



Rapport de recherche
Mai 2016

Les données ouvertes dans l'administration publique québécoise : Utilités, freins et pistes de solution

Par

Christian Boudreau, Ph. D., professeur, ENAP

Daniel J. Caron, Ph. D., professeur, ENAP

 École nationale
d'administration
publique

NetGouv Données ouvertes***Mesure de l'apport et de l'évolution du gouvernement ouvert au Québec***

Ce rapport de recherche s'inscrit dans le cadre du projet *NetGouv Données ouvertes* qui vise à élaborer, mettre en place et diffuser les résultats d'une nouvelle mesure de l'usage des données gouvernementales au Québec et à évaluer comment cette mesure favorise la croissance économique. Ce projet cherche à mesurer l'adhésion à la nouvelle approche d'ouverture des données gouvernementales au sein des différents ministères, organismes gouvernementaux et municipalités. Il s'intéresse également à la collaboration à l'interne avec d'autres ministères et organismes, ainsi qu'à la planification et au déploiement des données ouvertes.

Équipe de projet – CEFRIO

Claire Bourget, directrice principale de la recherche marketing

David Huynh Quan Suu, chargé de projet en recherche marketing

Guillaume Ducharme, directeur des communications

Julia Gaudreault-Perron, chargée de projet

Annie Lavoie, conseillère aux communications

Équipe scientifique

Christian Boudreau, Ph. D., professeur, École nationale d'administration publique

Daniel J. Caron, Ph. D., professeur, École nationale d'administration publique

© CEFRIO, 2016. Tous droits réservés.

L'information contenue dans les présentes ne peut être utilisée ou reproduite par une tierce partie à moins d'une autorisation écrite du CEFRIO.

TABLES DES MATIÈRES

INTRODUCTION	4
1. UTILITÉS POTENTIELLES LIÉES À LA (RÉ)UTILISATION DES DONNÉES OUVERTES	6
1.1 DÉVELOPPEMENT D'APPLICATIONS ET DE SERVICES	6
1.2 TRANSPARENCE DE L'ÉTAT ET SURVEILLANCE PAR LA SOCIÉTÉ.....	7
1.3 AVANTAGES POUR LES ORGANISMES.....	8
2. FREINS ORGANISATIONNELS À LA LIBÉRATION DES DONNÉES	10
2.1 D'AUTRES PRIORITÉS.....	10
2.2 DES OPÉRATIONS QUI PEUVENT ÊTRE EXIGEANTES	11
2.3 DÉSUÉTUDE ET ÉPARGILLEMENT TECHNOLOGIQUES.....	12
2.4 DES DONNÉES DE QUALITÉ INÉGALE.....	13
2.5 SENSIBILITÉ DE L'INFORMATION ET CULTURE DU SECRET	15
2.6 VENTE DE DONNÉES	16
3. CONSTATS ET PISTES DE SOLUTION	18
3.1 UNE GOUVERNANCE CENTRALE AUX ORIENTATIONS CLAIRES.....	18
3.2 UNE MISE EN COMMUN D'EXPERTISES ET DE RESSOURCES.....	20
CONCLUSION	22

INTRODUCTION

L'administration publique est sans conteste l'un des plus gros producteurs et détenteurs de données d'intérêt public. Ses données recouvrent plusieurs secteurs d'activités dans lesquels elle est engagée, comme le transport, la santé, la sécurité publique, l'économie, l'environnement, la culture, l'éducation, l'agriculture et bien d'autres secteurs. En libérant ses données et en les rendant facilement accessibles et exploitables, l'administration publique entend stimuler l'économie par le développement d'applications mobiles et, du coup, améliorer son offre de services à faible coût tout en favorisant la transparence de ses institutions. Les données ouvertes seraient porteuses de retombées multiples (ex : économique, démocratique et administrative) grâce à leur (ré)utilisation¹ par divers acteurs de la société (ex : citoyens, journalistes, entreprises, OBNL, administration publique). Par « données ouvertes », nous entendons des données numériques, le plus souvent granulaires² et gratuites, que des organismes publics rendent accessibles dans des formats et des conditions qui assurent leur libre utilisation, réutilisation et rediffusion par quiconque³.

Malgré l'engouement qu'elles suscitent depuis quelques années à l'échelle internationale, les données ouvertes demeurent, dans une large mesure, un phénomène en émergence. Au-delà des déclarations gouvernementales en la matière, les administrations publiques se heurtent à divers obstacles qui freinent la libération des données, à savoir la production, la mise en forme et la diffusion de données ouvertes. Si la majorité des grandes villes occidentales et des démocraties libérales se sont dotées de portails de données ouvertes, plusieurs de ces sites contiennent des jeux de données dont la qualité, la pertinence et le degré d'ouverture laissent parfois à désirer. En effet, il n'est pas rare d'y retrouver des données tantôt obsolètes, tantôt peu documentées, tantôt dans des formats difficilement réutilisables (ex : PDF).

À l'image de l'administration Obama en 2008, le gouvernement du Québec a convié les principaux ministères et organismes provinciaux (ci-après organismes) à libérer en 2012 au moins trois jeux de données afin de les déposer dans le nouveau portail de données ouvertes du Québec⁴. 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. Ici comme ailleurs, la libération des données est un mouvement qui ne semble pas aller de soi pour l'administration publique. Si plusieurs représentants d'organisme, du moins ceux que nous avons rencontrés dans le cadre de cette étude, ont une conception assez juste de ce qu'est une donnée ouverte, la mécanique des données ouvertes et leurs retombées leur apparaissent nébuleuses :

¹ La réutilisation des données ouvertes et leur utilisation consistent en des activités distinctes : la première renvoie à l'exploitation originale des données ouvertes en vue de créer des applications et de services, alors que la deuxième porte sur la visualisation de ces données et leur interrogation. La réutilisation peut conduire au développement d'outils de visualisation et d'interrogation afin de simplifier l'utilisation des données ouvertes.

² La granularité est essentielle à l'exploitation d'un jeu de données. D'ailleurs les agrégations réalisées par les gouvernements dans leurs procédés d'anonymisation des données rendent caduque la réutilisation de plusieurs jeux de données publiés.

³ Définition adaptée de l'Open Data Charter.

⁴ Plus récemment, le gouvernement du Québec lançait, le 5 avril 2016, le portail *Données Québec* qui intègre les données ouvertes du précédent portail gouvernemental et des portails de cinq villes du Québec suivantes : Montréal, Québec, Sherbrooke, Laval et Gatineau. À sa création, ce portail provincial et municipal comptait 770 jeux de données.

Comment s'y prend-on pour libérer des données? Par où commencer? Quelles données doit-on libérer? Les retombées dépassent-elles les coûts qui y sont associés? Le jeu en vaut-il la chandelle? Voilà quelques-unes des interrogations fréquemment soulevées par ces représentants.

Prenant appui sur une vingtaine d'entretiens menés entre mai et octobre 2015 auprès de représentants (ci-après participants)⁵ d'une dizaine d'organismes provinciaux du gouvernement du Québec⁶, la présente recherche examine les données ouvertes sous trois angles. Premièrement, elle aborde les utilités potentielles des données ouvertes, telles que perçues par les participants de l'étude. Deuxièmement, elle s'attarde aux conditions organisationnelles qui tendent à freiner la libération des données au sein de l'administration publique québécoise. Troisièmement, elle propose des pistes pour faciliter la libération de données ouvertes dans l'administration publique.

⁵ Les propos des participants prennent la forme de citation dans le texte.

⁶ Les organismes approchés dans cette étude sont les suivants : Secrétariat du Conseil du trésor, Société de l'assurance automobile du Québec, Ministère de la Santé et Services sociaux, Ministère de l'Économie, de l'Innovation et des Exportations (devenu le Ministère de l'Économie, de la Science et de l'Innovation), Revenu Québec, Ministère de Culture et des Communications, Ministère de la Sécurité publique, SOQUIJ, Institut de la statistique du Québec.

1. UTILITÉS POTENTIELLES LIÉES À LA (RÉ)UTILISATION DES DONNÉES OUVERTES

1.1 DÉVELOPPEMENT D'APPLICATIONS ET DE SERVICES

Bien que les retombées des données ouvertes ne soient pas toujours d'une grande clarté pour les participants, ceux-ci leur reconnaissent une certaine valeur économique, voire commerciale, liée au développement d'applications, en particulier sur les appareils mobiles. Comme le signalent des participants, des données (ou jeux de données) peuvent paraître sans grand intérêt lorsqu'elles sont prises isolément alors que ces mêmes données (ou jeux) peuvent conduire à la production de nouvelles connaissances et de nouveaux produits ou services quand on les combine ensemble de façon originale.

Le réel bénéfice des données ouvertes, c'est de pouvoir les amalgamer ensemble et faire naître de nouvelles connaissances et applications [...] On fait naître quelque chose qui n'existait pas.

Si le potentiel de développement peut varier d'une donnée ouverte à une autre, les données géospatiales (ou géomatiques) semblent, quant à elles, faire consensus auprès des participants. Les données géospatiales permettent de localiser, c'est-à-dire situer géographiquement diverses informations pertinentes au moyen d'une interface cartographique. Pour un répondant, les potentialités des données géospatiales sont de nature transversale, au sens où un positionnement géographique permet de superposer des données de plusieurs domaines ou secteurs d'activités de l'administration publique.

Il peut être tentant de percevoir la géomatique comme une spécialité. Cependant, dans le contexte des données ouvertes, c'est presque une normalité, quel que soit le domaine qui caractérise les données.

Ainsi, les données géospatiales peuvent être d'intérêt pour les individus et les organisations qui se voient offrir la possibilité de superposer différentes couches de données sur une carte selon les besoins du moment.

Il est devenu normal qu'un citoyen ait accès à des données qui lui sont utiles près de chez lui ou peu importe où il se trouve. La localisation des CPE ou des autobus peut l'intéresser [...] tout comme la criminalité dans certains quartiers, à partir des données de la Sûreté du Québec [...] Si vous libérez tout ce qui est routes, adresses, édifices, pistes cyclables, pistes de motoneige, sentiers pédestres, lacs [...] il y aura forcément des gens qui vont créer de belles applications mobiles.

Les données géospatiales numérisées peuvent aussi être utiles pour assurer la sécurité civile à la suite de catastrophes naturelles (ex., inondation du Richelieu) ou de tragédies humaines (ex., Lac-Mégantic). Les autorités publiques y ont de plus en plus recours en situation de crise pour intervenir efficacement sur le terrain. Par exemple, le ministère de la Sécurité publique « fournit aux centres 911 une interface

cartographique qui leur permet d'afficher plusieurs couches d'information [...] dans le but de localiser des gens en difficulté ».

Par le développement d'applications, les données ouvertes peuvent contribuer au développement de l'économie ainsi qu'à l'amélioration de la prestation des services publics, et ce à un faible coût pour l'État, font remarquer plusieurs participants; que l'on pense aux nombreuses applications développées dans les domaines du transport en commun, de la météorologie, de l'agriculture, de l'emploi, de l'économie, de la santé, de la justice et de la sécurité publique ou civile. Pour certains participants, il est clair que le gouvernement ne peut rivaliser avec la créativité et l'agilité de l'entreprise privée et avec sa capacité à prendre des risques dans le développement d'applications.

Quand le gouvernement développe, il y va avec ceinture-bretelles [...] les coûts de développement sont beaucoup plus chers dans un contexte gouvernemental [...] que dans le privé... Si l'entreprise fait une erreur dans l'application [...] personne ne va s'en émouvoir. Par contre, si le ministère en fait une, ça va faire la FrontPage d'un journal [...] Dès que le ministère développe, c'est la grosse machine qui embarque.

Des initiatives gouvernementales, comme la création de portail de données ouvertes et la tenue d'*hackathons*⁷, visent à stimuler le développement d'applications de toutes sortes, aux utilités parfois insoupçonnées, par divers acteurs économiques et autres acteurs (ex : OBNL) de la société, soutiennent des répondants. Par ces initiatives, l'État entend solliciter des expertises créatives, à l'extérieur de ses frontières, pour contribuer à l'amélioration de ses services sans que cela ne lui coûte trop cher.

1.2 TRANSPARENCE DE L'ÉTAT ET SURVEILLANCE PAR LA SOCIÉTÉ

La libération des données est aussi associée à des retombées dites démocratiques. En effet, elle peut constituer un mécanisme de transparence et d'imputabilité gouvernementales en mettant en lumière la performance de l'administration publique et, plus précisément, la qualité de ses services, l'efficacité de ses ressources et l'efficacité de ses programmes. Dans un contexte de restriction budgétaire qui commande une gestion serrée des ressources, les organismes sont condamnés à être de plus en plus transparents. En libérant des données sur leur fonctionnement, leurs programmes et leurs services, les organismes « s'exposent » à une surveillance plus étroite par divers acteurs de la société qui peuvent suivre de près leurs activités et analyser leurs décisions. Les données ouvertes sont d'ailleurs souvent dépeintes comme une façon de diminuer les risques de corruption et de collusion en ce qu'elles permettent d'assurer un meilleur suivi des marchés publics et, plus particulièrement, de la gestion des contrats d'approvisionnement avec les entreprises privées. La libération des données du système électronique d'appel d'offres du gouvernement du Québec (SÉAO)⁸ s'inscrit dans cette stratégie de transparence et de surveillance des marchés publics au Québec, affirment des participants. Avec les

⁷ Événement se déroulant habituellement sur plusieurs jours où des équipes de développeurs et d'utilisateurs conçoivent des applications en exploitant en particulier des jeux de données ouvertes.

⁸ Un organisme est tenu de publier certains renseignements relatifs aux contrats qu'il a accordés et dont le montant est égal ou supérieur à 25 000 \$.

données du SÉAO,

Une entreprise est capable de voir à quel prix les contrats ont été accordés dans sa région [...] Elle peut ensuite ajuster ses stratégies contractuelles en conséquence [...] ou dénoncer une situation anormale [...] surtout quand ce sont toujours les mêmes qui gagnent.

De l'avis de plusieurs participants, les journalistes sont parmi les acteurs les plus susceptibles d'utiliser les données ouvertes afin de suivre les activités de l'État, si l'on en juge par l'intérêt qu'ils portent aux affaires publiques et par le nombre élevé de demandes d'information qu'ils adressent aux organismes. « On reçoit à peu près mille demandes médias par année », affirme un responsable des communications d'un organisme. Quand les données permettent d'identifier un individu « on va leur suggérer de faire une demande d'accès à l'information », ajoute-t-il.

Les chercheurs sont, eux aussi, perçus par les participants comme d'importants utilisateurs potentiels de données ouvertes. Un accès gratuit à des quantités importantes de données brutes ne peut que faciliter la vie des chercheurs dans la réalisation de leurs travaux. Plusieurs participants affirment que leur organisation fait déjà appel à des chercheurs pour étudier des problèmes ou des thèmes qui la préoccupent. Selon l'étendue et la qualité des données ouvertes, les chercheurs peuvent être en mesure d'analyser le fonctionnement des programmes et des services publics et d'en évaluer la performance.

1.3 AVANTAGES POUR LES ORGANISMES

Il peut être exigeant pour les organismes qui détiennent beaucoup de données de répondre aux demandes d'accès à l'information par des journalistes, des chercheurs et autres utilisateurs. Le volume et la complexité des demandes tendent à augmenter avec l'informatisation des dossiers, souligne-t-on. « Certaines demandes peuvent être lourdes, très lourdes à traiter », indique un responsable des données statistiques d'un organisme. Or, les données ouvertes sont vues par les participants comme une façon de diminuer les demandes d'information en permettant aux gens de « faire leurs propres exploitations [...] c'est une professionnelle de moins qui passe du temps [...] à répondre à la demande ». Afin de simplifier l'exploitation des données ouvertes et de favoriser l'autonomie des utilisateurs, des participants suggèrent de mettre en place des outils de visualisation et d'interrogation simples : « Nous pourrions construire un beau petit tableau croisé dynamique [...] l'utilisateur n'aurait qu'à choisir les variables qu'il veut ».

La (ré)utilisation des données libérées n'intéresse pas que des acteurs non gouvernementaux. 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 à tous les ans [...] peut permettre au ministère de la Famille d'identifier des endroits pour installer des CPE.

Les organismes peuvent aussi se montrer intéressés par les données ouvertes produites par d'autres administrations publiques lors de l'élaboration ou révision de leurs propres programmes. Grâce à l'examen de données probantes sur des initiatives gouvernementales dans d'autres pays, il devient possible pour un organisme de prédire les effets sur sa clientèle de certaines modifications à ses programmes sans avoir à les expérimenter. Par exemple,

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.

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 métadonnées⁹. « D'autres provinces pourraient vouloir se comparer avec nous, même d'autres pays ». La comparaison ou l'étalonnage devient une activité d'autant plus accessible que les données sélectionnées respectent des standards internationaux, nous dit-on : « il y a maintenant des organismes internationaux, dont l'Open Government Partnership, qui émettent des standards [...] on est capable de se " benchmarker " ».

En somme, les utilités ou retombées potentielles des données ouvertes et leurs utilisateurs peuvent être nombreux et variés. Compte tenu des avantages économiques, administratifs, démocratiques et scientifiques qu'il procure, le mouvement de libération des données « ne peut aller qu'en s'accroissant », affirme l'un des participants. « C'est le début de quelque chose d'important », signale un autre. Cependant, la libération des données et la transparence gouvernementale qui l'accompagne comportent d'importants défis de gestion pour l'organisation, que les discours gouvernementaux ont tendance à occulter au profit d'une rhétorique principalement orientée sur les retombées. La prochaine section s'attardera à montrer, toujours à partir de l'administration publique québécoise, les freins organisationnels à la libération des données ouvertes.

⁹ Une métadonnée définit les caractéristiques ou attributs d'un ensemble de données. Elle précise, entre autres, le contenu de cet ensemble, son format, sa mise à jour et son ou ses auteur(s) ou contributeur(s).

2. FREINS ORGANISATIONNELS À LA LIBÉRATION DES DONNÉES

2.1 D'AUTRES PRIORITÉS

En dehors de l'année de création du portail en 2012, les efforts déployés par plusieurs organismes dans la libération des données semblent être au ralenti. La majorité des participants signale en effet que les données ouvertes arrivent bien loin dans l'ordre des priorités de leur organisation : « Les deux dernières années et demie [...] la priorité n'était vraiment pas là-dessus », affirme un participant. Deux éléments contextuels tendent à expliquer cette situation. D'une part, un changement de gouvernement, quelques mois après la création du portail de données ouvertes à l'été 2012, a bouleversé les priorités gouvernementales en cette matière.

Quand c'est sorti en juin 2012, la volonté du Gouvernement libéral de l'époque et de Madame Courchesne était très claire [...] Mais, dans les années qui ont suivi [...] on a moins senti cette volonté-là du Gouvernement péquiste.

D'autre part, plusieurs participants soutiennent que leur organisation ne dispose pas des ressources nécessaires pour commencer des projets qui requièrent des nouveaux investissements et dont les retombées ne sont pas immédiates, comme la libération des données, compte tenu du contexte de restriction budgétaire et de la surcharge de travail que cela entraînerait. « On est plus en coupure d'effectifs qu'en expansion [...] on n'a pas les effectifs pour supporter des choses comme ça ». Les priorités de l'organisation sont ailleurs, affirme-t-on, principalement dans le maintien et la consolidation des opérations courantes. Les directions des organismes rencontrés mettent d'ailleurs peu de pression sur les personnes désignées comme responsables des données ouvertes, communément appelées les répondants.

En dehors de l'année de création, on ne m'a pas dit : vas-y à fond [...] J'avais le mandat de collaborer, de publier des données, mais de ne pas en faire une priorité numéro un. On avait bien d'autres chats à fouetter.

Le mouvement de libération des données est le résultat d'initiatives individuelles ou de groupes restreints plutôt que d'une stratégie organisationnelle structurée, soutiennent la majorité des participants. « Il y a une volonté, mais qui est plus individuelle ou par équipes », soutient l'un d'eux; « [o]n est pas mal dans les initiatives et la liberté de professionnels qui s'occupent de ça », ajoute un autre. Bien que la majorité des responsables de données ouvertes rencontrés se soient dotés d'un comité¹⁰ pour les appuyer dans leurs tâches, c'est en bonne partie sur leurs épaules que reposent le projet de libération des données dans leur organisation et la conduite des activités en cette matière. Dans cette mouvance d'initiatives individuelles, un responsable des données ouvertes a dû développer sa propre procédure manuelle et se procurer lui-même sur Internet les outils pour libérer des jeux de

¹⁰ Les comités sont habituellement constitués de représentants de divers secteurs de l'organisation : communication, technologie, lignes d'affaires, accès à l'information, direction, etc.

données géospatiales détenus par son organisation.

C'est pas fait automatiquement [...] la dernière fois, la mise à jour des sept fichiers m'a pris trois jours [...] Il fallait parfois que je passe chez nous parce qu'ici, je n'avais pas Google Earth [...] quand t'es un gouvernement, ça prend une licence [...] Je le faisais chez nous le soir [...] j'ai écrit une procédure qui fait huit pages dans Word.

Des responsables de données ouvertes perçoivent leur travail avant tout comme un exercice d'évangélisation qu'ils mènent le plus souvent seuls, sans grand appui formel de la haute direction. Pour plusieurs d'entre eux, cet exercice de sensibilisation est à recommencer chaque fois qu'ils approchent de nouveaux membres de leur organisation pour les initier aux données ouvertes ou pour les faire participer d'une quelconque façon à la libération de données.

2.2 DES OPÉRATIONS QUI PEUVENT ÊTRE EXIGEANTES

La libération de données peut être une opération exigeante, particulièrement quand il s'agit de sélectionner des jeux de données au potentiel de (ré)utilisation. Cela présuppose d'abord une bonne connaissance de l'information que détient l'organisation. Un inventaire des données est un préalable à leur libération; « c'est la première chose qu'il faut faire avant de déterminer les données à libérer ». Une fois l'inventaire complété, la sélection des données demeure une opération complexe en raison des nombreuses discussions et négociations qu'elles peuvent susciter à l'interne. Pour plusieurs participants, la sélection constitue de loin l'étape la plus « la plus sensible [...] Si tu dépenses 100 000 \$, tu vas en dépenser 90 000 \$ en débats et 10 000 \$ en technologies pour la publier ». Cette étape est d'autant plus sensible que les données à libérer servent à rendre des comptes, affirme un participant qui a participé à la libération de ce type de données.

Parce que c'est de la donnée de reddition de comptes [...] et pas des données statistiques sur le vieillissement de la population [...] c'est fou le temps que ça prend [...] Ça pris de nombreuses discussions au sommet [...] Puis comme c'est de la donnée qui est sujette à interprétation par la suite [...] c'est négocié [...] avec le Contentieux.

La préparation et la mise à jour de certains jeux de données peuvent aussi être une opération d'une grande complexité, entre autres, quand les données ouvertes couvrent l'ensemble des activités gouvernementales, comme les dépenses publiques, et que ces données nécessitent de fréquentes mises à jour.

Le budget de dépenses du gouvernement, c'est cinq livres [...] La mise à jour de ça se fait au quotidien par les décisions du Conseil du Trésor. On ajuste les enveloppes, les budgets, les éléments programmes [...] à coups de 1800 décisions par année ... mettre ça en ligne, c'est des efforts colossaux [...] surtout que les données ne sont pas dans un format facilement exploitable a priori ... on les a en format PDF.

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 support à la clientèle. Par exemple, les utilisateurs peuvent détecter des erreurs dans les jeux de données et les signaler à l'organisme détenteur. « Si le jeu de données [...] a des erreurs – il faut s'attendre à avoir des appels et ça, ça demande du temps ». 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 peut en être de même quand les données sont produites par différentes instances dans diverses circonstances.

*C'est pas tout de mettre les données en ligne, le téléphone sonne après ça!
« Comment ça que il y est écrit ça [...] à l'étude des crédits [...] dans votre rapport annuel de gestion [...] lors de l'interpellation en chambre, ce n'est pas ça que vous avez dit [...] De gros efforts sont requis pour le support postpublication [...] Ça donne lieu, dans certains cas, à des correctifs [...] tu dois parfois remonter jusqu'à la source.*

Enfin, des participants craignent de voir des utilisateurs s'adonner à des interprétations erronées de données ouvertes en raison de compétences méthodologiques limitées, notamment dans l'interprétation des statistiques. Selon eux, une libération des données amènera leur organisation, en particulier leur service de recherche et de production de statistiques, à répondre à diverses interrogations des utilisateurs lorsqu'il y aura des incompréhensions ou des écarts dans l'interprétation des données.

*Si les gens deviennent plus autonomes [...] ils risquent d'arriver à des conclusions différentes des nôtres [...] Ce n'est pas tout le monde qui est à l'aise à faire ça [...] Ça nécessitera de notre part un arbitrage méthodologique et surtout d'interprétation :
« Comment ça se fait que vous, vous sortez tel type de données puis que moi, quand je l'exploite moi-même, je n'arrive pas au même résultat! »*

2.3 DÉSUÉTUDE ET ÉPARILLEMENT TECHNOLOGIQUES

Dans les organismes à l'étude, comme dans l'administration publique québécoise en général, plusieurs infrastructures technologiques sont vieillissantes. Certains systèmes d'information, qui ont plus d'une trentaine d'années d'existence, n'ont pas été modernisés (ou l'ont été très peu) depuis leur création, dont plusieurs systèmes de mission qui gèrent les programmes et services publics des organismes. Outre le fait que « beaucoup de systèmes ont un cycle de désuétude qui arrive à sa fin », ces mêmes systèmes n'ont pas été architecturés au départ pour permettre une extraction facile des données : « Aller chercher la donnée, la préparer pour qu'elle devienne ouverte [...] peut prendre beaucoup de temps ». Qui plus est, le personnel en mesure d'exploiter ces vieux systèmes se fait de plus en plus rare :

Ça prend des gens qui ont des compétences quasi archéologiques [...] Les jeunes qui viendraient travailler là-dedans ... repartiraient au bout d'un mois.

La désuétude des systèmes n'est pas l'unique défi technologique auquel devra faire face le gestionnaire

lors de la libération des données. Celui-ci devra aussi composer avec une diversité de systèmes d'information et des plateformes technologiques, autre particularité technologique qui comporte son lot de difficultés¹¹.

Un de mes jeux de données est sur les technologies ORACLE, l'autre sur des technologies Microsoft [...] Rendre publiques les données me demande alors du temps [...] l'éparpillement technologique, c'est une réalité qui est bien de notre temps.

L'éparpillement technologique est, entre autres, lié au fait que les organismes publics ont eu tendance jusqu'ici à développer des systèmes d'information spécifiques à chacune de leur ligne d'affaires, indique-t-on. Par exemple, un participant fait état de « 22 environnements technologiques différents » dans son organisation, le plus souvent non intégrés, précise-t-il. Cette multiplicité de systèmes cloisonnés rend complexe non seulement le processus d'extraction des données, mais aussi leur inventaire : « Il n'y a pas d'architecte qui a une vision globale de tout ce qui est données au ministère ». La qualité des données s'en trouve aussi affectée, notamment en ce qui a trait à l'unicité de l'information. C'est le cas de certaines données administratives nécessaires à la gestion des programmes et des services, comme l'adresse des entreprises.

On a beaucoup d'adresses pour les entreprises [...] Il y a beaucoup de dédoublement [...] Il faut faire du ménage [...] Nos différents services gèrent parfois les mêmes variables... Ce sont des coûts exorbitants pour jumeler et intégrer tout ça.

2.4 DES DONNÉES DE QUALITÉ INÉGALE

La majorité des participants se sont montrés frileux à libérer des données lorsque celles-ci ne sont pas de grande qualité. Il faut dire que « plusieurs opérations de gestion, comme celles menées dans le cadre de la planification budgétaire ou de la reddition de compte, se font encore dans Excel », soutient un participant. La mise à jour des données est l'un des critères de qualité évoqué qui tend à freiner l'élan pour une libération de données. Ce critère apparaît particulièrement problématique pour les données de transactions à grand volume, compte tenu de la fréquence et de l'étendue des mises à jour. Des retards dans le processus de saisie peuvent donner lieu à des écarts importants entre le moment de la transaction et celui de la libération des données, comme en témoignent les propos suivants :

Au ministère [...] la mise à jour des données est parfois très discutable [...] Prenons l'exemple des aides financières qu'on traite. Il y a des gens chez nous [...] qui n'entrent pas [au fur et mesure] les données dans le système de mission [...] ils vont les rentrer à la fin [...] ces données ne reflètent pas la réalité [...] ça prend une synchronisation entre le transactionnel et la donnée qu'on va libérer.

La mise à jour peut être une opération d'autant plus problématique qu'un même ensemble de données peut

¹¹ Cependant, un participant signale qu'« il existe des technologies qui permettent la lecture des données dans les systèmes (même les vieux) sans avoir à les modifier ».

être alimenté par différents contributeurs extérieurs à l'organisation qui les reçoit. La contribution de plusieurs partenaires peut aussi entraîner des erreurs de saisie et, ainsi, affecter la fiabilité des données. Quand la donnée que détient l'organisme « nous vient de l'extérieur, elle [fiabilité de la donnée] est plus compliquée à vérifier et c'est en bout de piste que vous le découvrez ». Ce n'est le plus souvent qu'une fois l'erreur signalée par les utilisateurs que l'organisme est en mesure d'apporter les corrections requises. Par exemple,

... dans le cas du SÉAO [...] ce sont des gens dans différentes municipalités et ailleurs qui entrent manuellement les données [...] Il n'y a personne qui valide les milliers de données qui sont saisies.

Pour un autre participant, le fait que des représentants de l'État aient à saisir manuellement plusieurs données administratives lors d'incidents est propice aux erreurs :

Le rapport de police d'événements [...] est saisi à la main, à la suite de l'événement [...] dans un formulaire excessivement dense [...] parfois, mes gens y trouvent des incohérences.

Par ailleurs, dans la production de statistiques et d'indicateurs de performance sur leurs services et leurs clientèles, des organismes peuvent utiliser respectivement des méthodologies différentes qui peuvent confondre les utilisateurs qui consultent et confrontent ces différentes sources¹². L'unicité de l'information peut aussi être un enjeu pour une même organisation. À ce propos, un participant fait remarquer que son ministère a utilisé quatre méthodologies différentes pour calculer des délais d'accès à différents services. Enfin, certaines banques de données administratives que gèrent centralement l'État, dont l'effectif de la fonction publique, mériteraient d'être épurées, documentées et normalisées avant d'être libérées, signale un participant.

Si on veut aller plus loin sur une base de données comme celle de l'effectif de la fonction publique, il y aurait un travail d'épuration colossal à faire.

Cependant, il faut reconnaître que la majorité des grands organismes de services (RAMQ, SAAQ, Revenu, RRQ, CSST, etc.) disposent de mécanismes de validation pour assurer la qualité des données qu'ils collectent dans le cadre de transactions menées auprès de leurs clientèles : « Nos données sont filtrées, analysées pour toutes les incohérences, et comparées d'une année à l'autre », précise un représentant d'un de ces organismes.

Par ailleurs, des participants soutiennent que l'État n'a pas nécessairement à attendre que les données d'intérêt public qu'il détient soient d'une qualité quasi parfaite (ex. : exactes, exhaustives et à jour) avant d'être publiées, *a fortiori* quand la libération de celles-ci ne représente aucun risque pour la population.

¹² C'est ce souci de standardisation des données qui a amené à la création de la Banque de données des statistiques officielles (BDSO) sur le Québec et qui est le fruit d'« une collaboration de ministères et organismes partenaires sous la coordination de l'Institut de la statistique du Québec » (site de la BDSO).

Par exemple, si je veux mettre des données sur les pistes cyclables au Québec, pourquoi attendre de les avoir toutes avant de publier [...] Les mettre à 75 %, c'est déjà ça [...] surtout quand [...] il n'y a pas d'effet négatif qu'elle ne soit pas complète à 100 %.

Pour ces participants, le fait de rendre disponibles le plus tôt possible des données peut être une façon d'en améliorer la qualité par le biais des rétroactions de ses utilisateurs. Ainsi, la qualité d'un jeu de données tend à croître avec l'usage. Cette forme d'externalisation ouverte (*crowdsourcing*) sur les utilisateurs a eu des effets bénéfiques sur la qualité d'une importante banque de données gouvernementales :

Le meilleur exemple c'est [...] la banque de données Adresse-Québec [...] On a commencé avec quelque chose complet, disons, à 65 % [...] Les utilisateurs, avec les gens d'Adresse-Québec, ont travaillé pour améliorer leur produit [...] On est rendu avec une banque de données au gouvernement du Québec qui contient 3 400 000 adresses [...] Elles y sont presque toutes [...] Bref, il ne faut pas attendre qu'elle soit exhaustive sinon on ne ferait pas grand-chose.

2.5 SENSIBILITÉ DE L'INFORMATION ET CULTURE DU SECRET

En vertu de différentes lois et règlements, plusieurs données gouvernementales revêtent un caractère confidentiel, en particulier les renseignements personnels, qui convie les organismes à mettre en place des mesures de sécurité qui protègent l'identité de leurs clientèles (ex. : citoyens, de professionnels ou d'entreprises). Dans la libération des données de nature transactionnelle entre l'État et ses clientèles, le droit de savoir se bute rapidement au droit à la vie privée des administrés. Face à ces données, les organismes se voient plus comme des gardiens que comme des diffuseurs d'information¹³.

Je suis plus dans une dynamique de protéger les données [...] On est plus dans le contrôle de l'information [...] La donnée est quelque chose qui doit être protégé à tout prix.

Même lorsque les données confidentielles sont rendues anonymes, des participants demeurent prudents face à leur libération, évoquant les possibilités de ré-identification grâce aux recoupements de données ouvertes de plus en plus nombreuses sur les administrés. La confidentialité peut aussi s'étendre à des collectivités lorsque les données libérées peuvent leur porter préjudice par des effets de stigmatisation sociale, ajoute un participant :

Les données peuvent stigmatiser [...] certaines populations [...] comme une information sur l'indice de défavorisation sociale et matérielle sur un territoire donné.

¹³ Dans la mesure où les données opérationnelles contiennent des renseignements personnels, leur transfert d'une organisation à une autre doit faire l'objet d'entente officielle, généralement approuvée par la Commission d'accès à l'information en vertu de l'article 70 de la Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels (aussi appelée Loi sur l'accès).

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.

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 [...] Je ne vais pas diffuser des données brutes qui auraient pour effet de dévoiler ma stratégie de négociation [...] Je ne vais pas non plus dévoiler mes jeux de données [...] concernant l'application d'une politique budgétaire qui va avoir une incidence sur l'économie.

Par ailleurs, les participants remarquent une frilosité des organisations à diffuser des données sur leur performance. Pour eux, la libération de ce type de données est susceptible d'être soumise à un processus serré de validation et d'autorisation par leurs autorités administratives et politiques¹⁴. L'inclination des organismes à contrôler le message et l'image les incite à la prudence, voire au secret, surtout quand il est question de libérer des données (ex : indicateurs de performance) qui permettent une reddition de comptes sur la place publique.

Tout le monde a sa culture du secret [...] pourquoi sortir des données et aller me flageoler sur la place publique [...] Il y a un réflexe de prudence [...] Je ne veux pas être le dernier de classe.

Pour un participant, la culture du secret et la crainte de la critique n'expliquent pas à elles seules la frilosité des organismes à libérer des données sur la performance organisationnelle.

Au-delà de la culture du secret, la culture de la mesure à des fins d'amélioration par la gestion axée sur les résultats est peu développée.

On invoque alors la difficulté des organismes à intégrer de réelles mesures d'impact dans leur processus de gestion et de révision des programmes et services. « On met beaucoup plus l'accent sur l'éventualité d'avoir à justifier un dépassement de coûts, que d'expliquer pourquoi on n'a pas atteint nos cibles ».

2.6 VENTE DE DONNÉES

Les organismes ont tendance à retenir des données qui mettent en lumière leurs opérations et leur performance, à moins que le gouvernement ne les y oblige, comme dans le cas du Système électronique d'appel d'offres du gouvernement du Québec ou du Tableau de bord sur l'état de santé des projets en ressources informationnelles du gouvernement du Québec. Mais, ce ne sont pas toutes les données gouvernementales qui conviennent aux organismes à la retenue, au contraire. Il suffit d'aller sur les sites des organismes pour constater que plusieurs données et documents d'intérêt public y sont diffusés. Certains

¹⁴ « Quand on publie une donnée, on l'a publiée aux politiciens en premier [...] si un journaliste la sort, au moins on avait informé préalablement le politicien ».

organismes, comme la SAAQ, sont passés maîtres dans la diffusion de statistiques, entre autres, pour sensibiliser la population sur divers problèmes afin d'amener celle-ci à adopter de conduites responsables. Par exemple, « en sécurité routière [...] on a tout intérêt à ce que les gens connaissent et s'approprient le problème ».

Cependant, la diffusion des données peut revêtir une valeur marchande et être vendue par les organismes eux-mêmes, le plus souvent à des utilisateurs externes¹⁵. Par la vente de données, des organismes cherchent à couvrir, en totalité ou en partie, les coûts d'acquisition, de préparation et de diffusion de ces données dans un contexte d'autofinancement. « Les revenus [de nos ventes] ... financent notre fonds d'information ». Par exemple, l'article 12 de la *Loi du ministère des Ressources naturelles et de la Faune*¹⁶ autorise ce ministère à « produire de l'information géographique et 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.

Adresses Québec [...] est vendue au coût du marché... Ça a été demandé en données ouvertes. On comprend que ça peut être un soutien au développement économique du Québec, mais on ne serait plus capable de la produire si on n'était pas autofinancé pour le faire.

Les revenus de la vente peuvent aussi servir à financer des services visant à préparer en amont la donnée (ex : normalisation, épuration, intégration, anonymisation et indexation) ou à y ajouter de la valeur.

Notre banque contient 300 champs indexés [...] On caviarde [...] on agence différents formats [...] Jusqu'à tout récemment, on recevait encore des WordPerfect [...] On a des systèmes en mesure d'extraire certaines métadonnées [...] Les décisions qui ont une valeur sont identifiées puis y a tout un processus éditorial qui se met en branle [...] Il y a des résumés qui sont faits [...] Donc, il y a un gros travail qui est fait là en amont lors de la réception [...] ce n'est pas gratuit d'avoir cette information [...] elle a un coût assez important [...] Il y a 120 personnes qui s'en occupent.

De l'avis de représentants d'organismes qui vendent des données, 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.

Pour nous le principal enjeu [...] C'est la tarification [...] C'est sûr que demain matin, si y avait pas de problématique d'autofinancement, nos données, on les ouvrirait!

¹⁵ Les données vendues par des organismes peuvent toutefois être offertes gratuitement aux autres organismes, comme dans le cas du projet ACRIgéo (Approche de coopération en réseau interministériel pour l'information géographique) dans lequel on retrouve une zone gratuite de partage de données géographiques entre 22 ministères et organismes du gouvernement du Québec. Ces mêmes données sont commercialisées à l'extérieur du gouvernement.

¹⁶ Devenu le ministère de l'Énergie et des Ressources naturelles à la suite du changement de gouvernement en 2014.

3. CONSTATS ET PISTES DE SOLUTION

Au Québec, les organismes gouvernementaux diffusent déjà sur leurs sites et dans divers documents (ex : rapport annuel de gestion, rapport statistiques sur des populations et des services, rapport de recherche, etc.) de nombreuses données d'intérêt pour le public. Or, beaucoup de ces données sont agrégées et contenues à l'intérieur de tableaux au format difficilement réutilisable. L'ouverture de ces données afin de les rendre réutilisables serait un pas dans la bonne direction.

Mais, pour de véritables retombées en matière de transparence, les organismes et le gouvernement devront aller plus loin dans la libération de données mettant en lumière leur mode de fonctionnement, l'efficacité de leurs programmes, l'efficience de leurs services et bien d'autres volets stratégiques en lien avec la performance de l'État. Comme le montre notre étude, la libération des données, en particulier les données opérationnelles ou transactionnelles pouvant servir à une reddition de compte, se bute à des pratiques, à des systèmes, à une culture et à un cadre légal plus enclins à contrôler l'information qu'à la diffuser.

Dans cette dernière section, nous proposons des leviers sur lesquels le gouvernement et ses organismes peuvent intervenir pour encourager et faciliter la libération des données utiles au développement d'applications et de services ainsi qu'à une plus grande transparence gouvernementale.

3.1 UNE GOUVERNANCE CENTRALE AUX ORIENTATIONS CLAIRES

Pour plusieurs participants, en particulier les répondants des données ouvertes dans les organisations à l'étude, le mouvement de libération des données a besoin non seulement d'une impulsion politique de départ pour démarrer, mais aussi d'un encadrement gouvernemental pour l'orienter et le soutenir. Ce mouvement prendra son essor dans l'administration publique quand le gouvernement et les autorités centrales donneront « *des directives plus claires* » aux organismes, nous dit-on, le tout accompagné d'une vision et d'un plan de mise en œuvre. « Tant que ça ne viendra pas d'en haut, je ne bougerai pas », affirme un participant. Sans politique, directive et plan d'action du gouvernement¹⁷, tout laisse croire que les organismes continueront à placer ailleurs leurs priorités, étant déjà très occupés à assurer le maintien de leurs opérations courantes dans un contexte budgétaire qui limite grandement la disponibilité de leurs ressources. « Tant que le politique et le Trésor ne poussent pas plus fort [...] si je n'ai pas de directives [...] les données plus sensibles, on va dire non ». Selon certains participants, les autorités centrales seraient légitimées de désigner les programmes devant faire l'objet d'une libération de données dans un souci de transparence, notamment dans la foulée des travaux de la Commission et du Bureau de la révision permanente des programmes au sein du Secrétariat du Conseil du trésor.

Une gouvernance centrale forte dans la libération des données ne signifie pas une absence de collaboration, au contraire. Celle-ci est essentielle entre les organismes gouvernementaux (ex : Conseil du trésor, ministères et autres organismes provinciaux, municipalités, société d'État) pour produire et

¹⁷ À l'image du Gouvernement du Canada et des principaux pays membres de l'Open Government Partnership.

diffuser les données ouvertes,¹⁸ mais aussi entre l'État et les acteurs de la société (ex : citoyens, OBNL, groupes associatifs, développeurs, entreprises) pour élaborer les stratégies gouvernementales et pour (ré)utiliser les données à des fins de développement et de transparence. À ce propos, signalons que « le Canada a élaboré son dernier plan d'action national en concertation avec les citoyens, des organisations de la société civile et le secteur privé »¹⁹. Le Canada a aussi émis une directive sur le gouvernement ouvert qui, en plus de soutenir une politique d'ouverture par défaut au regard des données opérationnelles, « prévoit des exigences claires et obligatoires qui devront être respectées par les ministères » et qui visent plus particulièrement « tous les nouveaux projets d'acquisition, d'élaboration ou de modernisation des applications, systèmes ou solutions informatiques des ministères, à l'appui des programmes et services »²⁰. De telles directives gouvernementales doivent être accompagnées d'outils opérationnels pour guider les organismes dans la mécanique de la libération.

On nous a dit : « Il faudrait que tu libères des données », mais on ne nous a pas dit comment, comme si, finalement, c'était quelque chose de simple [...] Il y a eu une annonce d'une ministre, mais y a toujours pas de politique, y a rien qui vient encadrer ça.

Afin de guider les organismes dans l'ouverture de leurs données, l'équipe responsable du gouvernement ouvert au Secrétariat du Conseil du trésor du Québec a produit et diffusé à l'attention des organismes un cadre opératoire de libération des données ouvertes à haute valeur.

L'organisme producteur de données ouvertes aura, lui aussi, un important rôle à jouer en matière de gouvernance au sein de son organisation. D'une part, un appui formel et affirmé de la haute direction de l'organisme s'avère essentiel pour mobiliser les ressources nécessaires et les acteurs organisationnels concernés par la libération des données, dont les responsables des lignes d'affaires, des technologies de l'information, de l'accès à l'information, des communications et bien d'autres responsables. D'autre part, l'organisme doit se doter de mécanismes (ex : stratégies, comités, normes, procédures) pour coordonner et structurer efficacement la libération de ses données dans le respect des lois et des règles éthiques, notamment en matière de *protection* des renseignements personnels.

Il faut se structurer davantage puis où il faut impliquer toutes les boîtes concernées [...] les Communications [...] la gestion documentaire, le légal [...] Notre direction n'a pas l'autorité d'ouvrir ces données.

Enfin, 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 ouvertes et adopté par les pays membres de l'*Open Government Partnership*, dont le Canada, la France, le Royaume-Uni et les États-Unis.

¹⁸ En même temps, « c'est un peu difficile pour le Trésor ou pour les organismes centraux de rentrer dans le détail dans les organisations pour dire : vous devriez libérer tel jeu de données, parce qu'il y a tellement de contraintes et de conséquences aussi à faire ça ».

¹⁹ Gouvernement du Canada (2014), Plan d'action pour un gouvernement ouvert 2014-2016.

²⁰ Gouvernement du Canada (2014), Directive sur le gouvernement ouvert, en vigueur depuis le 9 octobre 2014.

3.2 UNE MISE EN COMMUN D'EXPERTISES ET DE RESSOURCES

Les organismes qui produisent des données d'intérêt public sont nombreux. Certains d'entre eux détiennent d'importantes banques de données sur les programmes et les services qu'ils administrent ainsi que sur les clientèles qu'ils desservent. Or, la majorité de ces données administratives, comme bien d'autres données gouvernementales, 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.

Il faut de la cohérence dans ce que tout le monde produit [...] Si tu ne mets pas ta métadonnée, tu peux penser qu'il y a plusieurs chiffres pour le même concept, la même variable [...] les métadonnées, c'est l'intelligibilité de la donnée. Si on ne donne pas ça, l'utilisateur est mal renseigné.

En assignant aux données ouvertes une signification précise dans un format particulier, les métadonnées peuvent contribuer à l'uniformisation de ces données et, du coup, faciliter leur (ré)utilisation, que ce soit pour les analyser ou pour les combiner avec d'autres. À cet égard, il convient de souligner que plusieurs données géospatiales au gouvernement du Québec ont été standardisées et formatées de manière à en faciliter le dépôt et la réutilisation dans le portail de données ouvertes du Québec.

Contrairement aux données géospatiales, les méthodologies, les normes et les outils utilisés pour produire des données transactionnelles ou financières peuvent varier d'un organisme, voire d'un système d'information ou ligne d'affaires, à un autre. Pour plusieurs participants, une standardisation de cette information est nécessaire dans un contexte d'intégration des services et de transparence gouvernementale. Certains précisent que cette standardisation passe par un cadre normatif qui balise les métadonnées et autres éléments méthodologiques dans la gestion de l'information « quelle que soit la source qui va la produire »²¹.

La libération des données peut aussi nécessiter qu'on rende celles-ci anonymes avant de les libérer, surtout quand il est question de transactions entre l'État et ses clientèles, en retranchant toute information permettant de remonter à l'identité de l'individu. L'anonymisation des données peut mobiliser d'importantes ressources, selon le volume et la nature des renseignements personnels à traiter.

On a une équipe pour le caviardage, qui ne sont pas loin de 15-20 personnes [...] c'est 100 000 décisions par année [...] On a aussi des robots qui passent là-dessus sur [...] Il y a un bout automatique... mais [...] ça prend une vérification de l'humain.

Enfin, les données ouvertes disponibles sur les portails gouvernementaux et leurs métadonnées sont

²¹ « Le cadre normatif indique comment faire ton système d'information [...] À la fin, t'as la recette [...] l'objectif étant d'arriver à une information unique ».

avant tout de la matière première pour développeurs ou concepteurs d'applications. Ceux-ci en sont les principaux réutilisateurs. Quant aux citoyens ordinaires et autres utilisateurs potentiels (ex : journalistes, chercheurs, OBNL, entreprises et organismes publics), les données ouvertes sur les portails ressemblent le plus souvent à de longues séries de chiffres inintelligibles. Pour rejoindre ces utilisateurs, les données ouvertes doivent être accompagnées d'outils pour en faciliter l'accès et le traitement. Dit autrement, la démocratisation des données ouvertes passe par le développement d'interfaces de visualisation le plus souvent gratuites et conviviales qui permettent aux utilisateurs de consulter et d'interroger les données ouvertes. Développés par des acteurs du secteur privé, public ou parapublic, plusieurs de ces outils se retrouvent sur les portails gouvernementaux, dont le nouveau portail *Données Québec* créé en avril 2016, sous la rubrique applications²². Grâce à ces outils, il est désormais possible de consulter et d'interroger, par exemple, les contrats octroyés par tous les organismes publics québécois, la liste des entrepreneurs bannis des contrats publics du gouvernement du Québec, les taux d'occupation des urgences au Québec, les condamnations des établissements alimentaires, la liste des prénoms les plus populaires et bien d'autres données. Ces outils de visualisation tendent à augmenter au fur et à mesure que se multiplient les jeux de données ouvertes d'intérêt public.

La libération des données est donc plus qu'une simple publication d'information sur un portail. Elle présuppose d'abord une mise en forme des données (ex : validation de la conformité, anonymisation, indexation et normalisation) pour faciliter et sécuriser leur réutilisation puis une visualisation de celles-ci pour assurer une large utilisation. À cet égard, l'administration publique québécoise peut compter sur des organismes qui ont su développer, avec le temps, des expertises permettant de normaliser et d'indexer les données ouvertes (ex : métadonnée), de les anonymiser, de les intégrer sur un même site ou de les consulter et de les interroger. C'est le cas de l'Institut de la statistique pour les données sur la population québécoise, de la Régie de l'assurance-maladie et du ministère de la Santé et des Services sociaux pour les données en santé, du ministère de la Sécurité publique et le ministère de l'Énergie et des Ressources naturelles pour les données en géomatique et de la SOQUIJ pour l'information juridique. Le gouvernement a donc intérêt à mettre à contribution ces partenaires dans le processus de libération des données.

²² Site consulté en mai 2016, <https://www.donneesquebec.ca/fr/>.

CONCLUSION

Les données ouvertes viennent bousculer des pratiques institutionnalisées en matière de gestion de l'information. Elles se heurtent en effet au fonctionnement bureaucratique de l'État, caractérisé par un contrôle hiérarchique de l'information. Habitué à protéger et à contrôler ses données dans des contextes opérationnels prévisibles, l'État est maintenant convié à libérer une partie de ses données pour permettre des (ré)utilisations qui lui échappe. Pour ce faire, il doit adopter une logique d'action associée au réseau qui, à l'opposé de la hiérarchie, se caractérise par la fluidité des échanges et par la diversité des collaborations et des contributions. Cela implique des changements dans la façon dont l'État gère l'information et développe ses systèmes ainsi que dans son rapport avec les citoyens, les chercheurs, les médias, les développeurs et 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, d'une part, pour faciliter le développement d'applications et de services à valeur ajoutée et à faible coût, d'autre part, pour rendre compte de sa performance et de sa probité.

La présente étude montre que diverses conditions organisationnelles et technologiques tendent à freiner la libération des données ouvertes au gouvernement du Québec. Cependant, ces freins ne doivent pas être vus comme des obstacles insurmontables et irréconciliables avec les principes d'ouverture et de transparence gouvernementale. Paradoxalement, le recours à certains mécanismes hiérarchiques et bureaucratiques peut s'avérer efficace pour diminuer ces freins. En effet, les représentants d'organismes que nous avons rencontrés attendent des autorités gouvernementales des directives claires et des outils pratiques avant d'aller de l'avant dans la libération des données, notamment en matière de transparence. Tout laisse croire que les organismes n'entreprendront pas ce virage d'eux eux-mêmes dans un contexte budgétaire qui les incite à concentrer leurs efforts sur le maintien des opérations courantes. Comme mécanisme bureaucratique, la normalisation des données par le recours à des méthodologies et des métadonnées uniformes se révèle, elle aussi, efficace pour assurer l'unicité de l'information et, du coup, faciliter la (ré)utilisation des données.

Si la libération des données gouvernementale a intérêt à être coordonnée centralement et à être normalisée, une collaboration *volontaire* des organismes ne peut être décrétée par une autorité centrale. Une telle collaboration est plus facile à obtenir quand les organismes tirent des avantages réels de la libération des données à partir d'expériences de (ré)utilisation concluantes. Bien que ces expériences soient encore peu nombreuses et peu documentées, le gouvernement peut compter sur un certain capital de sympathie auprès des organismes. En effet, les participants reconnaissent d'emblée que les données ouvertes peuvent être avantageuses pour des organismes qui souhaitent 1) faire développer des applications et des services à faible coût, 2) se comparer à d'autres organismes ou 3) utiliser de l'information gouvernementale, d'ici et d'ailleurs, afin de les appuyer dans leurs décisions stratégiques sans avoir à passer par des demandes d'accès à l'information (voir section 1.3).

En matière de géopositionnement, plusieurs organismes tirent déjà avantage à partager leurs données grâce à la création d'une plateforme de visualisation commune, appelée Infrastructure Géomatique

Ouverte (IGO)²³. Cette plateforme permet à un organisme de localiser géographiquement ses données ainsi que celles provenant d'autres organismes au moyen d'une interface cartographique ouverte. L'organisme peut ainsi profiter des données d'un réseau de contributeurs « pour une fraction du coût d'acquisition ». Pour l'organisme qui souhaite se familiariser avec la mécanique de la libération des données ou pour d'autres raisons, il lui est possible avec l'IGO d'ouvrir, dans un premier temps, ses données aux organismes gouvernementaux avant de les rendre publiques. « Si un ministère décide de transférer un jeu de données qui est dans une zone intranet vers le public, il n'a qu'à peser sur un piton et il n'a rien à changer ».

Enfin, l'expérimentation d'initiatives porteuses ou concluantes, menées par des acteurs gouvernementaux et non gouvernementaux, rend le phénomène de libération des données concret. La promotion de ces initiatives et leur mise en valeur, conjuguées à un leadership politique et administratif aux orientations claires, peuvent constituer pour l'administration publique un terreau propice à l'ouverture des données dans un souci d'amélioration des services et de transparence gouvernementale.

²³ « IGO a été conçue grâce à un réseau d'échange regroupant des spécialistes en géomatique provenant de plusieurs ministères et organismes de l'Administration publique québécoise. Elle permet de tirer profit d'une multitude de données géographiques (patrimoine culturel, territoires agricoles, écoles, casernes d'incendie, villages relais, etc.) grâce à une interface cartographique accessible par un navigateur Web » (Site de l'IGO, consulté en mai 2016, <http://igouverte.org/>)