

Parce que
j'aime le savoir

L'ACFAS

DONNÉES OUVERTES
Document de réflexion



Association francophone
pour le savoir

Acfas

DONNÉES OUVERTES – DOCUMENT DE RÉFLEXION

TABLE DES MATIÈRES

Mise en contexte	p. 3
1. Une donnée ouverte, qu'est-ce que c'est?	p. 3
2. Situation au Québec	p. 4
3. Bénéfices de l'ouverture des données	p. 4
4. Freins	p. 6
5. Enjeux	p. 7
6. Et ailleurs?	p. 10
Conclusion	p. 12
Annexes	p. 13
1. Entrevue avec Daniel J. Caron de l'École nationale d'administration publique	
2. Entrevue avec Alex Guindon de l'Université Concordia	
Bibliographie	p. 19
Contact	p. 21

La production de ce document a été rendue possible grâce au soutien financier de la Fondation Lucie et André Chagnon.

Dans le présent document, le masculin est utilisé sans aucune discrimination et dans le seul but d'alléger le texte.

MISE EN CONTEXTE

La volonté d'ouvrir les données a débuté dans le courant des années 2000 dans plusieurs pays précurseurs (Grande-Bretagne, États-Unis) pour devenir un véritable phénomène mondial qu'on ne peut plus ignorer. Les pays qui ont raté le premier train n'ont pas le choix de s'inscrire dans ce mouvement qui va devenir une norme internationale et une nouvelle façon de concevoir l'information. Le Québec a amorcé ce processus inéluctable, bien qu'il accuse un certain retard par rapport à son voisin américain ou à d'autres provinces canadiennes comme l'Ontario.

À l'heure où il apparaît incontournable de nous pencher comme société sur la question des données ouvertes, ce document de réflexion présente une analyse préliminaire de ses enjeux au Québec et ailleurs dans le monde. En permettant de mieux cerner les enjeux qui nous sont spécifiques comme société, ce document souhaite alimenter les discussions entre les chercheurs, les acteurs de la société civile et les décideurs.

1. UNE DONNÉE OUVERTE, QU'EST-CE QUE C'EST?

C'est une donnée numérique (non confidentielle) dont l'accès et l'usage sont laissés libres aux usagers. Elle peut être d'origine publique ou privée, produite notamment par une collectivité, un service public ou une entreprise. Elle est diffusée de manière structurée selon une méthode et une licence ouverte garantissant son libre accès et sa réutilisation par tous, sans restriction technique, juridique ou financière.

Il faut distinguer la donnée d'une publication. La donnée est l'élément de base (quantitatif, qualitatif ou expérimental) sur lequel repose une recherche ou une publication.

L'ouverture des données s'inscrit dans une tendance qui considère l'information comme un bien commun dont la diffusion est d'intérêt public. Les données ouvertes s'inscrivent aussi dans une démarche plus vaste de transparence, de participation des citoyens et d'avancement des connaissances.

Les données ouvertes peuvent concerner les données du secteur public. Elles peuvent aussi être des données issues d'instituts de recherche scientifique publics (ou privés bénéficiant de financements publics, ou souhaitant volontairement ouvrir et partager leurs données), et accompagnent souvent des résultats de recherche eux-mêmes en libre diffusion.

Les enjeux dépassent maintenant le simple droit d'accès aux documents administratifs. En effet, les données ouvertes s'étendent au secteur privé, avec certaines organisations à but non lucratif. À titre d'exemple, Wikipédia et les grands wikis de la même famille placent tout leur contenu sous licence libre CC-BY-SA¹, et le dernier projet de cette famille Wikidata est une base de données ouverte de bases de données libres.

¹ Cette licence permet aux autres de remixer, arranger, et adapter une œuvre, même à des fins commerciales, tant qu'on accorde à l'auteur le mérite en citant son nom et qu'on diffuse les nouvelles créations selon des conditions identiques.

2. SITUATION AU QUÉBEC

Même si depuis le 5 avril 2016, le gouvernement du Québec a lancé le portail *Données Québec*², qui réunit les données ouvertes de Québec, Montréal, Laval, Gatineau et Sherbrooke ainsi que de plusieurs ministères du gouvernement du Québec, la libération des données gouvernementales en est au stade des balbutiements au Québec, essentiellement parce qu'elle ne fait pas encore partie des priorités de l'ensemble des ministères, des organismes provinciaux et municipaux, et ce, malgré l'intention politique clairement énoncée de mettre en place un gouvernement ouvert et transparent.

La première étape avait eu lieu en 2012 lorsque le gouvernement du Québec avait convié les principaux ministères et organismes provinciaux à libérer au moins trois jeux de données afin de les déposer dans le nouveau portail de données ouvertes du Québec.

Bien que la majorité des organismes ait répondu à cet appel, peu de nouveaux jeux de données se sont ajoutés au portail depuis sa création.

Certains organismes, par contre, comme la Société de l'assurance automobile du Québec (SAAQ), sont passés maîtres dans la diffusion de statistiques, entre autres, pour sensibiliser la population à divers problèmes afin d'amener celle-ci à adopter des conduites responsables.

3. BÉNÉFICES DE L'OUVERTURE DES DONNÉES

Le réel bénéfice des données ouvertes, c'est de pouvoir les amalgamer et de faire naître de nouvelles connaissances et applications.

L'ouverture des données offre la possibilité à des développeurs d'utiliser ces données pour créer des applications répondant à des besoins particuliers, par exemple pour trouver son autobus ou les services de la ville plus facilement. Ainsi, début 2017, un *hackathon* a eu lieu à Sherbrooke. Cette compétition a réuni pendant quelques jours des développeurs qui, en se basant sur les données ouvertes répertoriées sur le portail *DonneesQuebec.ca*, ont dû imaginer et programmer une application web et mobile pour rendre la vie des citoyens plus facile.

Cela permet à des entreprises privées de développer des services à l'intention des citoyens à un coût abordable pour le gouvernement. En effet, le gouvernement ne peut rivaliser avec la créativité et la capacité à prendre des risques de l'entreprise privée en ce qui concerne le développement d'applications.

Actuellement, seules certaines données sont partagées entre les différents organismes publics, par exemple les données géographiques dans le cas du projet ACRIGéo (Approche de coopération en réseau interministériel pour l'information géographique) ou lorsqu'un enjeu précis le nécessite. De plus, les données administratives sont méconnues, voire

2 Près de 800 jeux de données touchant différents secteurs (environnement, transport, économie, infrastructures et sécurité publique) sont disponibles pour le moment.

inconnues, l'exploitation généralisée et optimale de ces données est inexistante. Ouvrir les données publiques permettrait de casser le système de silos qui prévaut actuellement entre les différents organismes publics. Ce qu'un organisme produit comme données peut être d'intérêt pour un autre et éclairer celui-ci dans sa prise de décision. Par exemple, l'indice de défavorisation dans les écoles, qui est un jeu de données publié par le ministère de l'Éducation tous les ans, peut permettre au ministère de la Famille d'identifier des endroits pour installer des Centres de la petite enfance (CPE). Il en résulterait une efficacité accrue des institutions publiques.

Les données ouvertes offrent aussi aux administrations publiques une façon de se comparer entre elles, particulièrement quand les données sont bien documentées du point de vue de la méthodologie et des métadonnées. Cela permettrait également aux provinces de se comparer entre elles. La comparaison devient d'autant plus accessible que les données sélectionnées respectent des standards internationaux, comme ceux émis par l'*Open Government Partnership*³.

« En acceptant de dévoiler ses données et de se comparer à d'autres villes, Montréal a par exemple réalisé qu'elle payait pas mal plus cher que Toronto pour la collecte des déchets. Elle a donc modifié ses appels d'offres, ce qui lui a permis d'épargner 7 millions d'un coup. »⁴

En recherche

« Le terme « données ouvertes » a fait son apparition en 1995 dans le monde scientifique. Le GCDIS (le Comité sur les données géophysiques et environnementales du Conseil national de la recherche aux États-Unis) publie cette année-là un document intitulé *De l'échange complet et ouvert des données scientifiques*.

Dans ce document, il est expliqué que l'atmosphère, la biosphère et les océans forment un ensemble intégré qui transcende les frontières et qui ne peut être étudié que par une recherche coopérative internationale, s'appuyant sur des données croisées provenant des différentes sources nationales. Et de conclure : « Les programmes internationaux de recherche sur le changement global et la surveillance environnementale dépendent du principe d'un échange complet et ouvert des données (i.e. les données et les informations sont rendues disponibles sans restriction, sur une base non discriminatoire, pour un coût réduit à celui des besoins de reproduction et de distribution) ».

Durant les 15 années qui suivent, le terme se répand dans le monde scientifique et fait l'objet de différentes initiatives : engagement de scientifiques pour des données du génome ouvertes en 2002; consortium pour les données géospatiales ouvertes en 2003; pétition pour des données ouvertes en cristallographie en 2005; création en 2006 de l'*Open Data Foundation* (ODaF) qui cherche à promouvoir la transparence dans les sciences sociales, économiques et du comportement... Il s'inscrit dans le mouvement plus large en faveur

3 Le Partenariat pour un gouvernement ouvert est un partenariat multilatéral visant à promouvoir un gouvernement ouvert, à savoir la transparence de l'action publique et son ouverture à de nouvelles formes de concertation et de collaboration avec la société civile, en faisant notamment levier sur le numérique et les nouvelles technologies.

Cette initiative a été lancée le 20 septembre 2011 lors d'une réunion des huit pays fondateurs : Brésil, Indonésie, Mexique, Norvège, Philippines, Afrique du Sud, Royaume-Uni et États-Unis.

4 http://plus.lapresse.ca/screens/5d7b0949-8bc1-41f5-907d-bd5741357931%7C_0.html (consulté le 16 octobre 2017).

d'une science ouverte qui vise à permettre aux résultats scientifiques en général, et pas uniquement aux données brutes, de circuler⁵.

Les données de recherche sont précieuses, car elles sont très souvent issues d'expériences complexes et coûteuses ou sont le fruit d'observations ponctuelles sur une longue période de temps. Or il n'est pas rare qu'une fois un projet terminé ou un article publié, ces données soient détruites ou, tout du moins, inaccessibles.

L'ouverture des données de recherche, c'est-à-dire les données recueillies par des chercheurs comme matière première à des études et des publications, permettrait un avancement important de la science et de la recherche, éviterait les doublons (deux recherches similaires) et procurerait à la communauté scientifique la possibilité de vérifier et de reproduire des résultats. De plus, les auteurs qui partagent leurs données voient augmenter leurs chances d'être cités⁶.

Dans les pays en développement, où les universités n'ont pas toujours accès aux périodiques scientifiques, les données ouvertes pourraient permettre à des chercheurs de mener leurs recherches de façon adéquate.

4. FREINS

La libération des données est un mouvement qui ne va pas de soi.

Les retombées que cela engendre sont souvent abstraites pour les organismes. La mise en avant d'exemples concrets de ce qu'il est possible de faire avec des données ouvertes serait un facteur de motivation. Au Royaume-Uni, par exemple, le croisement des données ouvertes a permis une diminution de 30 % de la facture de consommation d'énergie dans les bâtiments publics en seulement deux mois⁷.

Le manque de politiques concrètes encadrant les grandes déclarations du gouvernement.

Comment s'y prend-on pour libérer des données? Par où commencer? Quelles données doit-on libérer? sont autant de questions qu'il est légitime de se poser avant d'entamer le processus d'ouverture des données. Par exemple, l'ouverture des données géospatiales fait consensus. Elles permettent de situer géographiquement diverses informations pertinentes au moyen d'une interface cartographique. Libérer les données relatives aux routes, adresses, édifices, pistes cyclables ou pistes de motoneige permettra à des développeurs de créer des applications mobiles utiles aux citoyens.

La culture au sein des organismes gouvernementaux est souvent plus encline à contrôler l'information qu'à la diffuser.

Ouvrir les données implique des changements dans la façon dont l'État gère l'information et

5 http://www.academia.edu/373264/Les_enjeux_publics_%C3%A9conomiques_et_citoyens_de_louverture_des_donn%C3%A9es_lexp%C3%A9rience_britannique (consulté le 16 octobre 2017).

6 <https://peerj.com/articles/175/> (consulté le 16 octobre 2017).

7 Charles-Henri Schulz lors de l'atelier *Open data, open standards* lors de l'*Open World Forum* du 24 septembre 2011.

développe ses systèmes ainsi que dans son rapport avec les citoyens, les chercheurs, les médias, les développeurs et les autres acteurs de la société. En libérant ses données, l'État est ainsi appelé à se décentrer de son rôle de gardien de l'information et à revêtir celui de promoteur de données exploitables.

Un modèle de financement à revoir.

Certains organismes vendent leurs données, le plus souvent pour couvrir, en totalité ou en partie, les coûts de production et de diffusion de ces données. Ainsi, si le gouvernement souhaitait rendre gratuites ces données qui génèrent des revenus, il devrait penser à des mesures compensatoires pour les revenus perdus pour les ministères producteurs de données s'ils ne peuvent plus s'autofinancer.

La confidentialité de certaines données peut également faire craindre de les rendre publiques.

Même lorsque les données confidentielles sont rendues anonymes, certains organismes demeurent prudents face à leur libération, évoquant les possibilités de réidentification grâce aux recoupements de données ouvertes de plus en plus nombreuses sur les administrés.

Le principe de confidentialité ne se limite pas à la protection des renseignements sensibles sur des individus, des collectivités ou des entreprises à qui l'État offre des services ou avec qui celui-ci fait affaire. Des données sur des activités gouvernementales jugées stratégiques peuvent aussi être frappées d'un sceau de confidentialité, comme le prévoit la *Loi sur l'accès à l'information*⁸. D'autre part, certaines organisations craignent la critique en cas d'ouverture de leurs données sur leur performance.

5. ENJEUX

La mise à jour⁹, le format¹⁰, l'anonymisation, la qualité des données et des métadonnées, la mise sur pied d'une plateforme centralisée accessible et disposant d'une interface facile à utiliser sont autant d'enjeu dans le dossier. Bien des données n'ont pas été produites pour être diffusées et, encore moins, pour être réutilisées. Un important travail de préparation et de mise en forme est donc requis pour en assurer la réutilisation dans le respect des renseignements confidentiels. Afin de rendre les données ouvertes intelligibles, celles-ci doivent être accompagnées de métadonnées qui en définissent de façon précise le contenu, le format et les dates de mise à jour. Sans métadonnées, la donnée n'est pas intelligible et l'utilisateur ne sera pas renseigné adéquatement. La démocratisation des données ouvertes passe aussi par le développement d'interfaces de visualisation le plus souvent gratuites,

centralisées et conviviales qui permettent aux utilisateurs de consulter et d'interroger les données ouvertes. En effet, pour que les données que l'on cherche soient facilement

8 Les décisions que prend le Conseil du trésor sont confidentielles pour une durée de 25 ans. Les analyses qui précèdent la prise de décision le sont tout autant.

9 Les dépenses publiques, par exemple, nécessitent de fréquentes mises à jour. Les décisions du Conseil du Trésor sur les mises à jour sont quotidiennes. Cela représente 1800 décisions par année...

10 Beaucoup de données sont déjà accessibles sur le site d'organismes gouvernementaux sous forme de tableaux, de PDFs... mais ces formats sont difficilement réutilisables.

trouvables, le mieux serait que toutes les données disponibles se retrouvent sur un même site offrant un système de recherche simple et efficace.

L'ouverture des données, qu'elles soient gouvernementales ou issues de la recherche, demande un certain encadrement de la part du gouvernement. Au-delà des déclarations gouvernementales en la matière, il faudrait, à l'heure de la première Stratégie numérique du Québec, une politique – interministérielle, tant dans sa conception, son intégration que sa mise en œuvre, autrement dit, une politique pour l'ensemble du gouvernement et ses ministères – qui vienne rapidement encadrer les grandes orientations. Rendre les données ouvertes n'est pas quelque chose de simple qui va de soi pour la plupart des employés des organismes publics ou même des chercheurs. Le mouvement de libération des données a besoin non seulement d'une impulsion politique de départ, mais aussi d'un encadrement gouvernemental pour l'orienter et le soutenir.

La libération des données peut aussi susciter diverses questions de la part des utilisateurs qui s'attendent à une réponse des organismes et, donc, à un certain soutien à la clientèle, ce qui demande du temps. Par exemple, les utilisateurs peuvent détecter des erreurs dans les jeux de données et les signaler à l'organisme détenteur. Les utilisateurs sont aussi susceptibles d'appeler l'organisme qui libère des données quand ils observent des variations entre les données ouvertes et d'autres données comparables auxquelles ils ont eu accès.

Il serait souhaitable de mettre à contribution des partenaires¹¹ qui possèdent l'expertise permettant de normaliser et d'indexer les données ouvertes (ex. : métadonnées), de les anonymiser, de les intégrer sur un même site ou de les consulter et de les interroger.

Libérer les données publiques permettrait de limiter le nombre de demandes d'accès à l'information. L'ouverture des données ne devrait pas, pour autant, remplacer une réforme de la *Loi de l'accès à l'information*, selon la Fédération professionnelle des journalistes du Québec¹².

Un accès gratuit à des quantités importantes de données brutes ne peut que faciliter la vie des chercheurs dans la réalisation de leurs travaux. Si les données ouvertes sont nombreuses et de qualité, les chercheurs peuvent être en mesure d'analyser le fonctionnement des programmes et des services publics et d'en évaluer la performance.

Libérer des données génère des coûts, notamment de main d'œuvre. Dans un contexte de réduction budgétaire, les organismes tentent déjà de conserver leurs activités et n'ont pas forcément de budget à consacrer à des tâches supplémentaires.

Certains choisissent donc de vendre leurs données pour couvrir, en totalité ou en partie, les coûts d'acquisition, de préparation et de diffusion de ces données dans un contexte d'autofinancement. Par exemple, l'article 12 de la *Loi du ministère de l'Énergie et des Ressources naturelles* autorise ce ministère à « produire de l'information géographique et à la distribuer à titre onéreux ». L'élaboration et la diffusion de la banque de données la plus complète sur les adresses au Québec s'inscrivent aussi dans cette logique marchande.

11 Institut de la statistique pour les données sur la population québécoise, Régie de l'assurance-maladie et du ministère de la Santé et des Services sociaux pour les données en santé, ministère de la Sécurité publique et ministère de l'Énergie et des Ressources naturelles pour les données en géomatique et SOQUIJ pour l'information juridique.

12 <https://www.fpqj.org/quebec-lance-programme-de-donnees-ouvertes/> (consulté le 16 octobre 2017).

Si le gouvernement souhaite libérer et rendre gratuites des données qui génèrent des revenus, il doit prévoir des mesures compensatoires pour les pertes de revenu des organismes qui les vendent ou pour financer le travail de gestion des données.

Des données disponibles, mais payantes restent-elles des données ouvertes? La question mérite d'être posée.

Il faut faire de l'ouverture des données une étape naturelle dans la production de l'information. C'est en insérant le processus d'ouverture des données tôt dans le cycle de développement des systèmes d'information, et non à la fin de ce cycle, au moment de la diffusion des données que les organismes pourront aspirer au principe d'ouverture par défaut, tel que proposé par la Charte du G8 sur les données ouvertes¹³ et adopté par les pays membres de l'*Open Government Partnership*, dont le Canada, la France, le Royaume-Uni et les États-Unis.

En recherche

Dans plusieurs domaines de la recherche, malgré des appels à de bonnes pratiques de transparence et de partage des données, la donnée brute n'est pas communiquée par les auteurs, même aux relecteurs, ce qui a pu entraîner de nombreuses fraudes scientifiques, notamment dans le domaine de la psychologie où 50 % des expérimentations présentent des résultats non reproductibles¹⁴, ce qui pose de graves problèmes de confiance dans les résultats.

Selon Alex Guindon¹⁵, le Québec a pris du retard sur d'autres régions au Canada, mais aussi et surtout sur la Grande-Bretagne et les États-Unis qui comptent sur des politiques d'archivage et de partage des données¹⁶ et sur des centres d'expertise qui facilitent ce travail. Ce retard sera long à rattraper. Et le temps presse : les résultats de la consultation des agences subventionnaires ne sont pas encore connus, mais il y a fort à parier qu'on retiendra l'idée de demander aux chercheurs de déposer un plan de gestion de données comme condition de l'obtention de leurs subventions¹⁷. En l'absence de ressources pour aider à la rédaction de ces plans, les chercheurs perdront un temps précieux qu'ils préféreraient sûrement consacrer à leur recherche. De plus, dans de nombreuses disciplines, faute de pouvoir compter sur un dépôt de données situé au Québec ou même au Canada, les chercheurs seront forcés d'archiver leurs données sur des sites américains.

Pendant ce temps, en Ontario, les universités, soutenues par le gouvernement provincial, comptent sur l'excellent *Scholars Portal* qui permet, entre autres choses, aux chercheurs ou aux bibliothécaires responsables des données de déposer leurs ensembles de données dans un dépôt certifié. Il y a lieu de se demander si les universités québécoises, surtout

13 <http://ouvert.canada.ca/fr/charte-du-g8-sur-les-donnees-ouvertes-plan-daction-du-canada> (consulté le 16 octobre 2017).

14 <http://science.sciencemag.org/content/349/6251/aac4716> (consulté le 16 octobre 2017).

15 Alex Guindon est le bibliothécaire du service de données numériques et géospatiales de l'Université Concordia.

16 Des consultations pour établir la nouvelle politique de gestion des données de recherche regroupant les trois organismes fédéraux – Instituts de recherche en santé du Canada (IRSC), Conseil de recherches en sciences naturelles et génie du Canada (CRSNG) et Conseil de recherches en sciences humaines du Canada (CRSH) - devraient démarrer prochainement.

17 http://www.science.gc.ca/eic/site/063.nsf/fra/h_83F7624E.html (consulté le 16 octobre 2017).

celles qui comptent le moins d'étudiants, pourront relever les défis associés à la gestion des données de recherche sans se doter d'outils communs. La collaboration semble de mise et celle-ci serait grandement facilitée par l'appui du gouvernement du Québec.

La quasi-totalité des provinces canadiennes et les États-Unis¹⁸ offrent gratuitement leurs données géospatiales aux chercheurs, permettant ainsi des avancées majeures en science. Au Québec, depuis 2015, une entente entre le ministère de l'Énergie et des Ressources naturelles (MERN) et le Bureau de coopération interuniversitaire (BCI) permet aux universités de la province d'avoir accès à une grande quantité de données géospatiales à des coûts acceptables, ce qui constitue une avancée importante pour la communauté universitaire. Dans un avenir rapproché, des discussions fructueuses avec le MERN pourraient permettre un accès entièrement gratuit à ces données pour les universités et la mise en place d'une plateforme unique et conviviale. Une telle plateforme faciliterait l'accès et l'utilisation de ces données par les chercheurs¹⁹.

6 - ET AILLEURS?

Au Canada

À l'échelle nationale, c'est en 2001 que le gouvernement canadien a créé « GéoBase », une base de données géospatiales. Au niveau municipal, c'est en 2009 que Vancouver et Toronto ont rendu disponibles leurs données, alors que Montréal n'a lancé son portail qu'en 2011. Au niveau provincial, la Colombie-Britannique a créé son portail en 2011. Le gouvernement du Québec venait alors seulement de publier un rapport à ce sujet.

Au Canada, le ministre Stockwell Day, président du Conseil du Trésor a annoncé, le 17 mars 2011, le lancement d'un projet-pilote de 12 mois pour le site fédéral de données ouvertes data.gc.ca/²⁰. Le 11 avril 2012, le ministre Tony Clement, nouveau président du Conseil du Trésor, a annoncé le Plan d'action du Canada pour un gouvernement ouvert. Ce plan de 3 ans inclut :

- L'accroissement du nombre d'ensembles de données disponibles et la mise en place d'un cadre de recommandation pour les nouvelles générations de plateformes;
- Un nouveau design et une réorganisation du portail national data.gc.ca, de façon à améliorer le niveau de standardisation des données, lors de la deuxième et de la troisième année.

En plus du site principal data.gc.ca/, l'initiative fédérale comprend :

- [GeoGratis](#) du ministère des Ressources naturelles (données géospatiales) ;
- [Données environnementales](#) du ministère Environnement Canada.

Le Canada est devenu un membre du Partenariat pour un gouvernement ouvert (PGO) en 2011. Le Partenariat pour un gouvernement ouvert est une initiative internationale qui

18 Le USGS (*United States Geological Survey*) a ouvert l'accès gratuit aux données Landsat à la communauté internationale en 2008. Depuis, le nombre de publications scientifiques se basant sur ces données a suivi une courbe exponentielle.

19 <http://geoindex-plus.bibl.ulaval.ca/> (consulté le 16 octobre 2017).

20 Maintenant <http://ouvert.canada.ca> (consulté le 16 octobre 2017).

fournit une plateforme aux réformateurs à l'intérieur et à l'extérieur de leur gouvernement afin de développer des réformes qui encouragent la transparence, renforce le pouvoir des citoyens, lutte contre la corruption et exploite de nouvelles technologies pour consolider la gouvernance. Depuis sa création en septembre 2011, plus de 2000 engagements ont été pris par les 69 pays membres, couvrant un tiers de la population mondiale.

Afin de devenir membre, les pays doivent répondre à un ensemble de critères d'admissibilité de base et doivent accepter une déclaration de gouvernement ouvert. Une fois membres, les gouvernements doivent élaborer un plan d'action national avec la société civile dans leur pays sur une base bisannuelle. Le gouvernement doit rendre des comptes régulièrement quant à ses progrès et doit travailler avec la société civile pour réaliser les réformes convenues. Les progrès sont évalués à intervalles réguliers par un chercheur indépendant nommé par le Mécanisme de rapport indépendant du PGO. Le PGO souligne le partenariat entre le gouvernement et la société civile à tous les niveaux. Son comité directeur est formé d'un nombre égal de représentants du gouvernement et de la société civile, avec des coprésidents issus de chacun des groupes.

Dans le monde

L'*Open Data Index*²¹ est un indicateur qui liste des pays selon l'ouverture de leurs données publiques (statistiques, cadastres, données environnementales, législations, budget, etc.). Il est publié annuellement depuis 2013, par l'*Open Knowledge Foundation*.

Les exemples

Qu'est-ce qui explique l'avance de pays comme les États-Unis ou la Grande-Bretagne?

Valérie Peugeot, chercheuse au sein d'Orange Labs et présidente de l'association Vecam explique ce qui a contribué au succès de ces deux pays : un mouvement en faveur de la libération des données publiques²² est apparu de façon quasi concomitante aux États-Unis et au Royaume-Uni, mais aussi au Danemark, en Finlande, en Australie... au début des années 2000. Cela s'est traduit par les lancements successifs, en mai 2009, du site du gouvernement américain data.gov et, en janvier 2010, du site du gouvernement britannique data.gov.uk, deux portails mettant à disposition des bases de données ouvertes provenant des différents corps de l'administration.

Aux États-Unis et en Grande-Bretagne, deux catégories d'intérêts et plusieurs groupes d'acteurs se sont croisés pour permettre l'émergence de ces initiatives.

Le premier groupe, très visible, regroupe les acteurs qui militent en faveur d'une meilleure qualité démocratique et d'une plus grande transparence des pouvoirs publics. Cette approche en faveur de la transparence a été portée par trois groupes d'acteurs qui ont permis d'obtenir des résultats concrets : les « *civic hackers* »²³ en premier lieu qui ont, souvent sans attendre que des données soient libérées, créé des applications, fournissant ainsi une démonstration vivante de l'utilité démocratique de cette ouverture. Ils ont souvent reçu le soutien de fondations comme *The Sunlight Foundation* aux États-Unis ou de la presse. *The Guardian* a joué un rôle fondamental au Royaume-Uni.

21 <http://index.okfn.org/place/> (Le Canada arrive 17e ; Taiwan, 1er ; la Grande-Bretagne, 2e ; les États-Unis, 8e.). (consulté le 16 octobre 2017).

22 Nombre de maladies nosocomiales contractées dans les hôpitaux, horaires des transports publics, dépenses publiques, espèces plantées dans les jardins publics...

23 Josh Tauberer aux États-Unis a ouvert la voie en 2004 avec le site GovTrack.us ou le groupe de développeurs et d'activistes derrière TheyWorkForYou.com lancé la même année au Royaume-Uni.

Deuxième groupe, des décideurs politiques et des personnalités du Web : aux États-Unis, l'implication de Barak Obama et de son *Chief Information Officer*, Vivek Kundra, précédemment *Chief Technology Officer* du District de Columbia (Washington DC) où il avait déjà commencé à libérer des données a été déterminante ; au Royaume-Uni, la personnalité de Tim Berners-Lee, considéré comme le père du Web, a joué un rôle fondamental pour rassurer l'administration et convaincre Gordon Brown de donner l'impulsion politique indispensable.

Enfin troisième groupe souvent ignoré, mais qui a joué un rôle fondamental au Royaume-Uni, les fonctionnaires dans les administrations. Ils ont identifié les bases de données à ouvrir et les ont proposées aux coordinateurs de projets, parfois en contournant une hiérarchie peu au fait de l'intérêt de la démarche, voire hostile.

À ces trois groupes, il convient d'en ajouter un quatrième, transversal : les collectivités locales. Très vite, parfois avant les pouvoirs publics nationaux, certaines d'entre elles ont mis à disposition leurs données, contribuant ainsi à légitimer le mouvement : Columbia, mais aussi Portland, New York, Londres sont parmi les pionnières.

À côté de cette finalité démocratique, un second enjeu est apparu très tôt autour des données ouvertes : celui de leur valeur sociale et économique, en tant que facteur d'innovation. En mettant à disposition des entreprises des données, les pouvoirs publics aident ces dernières à construire de nouveaux services et ce faisant à simplifier la vie quotidienne de leurs utilisateurs, à créer de l'emploi et à générer globalement de l'activité.

CONCLUSION

La révolution de l'ouverture des données est en marche, mais, pour Timothy John Berners-Lee, fondateur de la *World Wide Web Foundation*, certains éléments sont nécessaires pour que ce mouvement progresse :

- un haut niveau d'implication politique pour continuer à rendre disponible des données publiques;
- des investissements substantiels et sur le long terme pour soutenir et former la société civile et les entrepreneurs;
- des réformes juridiques;
- et répondre aux besoins des communautés locales.

Nous vous invitons à partager ce document afin qu'il puisse servir de point de départ à des discussions sur la façon d'ouvrir les données et sur les moyens à mettre en place pour que le Québec rejoigne le train de ce phénomène inéluctable à l'échelle mondiale.

ANNEXE 1 - Entrevue avec Daniel J. Caron de l'ÉNAP

Daniel J. Caron est professeur à l'École nationale d'administration publique (ÉNAP) où il est titulaire de la Chaire de recherche en exploitation des ressources informationnelles. Il enseigne aussi l'évaluation de programmes à la *School of Public Policy and Administration* à l'Université Carleton. Ses champs de recherche et d'expertise incluent l'exploitation des ressources informationnelles, l'évaluation des politiques publiques et administratives et les questions autochtones. Il est présentement l'un des coprésidents du panel sur la transparence et le gouvernement ouvert à l'*International Research Society of Public Management*. Il concentre ses efforts de recherche sur les répercussions des technologies numériques sur l'usage de l'information dans le fonctionnement de l'État et des organisations publiques.

Il est l'auteur de nombreux travaux dans le domaine de l'administration publique. En 2011, il a publié *Web HT.O. Pour une société informée : la pertinence numérique et ses défis pour les sociétés démocratiques au 21^e siècle* aux éditions Hermann, à Paris. Son dernier livre, *l'Homme imbibé*, a été publié au printemps 2014 et porte sur les enjeux liés à l'environnement numérique et à son impact sur le développement des codes d'interprétation dans nos sociétés.

Avant de se consacrer entièrement à la recherche et l'enseignement, Daniel J. Caron a travaillé plus de trente années dans la fonction publique canadienne. Son dernier poste était celui d'administrateur général de Bibliothèque et Archives Canada. Il est titulaire d'un baccalauréat et d'une maîtrise en économie de l'Université Laval et d'un doctorat en sciences humaines appliquées de l'Université de Montréal.

Est-ce que le Québec est en retard au sujet de l'ouverture de ses données par rapport à d'autres provinces canadiennes ou d'autres pays?

« Non, pas du tout. Il y a beaucoup de grandes déclarations, de grandes orientations, mais on s'appuie surtout sur certains très bons coups comme Boston avec son *city score* et son label de ville intelligente. Après on en fait du *benchmarking*, ce qui n'est pas mauvais en soi puisque ça a quelque chose d'incitatif, ça force les gens à se questionner. Mais au final on a très peu d'études sur les retombées de l'ouverture des données et sur les moyens qu'on a pris pour la mettre en place. Concernant la transparence, on est allé chercher des pays très en retard dans ce domaine et on les cite maintenant en exemple. Au Québec, on n'en est plus là, c'est quelque chose d'intégré dans notre société et on devrait donc être loin devant, mais on stagne. L'Ontario²⁴ aussi a quelques expériences intéressantes à son actif, mais on en est encore aux balbutiements.

24 L'Ontario est l'un des 15 gouvernements infranationaux choisis pour participer au programme pilote de l'*Open government partnership*. Mis en place dans 70 pays, ce programme prône la transparence gouvernementale et son ouverture à de nouvelles formes de concertation et de collaboration avec la société civile, en faisant notamment levier sur le numérique et les nouvelles technologies. Un de ses quatre objectifs c'est d'accroître la disponibilité des informations. Les membres s'engagent à promouvoir la divulgation des activités gouvernementales à tous les niveaux, à recueillir et publier les données sur les dépenses publiques et la performance du gouvernement en matière de services essentiels, et à publier des informations sans délai et dans des formats standard, que le public peut aisément comprendre et qui sont faciles à réutiliser. L'Ontario s'est également engagé élaborer un outil numérique pour encourager les jeunes à partager leurs idées sur les programmes et services gouvernementaux et créer un programme de formation ouvert pour le personnel gouvernemental dans le secteur public de l'Ontario.

Derrière les grandes orientations politiques, on s'est rendu compte avec nos rapports qu'il existe des barrières qui dépassent ces belles orientations. Il faut repenser complètement l'information. On a encore des réflexes de produire sur papier, on gère des documents, pas du flux. Or on se dirige dans un monde de flux informationnel et il faut repenser complètement dès le départ ce qu'on va faire des données. Quelles données libère-t-on? Pour qui? À quelle fin? Comment? Il ne faut pas répondre à ces questions une fois les données produites, mais bien avant pour créer des bases de données qui y répondent. Il faut forcer l'administration à rendre la donnée compréhensible. »

D'où doit venir l'impulsion?

« L'impulsion doit venir à la fois du haut – par une volonté politique, des grandes orientations, et des appuis (financiers ou de mise en place des pratiques) – et du bas. Les grands oubliés, ceux qu'on ne consulte pas assez dans ce dossier, ce sont les utilisateurs, parce qu'à quoi bon ouvrir des données si personne ne les utilise? Il faut donc consulter les chercheurs, les développeurs, les groupes de citoyens, les journalistes et instaurer une coproduction pour développer les bases de données selon leurs besoins. En ce moment, ça tire un peu de partout et je pense que ce sont les usages et les usagers qui, par la force des choses, vont faire bouger l'État.

Le problème, à mon avis, c'est qu'on ne teste pas. Ça prend des projets-pilotes, ça suppose éventuellement des erreurs, mais il faut le faire pour se rendre compte de ce qui fonctionne ou pas plutôt que d'en rester aux mots. Ça prend un certain courage politique.

On devrait penser que les données sont un bien public, payé avec les taxes des citoyens et qu'il est donc normal qu'ils y aient accès, en respectant les limites de confidentialité et de sécurité publique bien sûr. Mais là encore, on pourrait autoriser, sous certaines conditions, des chercheurs qui travaillent sur des sujets sensibles, à avoir accès à ces données.

Aux États-Unis, par exemple, ils sont moins frileux sur la question des renseignements personnels que nous au Québec. Il n'y a qu'à prendre l'exemple du projet-pilote de la carte santé à Rimouski dans les années 1990. Ça n'a pas marché alors que dans d'autres pays comme le Danemark par exemple, toutes les informations médicales sont accessibles au niveau national grâce à un système central informatisé. Mais la différence, c'est que là-bas on se fie à l'État. »

Quels sont les obstacles à l'ouverture des données?

« Il y a d'abord la culture gouvernementale, institutionnelle. On pense hiérarchie, verticalité, approbation, filtrage. Il y a des craintes qu'un lâcher-prise, en ce qui a trait à l'information qu'on a l'habitude de contrôler, entraîne des critiques et des remises en question. On pense aussi beaucoup technologie, mais l'ouverture des données ne repose pas que sur la technologie. La technologie vient seulement appuyer l'information. Ce qui est dommage c'est qu'on parle beaucoup des erreurs engendrées par un algorithme pour appuyer le fait qu'il ne faut pas se reposer là-dessus, mais l'erreur humaine, elle existe aussi et on ne remet pas en question l'utilité des gens. À chaque scandale autour des données, on recule de 10 ans. »

Et les bénéfiques?

« Je pense que ça donnerait de meilleures politiques publiques. Les citoyens seraient mieux informés et pourraient dire à leur gouvernement ce qu'ils veulent de lui. »

Et les chercheurs là-dedans? Ils sont à la fois utilisateurs et fournisseurs de données...

« Tout à fait. Cependant, on sent une résistance de la part des chercheurs qui ne veulent pas forcément lâcher leurs données récoltées, qui sont une mine d'or avec laquelle ils peuvent publier trois, quatre, cinq articles. L'information c'est le pouvoir. Un des enjeux, c'est de savoir qui possède quoi. Parce que les chercheurs sont payés avec nos taxes aussi.

Ce serait un enrichissement pour la société que le flux informationnel aille dans les deux sens, que les institutions nourrissent les chercheurs et inversement. »

ANNEXE 2 - Entrevue avec Alex Guindon de l'Université Concordia

Alex Guindon est le bibliothécaire du service de données numériques et géospatiales de l'Université Concordia. M. Guindon a obtenu une première maîtrise en Sciences politiques de l'Université du Québec à Montréal et une seconde en Sciences de la Bibliothèque et de l'Information de l'Université de Montréal. Il travaille comme bibliothécaire à Concordia depuis 2000 et a été bibliothécaire de liaison pour les études françaises, la géographie et la science politique.

Ses intérêts de recherche actuels tournent autour de la gestion des données de recherche. Il est membre du groupe d'experts du Plan de gestion des données (DMP) de l'Association des bibliothèques de recherche du Canada (CARLABRC) qui est en train de créer un modèle qui servira à aider les chercheurs dans leurs propositions de subvention.

M. Guindon a été membre du Comité d'éducation de l'Initiative de libération des données et est membre du comité de gestion de la branche McGill-Concordia du Centre interuniversitaire québécois pour les statistiques sociales (CIQSS).

Alex Guindon s'intéresse depuis environ 6 ans déjà à l'ouverture des données, domaine auquel il s'est formé lui-même. Le 21 novembre 2017, il donnera une formation sur l'ouverture des données aux étudiants aux études avancées de l'Université Concordia.

Où en est-on au Québec, par rapport aux autres provinces, en termes d'ouverture des données de recherche?

« Les autres provinces sont généralement beaucoup plus avancées que le Québec pour permettre aux chercheurs de déposer des données de recherche. Les universités ont des fonds de dépôt permettant à n'importe quel chercheur de n'importe quelle discipline de déposer ses données et de les rendre accessibles.

On va dire que la réflexion s'est faite plus tardivement ici et il n'y a vraiment pas grand-chose de mis en place. Il n'existe pas de directives contraignantes de la part des agences de financement pour forcer les chercheurs à déposer leurs données. Je pense que ça va arriver d'ici quelques mois, ceci dit. Au Québec, c'est une minorité de chercheurs qui sont sensibilisés à l'ouverture des données de recherche et qui les déposent dans des fonds de dépôt. On parle de chercheurs dans quelques spécialités : informatique, mathématiques, santé... D'ailleurs en génétique, par exemple, les Instituts de recherche en santé du Canada (IRSC), qui sont l'organisme fédéral de financement de la recherche en santé au Canada, obligent les chercheurs qui reçoivent du financement à déposer leurs données. Mais pour ce que j'en sais, personne ne vérifie que c'est vraiment le cas. Certaines publications périodiques exigent maintenant que les chercheurs déposent leurs données avant publication, ça va sans doute être un incitatif, car c'est essentiel pour un chercheur de publier.

Quand on parle de données de recherche, mises à part les données en santé, qui sont souvent publiques, on ne parle pas forcément de données qui sont accessibles au grand public. Elles sont accessibles soit à la communauté universitaire, soit au gouvernement, mais ce sont surtout les chercheurs qui vont les utiliser, car les données sont très spécialisées et requièrent des connaissances poussées pour être compréhensibles.

Les données déposées peuvent aussi être disponibles sur demande, fournies avec un protocole et une licence d'utilisation qui régissent leur utilisation, ou alors sous embargo.

Si les autres provinces sont plus avancées, elles n'en sont encore qu'au début de l'ouverture de leurs données de recherche si on compare la situation à d'autres pays comme les États-Unis, bien sûr, ou surtout la Grande-Bretagne qui est LE modèle à suivre, le pays le mieux organisé, avec le plus de ressources et qui a pris le virage de l'ouverture des données de recherche depuis longtemps. »

Quels sont les enjeux de l'ouverture des données de recherche?

« L'enjeu, c'est de créer une masse critique de données ouvertes pour accélérer la recherche et éviter les doublons. Le fondement même des sciences, c'est la reproductibilité des résultats. Ça c'est dans la théorie, mais en pratique ça n'arrive pas souvent, car les données utilisées pour arriver à un résultat donné ne sont, en fait, pas accessibles. Il y a très peu de demandes de la part des chercheurs pour déposer leurs données donc cela crée peu d'incitatifs envers les bibliothèques à mettre en place des systèmes pour aider les chercheurs.

Il faut que les chercheurs pensent à l'ouverture des données en amont, avant leur collecte. Ça se planifie dès le départ, les sortes de données qu'ils vont générer. Si ce sont des chercheurs en sciences sociales, il faut faire signer une entente aux participants comme quoi ils acceptent que les données soient rendues publiques... L'idéal c'est de documenter les données au fur et à mesure de leur collecte, mais on en revient toujours à l'argent, ça nécessite une personne dans une équipe de recherche qui se dédie à ça. Pour que les données soient compréhensibles et réutilisables, il faut leur attacher des métadonnées qui expliquent la méthodologie employée pour la collecte, les instruments de recherche utilisés...

Aux États-Unis, au Michigan, *The Interuniversity Consortium for Political and Social Research* (ICPSR) est un grand centre d'expertise de l'ouverture des données en sciences sociales qui fait le travail pour les institutions membres. Mais il faut que les données soient documentées depuis leur collecte pour qu'une tierce personne puisse ensuite les traiter. Si je pense à long terme, n'hésitons pas à rêver, il faudrait modifier la façon dont on évalue les chercheurs et valoriser autant la publication des données que celle des articles scientifiques. »

Qu'est-ce qui pourrait accélérer le processus d'ouverture des données de recherche?

« Le gouvernement pourrait exiger des chercheurs qu'ils rendent publiques leurs données de recherche s'ils ont utilisé des données gouvernementales ouvertes par exemple. Parfois aussi les chercheurs ont signé des ententes restrictives, pour avoir accès à des données gouvernementales et pour les utiliser, qui les empêchent de publier les données de recherche produites avec ces données gouvernementales.

Dans mon travail, j'aide les chercheurs à obtenir les données gouvernementales dont ils ont besoin et elles ne sont pas souvent si faciles à obtenir. Il faut signer des ententes, les acquérir à grands frais... Aux États-Unis, toutes les données géospatiales sont gratuites

et ouvertes. Ça prendrait aussi des formations, pour les chercheurs comme pour les bibliothécaires ou archivistes. Pour le moment, les seules personnes spécialisées se sont formées toutes seules, par intérêt personnel. Ça va sans doute changer, car j'ai vu qu'à l'École de bibliothéconomie et des sciences de l'information de l'Université de Montréal ils recherchent un professeur pour donner un cours sur la gestion de données.

On sent quand même qu'il y a une ouverture. Je tiens à saluer le travail du gouvernement fédéral qui a ouvert toutes ses données statistiques et géospatiales. Au niveau provincial, on a une entente avec le MERN (ministère de l'Énergie et des Ressources naturelles) et le ministère collecte des données statistiques sur l'utilisation de leurs données géospatiales et recueille nos commentaires pour l'amélioration de leur service. »

BIBLIOGRAPHIE

La gestion des données de recherche en bibliothèque universitaire, Alex Guindon, Volume 59, numéro 4, octobre-décembre 2013, Association pour l'avancement des sciences et des techniques de la documentation (ASTED).

Rapport de recherche, mai 2016, *Les données ouvertes dans l'administration publique québécoise : Utilités, freins et pistes de solution*, Christian Boudreau, Ph. D., professeur, ÉNAP et Daniel J. Caron, Ph. D., professeur, ÉNAP.

http://recherche.enap.ca/cerberus/files/nouvelles/documents/La_recherche/NetGouv_Lesdonneesouvertesdansadministrationquebecoise_2016.pdf

« La démocratie 2.0 en format intelligent », Danny Raymond, *Agence Science-Pressé*, 12 mars 2013.

<http://www.sciencepresse.qc.ca/actualite/2013/03/12/democratie-20-format-intelligent>

« Les données ouvertes de 5 villes réunies dans un portail », *Radio-Canada*, publié le 5 avril 2016 à 7 h 14, mis à jour le 5 avril 2016 à 13 h 36.

<http://ici.radio-canada.ca/nouvelle/774084/donnees-ouvertes-portail-lancement-cinq-grandes-villes>

« Boston City Score, vers une gestion municipale en temps réel », *Le blog OuiHop*, 5 juillet 2016.

<http://smartattitude.ouihop.com/smart-attitude/smart-techno/boston-city-score-vers-une-gestion-municipale-en-temps-reel/>

Nouvelles initiatives pour augmenter la participation du public en Ontario, Secrétariat du Conseil du Trésor, 9 décembre 2016.

<https://news.ontario.ca/tbs/fr/2016/12/nouvelles-initiatives-pour-augmenter-la-participation-du-public-en-ontario.html>

« Ontario adopts new open government initiatives », Mandy Kovacs, *It World Canada*, 12 décembre 2016

<http://www.itworldcanada.com/article/ontario-adopts-new-open-government-initiatives/389177>

« L'open data pour démasquer les logements insalubres », Mathias Marchal, *Journal Métro*, publié le 19 février 2013, mis à jour le 20 février 2014.

<http://journalmetro.com/actualites/montreal/260093/lopen-data-pour-demasquer-les-logements-insalubres/>

« 'Open data': Why it's making waves across Canada », *CBC News*, 1^{er} mai 2016.

<http://www.cbc.ca/news/technology/open-data-movement-canada-1.3557562>

« Troisième Plan biennuel dans le cadre du Partenariat pour un gouvernement ouvert », publié en juillet 2016, mis à jour le 03 mars 2017.

<http://ouvert.canada.ca/fr/contenu/troisieme-plan-biennuel-partenariat-gouvernement-ouvert>

« Charte du G8 sur les données ouvertes, plan d'action du Canada », publié en juin 2014, mis à jour le 8 mars 2015.

<http://ouvert.canada.ca/fr/charte-du-g8-sur-les-donnees-ouvertes-plan-daction-du-canada>

Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique, Valérie Peugeot, 16 novembre 2010.

http://www.academia.edu/373264/Les_enjeux_publics_%C3%A9conomiques_et_citoyens_de_louverture_des_donn%C3%A9es_lexp%C3%A9rience_britannique

« The Open Data Divide: UK & USA lead the world, but others fall behind », Derek du Preez, 21 janvier 2015.

<http://diginomica.com/2015/01/21/open-data-divide-uk-usa-lead-world-others-fall-behind/>

« Out of the box », *The Economist*, 21 novembre 2015.

<http://www.economist.com/news/international/21678833-open-data-revolution-has-not-lived-up-expectations-it-only-getting>

« How does a country get to open data? What Taiwan can teach us about the evolution of access », Jonathan Stray, *Nieman Lab*, 10 avril 2013.

<http://www.niemanlab.org/2013/04/how-does-a-country-get-to-open-data-what-taiwan-can-teach-us-about-the-evolution-of-access/>

« Préserver les données de la recherche à l'ère du Big Data », Guillaume Garvanèse, *CNRS Le Journal*, 9 septembre 2016.

<https://lejournel.cnrs.fr/articles/preserver-les-donnees-de-la-recherche-a-lere-du-big-data>

CONTACT

ASSOCIATION FRANCOPHONE POUR LE SAVOIR – ACFAS

425, rue De La Gauchetière Est
Montréal (Québec) H2L 2M7

www.acfas.ca

Coordination du document :
Isabelle Gandilhon
isabelle.gandilhon@acfasc.ca
514 849-0045, poste 222

Rédaction du document :
Alexia Malarmey

Rédaction : mars 2017
Diffusion : octobre 2017

© ACFAS