ajoutant aux 4 précédentes.

Bien que les données ouvertes s'adressent avant tout aux intermédiaires, les citoyens ne sont pas en reste. Des administrations ont développé des outils de visualisation pour démocratiser l'accès aux données ouvertes. La ville de Montréal a été particulièrement active à ce chapitre en déployant les outils de visualisation suivants: *Vue sur les contrats*; *Vue sur les indicateurs de performance*; *Vue sur le budget* et *Vue sur la sécurité*.

Des stratégies visant la participation de la communauté. Parmi les administrations étudiées, la ville de Montréal a été la plus active dans les stratégies visant à faire participer la communauté d'intermédiaires (ci-après communauté). Dès 2012, la ville de Montréal créait une première table de concertation, constituée de membres de la communauté et de représentants municipaux, dont le principal mandat consistait à sélectionner les jeux de données à haut potentiel de réutilisation. Une deuxième table est créée en 2013, dont le mandat consistait cette fois-ci à organiser des événements, en particulier des hackathons. La ville de Montréal a aussi permis à la communauté de commenter en ligne la refonte de sa politique sur les données ouvertes, en octobre 2015. La ville a d'ailleurs enchâssé dans cette politique l'obligation de consulter la communauté pour la publication de ses données ouvertes. Pour le gouvernement provincial, les mesures de participation et d'accompagnement déployées par l'équipe responsable des données ouvertes ont surtout visé les organismes publics. À ce propos, des *Lignes directrices sur la diffusion de données ouvertes* (2016) et un *Guide de priorisation et diffusion des données ouvertes* (2017) ont été produits à l'attention des répondants en gouvernement ouvert au sein des organismes susceptibles de publier des données dans *Données Québec*. Aucune modalité de participation de la communauté n'y est prévue. Quand il s'agit de communiquer avec la communauté, les administrations québécoises privilégient les médias sociaux. Par exemple, Twitter est utilisée, depuis 2013, par l'équipe des données ouvertes de la ville de Montréal et, depuis 2015, par l'équipe du gouvernement ouvert au gouvernement du Québec, pour informer la communauté sur la libération de nouveaux jeux de données, sur le développement d'applications ou d'outils de visualisation et sur la tenue d'événements ou de consultations.

Des événements misant sur la créativité de la communauté. Les mesures visant la réutilisation des données s'inscrivent aussi dans des activités de création, en particulier des hackathons, tel que mentionné plus haut. *Montréal Ouvert* a été le premier groupe de citoyens à organiser un tel événement, en décembre 2010, puis un deuxième, en janvier 2011. Le portail de données ouvertes de Montréal n'étant pas encore créé, les organisateurs ont dû, eux-mêmes, extraire des données du site de la ville et les convertir en formats réutilisables. Ce n'est qu'au troisième hackathon de *Montréal Ouvert*, tenu dans les bureaux de la ville en novembre 2012, que Montréal publie des données ouvertes. Ces activités de création marquent le coup d'envoi d'une série d'hackathons organisés tantôt par la communauté, tantôt par la ville de Montréal ou conjointement. L'engouement pour les hackathons gagne d'autres villes québécoises, dont les villes de Québec, de Gatineau, de Sherbrooke et de Laval. Le gouvernement du Québec décide, à

son tour, de s'impliquer dans l'organisation d'hackathons à partir de 2014. Il se joint alors aux villes de Québec, de Montréal, de Gatineau et de Sherbrooke pour coordonner et financer l'évènement *HackQc* qui s'est tenu simultanément dans ces 4 villes, en mars 2014, et dont la finale provinciale s'est terminée à Québec. Depuis, deux autres hackathons d'envergure provinciale ont été organisés par le gouvernement du Québec: l'un tenu à Sherbrooke, en mars 2017, l'autre à Montréal, en mai 2018.

L'histoire des données ouvertes au Québec montre que les hackathons ne servent pas qu'au développement de prototypes. C'est aussi l'occasion de poursuivre des projets de développement en cours. Par exemple, *Nord-Ouvert*, un OSBL de Montréal spécialisé dans le développement d'applications civiques, y a eu recours pour faire progresser certaines de ses applications (ex: Represent. opennorth.ca, MaMairie.ca ou BudgetPlateau.com). De plus, les hackathons ont servi de vitrine pour la promotion des données ouvertes auprès des fonctionnaires, des élus et des citoyens en montrant l'utilité potentielle de ces données et le dynamisme de la communauté. Les administrations ont aussi profité de ces événements pour publier de nouvelles données ouvertes. Enfin, les hackathons se sont révélés être une façon conviviale pour des étudiants, des activistes et des journalistes de s'initier à la réutilisation des données et de faire connaissance et d'interagir en personne avec des membres de la communauté.

La réutilisation des données ouvertes par la communauté

Une communauté virtuelle qui s'entraide et qui collabore avec les autorités. La communauté s'est montrée beaucoup plus active sur ses propres médias sociaux, en particulier dans le *Google Groupe Montréal Ouvert*, que dans les médias sociaux gérés par les administrations. Entre 2011 et 2016, plusieurs développeurs, entrepreneurs de jeunes pousses, activistes, journalistes, représentants gouvernementaux ou d'OBSL, chercheurs, professeurs et étudiants y ont échangé des conseils, des données et des programmes. Ces échanges ont permis aux membres de la communauté de repérer des données sur le Web, de les extraire, de les nettoyer, de les structurer et de les convertir dans des formats non propriétaires pour faciliter leur réutilisation. Les membres de la communauté ont aussi été nombreux à offrir leur aide à d'autres membres ou à en demander afin de répondre à des besoins aussi variés que le repérage, l'extraction, la conversion et le partage des données, le développement d'outils de visualisation et l'organisation d'événements. Le nombre élevé des échanges et la rapidité des réponses témoignent du dynamisme de cette communauté. Comme lieu d'échanges et d'entraide, le *Google Groupe Montréal Ouvert* s'est avéré être un outil essentiel dans la constitution de la communauté, particulièrement entre 2011 et 2016.

Quant à l'action politique de la communauté, elle visait au départ à faire pression sur les administrations pour qu'elles libèrent leurs données. L'utilisation des médias traditionnels et sociaux par la communauté, en particulier par les groupes *Montréal Ouvert*, *Capitale Ouverte* et *Gatineau Ouvert*, conjuguées à l'approche collaborative de ces groupes, ont eu un effet persuasif sur les trois villes concernées. Lancé en 2011, le groupe *Montréal Ouvert* s'est donné comme mission

d'amener la ville de Montréal à adopter une politique de données ouvertes. Une première politique fut adoptée en 2013, à la suite des nombreuses apparitions de *Montréal Ouvert* dans les médias et de ses multiples interventions auprès d'employés, de gestionnaires et de politiciens municipaux. Quant à *Capitale Ouverte*, c'est à la suite d'un événement, tenu à l'été 2011 et auquel les journalistes ont été conviés, que la ville de Québec a décidé de publier des données ouvertes et d'organiser des hackathons conjointement avec ce groupe. La popularité des premières applications mobiles (ex: Patinoire Montréal, Resto-net, Zonecône.ca et Doctr) développées à partir d'extractions de données sur le Web, combinée au succès d'hackathons organisés par la communauté, ont aussi contribué à la publication de données ouvertes au Québec. Pour rassurer les autorités de l'époque, affirme un ancien représentant de *Montréal Ouvert*, devenu entrepreneur d'une jeune pousse, « on a d'abord décidé d'utiliser des données qui n'allaient pas faire peur aux gens, donc ne pas commencer avec des données sur les finances ou les contrats. On a commencé par une application sur les patinoires. »

Des développements d'applications surtout civiques. Si les hackathons ont servi de bougies d'allumage à plusieurs projets de développement, seule une minorité de ces projets a su dépasser le stade du prototype. Plus rares encore sont les projets de développement qui, initiés lors d'hackathons, ont conduit à la création d'applications commerciales pérennes. Les données ouvertes et les activités de création qui les réutilisent semblent insuffisantes pour assurer la longévité des applications. Sauf quelques exceptions, comme *Transit*, *Doctr*, *Local Logic* et *Cyclemap*, la très grande majorité des applications mobiles développées à partir de données ouvertes n'atteignent pas une maturité commerciale suffisante pour devenir rentables et exportables. C'est le cas de *Prkng*, une application utilisée pourtant par plus de 70 000 personnes pour trouver des stationnements gratuits à Montréal, qui a cessé de fonctionner en février 2018. [M]ême si l'information traitée par *Prkng* avait de la valeur, l'entreprise n'a pas été capable de démontrer aux investisseurs que le succès montréalais pouvait facilement être reproduit dans d'autres villes.⁷

Alors que la création d'applications commerciales et de jeunes pousses à partir de données ouvertes ne semble pas être à la hauteur des attentes des administrations municipale et provinciale, les applications civiques qui permettent d'exercer une surveillance sociale connaissent un certain succès, notamment en ce qui a trait à la réutilisation des données du SEAO. Depuis l'ouverture des données du SEAO, différents applications ont vu le jour, comme le *Vérificateur Citoyen*, l'*Analyseur de contrats publics québécois*, l'*Espace Data* et le *Contrac-Actuel*. Rappelons que la ville de Montréal a développé son propre outil de visualisation des contrats qu'elle octroie, *Vue sur les contrats*, à partir de l'exploitation de ses données ouvertes.

Des données ouvertes de qualité variable. Malgré les tentatives gouvernementales de normaliser et d'intégrer les données ouvertes, certains problèmes de qualité

⁷ Johnson, M., "La vie après la mort de Prkng", *L'actualité*, 15 février 2018.

persistent. C'est le cas des données du SEAO. Des hactivistes et des journalistes y ont repéré de nombreuses erreurs de saisie.

... plus de 275 entreprises portent le numéro d'entreprise Québec (NEQ) 123456789 ou 111111111 ... Plus de 30 000 champs d'adresse de fournisseurs sont vides. Certains contrats étaient listés en double, d'autres n'y existaient pas ... Plusieurs dates sont visiblement erronées ... Les noms de fournisseurs sont écrits de plusieurs manières différentes (« Bell », « BELL CANADA » ou « Compagnie de téléphone Bell du Canada », etc.), ce qui complique considérablement les efforts de compilation de l'argent reçu par chaque fournisseur.⁸

Ces erreurs s'expliquent par le fait que « [c]e sont des gens dans différentes municipalités et organismes qui entrent manuellement les données [...] et qu'on ne valide pas ces données au Secrétariat du Conseil du trésor ». (Professionnel du SCT) Les erreurs de saisie ne se limitent pas aux données du SEAO. La ville de Montréal reconnaît l'existence de problème de qualité dans d'autres jeux de données ouvertes, dont les *Déclarations des exterminations de punaises de lit*. Selon la fiche descriptive de ce jeu : « Les données ont un faible degré de fiabilité car elles sont consignées manuellement par des tiers, soit les gestionnaires de parasites et ne font l'objet d'aucune validation de la part de la Ville de Montréal ».⁹ Des représentants de la ville de Montréal reconnaissent que la saisie manuelle peut retarder la mise à jour des données ouvertes : « [L]a fréquence de mise à jour des ensembles de données n'est pas toujours au rendez-vous. Ce manque de fraîcheur découle du fait que la mise à jour des données est encore trop souvent manuelle ».¹⁰

Les contrats publics : Un terrain propice à la surveillance par la communauté. La réutilisation des données prend aussi la forme d'une surveillance des activités gouvernementales, en particulier des marchés publics, exercée par la communauté. L'évènement *Hackons la corruption*, tenu en novembre 2012 à Montréal, dans un contexte secoué par de graves allégations de corruption et collusion, est une première action dans ce sens. Organisé par *Montréal Ouvert*, cet évènement hautement médiatisé a réuni des membres de la communauté, des politiciens et des journalistes. Quelques mois plus tard, des membres de *Québec Ouvert*¹¹ lancent une pétition électronique sur le site de l'Assemblée nationale pour libérer les données du SEAO. Devant cette pression politique et sachant que les données du SEAO font l'objet d'extraction sur le Web par la communauté, le gouvernement du Québec décide d'ouvrir les données du SEAO en 2013. De nouvelles applications sont alors développées par la communauté pour interroger le SEAO en vue de détecter des modes opératoires de collusion et de corruption dans l'octroi des contrats publics. La communauté se montre d'autant plus disposée à exercer un rôle de surveillant que les administrations ne semblent pas l'exercer.

⁸ Gagnon, K. et coll., Question simple, réponse introuvable, *La Presse*, 2 novembre 2017.

⁹ <http://donnees.ville.montreal.qc.ca/dataset/declarations-exterminations-punaises-de-lit>

¹⁰ <http://donnees.ville.montreal.qc.ca/portail/nouvelle-politique-et-avancement-des-travaux/>

¹¹ Né d'une fusion de Montréal Ouvert, Capitale Ouverte et Gatineau Ouvert.

Lionel Perez et Pierre Desrochers, président du comité exécutif de la Ville [de Montréal], ont admis vendredi que la Ville ne tente pas de détecter des modes opératoires dans les soumissions ou pour les contrats offerts en sous-traitance à des soumissionnaires rivaux.¹²

Le SEAO est rapidement devenu le jeu de données ouvertes le plus téléchargé au Québec. Or, en réutilisant ces données, des membres de la communauté constatent que des problèmes de qualité les empêchent d'exercer pleinement leur rôle de surveillant. Par exemple, le fait qu'un même fournisseur puisse être saisi de diverses façons rend difficile le regroupement des contrats par client. Certains membres proposent qu'on inscrive dans le SEAO le numéro du Registre des entreprises, ce qui fut fait en février 2014. Si le SEAO est le jeu de données ayant fait l'objet du plus grand nombre de vérifications par la communauté, Pascal Robichaud est sans conteste le membre le plus dédié à cette tâche. Cet hactiviste¹³s'est montré particulièrement déterminé dans la détection des erreurs de saisi et des délais de publication non respectés et dans leur signalement aux administrations concernées pour que celles-ci apportent les corrections nécessaires. Entre 2013 et 2015, «[e]n moyenne, il en a trouvé pas moins de 9000 par mois». (Journaliste de données)

Discussions et conclusion

Les administrations ont un apport essentiel dans la réutilisation des données ouvertes, dont celui de publier des données de qualité ayant un potentiel de réutilisation. Les mesures de standardisation et d'intégration relatifs au portail *Données Québec* sont un pas dans cette direction. Bien que saluées par plusieurs membres de la communauté, ces mesures gouvernementales se sont montrées insuffisantes pour prévenir les problèmes d'exactitude, de mise à jour et de complétude de certains jeux données, particulièrement quand la saisie est réalisée manuellement. Pour assurer la qualité des données, d'autres mesures d'encadrement s'avèrent nécessaires, comme le recours à des listes de vocabulaire contrôlé, à des systèmes de validation de la saisie et à des procédures d'inscription automatisées. Si les administrations et la communauté peuvent procéder à des vérifications *a posteriori* pour évaluer la qualité des données et pour les nettoyer, cette procédure peut s'avérer lourde dans la mesure où elle peut être requise lors de chacune des mises à jour d'un jeu de données.

Quant aux retombées commerciales des données ouvertes que les administrations faisaient miroiter lors du lancement de leur portail, en particulier les municipalités qui y voyaient un levier de développement économique, elles ne sont toujours pas avérées si l'on en juge par le faible nombre d'applications commerciales développées à partir de ces données (voir aussi Lee et coll., 2016). Pour ce qui est des hackathons, outils privilégiés par les administrations pour stimuler

¹² Gyulai, L. et coll., "transparence dans les dépenses et l'attribution des contrats", *La Presse*, 1 novembre 2017.

¹³ Citoyen versé en informatique et dont l'engagement civique se traduit par une réutilisation des données ouvertes à des fins de transparence et d'intégrité gouvernementales.

le développement d'applications et la création de jeunes pousses locales, l'apport de ces événements se mesure plus en termes de formation, de socialisation et de visibilité que d'innovation commerciale. Devant ce succès économique mitigé, doit-on conclure à l'inutilité des données ouvertes? Dans quelle mesure ces données constituent-elles un outil efficace de transparence et de surveillance sociale des administrations? Qu'en est-il des autres retombées?

Notre étude offre des éléments de réponse à ces questions. Tout d'abord, elle montre que les retombées des données ouvertes débordent l'innovation commerciale et la création de jeunes pousses. Les nombreuses demandes de jeux de données, dont celles adressées à la ville de Montréal sur son portail, témoignent de l'intérêt de la communauté pour des réutilisations autres qu'économiques. Si les retombées économiques des données ouvertes apparaissent jusqu'ici timides, leurs retombées en termes de transparence gouvernementale semblent plus porteuses et mobilisatrices au sein de la communauté. Les données ouvertes intéressent en effet des hactivistes, des journalistes et des entrepreneurs civiques qui désirent exercer une surveillance des activités gouvernementales afin de détecter des problèmes ou irrégularités, d'en informer les autorités et d'alerter le public. Pour exercer efficacement leur rôle de chien de garde et de courroie de transmission auprès du public, ces membres de la communauté font de la qualité des données une de leur principale revendication.

L'étude montre aussi que la capacité de la communauté à surveiller les activités des administrations ne dépend pas seulement de la qualité des données ouvertes, mais aussi de leur pertinence en termes d'imputabilité gouvernementale. Or, au Québec comme ailleurs¹⁴, les administrations se montrent réticentes à libérer leurs données, particulièrement lorsque celles-ci sont de mauvaise qualité ou jugées politiquement sensibles (Boudreau et Caron, 2016; voir aussi Zuiderwijk et coll. 2014). Le SEAO vient une fois de plus appuyer notre propos. Pour exercer une meilleure surveillance des marchés publics, des hactivistes et des journalistes ont manifesté le souhait de croiser les données du SEAO avec d'autres ensembles de données, en particulier le registre des entreprises. À la suite des demandes répétées de la communauté, le gouvernement du Québec décidait, en 2015, de publier ce registre en données ouvertes, à l'exception du nom des actionnaires, invoquant la protection des renseignements personnels. Or, sans le nom des actionnaires, il est difficile d'établir un lien entre les firmes qui obtiennent les contrats et leur propriétaire, affirment des membres de la communauté. Quant à l'outil de visualisation de la ville de Montréal, *Vue sur les contrats*, il diffuse moins de renseignements sur l'octroi des contrats que ne le font certaines villes nord-américaines sur leur portail¹⁵:

Des villes comme New York et Chicago, en plus de mettre en ligne le montant d'un contrat et le soumissionnaire choisi, divulguent aussi le prix des soumissions non

¹⁴ On observe le même phénomène en France (voir Goeta et coll., 2014) et aux Pays-Bas (voir Zuiderwijk et coll., 2014).

¹⁵ En revanche, Montréal se compare aux autres grandes villes canadiennes qui, comme elle, ne rendent publiques que le nom de l'entreprise gagnante et le montant de sa soumission.

*retenues, les paiements faits à des sous-traitants et le coût total du projet réalisé. On ne trouve rien de tout cela dans Vue sur les contrats.*¹⁶

Un changement dans les pratiques de gestion de l'information des administrations s'avère nécessaire pour accélérer le virage vers une meilleure transparence gouvernementale et une démocratisation de la gouvernance par le numérique. Pour ce faire, les données ouvertes devront être vues par les administrations comme un mécanisme d'amélioration de la performance, des programmes et des services publics, et non plus comme un mécanisme de condamnation publique qui incite à la prudence. L'adoption d'une politique d'ouverture par défaut des données publiques peut être un premier jalon institutionnel à poser dans ce virage vers une plus grande transparence et performance des administrations.

Quoiqu'il en soit, la communauté ne se limite pas aux seules données ouvertes lorsqu'elle veut surveiller les activités gouvernementales. Elle s'adonne aussi à des opérations d'extractions de données sur les sites Web gouvernementaux et, lorsque les données sont inaccessibles en ligne, elle recourt aux traditionnelles demandes d'accès à l'information. Comme le signale un journaliste de données : « *Les données ouvertes ne sont qu'une minuscule fenêtre sur l'ensemble des données possédées par l'État* » alors que l'extraction de données procure « *l'exclusivité de la nouvelle ainsi que le plaisir d'écrire un script d'extraction* ». N'empêche qu'un bon nombre de journalistes québécois utilisent les données ouvertes pour appuyer la trame narrative de leurs articles portant sur la qualité des services publics et autres thèmes d'intérêt public, comme la gestion des contrats publics, la criminalité, les terrains contaminés, la qualité de l'air, les punaises de lits et la rapidité des opérations de déneigement.

L'étude montre aussi que la réutilisation des données déborde le développement d'applications, la surveillance de l'intégrité des administrations et la publication d'articles d'intérêt public. Elle s'étend aussi à des activités d'évaluation et de planification des programmes et services publics. Les données ouvertes de type transactionnel¹⁷, issues des systèmes de mission, permettent à des fonctionnaires ou à des chercheurs d'évaluer l'efficacité des programmes et la qualité des services (Boudreau et Caron, 2016). C'est le cas de la Société d'assurance automobile du Québec (SAAQ) qui, depuis 2017, publie d'importants ensembles de données ouvertes sur ses services et ses clientèles (ex : accidents de la route), lesquels peuvent appuyer des activités de recherches évaluatives et d'élaboration de programmes. Les administrations peuvent aussi profiter de données ouvertes publiées par d'autres administrations pour prédire les effets sur leurs clientèles de certaines modifications à leurs programmes, sans avoir à les expérimenter. Les États-Unis ont une base de données ouvertes qu'ils appellent FARS [...] sur les accidents mortels de la route [...] Ce jeu de données nous a permis de faire notre propre analyse [...] et de mesurer l'impact de l'augmentation des limites de vitesse sur les autoroutes. (Gestionnaire de la SAAQ)

¹⁶ Gyulai, L. et coll, 1 novembre 2017, *op. cit.*

¹⁷ En particulier les données administratives sur l'utilisation des services publics par les usagers.

En libérant des données sur leurs programmes et services, les administrations peuvent se comparer entre elles en vue d'améliorer leur performance. Par exemple, l'outil *Vue sur les indicateurs de performance* permet « de consulter des indicateurs de performance portant sur la Ville de Montréal seule, ou comparée à Calgary, Toronto et Winnipeg »¹⁸. Grâce à cet outil de visualisation, les autorités de Montréal ont constaté « que l'entretien de ses rues lui coûte nettement plus cher, chaque kilomètre de voie lui coûtant 27 600 \$ par an, contre 11 500 \$ à Toronto ».¹⁹ Quant aux chercheurs québécois, 13 d'entre eux ont réclamé, dans une lettre publiée dans le quotidien *Le Devoir*, de meilleurs accès aux données gouvernementales non seulement pour renforcer leur capacité de production de recherches scientifiques, mais aussi pour permettre une « évaluation objective de l'efficacité des politiques publiques dans des domaines comme l'éducation, la santé, l'emploi, la langue ou l'immigration ».²⁰

Enfin, en termes de transparence des activités gouvernementales, les données ouvertes pourraient devenir une arme redoutable que les administrations devront manier avec précaution pour éviter des dérives : le profilage de groupes ou de territoires jugés à risque, la réidentification d'individus à partir du recoupage de données anonymisées et la gouvernance algorithmique sur la base de données erronées ou des calculs biaisés sont autant de risques sociétaux que les administrations devront bien encadrer. Ainsi, dans cette marche vers une plus grande transparence, les administrations devront non seulement publier des données ouvertes de meilleure qualité et à haut potentiel de réutilisation, mais aussi s'assurer que la réutilisation de leurs données ne portera pas préjudice aux citoyens qu'elles souhaitent mieux (des)servir par le numérique.

Identifiant ORCID : Christian Boudreau : <https://orcid.org/0000-0003-2220-9246>

Références bibliographiques

- Beck, U (1992) *Risk Society: Towards a New Modernity*, Londres: Sage Publications.
- Boudreau C (2006) Multipolarité de la surveillance et gestion des médicaments au Québec, *Recherches sociographiques*, 47(2): 299-320.
- Boudreau C et Dj Daniel (2016) *Les données ouvertes dans l'administration publique québécoise: utilités, freins et pistes de solution*. Québec: Cefrio/ENAP.
- Castells, M (2001) *La société en réseaux*, Paris: Fayard.
- Deglise, F. (2013) 13 chercheurs réclament un meilleur accès aux données, *La Presse*, 8 Avril.
- Gagnon, K., Levesque, K., Gyulai, L. et De Lorimier T. (2017) Question simple, réponse introuvable, *La Presse*, 2 Novembre.
- Goëta S et C Mabi (2014) L'open data peut-il (encore) servir les citoyens? *Mouvements*, 3(79): 81-91.
- Gonzalez-Zapata F et R Heeks (2015) The multiple meanings of open government data: Understanding different stakeholders and their perspectives. *Government Information Quarterly*, 32 (4): 441-452.

¹⁸ Site *Vue sur les indicateurs de performance*, <http://ville.montreal.qc.ca/vuesurlesindicateurs/index.php>.

¹⁹ Normandin, P.-A., "Montréal s'améliore, mais continue à traîner de la patte", *La Presse*, 18 novembre 2018.

²⁰ Deglise, F. "13 chercheurs réclament un meilleur accès aux données", *La Presse*, 8 avril 2013.

- Gyulai, L. et al. (2017) Transparence dans les dépenses et l'attribution des contrats, *La Presse*, 1 Novembre.
- Harrison TM, Pardo TA et M Cook (2012) Creating Open Government Ecosystems: A Research and Development Agenda. *Future Internet* 4: 900-928.
- Janssen M, Charalabidis Y et A Zuiderwijk (2012) Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4): 258-268.
- Johnson, M. (2018) La vie après la mort de Prkng, *L'actualité*, 15 février.
- Lee M, Almirall E et J Wareham (2016) Open Data and Civic Apps: First-Generation Failures, Second-Generation Improvements. *Communications of the ACM*, 59(1): 82-89.
- Normandin, P.-A. (2018) Montréal s'améliore, mais continue à traîner de la patte, *La Presse*, 18 Novembre.
- O'Reilly T (2010) *Government as a Platform*, O'Reilly Media.
- Paillé, P et A Mucchielli (2008) *L'analyse qualitative en sciences humaines et sociales*, 2^e édition, Paris: Armand Colin.
- Secrétariat du conseil du trésor (2016) Lignes directrices sur la diffusion de données ouvertes.
- Secrétariat du conseil du trésor (2017) Guide de priorisation et diffusion des données ouvertes.
- Strauss, A.L., & Corbin, J. (1990). *Basic of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Susha I, Grönlund Å et M Janssen (2014) Organizational measures to stimulate user engagement with open data. *Transforming Government: People, Process and Policy*, 9(2): 181-206.
- Worthy, B. (2016) The impact of open data in the UK: Complex, unpredictable, and political. *Public Administration* 93(3): 788-805.
- Zuiderwijk A et M Janssen (2014) Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31: 17-29.

Christian Boudreau est professeur à l'École Nationale d'administration Publique du Québec (Canada). Ses travaux de recherche portent sur le gouvernement ouvert, la gouvernance collaborative et la surveillance. Il est l'auteur de plusieurs publications scientifiques sur la gestion de projets multipartenaires et sur le déploiement des caméras de surveillance dans les lieux publics au Canada et au Québec. Ses plus récents travaux portent sur la réutilisation des données ouvertes ainsi que sur la participation citoyenne de type *crowdsourcing*.