

LES RÉUTILISATEURS ET LES RÉUTILISATRICES DES DONNÉES PUBLIQUES OUVERTES

Le cas de Montréal

Jérémy Diaz et Sandra Breux



Institut national
de la recherche
scientifique

LES RÉUTILISATEURS ET LES RÉUTILISATRICES DES DONNÉES PUBLIQUES OUVERTES

Le cas de Montréal

Jérémy Diaz et Sandra Breux

Rapport de recherche remis au laboratoire de l'innovation
urbaine de Montréal

Institut national de la recherche scientifique
Centre Urbanisation Culture Société

Janvier 2021

Responsabilité scientifique : Sandra Breux

sandra.breux@ucs.inrs.ca

Institut national de la recherche scientifique

Centre - Urbanisation Culture Société

Diffusion :

Institut national de la recherche scientifique

Centre - Urbanisation Culture Société

385, rue Sherbrooke Est

Montréal (Québec) H2X 1E3

Téléphone : (514) 499-4000

Télécopieur : (514) 499-4065

<https://inrs.ca/>

ISBN 978-2-89575-414-5

Dépôt légal : - Bibliothèque et Archives nationales du Québec, 2021

TABLE DES MATIÈRES

| | |
|--|-----------|
| SOMMAIRE EXÉCUTIF | 1 |
| INTRODUCTION | 3 |
| 1. DONNÉES PUBLIQUES OUVERTES : COUP D’ŒIL SUR LES PRINCIPAUX ÉCRITS | 5 |
| 1.1 <i>Les politiques de données ouvertes : deux grandes promesses</i> | 5 |
| 1.2 <i>L'évolution des politiques municipales de données ouvertes : entre offre et demande</i> | 7 |
| 1.2.1 Des politiques en faveur de l'offre | 7 |
| 1.2.2 Des politiques en faveur de la demande | 8 |
| 1.3 <i>Les réutilisateurs</i> | 9 |
| 1.3.1 L'utilisateur non-professionnel | 9 |
| 1.3.2 Les entreprises, les entrepreneurs et les développeurs | 10 |
| 1.3.3 Les journalistes de données | 10 |
| 1.3.4 Les chercheurs, les professeurs et les étudiants | 11 |
| 1.3.5 Les groupes communautaires et ONG..... | 11 |
| 1.4 <i>La donnée, un enjeu politique</i> | 11 |
| 1.4.1 Des enjeux politiques | 12 |
| 1.4.2 Vers des politiques de données ouvertes axées sur la valeur ? | 12 |
| 1.4.3 Des connaissances encore lacunaires | 13 |
| 2. DÉMARCHE MÉTHODOLOGIQUE | 14 |
| 3. PRÉSENTATION DES RÉSULTATS | 17 |
| 3.1. <i>Présentation des messages reçus</i> | 17 |
| 3.1.1 Évolution du nombre de messages..... | 17 |
| 3.1.2 Les demandes à l'interne | 17 |
| 3.1.3 Les demandes externes | 18 |
| 3.2 <i>Les types d'utilisateurs</i> | 19 |
| 3.2.1 Les demandes des usagers non-professionnels..... | 19 |
| 3.2.2 Les demandes des entreprises, entrepreneurs et développeurs | 19 |
| 3.2.3 Les demandes du milieu académique | 20 |
| 3.2.4 Les journalistes de données | 20 |
| 3.2.5 Les demandes des organismes (communautaires, ONG) | 20 |
| 3.2.6 Les demandes des fonctionnaires d'autres entités | 20 |
| 3.2.7 Les thématiques et jeux de données récurrents..... | 20 |
| 3.2.8 Récapitulatif de l'analyse des messages..... | 20 |
| 3.3. <i>Présentation des entrevues avec les réutilisateurs du portail des données ouvertes</i> | 21 |
| 3.3.1 Les usagers non-professionnels | 21 |
| 3.3.2 Les entreprises, entrepreneurs et développeurs | 22 |
| 3.3.3 Les chercheurs et les étudiants..... | 24 |
| 3.3.4 Les journalistes de données | 25 |
| 3.3.5 Les autres répondants (ONG et groupes communautaires)..... | 26 |
| 3.3.6 Synthèse de l'analyse des entretiens | 27 |
| 4. TENDANCES ET CONSTATS PRINCIPAUX SUR LES RÉUTILISATEURS ET LEURS DÉMARCHES | 28 |
| 4.1 Une asymétrie parmi les usagers dans la chaîne de traitement de données | 28 |
| 4.2 Les démarches de type exploratoire et de type exploitation..... | 29 |
| 4.3 Les types d'impacts directs et indirects..... | 30 |
| CONCLUSION | 31 |
| <i>Références bibliographiques</i> | 35 |
| ANNEXE 1 | 40 |

Table des tableaux

| | |
|--|----|
| Tableau 1. Synthèse des connaissances manquantes sur les réutilisateurs des données ouvertes des portails municipaux et l'impact de ces utilisations | 13 |
| Pyramide d'impacts des portails de données ouvertes de Pichot-Damon (2019b, en ligne) | 15 |
| Tableau 2. Catégorie et nombre d'utilisateurs ayant accepté de participer à une entrevue | 16 |
| Tableau 3. Nombre de messages reçus par année | 17 |
| Tableau 4. Nombre et catégories d'acteurs qui ont déposé une demande au portail de données ouvertes de 2016 à 2019 | 18 |
| Tableau 5. Nombre d'utilisateurs ayant envoyé au moins un message entre 2016 et 2019 | 19 |
| Tableau 6. Types d'utilisateur-réutilisateur du portail des données ouvertes la Ville de Montréal ayant accepté de participer aux entrevues | 27 |

Sommaire exécutif¹

- L'objectif du présent rapport est de dresser un portrait des réutilisateurs du portail des données publiques ouvertes de la ville de Montréal. Dans ce rapport, les données publiques ouvertes se rapportent aux données provenant d'une institution publique mise à la disposition du public à travers un portail web.
- Les écrits ont permis d'identifier cinq catégories de réutilisateurs de ces données : l'utilisateur non-professionnel, les entreprises, entrepreneurs et développeurs, les journalistes de données, les chercheurs, professeurs et étudiants, les groupes communautaires et les ONG.
- Bien que réagissant à la demande, le portail des données ouvertes de la ville de Montréal se situe majoritairement du côté de l'offre. L'analyse des demandes réalisées au portail entre 2016 et 2019 montre qu'un bassin de 200 personnes gravitent chaque année autour du portail et que les catégories d'utilisateurs sont les mêmes que celles identifiées dans les écrits. Dans ces demandes, l'utilisation prévue des données est rarement expliquée. Le profil des utilisateurs et l'impact ou la valeur créés grâce aux données ne sont pas connus.
- L'analyse de 19 entrevues semi-dirigées avec des volontaires, utilisateurs des données ouvertes, a permis de répondre en partie à ce manque de connaissances. Parmi les personnes interviewées, deux profils se dégagent : les explorateurs – c'est-à-dire ceux qui utilisent les données dans un but personnel pour développer leurs compétences et voir ce qu'il est possible de faire avec ces données – et les exploitateurs – c'est-à-dire ceux qui interagissent avec les données dans le but de créer un produit éventuellement monnayable.
- Les utilisateurs interviewés possèdent tous les compétences nécessaires à la lecture, à la compréhension, à l'analyse et au traitement des données ou sont entourés de personnes à même de les épauler dans ces tâches. Les impacts et la valeur d'usage créés ne sont pas mesurés, ils sont, au mieux, estimés. Les utilisateurs déplorent a) la qualité des données et le long travail de nettoyage et de restructuration qui s'ensuit; b) le manque de reconnaissance et de suivi associés aux activités en lien avec les produits qui découlent de leur exploitation des données; c) l'intérêt variable des données publiées. Les utilisateurs interviewés mettent également de l'avant les avantages du portail des données ouvertes. Ils considèrent que la publication des données ouvertes constitue une valeur ajoutée pour leurs pratiques personnelles et professionnelles (notamment parce que ces données offrent une source d'information publique additionnelle permettant d'enrichir les bases de données existantes, offrent la possibilité d'établir une sériation des données, etc.). Ces données représentent également un gain de temps (accès direct aux données) et un gain monétaire (pas de besoin d'une demande d'accès à l'information, amélioration rapide d'un service, etc.)

¹ Ce rapport se focalise sur l'échelle individuelle. L'écriture épiciène nous semblait ne pas refléter adéquatement cette échelle. De plus, l'écriture inclusive des termes réutilisateur, réutilisatrice; usager, usagère nous semblait apporter une lourdeur à la lecture. Ce sont les raisons pour lesquelles nous avons opté pour l'écriture au masculin. Néanmoins, les termes « réutilisateur », « usager » doivent se lire « réutilisateur, réutilisatrice », « usager, usagère ».

- L'analyse des messages reçus par l'équipe du portail des données ouvertes et les entrevues réalisées avec des usagers volontaires permettent de soulever plusieurs pistes d'action : 1) des pistes d'actions *réalisables* par l'équipe qui sous-tend le portail, 2) des pistes d'actions *souhaitables* qui engendreraient des moyens humains et financiers plus conséquents, et 3) des pistes d'actions *idéales* qui exigeraient une redéfinition du portail des données ouvertes et des acteurs impliqués.
- Les pistes d'actions *réalisables* visent à améliorer le contexte d'utilisation de l'utilisateur, à augmenter le nombre de réutilisateurs potentiels, et à se doter à plus long terme d'une charte de qualité de la donnée. Plus précisément, il s'agirait de disposer de sections pédagogiques sur le portail des données ouvertes, expliquant la définition, le but, la construction des données ouvertes et leurs potentiels tout en informant sur les types de fichiers et formats fournis et leurs objectifs. Une traçabilité de la donnée pourrait également être prévue, en publiant les différentes versions d'un jeu de données et leurs erreurs éventuelles. À plus long terme, il s'agirait de se doter d'une charte de qualité des données qui conviendrait à la réalité des 19 arrondissements de la municipalité.
- Les pistes d'actions *souhaitables* ont pour but de dépasser la vision de l'utilisateur-consommateur de données et à faire du portail des données ouvertes un espace de collaboration. Il s'agirait de faire du portail des données ouvertes un tremplin pour a) créer une interdépendance et un suivi entre usagers; b) créer une plus-value qui découle directement de ces relations; c) indiquer les autres fournisseurs de données qui pourraient être utiles aux usagers; d) offrir sur la plateforme des outils et des technologies dont les usagers pourraient avoir besoin pour utiliser ces données et innover. Ces pistes d'actions nécessitent cependant d'élargir le nombre d'acteurs impliqués dans la gestion du portail, de questionner le rôle et la position de chacun et les moyens associés pour parvenir à cette nouvelle vision.
- Les pistes d'actions *idéales* ont pour objectif de faire du portail des données de la ville de Montréal un élément d'un projet politique plus vaste visant à favoriser la participation d'acteurs pluriels, en s'assurant a) que les usagers s'engagent sur le long terme, b) qu'ils sont représentatifs d'une diversité d'usagers; c) que le portail constitue une plateforme d'apprentissage continue en lien avec la mise à disposition des outils et des technologies nécessaires; e) que la ville dispose du leadership pour s'assurer de l'épanouissement et de la pérennité des interactions créées; f) de créer un espace de production participative des données, réalisées par et pour les citoyens. Cette redéfinition exige une fois encore de réfléchir à l'implication de divers acteurs et à leurs différentes missions.

Introduction

Ce rapport a pour objectif de dresser un portrait des réutilisateurs² du portail des données publiques³ ouvertes de la Ville de Montréal. Il s'agira, d'une part, de revenir sur les principaux écrits portant sur les données ouvertes en général et leur utilisation, et d'autre part, de détailler la nature, le contenu et le portrait des différentes demandes réalisées en lien avec le portail des données ouvertes entre 2016 et 2019. Il sera également question de réaliser des entrevues semi-dirigées avec les réutilisateurs, de creuser le type d'impacts et la valeur d'usage des données ouvertes pour chaque catégorie de réutilisateurs, ainsi que de comprendre la chaîne de traitement des données opérée.

La Ville de Montréal a mis en place son portail de données ouvertes en 2011, suivi en 2012 d'une première politique des données ouvertes (Dickner 2017), révisée en 2015. La politique d'ouverture des données de la Ville de Montréal est une ouverture « par défaut » (Ville de Montréal, 2015, 5) : toute donnée est ouverte d'office, choix récurrent dans de nombreuses administrations municipales à l'échelle internationale. Néanmoins, au Québec, seule la Ville de Montréal met en œuvre une politique d'« ouverture par défaut » assurée par le Laboratoire d'Innovation Urbaine de Montréal (LIUM). Depuis mai 2018, le portail des données ouvertes est sous la responsabilité du LIUM dont l'un des champs d'actions est d'« assurer la transparence de l'appareil gouvernemental, notamment par l'ouverture et l'utilisation, interne et externe, des données de manière humaine » en soutenant « les utilisateurs de données et [en animant] la communauté pour générer le maximum de valeur sociale et économique à partir des données de la Ville » (Site Internet du LIUM, 2020⁴)

En 2020, le nouveau portail des données ouvertes a vu le jour. Parallèlement à cela, une Charte des données numériques⁵ a été publiée. Indépendante de la politique des données ouvertes, celle-ci établit un certain nombre de valeurs et de principes généraux entourant l'utilisation et la diffusion des données numériques. La Charte vise à orienter l'usage éthique des données.

Nous avons procédé en quatre temps. Dans un premier temps, à partir d'une base de données des écrits scientifiques anglophones et francophones, nous avons réalisé un survol de la recherche actuelle existante sur les données ouvertes et leur utilisation et déterminé une démarche méthodologique (parties 1 et 2). Dans un deuxième temps, nous avons brossé un portrait et une analyse des demandes reçues par le laboratoire d'innovation urbaine. Cela nous a permis de réfléchir aux questions qu'il serait pertinent de poser aux réutilisateurs du portail des données ouvertes de la Ville de Montréal, ce que nous avons réalisé à l'aide d'entrevues semi-dirigées

² À la suite d'autres écrits (Dymytrova et Paquienséguy, 2019), nous parlerons dans le cadre de ce rapport de « réutilisateurs » dans la mesure où la donnée existe, est utilisée par l'administration qui l'a créée, et ensuite publiée pour être ré-utilisée.

³ Dans le cadre de ce rapport et comme précisé dans le sommaire, nous parlons de données « publiques » ouvertes, dans le sens où celles-ci émanent d'une institution publique. Il s'agit de distinguer ces données de celles produites par un acteur privé et qui peuvent être « ouvertes ».

⁴ <https://laburbain.montreal.ca/propos>, consulté le 13 décembre 2020.

⁵ Cette charte est accessible à l'adresse suivante :

https://laburbain.montreal.ca/sites/default/files/charte_donnees_numeriques_1_0.pdf

Consultée le 28 octobre 2020.

(parties 3 et 4). Puis, après avoir analysé les réponses recueillies auprès des réutilisateurs des données ouvertes, nous avons identifié les forces et les limites du portail des données ouvertes et de son utilisation actuelle, tout en suggérant plusieurs pistes d'actions (conclusion).

Ce rapport comporte des limites. Premièrement, l'analyse des messages reçus est susceptible de comporter des erreurs, car une même personne peut écrire avec différentes adresses de messagerie. Deuxièmement, les entrevues réalisées ne reflètent la réalité que d'une partie des réutilisateurs du portail, puisque leur participation aux entretiens reposait sur une base volontaire. De plus, ceux-ci sont majoritairement des usagers de l'ancien portail des données ouvertes (qui était en vigueur jusqu'en décembre 2020). Troisièmement, cette analyse ne prend pas en considération les réutilisateurs qui téléchargent les données ouvertes sans interagir avec l'équipe du portail des données ouvertes. Par conséquent, ce rapport doit être davantage vu comme une photographie d'une réalité à un moment donné.

1. Données publiques ouvertes : coup d'œil sur les principaux écrits

Les écrits sur les données ouvertes sont relativement nombreux et s'inscrivent dans l'ensemble des réflexions sur la ville intelligente qui abondent depuis le milieu des années 2000. La définition de ce que peuvent être les données ouvertes demeure relativement consensuelle selon les contextes : « par donnée ouverte, on entend une donnée qui peut être 'librement utilisée, réutilisée et redistribuée par quiconque – sujette seulement, au plus, à une exigence d'attribution et de partage à l'identique' [...] Les données ouvertes peuvent, en somme, décrire toutes les dimensions d'une administration publique – sauf, naturellement, lorsque la sécurité publique, le respect de la vie privée ou les revenus de l'État sont en jeu » (Dickner 2017, 1)⁶. Les analyses sur les données ouvertes peuvent être divisées en trois grandes catégories. Il y a tout d'abord les écrits qui documentent l'émergence des portails des données ouvertes et leurs évolutions. On retrouve ensuite une série de travaux qui examinent la nature, les forces et les limites des portails créés. Enfin, une dernière série d'analyses, somme toute encore assez rares, dresse un portrait des réutilisateurs de ces portails. C'est cette dernière catégorie qui sera ici plus spécifiquement détaillée. Parallèlement aux écrits scientifiques, plusieurs réflexions issues de personnes impliquées dans les démarches d'ouverture de données complètent ces analyses en offrant des regards souvent davantage axés sur la pratique, permettant de pousser plus loin les enjeux autour des politiques de données ouvertes. Dans ce cadre, nous reviendrons tout d'abord sur les grands principes qui guident les politiques de données ouvertes, avant de résumer l'évolution et les différentes formes qu'elles peuvent prendre. Puis, nous détaillerons les types de réutilisateurs identifiés. Enfin, nous terminerons cette réflexion en abordant les dimensions politiques de la donnée.

1.1 Les politiques de données ouvertes : deux grandes promesses

D'un point de vue théorique, les projets d'ouverture des données reposent sur deux grandes promesses : une promesse démocratique d'une part et une promesse économique d'autre part⁷.

⁶ Selon le Secrétariat du Conseil du trésor du Québec, « les données ouvertes sont des données brutes, structurées et diffusées sous format ouvert sous forme de fichiers numériques, auxquelles est associée une licence d'utilisation garantissant leur libre accès et précisant leurs conditions d'utilisation » (Secrétariat du Conseil du trésor 2019). Nous préférons la définition de Dickner qui s'appuie sur l'Open Knowledge Foundation et qui ne met pas l'accent sur les données dites « brutes », élément que l'on retrouve également dans la Charte Internationale des données ouvertes (2018). Dickner souligne par ailleurs que dans le cas de Montréal, le terme données est utilisé « comme un équivalent du terme document » (Dickner, 2017, p.2).

⁷ « Nous n'abordons pas ici les questions relatives à « l'Open source » dont découlent les données ouvertes, ni la conception de l'information en tant que telle, et ni le lien avec la façon de concevoir les réutilisateurs. Un aperçu des débats existants est disponible sur ce site: <https://www.gnu.org/philosophy/open-source-misses-the-point.fr.html>

La promesse démocratique s'articule autour de deux concepts centraux que sont la transparence des pouvoirs publics et la participation citoyenne (Silberzahn et Saujot 2018). Dans les faits, « la transparence est perçue comme moyen de renouveler la légitimité des institutions et de leurs

Les données ouvertes : une certaine vision de la citoyenneté ?

Les politiques des données ouvertes tendent à véhiculer une certaine vision de la citoyenneté arrimée autour de deux pôles : 1) le citoyen doit être actif ; 2) le citoyen doit être innovant: "Citizens no longer want to be passive recipients of legislation that is considered 'inflicted' upon them but rather, seek constructive ways to engage-contribute-use the formation of public policy as a means to enhance their civic responsibilities" (Weerakkody et al. 2017, 285). Cette vision de la citoyenneté est associée à un certain nombre de valeurs exprimées par Cardullo et Kitchin (2019, 11, notre traduction) : « l'autonomie personnelle, le choix du consommateur, l'entrepreneuriat, la participation, la rétroaction, la négociation et la création ».

« bien commun » et d'entrevoir de nouvelles possibilités de gestion et de gouvernance » (Meszaros et al. 2015). Parallèlement à cette promesse démocratique se profile une promesse économique arrimée à deux concepts majeurs : celui de l'innovation et de la création de services et d'entreprises. L'idée sous-jacente est que la libération de données publiques va susciter la création de nouveaux services. Une telle idée n'est pas sans lien avec la notion de rétroaction déjà présente dans la conception de la ville intelligente (Janssen, Charalabidis, et Zuiderwijk 2012, 259). Le principe d' « ouverture par défaut » est le reflet de l'« idéologie de la communication » dont l'idée centrale est que la libre circulation de l'information constitue le fondement de toute transformation de la société. Plusieurs auteurs ont décrit l'origine de cette croyance au sein de la cybernétique et des mouvements des contre-cultures américaines (Turner, 2006) jusqu'à l'alliance récente entre les libéraux et libertaires dont l'un des désirs est d'avoir un accès libre et gratuit à un maximum d'informations dans tous les domaines publics et privés (Barbook et Cameron, 1996 ; Breton et Proulx, 2006), en passant par le projet de société globale de l'information à la fin des années 1990 (Ouellet, 2016). L'ouverture « par défaut » des données ouvertes, couplées aux deux promesses démocratique et économique, s'inscrit dans une vision de la démocratie qui fait de celle-ci un marché, reposant sur trois croyances : 1) la croyance en la distribution optimale de l'information

représentants » (p.4). Il s'agit d'une question d'imputabilité des pouvoirs publics, c'est-à-dire de « rendre des comptes sur le fonctionnement des institutions » (p.4), mais également de regagner la confiance de la population envers les institutions (Janssen, Charalabidis, et Zuiderwijk 2012).

Du côté de la participation, en théorie, l'ouverture des données est associée à l'idée d'un approfondissement de la démocratie, permettant à tout individu de faire entendre sa voix. Il s'agit de créer du

Les données ouvertes : retour à l'utopie libertarienne ?

L'ouverture des données fait écho aux idées déjà présentes dans les premiers balbutiements du cyberspace : 1) contribution, coopération et partage entre réutilisateurs (chaque utilisateur peut contribuer et participer) ; 2) absence de hiérarchie (antiétatisme notamment) ; 3) évacuation du politique (Turner, 2013 ; Flichy, 2001, Musso, 2010 ; Peugeot, 2018).

et la promotion d'une démocratie qui s'apparente à un marché où chacun choisit librement sa consommation d'informations ; 2) la croyance que les problèmes sociaux résultent entre autres choses d'un manque de communication et que ceux-ci pourront être résolus par le recours aux technologies de communication et d'information ; 3) la croyance que chaque individu est une source d'information qui peut ajouter de la valeur aux communautés et à l'économie.

Ces deux promesses sont au cœur des différentes politiques adoptées dans divers contextes, notamment à l'échelle municipale.

1.2 L'évolution des politiques municipales de données ouvertes : entre offre et demande

Un regard sur les différentes politiques municipales adoptées depuis une dizaine d'années met en évidence l'adoption de deux stratégies spécifiques, non exclusives l'une de l'autre : des politiques de données ouvertes axées sur l'offre et la publication massive de données d'une part et des politiques de données ouvertes tournées vers la demande d'autre part.

1.2.1 Des politiques en faveur de l'offre

Dans les années 1980, la donnée constituait un produit de luxe : son coût de production était élevé de même que celui de son stockage et de son analyse. Les différentes avancées techniques dans le domaine de l'informatique et des télécommunications ont permis de multiplier la production des données, tout en réduisant drastiquement les coûts de stockage et d'analyse, amenant certains à parler d'un véritable « déluge de données » (The Economist, 2010).

Cette surabondance de données a contribué à modifier les rapports de force entre les différents acteurs : le contrôle et la maîtrise de la donnée sont devenus un enjeu, notamment pour les administrations publiques. C'est ainsi que sont apparues les premières initiatives de plateformes numériques municipales : les portails de données ouvertes.

Ces premières politiques axées sur l'offre ne sont pas sans comporter un certain nombre de limites. Parmi celles-ci figure la faible qualité de certaines données publiées. Les chercheurs rappellent que le volume de données n'est pas

corrélé à leur qualité : « La multiplication des données ne donne pas forcément lieu à un surcroît de savoir » (Silberzahn et Saujot, 2018, 11; Janssen, Charalabidis, et Zuiderwijk, 2012). Sur les portails municipaux, il n'est pas rare de trouver des jeux de données sur les sites dont la qualité, la pertinence ou le degré d'ouverture laissent à désirer (Kubler et al. 2018). Certains jeux de données sont parfois obsolètes, peu documentés ou dans des formats difficilement réutilisables. Pour l'expliquer, certains pointent le peu d'attention accordée à l'utilisateur (Gascó-Hernández et al. 2017; Dymytrova et Paquienséguy 2017) tandis que d'autres notent et classent la qualité des jeux de données (Tim Berners-Lee, 2009 ; Martin et al., 2013) pour encourager une montée en gamme des

Les données ouvertes : quelques dates clés à l'international

2001 : Open Access Initiative
2002 : Creative Commons
2004 : Open Knowledge Foundation
2007-2009 : Open Government Act
2010 : Open Government Licence
2013 : Charte pour l'ouverture des données publiques
2018 : Refonte de la précédente Charte

D'après Meszaros et al., (2015, 13-14)

catalogues et le recrutement de nouveaux profils professionnels au sein des administrations publiques (architecte réseau, analyste de données, etc.). En effet, les administrations publiques se heurtent à divers obstacles qui freinent la libération des données tels que la production, la mise en forme ou la diffusion (Conradie & Choenni, 2014). Le Québec ne fait pas exception (Boudreau & Caron, 2016; Dickner, 2017). À cela s'ajoute l'idée que les données publiées sont peu sensibles (Silberzahn et Saujot, 2018), ne permettant pas réellement de constituer un contre-pouvoir, mettant à mal l'une des promesses sous-tendues par les projets d'ouverture de données.

Ces premières limites vont remettre en cause un présupposé majeur, à savoir qu'il y aurait forcément des personnes intéressées par les données publiées et des « publics » spécifiques. En d'autres termes, il « suffirait » de publier les données pour créer un intérêt et des bénéfices. C'est une des raisons pour laquelle, certaines politiques se sont davantage tournées vers la demande plutôt que vers l'offre.

1.2.2 Des politiques en faveur de la demande

Depuis le milieu des années 1990, la demande en données est en progression constante (Chignard et Benyayer, 2015). Le secteur privé a développé des modèles d'affaire qui tirent parti de l'exploitation des données ouvertes pour fournir des services à valeur ajoutée. Parallèlement à cela, du côté de la société civile, de nouvelles institutions locales et internationales se sont constituées

Quelle valeur pour la donnée ?

« En fait l'ouverture d'un certain jeu de données spécifique a peu de valeur en elle-même : c'est dans les possibilités de croisement entre différents jeux de données visant la production de nouvelles informations que l'open data puise sa force. En conséquence, l'importance des 'interprètes' (qui trient, analysent, simplifient et créent des visualisations) croît » (Silberzahn et Saujot, 2018, 7).

pour revendiquer plus de transparence des gouvernements, pour lutter contre la corruption, pour favoriser l'engagement citoyen et l'accès à l'information.

Dans ce contexte, un défi apparaît toutefois : celui d'intéresser des publics (Ubaldi 2013; Weerakkody et al. 2017). Pour d'autres, la demande est réelle, mais le mouvement des données ouvertes souffre d'une inadéquation entre l'offre (ou la disponibilité) de données et la demande réelle (et les

utilisations) (Verhulst et Young 2016). L'absence d'intermédiaire et de collaboration autour de ces données l'expliquerait en partie (idem). Selon plusieurs auteurs (Johnson et Robinson, 2014; Turki et al., 2017; Landry, 2019), l'organisation d'événement et de concours (*hackathons*) aurait néanmoins permis, de combler l'écart entre l'offre et la demande en stimulant l'intérêt d'une diversité d'acteurs autour des données ouvertes.

Une des limites les plus souvent identifiées est la faible participation des publics. Pour certains, il s'agit tant d'une question d'information, que d'un manque de capacités des réutilisateurs pour réutiliser et analyser les données (Gascó-Hernández et al. 2017). Pour Silberzahn et Saujot (2018), c'est l'accès qu'il convient de regarder de plus près : « L'accès au sens large implique : de disposer de ressources physiques pour accéder aux contenus (ordinateur, smartphone) ; la capacité de le comprendre et de rendre les contenus utilisables (littéracie des données ou datalalphabétisation), de disposer de ressources institutionnelles ou collectives (pour disséminer le message et engager un maximum de personnes » (p.6). Plusieurs mettent ainsi l'accent sur la nécessité de former les

citoyens aux données (Goëta et Mabi 2014) et à la programmation informatique dont la maîtrise permet de mieux vivre dans le monde numérique (Proulx, Garcia et Heaton, 2014).

Plusieurs enquêtes soulignent toutefois que les publics demeurent mal identifiés. Foulonneau et al. (2014) notent que les données ouvertes qui privilégient la demande sur l'offre risquent de réduire l'ouverture des données à un simple service ôté de toute dimension socio-politique. Inversement, Johnson et Robinson (2014) soutiennent que les données ouvertes fournies dans le cadre des hackathons peuvent promouvoir une forme de sous-traitance de fonctions gouvernementales avec des résultats potentiellement négatifs pour la société. Enfin, les écrits scientifiques soulèvent les risques et les inconvénients de l'ouverture des données tels que la violation de la vie privée, l'interprétation erronée des données, les formes d'exclusion (Gurstein, 2011 ; Zuiderwijk et al., 2014; Peeters et Widlack, 2018; Kobayashi et al., Brandusescu et Nwakanma, 2019) et plus largement le droit à l'information (Johnson, 2014).

1.3 Les réutilisateurs

Quelques analyses ont tenté de brosser un portrait des réutilisateurs réels des données ouvertes. Plutôt que d'utilisateurs, les chercheurs tendent à parler de réutilisateurs : « [...] les réutilisateurs qu'ils soient professionnels ou amateurs, valorisent l'open data via des applications qu'ils créent en transformant les données en services à destination des usagers, des clients, des consommateurs, des citoyens, du territoire » (Dymytra et Paquiénéguy 2019, 2). Ces analyses convergent toutes dans le même sens et indiquent premièrement l'existence de six grands types de réutilisateurs : les individus, les entreprises, le milieu académique, les développeurs, les organismes non gouvernementaux et les journalistes. Néanmoins, comme le soulignent Françoise Paquiénéguy et Valentyna Dymytra : « Notre enquête a révélé une variété de profils des réutilisateurs des données ouvertes qui pose problème car elle regroupe sous un intitulé théoriquement fédérateur, des métiers, des compétences et des positions hiérarchiques différents » (Dymytra et Paquiénéguy 2017, 8). Deuxièmement, un autre élément mentionné par les écrits est l'absence d'« utilisateur moyen », notamment parce que les motivations sont diverses parmi les différents groupes, mais également parfois également au sein même des groupes (Worthy 2015, 794). Une autre précision doit être apportée : certains réutilisateurs identifiés sont aussi catégorisés comme étant des « infomédiaires ». Dit rapidement, “ infomediaries, both human and nonhuman, negotiate the gap between open data providers and end-users, and can take the form of service providers, portals, and platform” (Fast et Rinner, 2017, p.7). En d'autres termes, un infomédiaire rend les données plus ou moins accessibles dépendamment du modèle d'affaires, en transformant les données brutes (Ferro et Orsella, 2013; Janssen et Zuiderwijk, 2014). Selon Johnson et Greene (2017), les journalistes, les groupes communautaires, les ONGs, les entreprises – auxquels s'ajoute le milieu académique – peuvent être considérés comme des infomédiaires. Dymytra et Paquiénéguy (2017) parlent ainsi de « réutilisateurs professionnels » lorsqu'elles parlent des infomédiaires.

1.3.1 L'utilisateur non-professionnel

L'appellation de ce public est délicate : elle désigne les personnes qui font une demande d'ordre personnel, sans lien avec une entreprise ou un organisme. Il est difficile de parler de « citoyen » (puisque cela confère une dimension politique à l'acteur désigné), de « résident ou d'habitant », puisqu'*a priori* le lieu de résidence importe peu. Les écrits tendent à parler de « publics » (Worthy, 2015) ou plus simplement de société civile. Par usager « non-professionnel », nous entendons un

usager qui utilise les données ouvertes à des fins autres que professionnelles. Les écrits soulignent l'extrême diversité de ce groupe. Selon l'enquête menée par Worthy, "The survey and the statistics above point to Open Data users as 'specific' individuals with 'specific interests'. It is likely that they are, as with FOI at local level, a core political activists, and a wider less committed group of the curious, often seeking micro or non-political information" (2015, 795). Peu de travaux ont été réalisés sur ce type d'utilisateurs, si bien que nos connaissances sont encore lacunaires : que fait l'utilisateur non-professionnel des données obtenues ? Quels impacts ont eu ces données ? Par ailleurs, il peut exister plusieurs types d'utilisateurs non-professionnels. On peut penser ainsi à l'utilisateur débutant ou amateur, c'est-à-dire celui qui est peu familiarisé avec les bases de données et outils informatiques, mais également à l'utilisateur expert dont les connaissances dans ces deux domaines peuvent faciliter la compréhension, le traitement voire la diffusion de ces données. Certains utilisateurs experts pourraient être également des infomédiaires. À notre connaissance toutefois, ce type d'utilisateur reste non documenté au sein des écrits. On ignore également le traitement des données qui est réalisé par ce type d'utilisateurs.

1.3.2 Les entreprises, les entrepreneurs et les développeurs

Les entreprises, les entrepreneurs et les développeurs d'applications sont parmi les groupes qui ont un intérêt spécifique pour les données ouvertes, notamment au regard des dépenses, des appels d'offres publiés, ou pour les développeurs des données accessibles (Worthy 2015; Magalhaes et Roseira, 2018). Selon Dymytrova, Larroche et Paquieséguy (2018), ce type d'utilisateur fait le traitement suivant des données : « 1. Récupération des données ; 2. Intégration dans une base de données interne ; 3. Transformation des formats ; 4. Vérification des données ; 5. Intégration dans une application web et/ou mobile » (p.23⁸). Cette catégorie d'utilisateurs peut aussi comprendre des « data scientists » dont Dymytrova, Larroche et Paquieséguy définissent ainsi les fonctions : « Les data scientists mobilisent des modèles statistiques et mathématiques pour produire de l'information avec une visée d'aide à la décision afin de répondre précisément aux demandes de leurs clients, qui sont principalement des grandes entreprises » (2018, 28). Précisons toutefois que de telles personnes peuvent être présentes au sein des municipalités mais également au sein du milieu académique. Selon ces mêmes auteurs, la chaîne de traitement est alors la suivante : « 1. Récupération; 2. Extraction et filtrage des données; 3. Nettoyage et structuration; 4. Intégration au sein d'un data lake⁹; 5. Calculs et analyses; 6. Couche de restitution » (p.28). Une fois encore, bien que la chaîne de traitement des données soit connue, il manque encore des informations sur l'impact de ces données, les publics touchés, etc.

1.3.3 Les journalistes de données

Comme mentionné plus tôt, la connaissance des réutilisateurs de données est encore jeune. Concernant les journalistes, considérés comme des infomédiaires, il semblerait que les données ouvertes soient utilisées par un type spécifique de journalistes, les journalistes de données (Worthy 2015; Dymytrova, 2018). Il s'agit d'une nouvelle forme de journalisme d'investigation qui consiste à analyser de manière systématique d'immenses bases de données pour lui donner du sens (voir par exemple la collaboration de WikiLeaks avec de grands quotidiens internationaux). Or,

⁸ Il est à noter que Dymytrova et al. (2018) ne parlent que des développeurs lorsqu'elles présentent la chaîne de traitement des données.

⁹ Un data lake (lac de données) est une méthode de stockage des données.

Rodrigo Zamith (2019) constate que le journalisme de données s'appuie principalement sur des sources institutionnelles et offre peu de collecte de données originales. La plupart du temps les données ouvertes servent aux journalistes à documenter un propos : « [...] les données publiées en open data permettent un journalisme 'de contexte' comme le projet Les Décodeurs qui vise à 'rétablir les faits' par notamment la cartographie et la visualisation de données. Dans le cas du journalisme de données, l'open data sert peu « de machine à scandale » mais plutôt de point d'appui à l'argumentation » (Goëta et Mabi 2014, 88). Selon Dymytrova, Larroche et Paquienséguy (2018), la chaîne de traitement des données par ce type de réutilisateurs est la suivante : « 1. Récupération; 2. Nettoyage et conversion des formats; 3. Analyses et calculs; 4. Vérification de la donnée; 5. Préparation d'une production finale » (p.35). À notre connaissance toutefois, il n'existe pas de portrait de ces journalistes de données et des types de médias (presse écrite, télévision, internet etc.) qu'ils nourrissent.

1.3.4 Les chercheurs, les professeurs et les étudiants

Les données ouvertes représentent des ressources pédagogiques inexploitées. Toutefois, le milieu académique utilise de manière marginale ces ressources accessibles dans le cadre de l'enseignement ou de la formation : « certains enseignants construisent des séquences pédagogiques mettant en œuvre ce nouvel apprentissage d'usage et de manipulation de données ouvertes » (Capelle, Jutand et Morandi, 2018, 20). De même, on ignore si le milieu académique utilise les données ouvertes dans le cadre de leurs recherches. Pour d'autres, le milieu académique en publiant des rapports, des articles ou autres à partir des données ouvertes, constituent des infomédiaires. La chaîne de traitement des données par ces réutilisateurs n'a pas été documentée à notre connaissance.

1.3.5 Les groupes communautaires et ONG

Les organisations et les militants de la société civile (ONG, OBNL, associations) sont étroitement liées à la production, à la distribution et à l'utilisation de données ouvertes (Church 2017). Premièrement, elles publient régulièrement leurs propres données de recherche (par exemple, l'état de la corruption dans chaque pays). Deuxièmement, elles intègrent des données accessibles au public pour les rendre plus accessibles aux réutilisateurs (par exemple, les ventes d'armes ou la santé environnementale) (Juanals et Pascual-Espuny, 2016). Troisièmement, elles révèlent et rendent publiques des informations d'utilité publique (par exemple, les fichiers bancaires des comptes offshores). En ce sens, elles peuvent également jouer le rôle d'infomédiaires.

1.4 La donnée, un enjeu politique

Ce rapide portrait indique l'existence de nombreuses zones d'ombre, notamment au regard de la valeur de la donnée et de ses dimensions politiques, soit autant de questionnements qui mettent à mal les promesses démocratique et économique qui sous-tendent les projets d'ouverture des données.

1.4.1 Des enjeux politiques

Un invariant réside dans les stratégies des différentes politiques de données ouvertes mises en place : la donnée est un enjeu politique (Lehmans, 2018). Plusieurs éléments sont à considérer. Premièrement, la donnée tend à redistribuer les rôles entre les différents acteurs qu'elle concerne. La donnée interroge également le rôle des administrations publiques, au regard de leur décision ou non de publier certaines données, mais également au regard de leur rôle de régulateur. Deuxièmement, plusieurs constatent l'existence d'une transparence et d'une imputabilité limitées (Goëta et Mabi, 2014). Celles-ci restent souvent liées aux données concernant les dépenses des administrations publiques. De même pour Verdier, « la démarche de transparence enferme les citoyens dans les données qu'on veut bien leur transmettre » (2016, en ligne). Ruijer et al. (2020) vont plus loin en soulignant : “This discussion revolves around the question whether transparency can exist if there is no one to make use of it. Effective transparency requires receptors capable of processing, digesting and using information” (p.14). Troisièmement, il existe une géopolitique et une diplomatie de la donnée : de la bataille en cours concernant l'utilisation des données personnelles aux mouvements internationaux en faveur de l'ouverture des données et le gouvernement ouvert, la donnée est aussi un enjeu de souveraineté. Étant donné que la collecte, le maintien et la diffusion de données ouvertes (parfois sensibles politiquement) dans des formats faciles à comprendre et à utiliser pour le grand public ont un coût, certains auteurs (Hess et Ostrom 2007) se posent la question de l'institution la mieux adaptée et la plus légitime pour gérer ce patrimoine commun. Dans un même ordre d'idées, une troisième vague de politiques de données ouvertes semble s'amorcer.

1.4.2 Vers des politiques de données ouvertes axées sur la valeur ?

Les politiques d'ouverture des données qu'elles soient axées sur l'offre ou sur la demande – voire les deux – comportent des limites. En 2018, la refonte de la Charte internationale des données ouvertes marque un tournant, en se basant notamment sur les limites des politiques des données ouvertes précédentes. La Charte propose de s'orienter vers une stratégie de publication ciblée pour « résoudre des problèmes politiques spécifiques », « hiérarchiser les actions et « intégrer le changement institutionnel ». Dans leur ouvrage, Chignard et Benyayer (2015) proposent une approche de la valeur de la donnée pour comprendre les opportunités et les risques liés au nouveau paysage des données. Ils rappellent que les trois fondements de la valeur de la donnée sont : le travail (la donnée vaut ce qu'elle coûte à produire et à obtenir), la rareté (de l'accès à la source des données), l'utilité (la donnée vaut ce que l'on peut en faire). Ce triptyque fait écho au constat de l'OCDE (2019, voir encadré). Ces mêmes auteurs précisent que les données ont également une valeur d'option, « c'est-à-dire une valeur attribuée à un usage potentiel futur » (Chignard et Benyayer, 2015, 45). Autrement dit, les utilisations directes ou indirectes des données ouvertes n'apparaissent pas immédiatement : soit parce que des seuils quantitatifs et qualitatifs doivent être atteints pour des usages émergents, soit parce qu'une collaboration entre des acteurs de secteurs différents et très éloignés est parfois requise pour être en mesure d'identifier et d'en extraire de la valeur.

1.4.3 Des connaissances encore lacunaires

Ces éléments appellent à une meilleure connaissance des réutilisateurs existants, leurs usages et leurs pratiques professionnelles en lien avec les données ouvertes, mais également une exploration des réutilisateurs potentiels et indirects de ces ressources disponibles. Comme le souligne Pichot Damon (2019a), la seule connaissance des jeux de données téléchargées ne peut suffire à dresser un portrait précis des réutilisateurs des données ouvertes.

Pour comprendre notamment la valeur

des données utilisées, il convient de saisir plus finement leurs impacts (Pichot Damon, 2019b). Le tableau ci-dessous fait la synthèse des connaissances manquantes actuellement.

Créer de la valeur à partir des données ouvertes

« Créer de la valeur à partir des données publiques ouvertes (DPO) implique d'augmenter la quantité de données et d'apprendre à mieux repérer les données de grande valeur, afin d'accroître la réutilisation. Cela nécessite également de comprendre les obstacles en présence, afin de mettre les données publiques au service de la chaîne de valeur du service public » (OCDE, 2019, 16).

TABLEAU 1. SYNTHÈSE DES CONNAISSANCES MANQUANTES SUR LES RÉUTILISATEURS DES DONNÉES OUVERTES DES PORTAILS MUNICIPAUX ET L'IMPACT DE CES UTILISATIONS

| | QUI ? | | | | |
|----------------------------------|--|---|---|---|---|
| | Usager non-professionnel | Entreprises, entrepreneurs Développeurs | Journalistes | Milieu académique | Groupes communautaires et ONG |
| Type de demande | Information non politique/information spécifique | Base de données spécifiques |
| Chaîne de traitement des données | Non-documenté dans les écrits | Documenté | Documenté | Peu documenté dans les écrits | Peu documenté dans les écrits |
| Type d'utilisations | Non-documenté dans les écrits | Application par exemple | Étayer un contexte par exemple | Documenter une réalité par exemple | Peu documenté dans les écrits |
| Type d'impacts | Non-documenté dans les écrits | Utilisation par un public ? Partiellement documenté dans les écrits | Utilisation par un public ? Partiellement documenté dans les écrits | Utilisation par un public ? Partiellement documenté dans les écrits | Utilisation par un public ? Partiellement documenté dans les écrits |
| Valeur d'usage créée | Non-documenté dans les écrits | Non-documenté dans les écrits | Non-documenté dans les écrits | Non-documenté dans les écrits | Non-documenté dans les écrits |

2. Démarche méthodologique

Notre démarche méthodologique s'est déroulée en deux temps. Dans un premier temps, nous avons procédé à l'analyse globale des messages reçus par le Laboratoire d'Innovation urbaine de Montréal de 2016 à 2019, soit 4 980 messages¹⁰. L'objectif n'était pas de réaliser une analyse fine et détaillée de chaque message, mais bien plus de saisir le type d'utilisateur et le type de demande. Il convient par ailleurs de rappeler que cette analyse sommaire avait pour but principal de mieux cibler les questions pertinentes à poser lors de la phase suivante, c'est-à-dire, la réalisation d'entrevues semi-dirigées avec des usagers du portail des données ouvertes de la ville.

Certains courriels ont été supprimés : les spams avérés, les courriels d'absence, les courriels qui indiquaient juste un mot de remerciement, les prises et reports de réunions, les courriels automatiques concernant la fréquentation du site, les mises à jour, etc. Les courriels automatiques sont fort nombreux, ce qui a contribué à une réduction subséquente du nombre de courriels à analyser. De même, comme les tickets comprennent également les messages envoyés sur twitter ou via l'adresse générique des données ouvertes, cela a coloré l'analyse : les messages sur twitter étant souvent de l'ordre de l'information ou de l'invitation et ils n'ont pas fait l'objet d'analyse. Dans un même ordre d'idées, nous n'avons pas retenu les courriels d'inscription à l'infolettre. Nous estimons à environ 2000 le nombre de messages non retenus pour l'analyse.

Les messages reçus étaient pour la plupart d'entre eux déjà pré-classés en plusieurs catégories : question interne, question externe, question pour une autre équipe, demande d'ouverture. Nous avons donc dans un premier temps rassemblé les messages appartenant à la même catégorie¹¹.

Puis, afin de dresser le portrait des réutilisateurs, des catégories d'analyse préliminaires ont été élaborées à partir de la revue des écrits réalisée :

- 1) Type d'utilisateur (usagers non-professionnels, journalistes, milieu académique, milieu communautaire, milieu de l'entreprise)
- 2) Objet de la demande (ouverture de données, plainte, etc.)
- 3) Type de demande (selon la classification offerte par le formulaire)
- 4) Nature de la demande, en référence aux jeux de données ouverts.

Ces catégories ne sont pas statiques. La lecture des messages a permis leur précision et ajustement. Nous avons d'abord traité les données par année. Ce traitement a permis de retracer en quelque sorte l'histoire de l'intérieur du portail des données ouvertes. Il convient de mentionner qu'une même personne peut écrire plusieurs messages : notre décompte vise donc les acteurs plus que le nombre de messages. Par ailleurs, nous ne pouvons avancer que des tendances : en effet, certains courriels font l'objet de transfert auprès de plusieurs personnes et parfois le courriel original n'est pas transmis, si bien que nous ne pouvons garantir que des courriels n'aient pas été

¹⁰ Parmi ces courriels, les 33 courriels reçus en 2020 n'ont pas été analysés.

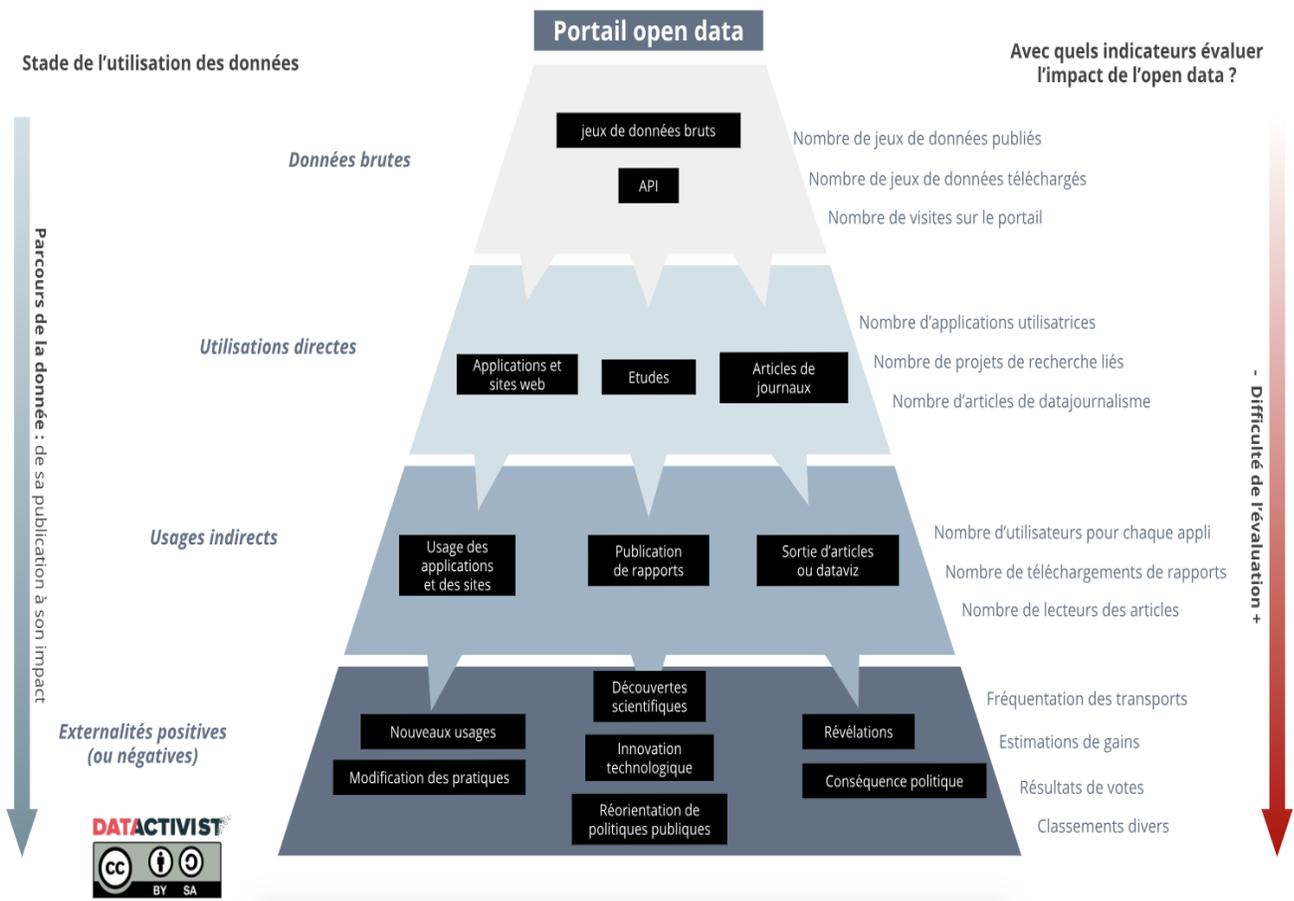
¹¹ En de rares occasions, nous avons déplacé des messages dans certaines catégories, considérant que le classement initial n'était pas adéquat.

comptabilisés deux fois. D'autres messages, faisant allusion à des rencontres en personne, étaient parfois sibyllins.

Cette première analyse nous a permis d'esquisser une grille d'entrevue à destination des personnes composant la liste de diffusion du Portail de données ouvertes. Cette grille d'entrevue s'inspire des travaux de Paquienseguy et al. (2016) et de la pyramide d'impact de Pichot-Damon (2019b). Elle avait pour but principal de retracer le parcours de l'utilisateur, la valeur et l'utilité des données ouvertes (voir la grille en Annexe). L'appel à participer aux entrevues a été diffusé au sein de la liste de diffusion du Portail, qui comporte quelques 300 abonnés et lors d'un séminaire présentant le nouveau site web des données ouvertes de Montréal qui s'est déroulé le 17 juin 2020 et où une soixantaine de personnes étaient présentes. La présente recherche a été présentée lors de ce séminaire.

Les personnes intéressées remplissaient un formulaire en ligne distribué via le Portail de données ouvertes. Ce type de recrutement ne peut ainsi nullement être considéré comme étant représentatif : les personnes visées sont des personnes a priori déjà intéressées par le portail des données ouvertes puisqu'inscrites volontairement sur la liste de diffusion du portail. Il ne s'agit donc là que d'une incursion ponctuelle auprès de réutilisateurs potentiels de ce portail.

PYRAMIDE D'IMPACTS DES PORTAILS DE DONNÉES OUVERTES DE PICHOT-DAMON (2019B, EN LIGNE)



Dix-neuf personnes ont répondu à cet appel et ont été interviewées entre le mois de juin et le mois d'août 2020. En raison du contexte de la pandémie de la COVID-19, ces entrevues ont été réalisées à distance via un outil de réunion à distance. Elles ont duré entre 30 minutes et 1 h30, pour une moyenne de 45 minutes. Ces personnes se répartissent selon différentes catégories (voir tableau 2).

TABLEAU 2. CATÉGORIE ET NOMBRE D'USAGERS AYANT ACCEPTÉ DE PARTICIPER À UNE ENTREVUE

| Catégorie d'usagers | Nombre de personnes ayant accepté de réaliser une entrevue |
|---|---|
| Usagers non-professionnels | 6 |
| Milieu académique | 4 |
| Entrepreneurs | 5 |
| Fonctionnaires et organismes communautaires | 2 |
| Journalistes | 2 |
| Total | 19 |

La partie suivante présentera l'analyse des demandes reçues par le portail des données ouvertes, puis présentera l'analyse des entretiens.

3. Présentation des résultats

3.1. Présentation des messages reçus

Le propos qui suit présente les grandes tendances de cette analyse. Nous débuterons par l'évolution du nombre de messages et les types de demandes en fonction des usagers.

3.1.1 Évolution du nombre de messages

Le nombre de messages reçus est relativement stable d'année en année avec une légère diminution en 2017 et une légère augmentation en 2019. Ces augmentations et diminutions ne signifient rien en tant que telles, car elles ne reflètent pas le nombre de messages réellement analysables : à titre d'exemple l'expéditeur Script_ckan_upload a généré 167 messages automatiques en 2016. Ces messages automatiques ne sont pas analysables. En 2017, ce même expéditeur a généré 210 messages et est utilisé à l'interne pour échanger des informations. En 2018 et en 2019, cet expéditeur disparaît. De même, l'expéditeur Vue sur les contrats génère 39 messages en 2016, 54 messages en 2017, 697 messages en 2018 et 854 en 2019. En somme, le nombre de courriels reçus n'est pas réellement significatif en raison du nombre des courriels automatiques présents. D'une année sur l'autre, il y a environ entre 700 et 800 messages qui peuvent faire l'objet d'une analyse.

TABLEAU 3. NOMBRE DE MESSAGES REÇUS PAR ANNÉE

| | 2016 | 2017 | 2018 | 2019 |
|----------------|------|------|------|------|
| Messages reçus | 1240 | 985 | 1216 | 1539 |

3.1.2 Les demandes à l'interne

Les messages échangés à l'interne s'arriment autour de plusieurs thématiques, qui toutes témoignent de l'histoire du portail. Ainsi en 2016, parmi les éléments abordés on retrouve :

- a) la mise en place de coordonnateur de données ouvertes au sein des arrondissements
- b) la création d'inventaires de données dans un ensemble de secteurs
- c) la question du format
- d) la question de la fiabilité, de la possibilité d'ouvrir des données
- e) l'ajustement avec d'autres données publiées par d'autres organismes (CMM, Données Québec, etc)

Ces différents éléments témoignent de la construction du portail. Puis les années suivantes, on constate que les informations échangées à l'interne visent des demandes de précision de données, des contacts en vue de l'ouverture de certains jeux de données, des mises à jour de données et des suivis plus globaux. Des demandes externes font aussi l'objet d'échanges à l'interne. De façon générale, le nombre de courriels échangés à l'interne a diminué entre 2016 et 2019.

3.1.3 Les demandes externes

Les demandes extérieures à la ville sont de plusieurs natures. On note plusieurs types d'acteurs :

- 1) Les usagers non-professionnels
- 2) Les développeurs, les entrepreneurs et les entreprises
- 3) Le milieu académique (chercheurs, étudiants etc.)
- 4) Les journalistes de données
- 5) Les groupes communautaires et ONG
- 6) Les fonctionnaires d'autres organismes : Données Québec, Ville d'Ottawa, etc.

Il est à noter qu'il est parfois difficile d'identifier avec certitude la provenance d'une demande. Nous avons croisé l'objet du courriel avec la nature de l'adresse courriel utilisée, mais la fiabilité de cette analyse ne peut être assurée : certaines personnes écrivent avec plusieurs adresses différentes, d'autres utilisent leurs adresses professionnelles pour réaliser une demande personnelle.

TABLEAU 4. NOMBRE ET CATÉGORIES D'ACTEURS QUI ONT DÉPOSÉ UNE DEMANDE AU PORTAIL DE DONNÉES OUVERTES DE 2016 À 2019

| | 2016 | 2017 | 2018 | 2019 |
|---|------|------|------|------|
| Usagers non-professionnels | 95 | 70 | 70 | 82 |
| Développeurs/Entrepreneurs /entreprises | 10 | 64 | 65 | 48 |
| Milieu académique | 21 | 39 | 38 | 37 |
| Journalistes de données | 8 | 5 | 11 | 11 |
| Groupes communautaires et ONG | 9 | 9 | 9 | 4 |
| Non-identifiés | 19 | 92 | 49 | 0 |
| Totaux | 162 | 279 | 242 | 182 |

Ce tableau est susceptible de contenir des erreurs, car une même personne peut écrire sous différentes adresses/chapeaux. Par exemple : Sophie écrit un premier message avec son adresse Gmail, elle signe de son nom Sophie X. Elle écrit par la suite un courriel de remerciement à partir d'une adresse professionnelle d'un magazine. Elle signe de son nom Sophie X. Puis, nous la retrouvons avec une autre adresse courriel professionnelle associée à un journal de presse quotidien reconnu. Dans ce cas-ci, il nous est extrêmement difficile de savoir si la première demande de Sophie était une demande d'ordre personnel ou en lien avec son emploi ; de même il est difficile de savoir s'il y avait un lien entre les deux autres messages et la fonction de la personne. Enfin, comment s'assurer que Sophie X ne désigne qu'une seule personne ?

Également, le tableau 3 reflète les messages reçus année par année, et donc ne précise pas les personnes qui écrivent chaque année. Nous avons donc par la suite établi le nombre d'acteurs ayant envoyé au moins un courriel entre 2016 et 2019 (Tableau 4). Cela permet de voir que 834 acteurs gravitent autour du portail des données ouvertes durant cette période avec une moyenne de 200 usagers par année.

TABLEAU 5. NOMBRE D'USAGERS AYANT ENVOYÉ AU MOINS UN MESSAGE ENTRE 2016 ET 2019

| | |
|---|-------|
| Nombre d'usagers ayant envoyé au moins un message | 1 109 |
| Nombre d'usagers ayant envoyé au moins un message (sans les fonctionnaires) | 834 |

Les demandes, quels que soient les acteurs, sont classables en différentes catégories : 1) les demandes de données ou d'ouverture ; 2) les précisions sur des données spécifiques ; 3) des demandes d'information ; 4) des erreurs dans un jeu de données ; 5) des prises de contact ; 6) les demandes de produits personnalisés ; 7) les demandes qui n'ont rien à voir avec les données ouvertes ; 8) les plaintes (lien brisé, dossier qui ne se télécharge pas). Ces demandes nous ont permis d'identifier plusieurs types d'usagers.

3.2 Les types d'usagers

3.2.1 Les demandes des usagers non-professionnels

Les demandes des usagers non-professionnels sont de plusieurs natures. Une partie de ces demandes n'a pas de lien avec le portail des données ouvertes, mais demeurent somme toute très minoritaire. Le portail est alors utilisé comme une sorte de guichet unique. De telles demandes sont en lien avec des problèmes de la vie quotidienne (camion dans la rue, durée du feu de signalisation trop courte, etc.). Les autres demandes sont de nature extrêmement précise (par exemple, une demande d'archive, une demande d'une photo aérienne de sa maison, etc.).

Souvent, en 2016, les demandes questionnent l'existence de données plutôt que de demander leurs ouvertures. Les années subséquentes, on parle plus de demandes d'ouverture. Il existe également quelques demandes en lien avec le format, un lien qui ne fonctionne pas ou une plainte, mais cela reste rare. Parmi les plaintes des citoyens, quelques demandes s'interrogent sur la fiabilité des données et le niveau de transparence des informations transmises.

Si la lecture des courriels envoyés par les usagers non-professionnels montre souvent que ce sont des usagers amateurs, parfois les demandes ou la fréquence des courriels envoyés témoignent de connaissances spécifiques chez ces usagers, invitant ainsi à parler davantage d'usagers experts. Un usager a ainsi pu être identifié clairement en tant qu'expert et son action a même été rendue publique dans les médias. Cette personne travaille dans le domaine de l'informatique.

3.2.2 Les demandes des entreprises, entrepreneurs et développeurs

Pour ce type de publics, une partie des demandes est liée à des demandes d'API, principalement pour les opérations de déneigement¹². L'objectif de l'obtention d'un jeton d'accès à l'API n'est pas spécifié. D'autres concernent des données de base (cartographie, nombre de véhicules, etc.) ou des précisions sur l'exploitation commerciale des photographies. Plusieurs sollicitations d'entrepreneurs et d'entreprises concernent des activités de démarchage. Ce sont rarement les mêmes entrepreneurs ou développeurs qui écrivent d'une année sur l'autre.

¹² Les données ouvertes pour les opérations de déneigement sont diffusées via un API et il faut obtenir un jeton du Service des technologies de l'information pour y accéder. Ceci permet une interaction par courriel avec les utilisateurs quand il y a un changement dans les attributs de données de l'API.

3.2.3 Les demandes du milieu académique

Les demandes du milieu académique sont souvent le fruit de personnes qui se présentent comme des étudiants, bien que certaines demandes émanent d'enseignants ou de bibliothèques universitaires. Elles visent des données précises (par exemple en lien avec les données électorales) ou l'accessibilité de certaines données. Il y a également des demandes d'entrevues, en lien le plus souvent avec des comparaisons de portails de données ouvertes à l'international.

D'autres demandes requièrent des précisions sur les conditions de création des données. Quelques demandes récentes du milieu académique révèlent l'utilisation des données ouvertes comme ressources pédagogiques dans le cadre de cours universitaires. Tout comme pour les entreprises, ce sont rarement les mêmes personnes qui écrivent d'une année sur l'autre.

3.2.4 Les journalistes de données

Ce sont souvent les mêmes journalistes qui écrivent d'une année sur l'autre ou du moins les mêmes médias. Leurs questions visent des demandes de données ou des demandes de précision vis-à-vis de certains jeux de données.

3.2.5 Les demandes des organismes (communautaires, ONG)

À l'instar des demandes des entreprises, celles des organismes associatifs et communautaires visent des demandes extrêmement précises. Nous avons également identifié la prise de contact de Wikidata avec le LIUM.

3.2.6 Les demandes des fonctionnaires d'autres entités

Les demandes des fonctionnaires d'autres entités sont des échanges concernant des jeux de données (avec le Ministère des Transports par exemple, la CMM etc.). Il est alors souvent question d'ajustement de données ou de poursuivre des échanges ayant eu lieu de vive voix. D'autres demandes émanent de villes ou autres institutions et souhaitent en savoir plus sur l'expérience montréalaise ou requièrent de l'aide.

Les messages des fonctionnaires externes à la Ville de Montréal ont été comptabilisés au sein des messages internes, car entre 2016 et 2020, ce sont les mêmes organismes qui reviennent et la lecture de ces messages laissent penser qu'il existe des contacts plus réguliers entre la Ville et ces entités.

3.2.7 Les thématiques et jeux de données récurrents

Ce sont souvent les mêmes jeux de données qui font l'objet de discussions : ceux ayant trait à l'environnement, au transport, aux infrastructures, au gouvernement et aux finances. Les thématiques abordées sont souvent similaires et en lien avec la publication des jeux de données (remorquage, parcs, borne-fontaine, stationnement, rôle évaluation, inspection sanitaire, comptage véhicule, piéton, cycliste, patinoires, arbres, etc.).

3.2.8 Récapitulatif de l'analyse des messages

Cette analyse sommaire permet de montrer :

- 1) Qu'il existe un bassin de 200 personnes environ qui chaque année prend contact avec le portail des données ouvertes pour effectuer une demande spécifique.
- 2) Que la lecture des messages ne nous a pas permis de trouver d'autres types de réutilisateurs que ceux recensés dans les écrits. Néanmoins, s'intéresser aux fonctionnaires de la municipalité – autres que ceux impliqués dans l'alimentation du portail – pourrait être intéressant.
- 3) Les demandes des usagers non-professionnels, des entreprises/développeurs, des organismes, des journalistes, reflètent ce que les écrits ont montré dans d'autres contextes.
- 4) Que l'utilisation immédiate ou future des données ouvertes téléchargées est rarement explicitée (particulièrement le cas pour les entreprises/développeurs).
- 5) On ignore la localisation géographique des personnes qui écrivent.

Ces éléments, en lien avec le tableau 1 qui clôturait la revue des écrits nous ont permis d'écrire une grille d'entrevue à destination des réutilisateurs (voir Annexe 1).

3.3. Présentation des entrevues avec les réutilisateurs du portail des données ouvertes

Au fil des entrevues, plusieurs points communs sont apparus dans le propos des réutilisateurs. Nous détaillerons les réponses obtenues en présentant (1) les répondants (genre, profil et compétences techniques, lieu de travail ; 2) leurs attentes vis-à-vis du portail ; 3) le type de demandes réalisées ; 4) les points positifs et négatifs exposés en lien avec les thématiques abordées dans notre grille d'entrevue.

3.3.1 Les usagers non-professionnels

Les usagers non-professionnels qui ont répondu à notre appel sont au nombre de six. Leur profil présente plusieurs similitudes. Ce sont des hommes, qui en raison de leur profession ou de leur scolarité, savent lire et analyser des données. À titre d'exemple, on retrouve ainsi un agent de recherche à la retraite, un développeur web, un analyste de données, etc. L'origine géographique de nos répondants est majoritairement montréalaise, un seul répondant résidant à Québec.

Leur prise de contact avec les données ouvertes s'est réalisée pour certains de façon aléatoire, tandis que d'autres attendaient celles-ci avec impatience. Généralement, ces personnes ne consultent pas d'autres sites de données ouvertes. Le type de demandes qu'ils font varie et amène à distinguer deux profils d'usagers non-professionnels qui peuvent être vus comme les deux pôles d'un même continuum, le passionné d'une part et l'aspirant professionnel d'autre part.

Pour le passionné, les motivations pour utiliser les données ouvertes sont personnelles et témoignent à des degrés divers d'un certain engagement politique. Ainsi, une personne souhaitait rendre disponibles le nombre et la qualité des ruelles vertes à Montréal, une autre voulait documenter le nombre de rongeurs nuisibles dans un arrondissement, etc. Ces personnes souhaitent par ce travail d'exploration des données redonner à la communauté ce qu'elles contribuent à créer grâce aux données ouvertes. L'aspirant professionnel est, de son côté, davantage un exploitateur de données qu'un explorateur¹³. Il travaille avec les données ouvertes pour parfaire ses compétences

¹³ March (1991).

(formation par ailleurs largement autodidacte) dans le but notamment d'augmenter son employabilité. Pour plusieurs de ces usagers non-professionnels, la frontière reste poreuse entre les motivations personnelles, voire politiques, et le monde professionnel.

Ces usagers entrent en contact avec l'équipe du portail des données ouvertes lorsqu'ils débutent leurs projets : ils ont alors des demandes très spécifiques. Ils sont satisfaits des réponses obtenues. Le parcours d'utilisation des données par ces personnes est le même. Il commence par un téléchargement, une visualisation, un nettoyage et une restructuration. Le degré de difficulté que rencontrent les usagers interrogés n'est pas le même selon l'interlocuteur. Pour certains, ce qui prend le plus du temps c'est la phase de nettoyage, pour d'autres c'est la phase de restructuration.

Trois points négatifs ont été mentionnés. D'une part, la phase de nettoyage a souvent été décrite comme étant longue et fastidieuse. Aucun format de fichier n'apparaît comme étant idéal, tout dépendant à chaque fois de la nature du projet. Certains usagers ont créé des applications, une visualisation d'un phénomène ou un outil. Néanmoins, un constat général ressort : les impacts de ces créations ne sont pas mesurés, seulement - dans le meilleur des cas - estimés. Généralement, ces créations sont utilisées par moins d'un millier de personnes. Une personne qui a obtenu un contrat grâce à une application qu'elle avait développée, a estimé l'utilisation de celle-ci à 3 000 personnes. D'autre part, quelques répondants ont manifesté une sorte d'insatisfaction quant à l'appui des institutions publiques lorsqu'ils développent des projets. Selon eux, cela n'intéresse pas les institutions, ce qui conduit à une forme de démotivation. Enfin, tous conviennent que la qualité des données est variable et que plusieurs données sont manquantes.

Néanmoins, les données ouvertes constituent une valeur ajoutée pour leurs pratiques. Elles leur permettent notamment de réaliser un gain de temps considérable. Certains ont également mentionné que cela constituait un gain monétaire puisqu'ils n'avaient pas besoin de faire une demande d'accès à l'information qui pourrait être payante en fonction de la nature des données demandées. Aucun des répondants ne télécharge des jeux de données en vue d'une utilisation potentielle future, mais un répondant nous a indiqué conserver systématiquement une copie des jeux de données qu'il télécharge.

En bref, le profil des usagers non-professionnels interrogés peut être résumé ainsi :

- Ce sont des hommes
- Ils ont les compétences pour lire et analyser les données
- Ils présentent un profil hybride allant de l'utilisateur passionné à l'aspirant professionnel, ce sont donc tant des explorateurs que des exploitants de données
- Leurs attentes envers le portail des données ouvertes sont élevées
- Ils identifient trois limites au portail actuel : 1) les données nécessitent un nettoyage trop long ; 2) le manque de suivi des institutions publiques sur les projets créés à partir des données ; 3) la qualité variable des données
- Ils identifient également trois points forts : 1) une valeur ajoutée pour leurs pratiques, 2) le gain de temps et 3) le gain monétaire

3.3.2 Les entreprises, entrepreneurs et développeurs

Les entreprises, entrepreneurs et développeurs qui ont accepté de participer à nos entretiens sont au nombre de cinq. Leur profil était similaire : des hommes qui possédaient une grande capacité

d'analyses des données. On retrouve ainsi deux dirigeants d'entreprises spécialisées en analyse de données, un consultant pour une firme de service-conseil, un programmeur à son compte et un ingénieur en génie civil. Étant donné leur profil, leurs motivations étaient professionnelles à l'exception du programmeur qui utilisait les données ouvertes comme ressource pédagogique pour former ses stagiaires. Quatre d'entre eux résident dans le Grand Montréal et un à Québec.

L'ensemble de ces usagers connaît l'existence du portail des données ouvertes de Montréal depuis sa création. À l'exception de deux personnes, ils téléchargent les jeux de données de manière fréquente (une fois par mois) ou très fréquente (une fois par semaine). L'ensemble des membres de cette catégorie sont des réutilisateurs d'autres jeux de données issues d'autres institutions publiques, d'organismes communautaires ou associatifs ou des entreprises. L'exploitation des jeux de données ouvertes par les membres de cette catégorie sert à enrichir une base de données constituée de plusieurs jeux issus d'horizons divers. Leur chaîne de traitement des données est semblable à celle des usagers non-professionnels. Majoritairement, ces individus produisent des livrables à leurs clients illustrés de cartes, d'analyses spatiales et de statistiques. Ce sont clairement des exploitants des données.

Trois limites ont été soulevées par ces usagers. Une personne a arrêté d'utiliser les jeux de données de la Ville de Montréal quelques mois après que le service a été rendu disponible à la population parce que la qualité des données était insuffisante. Le travail de préparation nécessitait trop de temps et d'efforts pour obtenir un jeu exploitable de données. Une autre personne souligne que la documentation est parfois insuffisante pour comprendre la façon dont les données ont été créées. La présence d'incohérences ou d'erreurs dans le jeu de données n'est pas un problème pour ces usagers qui possèdent la compétence d'outrepasser ces difficultés soit en nettoyant le jeu de données soit en l'enrichissant d'autres jeux collectés dans d'autres bases de données pour créer l'information souhaitée. Un des développeurs a manifesté une sorte d'insatisfaction quant à l'absence d'intérêts que semblent avoir les institutions publiques lorsqu'ils développent des applications. Il aimerait des signaux plus forts des institutions publiques pour soutenir la création d'application dans un paradigme d'innovation ouverte. Une vitrine des applications développées, un guichet et/ou l'ouverture d'une ligne directe seraient appréciés. Deux autres propositions pour améliorer le service des données ouvertes ont été avancées : plus de visualisations et ôter les accents dans les jeux de données.

L'accès libre des jeux de données ouvertes présente plusieurs avantages pour ces usagers. Une personne estime que l'existence de ces données représente un gain de temps et une réduction des coûts à hauteur de 5 à 10% dans le cadre de son entreprise. La centralisation des informations permet d'obtenir plus rapidement les jeux de données sans effectuer des demandes aux différents services ce qui exige du temps et de l'énergie. Plusieurs mois sont parfois nécessaires pour obtenir les données. Pour deux autres personnes, l'accès rapide à ces informations leur permet de bonifier les analyses en enrichissant les données. Ils réduisent le temps passé à obtenir les données et à les nettoyer pour augmenter le temps passé sur l'analyse des données. Ils augmentent la qualité de leurs services et approfondissent leur connaissance de la situation à l'étude. Enfin, pour d'autres, l'accès à ces données a un effet d'aubaine. Ils s'en servent pour produire un service à valeur ajoutée à une clientèle ciblée.

En bref, le profil de ces usagers est le suivant :

- Ce sont des hommes
- Ils ont les compétences pour lire et analyser les données
- Ce sont des exploiters de données
- Ils identifient trois limites au portail actuellement : 1) la qualité des données ; 2) le manque de suivi/soutien des institutions publiques lorsqu'une application est développée grâce aux données ouvertes ; 3) le manque d'information sur la façon dont la donnée est créée
- Ils identifient également trois points positifs : 1) le gain de temps, 2) le gain monétaire, 3) l'amélioration de leurs services

3.3.3 Les chercheurs et les étudiants

Deux chercheurs et deux étudiants ont accepté de répondre à nos questions. Ce sont deux femmes et deux hommes. Ces individus présentent une très grande autonomie technique dans leur utilisation des données ainsi qu'une grande compréhension des enjeux liés à la diffusion et à l'utilisation pour un large public des données ouvertes. Les deux chercheurs travaillent dans des structures de recherche dans le secteur de la santé. Un des étudiants, au troisième cycle, évolue au sein d'une chaire de recherche sur la mobilité tandis que l'autre a entrepris des études en informatique. Ils sont tous basés à Montréal. Les profils de ces personnes présentent tous une trajectoire d'usage très similaire aux autres usagers.

Leurs motivations pour utiliser les données ouvertes sont professionnelles pour les chercheurs et académiques et personnelles pour les étudiants. Trois d'entre eux utilisent les données ouvertes pour caractériser différents milieux urbains à Montréal afin de comprendre l'évolution physique de l'environnement urbain et ses impacts sur les comportements de la population. L'un des étudiants utilise, en revanche, les données ouvertes pour des motivations personnelles dont le but est d'apprendre des compétences transversales à son domaine d'étude (l'informatique), particulièrement la maîtrise de l'analyse spatiale. L'ensemble des personnes utilisent un très grand nombre de jeux de données ouvertes. Ce sont tant des exploiters que des explorateurs. Ces jeux de données enrichissent une base de données personnelle aux différentes structures de recherche alimentée par d'autres sources de données.

Ces personnes ont identifié plusieurs limites aux données ouvertes actuelles. Ainsi, dans la phase de nettoyage des données ouvertes, les chercheurs ont indiqué la nécessité, parfois, de demander au service des données ouvertes de libérer certaines colonnes dont l'information était restreinte pour des motifs de respect de la vie privée. Ces actions sont permises grâce à des ententes bilatérales entre l'institution de recherche et la Ville. Parfois, c'est la nature des données qui posent problème : une des personnes regrettait notamment que certains jeux de données n'étaient disponibles qu'à l'échelle des arrondissements de la Ville de Montréal et ne couvraient pas l'ensemble de l'île de Montréal. Une autre regrettait l'absence d'un format commun entre tous les portails des villes du Canada ce qui empêche leur comparaison. Cependant, ces membres sont particulièrement sensibles à l'absence de documentation liée à chaque jeu de données. Ils regrettent que certains jeux plus anciens ne soient plus disponibles sur le portail. Enfin, deux discussions avec les étudiants ont porté sur la capacité de la Ville à permettre la mise en œuvre d'un service visant à soutenir la création de données collaboratives appréhendées comme des jeux de données créés par et pour les citoyens (crowdsourcing ou production participative).

Plusieurs points positifs ont également été soulevés. Concrètement, l'accès libre à des jeux de données ouvertes permet aux chercheurs d'enrichir leurs bases de données. Ces jeux soutiennent la production de connaissances (rédaction de rapports, d'articles scientifiques, de l'aide à la décision publique) même si, d'après leurs témoignages, ces chercheurs auraient été quand même en mesure d'obtenir ces jeux de données sans l'existence de ce portail. Si l'accès libre à des données n'est pas l'intérêt principal pour les chercheurs, l'existence du portail des données ouvertes présente un autre avantage : l'archivage et la centralisation des données par thématique permettent d'identifier des jeux de données similaires (par exemple, des données autour du vélo ou du verdissement). C'est davantage la sériation des jeux de données qui présente un fort intérêt pour les chercheurs en leur offrant un gain de temps.

En bref :

- Ce sont deux hommes et deux femmes
- Ils disposent des compétences pour lire et analyser les données
- Ce sont tant des explorateurs que des exploiters de données
- Les limites identifiées des données ouvertes actuelles sont : 1) la nature des données ; 2) le format des données ; 3) l'absence d'information sur la façon dont la donnée est construite
- Les points positifs identifiés des données ouvertes actuelles sont : 1) l'enrichissement de base de données existantes; 2) la possibilité de sériation des données

3.3.4 Les journalistes de données

Les usagers journalistes qui ont accepté de se prêter au jeu de l'entretien sont au nombre de deux. Ce faible nombre s'explique notamment, comme le précise l'une des personnes, par le faible nombre de journalistes spécialistes des données au Québec et au Canada. Les deux journalistes sont un homme et une femme. Leurs compétences en analyse de données sont largement autodidactes et ils s'appuient tous deux sur leurs collègues. Leurs motivations pour utiliser les données ouvertes sont professionnelles. La chaîne de traitement des données est exactement la même que celle des usagers non-professionnels. Il s'agit parfois de documenter une thématique spécifique grâce aux données ouvertes ou bien de partir des données ouvertes pour créer un « sujet ». Ce sont tant des explorateurs que des exploiters de données. Leur lieu de travail est situé à Montréal.

Leur contact avec les données ouvertes s'est réalisé dès l'ouverture du portail des données ouvertes. Le portail des données ouvertes de la ville de Montréal n'est pas le seul portail qu'ils utilisent, mais ils attendaient celui-ci avec impatience. Une des deux personnes indique qu'elle vérifie toujours les données proposées, notamment pour saisir leur degré de transparence. Une des deux personnes souligne qu'au début elle le consultait toutes les semaines, mais qu'ensuite sa consultation a baissé en raison de la nature des données présentées.

Plus précisément, trois limites principales ont été décrites. Premièrement, les données ont besoin d'être beaucoup retravaillées avant d'être exploitables, ce qui se conjugue mal avec le peu de temps dont disposent les journalistes pour écrire des articles. Deuxièmement, les données sont présentées comme n'étant pas toujours « intéressantes » pour écrire un article, mais une des deux personnes nuance ce point en soulignant que les données publiées tendent à devenir de plus en plus pertinentes. Néanmoins, les deux journalistes trouvent qu'il n'y a pas suffisamment de données qui soient « vraiment intéressantes », mais cette réalité n'est pas propre aux données du portail de

la Ville de Montréal : ce constat est généralisable selon eux à l'ensemble du Québec. Troisièmement, la qualité des données est remise en question notamment en raison de leur caractère incomplet.

Les deux personnes soulignent qu'il est important de conserver ces données ouvertes, même si elles n'ont pas une utilisation claire et immédiate, car celles-ci pourraient être utiles à plus long terme. Les interviewés considèrent qu'en ce sens le portail des données ouvertes va dans la bonne direction. Une des deux personnes précise également que ces données permettent un gain de temps : s'adresser au portail des données ouvertes permet d'obtenir rapidement une information en lien avec les données.

En bref, le profil de ces usagers est le suivant :

- Ce sont un homme et une femme
- Ils sont autodidactes dans leur formation et l'amélioration de leurs compétences dans le traitement et l'analyse des données. Ils sont entourés de personnes qui peuvent pallier leur manque de connaissances éventuelles
- Ce sont autant des exploiters que des explorateurs de données
- Les limites identifiées des données ouvertes actuelles sont : 1) le degré « d'intérêt » des données ; 2) la qualité des données ; 3) le long travail de traitement des données avant leur exploitation
- Les points positifs identifiés des données ouvertes actuelles sont : 1) données importantes pour l'avenir; 2) gain de temps.

3.3.5 Les autres répondants (ONG et groupes communautaires)

Cette dernière catégorie rassemble trois répondants qui disposent d'un profil académique spécifique, bien que travaillant ou ayant travaillé dans des institutions publiques à titre de fonctionnaire ou d'un OBNL. Ces derniers entretiens ont confirmé les propos des autres catégories d'usagers, c'est la raison pour laquelle nous décrivons plus rapidement la réalité dont ils nous ont fait part.

Les trois répondants sont deux hommes et une femme. Ils ont utilisé les données ouvertes dans le cadre de leur travail, la prise de contact avec le portail des données ouvertes ne s'est pas réalisée de façon identique. Pour une des deux personnes, le contact avec le portail est récent (mai 2020) et n'a pas donné accès directement aux données nécessaires. Si cette personne semble détenir des compétences en traitement de données, elle a à sa disposition toute une équipe de support technique. L'autre personne a été embauchée dans une institution publique fédérale en raison de son profil académique et de ses compétences. Elle détient une vaste connaissance des jeux de données ouvertes à tous les échelons politiques du Québec et du Canada. Ces deux répondants ont travaillé l'un à Montréal et l'autre à Québec. Le traitement des données est le même que pour les autres usagers.

Dans les trois cas, l'objectif était de créer de nouvelles connaissances au sein des institutions visées. Ce sont majoritairement des exploiters de données. Les limites et forces identifiées vis-à-vis du portail des données ouvertes sont les mêmes que celles des autres répondants. Deux différences dans leurs propos ont été notées par rapport aux autres usagers. Dans un cas, le portail

des données ouvertes a permis d'identifier où se trouvaient les données recherchées. Dans un autre cas, un des deux interlocuteurs a souligné qu'en dehors des impacts positifs des données ouvertes, il peut exister un impact négatif qui consiste en la mauvaise interprétation des données ou leurs croisements et le sens qui leur sont donnés.

En bref, le profil de ces usagers est le suivant :

- Ce sont deux hommes et une femme
- Ils ont les compétences pour lire et analyser les données
- Ce sont des exploiters de données
- Les limites identifiées des données ouvertes actuelles sont les mêmes que les usagers précédents.
- Les points positifs identifiés des données ouvertes actuelles sont les mêmes que les usagers précédents.

3.3.6 Synthèse de l'analyse des entretiens

Ces résultats nous permettent de dresser un premier portrait général de l'ensemble de nos répondants. En effet, en dépit de quelques différences entre les profils, des points communs ressortent entre tous les répondants et l'utilisation qu'ils font des données ouvertes (voir Tableau 6).

TABLEAU 6. TYPES D'USAGER-RÉUTILISATEUR DU PORTAIL DES DONNÉES OUVERTES LA VILLE DE MONTRÉAL AYANT ACCEPTÉ DE PARTICIPER AUX ENTREVUES

| | Usager non-professionnel | | Usager professionnel | | | |
|----------------------------------|--|---|---|--|--|---|
| | Passionné | Aspirant professionnel | Entreprises, entrepreneurs, développeurs | Journalistes | Chercheurs, professeurs, étudiants | Groupes communautaires et ONG |
| Type de demande | Demande d'un produit personnalisé / Base de données spécifiques | Base de données spécifiques | Base de données spécifiques / Demande d'un produit personnalisé | Base de données spécifiques / Précisions sur les jeux de données | Base de données spécifiques / Précisions sur les jeux de données | Base de données spécifiques |
| Chaîne de traitement des données | Souhaite apprendre en résolvant les obstacles. | Souhaite apprendre en résolvant les obstacles. | Sans difficulté. Souhaite gagner du temps sans nettoyer les bases de données. | A appris sur le tas. Donne une importance à la documentation. | Importance donnée à la documentation. | Sans difficulté. Souhaite gagner du temps sans nettoyer les bases de données. |
| Type d'utilisations | Démarche exploratoire dominante | Démarche exploratoire dominante | Démarche exploitation dominante | Démarche exploitation et exploratoire | Démarche exploitation et exploratoire | Démarche exploitation dominante |
| Type d'impacts | Plaisir de faire. Auto-apprentissage. Développement de compétences personnelles. | Développement de compétences personnelles. Augmenter son employabilité. Développement d'un service ou d'une application | Développement d'un service ou d'une application. Aide à la prise de décision. | Révélation d'information. Production d'opinion. | Création de connaissances. Aide à la prise de décision. | Renforcer leur connaissance sur un secteur d'activité. |

De façon générale, les ré-utilisateurs qui ont accepté de participer à notre enquête sont :

- Majoritairement des hommes (15/19)
- Ils disposent des compétences pour lire et analyser les données ou, à défaut, sont capables de s'appuyer sur d'autres personnes pour réaliser ces tâches.
- Les entreprises, entrepreneurs et développeurs, ainsi que les groupes communautaires et ONG sont majoritairement des exploiters de données.
- Les autres réutilisateurs (usagers non-professionnels, journalistes, chercheurs et étudiants) sont tant des explorateurs que des exploiters de données.
- Les limites principales identifiées par les répondants sont au nombre de trois :
 - La qualité des données (comprend la nature, le format, l'information),
 - Le suivi des institutions publiques lorsqu'un produit est créé à partir des données ouvertes
 - L'intérêt variable des données publiées
- Les forces du portail des données ouvertes sont au nombre de trois :
 - Une valeur ajoutée pour leurs pratiques personnelles et professionnelles (données utiles plus tard, amélioration des services, etc)
 - Le gain de temps
 - Le gain monétaire

Quelques répondants, comme les entrepreneurs et les étudiants ont également émis quelques idées sur ce que serait selon eux un portail de données ouvertes idéal :

- Une vitrine des applications et des projets développés utilisant des données ouvertes
- Un guichet unique ou une ligne directe vers l'ensemble des données ouvertes existantes (incluant celles d'autres organismes)
- Création de données collaboratives (créées par et pour les citoyens)

4. Tendances et constats principaux sur les réutilisateurs et leurs démarches

L'analyse des messages reçus par le portail et l'analyse des réponses de quelques réutilisateurs permet d'esquisser quelques tendances quant à la chaîne de traitement de données, aux démarches entreprises et à leurs objectifs et aux impacts qui en découlent. Trois tendances principales se dessinent :

- L'existence d'une asymétrie parmi les usagers dans la chaîne de traitement des données
- Des démarches de type explorateur et de type exploratoire
- Des types d'impacts directs et indirects

4.1 Une asymétrie parmi les usagers dans la chaîne de traitement de données

La réutilisation des données consiste à extraire des données de leur contexte initial de production pour leur offrir un autre cadre d'interprétation et de traitement au sein de nouveaux contextes sociaux. Entre ces deux actions d'extraction et de réinterprétation, la chaîne de traitement des données assure la réutilisation des données. Or, certains usagers ne dépassent pas le stade de la consultation. Nos résultats ont montré que certains usagers consommaient directement

l'information à travers la consultation d'un plan ou le téléchargement d'une photographie aérienne. Dans le cas d'une réutilisation des données au sens où nous l'avons défini, nos résultats confirment que la réutilisation à travers une chaîne de traitement de données semblables pour l'ensemble des usagers « professionnels » est conditionnée par la qualité (mise à jour régulière), la fiabilité (vérification de la donnée, présence d'une documentation explicitant ses conditions de productions et ses limites) et l'interopérabilité des données (Paquienséguy, 2017).

Des disparités dans la capacité de juger de la réutilisation des jeux ont été observées. Il y a clairement une différence forte entre les non-professionnels et les professionnels dans leur capacité à identifier un jeu de données valorisable et à le valoriser en produisant une information compréhensible à partir des jeux de données de la Ville. L'observation des différences dans l'extraction et le traitement des données montrent que l'ensemble des usagers présentent de très grandes similarités dans leur trajectoire d'usage avec les données. Ils disposent des mêmes outils numériques et de compétences techniques élevées. Les différences sont davantage liées aux ressources organisationnelles et à la compréhension plus fine de la nature d'une information. Par ailleurs, les données que nous avons analysées ne nous permettent de saisir le nombre de personnes qui « ne font que » télécharger les données sur le portail sans entrer en communication avec l'équipe des données ouvertes¹⁴.

Enfin, les répondants qui ont accepté de nous répondre sont majoritairement des hommes. La taille restreinte de notre échantillon ne nous permet pas de tirer des constats plus généraux sur un éventuel déséquilibre entre hommes et femmes sur l'utilisation des données ouvertes et les compétences nécessaires qu'elle exige. Néanmoins, ce constat invite à poursuivre plus en profondeur cette réalité avec des données plus importantes (Brandusescu et Nwakanma, 2019).

4.1.2 Les démarches de type exploratoire et de type exploitation

La compréhension des disparités au sein de la chaîne de traitement de données nous amène à penser que l'amélioration des conditions de valorisation des données ouvertes n'est pas uniquement liée aux types d'usagers, mais aussi à la prise en compte des démarches entreprises par l'utilisateur. Les utilisations des données ouvertes des usagers peuvent se classer selon deux types (March, 1991) : les démarches de type exploratoire et celles de type exploitation.

Les utilisations de type exploratoire présentent des profils hybrides en matière de motivations des usagers (tantôt personnelle, tantôt professionnelle, parfois les deux) et de buts à atteindre. Des

¹⁴ À noter que le nouveau site web de données qui remplace l'ancien Portail offre une visualisation du nombre de téléchargements et des jeux de données téléchargés sur le portail des données ouvertes de la ville de Montréal. Une analyse plus fine de la fréquence de ces téléchargements reste à faire : <https://donnees.montreal.ca/ville-de-montreal/frequentation-du-portail-de-donnees-ouvertes>

usagers non-professionnels, des développeurs, des étudiants, des entrepreneurs ou même des chercheurs soulignent l'importance de leur apprentissage dans leur utilisation des données ouvertes. Ces démarches sont liées davantage à l'expérimentation, à la créativité mais aussi à l'acceptation de l'échec. Ce qui est recherché, c'est l'apprentissage sur le tas et le développement de savoirs et de compétences transversales à leur noyau d'expertise. Pour ces usagers, le temps passé à traiter les jeux de données importe peu. Le processus est plus important que l'objectif.

Les utilisations de type exploitation sont identifiables uniquement auprès de ceux qui utilisent certaines données dans le cadre professionnel de la réalisation d'un projet rémunéré. La chaîne de traitement de la réutilisation des données doit être la plus courte possible. Plusieurs interrogés évoquent le même cas de la réception d'un document en format PDF en guise de réponse à leur demande d'information. Le temps passé à retravailler l'information en sein d'un fichier utilisable a découragé plusieurs d'entre eux. Certains se sont alors tournés vers d'autres organisations pour obtenir des informations similaires.

4.1.3 Les types d'impacts directs et indirects

Qu'ils soient professionnels ou non, les usagers valorisent les données ouvertes via la transformation des données en services à destination d'eux-mêmes, de clients, de citoyens ou de la communauté. Majoritairement, les impacts directs et indirects de ces valorisations n'ont pas été mesurés par les usagers.

À l'échelle des individus, l'impact identifié est rattaché à une quête de performance et d'efficacité individuelles. Les impacts positifs visent alors l'acquisition de nouvelles compétences dans l'analyse des données et le traitement statistique et cartographique. Pour certains, cette volonté d'apprendre sur le tas s'inscrit dans une volonté d'augmenter son employabilité et de devenir plus productif. Certains amateurs ont alors publié des billets de blog avec leur découverte, construit des sites internet pour diffuser leur connaissance ou développer une application.

Conclusion

Loin des premières aspirations militantes des partisans du logiciel libre qui reprochaient au mouvement de l'*open source* de n'être qu'un mode de développement sans vision politique (Stallman, 2007)¹⁵, le portail des données ouvertes est structuré comme un service rendu à la population dont la satisfaction des besoins prime sur la réalisation d'un projet plus vaste. La relation avec les usagers est un rapport utilitaire basé principalement sur la demande d'un produit non disponible, la résolution de problèmes identifiés ou la requête d'information supplémentaire sur l'origine des données. En de rares occasions, les échanges dépassent ce cadre d'une relation opérationnelle et utilitaire entre un représentant du portail et un individu. La valorisation des données ouvertes ne semble pas être au cœur de ces échanges. Difficile dans ces conditions de voir se réaliser la vision techniciste de ceux qui voyait les technologies numériques comme un moyen de moderniser les institutions et revitaliser la démocratie (Noveck, 2009; 2012). Comme le rapporte Laëticia Prido (2017), les partisans de l'ouverture des données ouvertes défendaient l'idée que celle-ci devait permettre l'accès aux citoyens à de plus en plus d'informations issues des gouvernements afin de leur permettre une meilleure compréhension des actions de l'État. Cette ouverture les inciterait à collaborer et à s'impliquer dans les décisions politiques.

Actuellement, à l'échelle du réutilisateur, le portail des données ouvertes à Montréal ne semble pas encore capable de remplir toutes les conditions pour favoriser la transparence, la collaboration et la participation. Ce constat doit cependant être nuancé. En effet, aux yeux des usagers interviewés, mais également aux yeux de ceux qui envoient des messages à l'équipe du portail des données ouvertes, le portail des données ouvertes est d'une part un projet prospectiviste : il suscite beaucoup d'attentes. Par conséquent, ses réalisations apparaissent – dans l'esprit des usagers – souvent peu à la hauteur de leurs attentes. De plus, la majorité des usagers interviewés sont des utilisateurs de l'ancienne version du portail : la mise en ligne d'une nouvelle version du portail web en 2020 témoigne d'une volonté d'ajustement continue de celui-ci.

Néanmoins, le mandat du portail ne semble pas toujours clair à celles et ceux qui s'adressent à lui. Ce déséquilibre entre les attentes des usagers et les capacités de réponse du portail des données ouvertes nécessite de 1) remettre le portail des données ouvertes dans le contexte de la municipalité et de ses compétences et de 2) réfléchir à des pistes d'actions en trois temps : les pistes d'actions réalisables par l'équipe qui soutient le portail, les pistes d'actions souhaitables qui exigeraient des moyens humains et financiers plus conséquents et enfin, les pistes d'actions idéales qui permettraient de répondre tant aux limites identifiées que la valorisation des forces de ce dernier.

¹⁵ <https://www.gnu.org/philosophy/open-source-misses-the-point.fr.html>, consulté le 14 décembre 2020.

Piste d'actions réalisables par l'équipe qui soutient le portail des données ouvertes et visant l'ensemble des réutilisateurs

Actuellement, le portail des données ouvertes est conçu comme un service public. Le nombre d'utilisateurs reste cependant assez faible (si on se fie au nombre de messages envoyés et aux entrevues réalisées) et se caractérise par sa capacité à disposer de certaines compétences en lecture et en analyse de données. Il pourrait donc dans un premier temps être envisagé de travailler sur l'amélioration du service proposé tout en permettant à un plus grand nombre d'utilisateurs potentiels de profiter de ce service.

- **Piste d'action n°1** : À court terme, une page sur le nouveau site web présentant en des termes vulgarisés ce que sont les données ouvertes et leurs buts pourraient être ajoutés. La façon dont l'ouverture des données se réalise, les thématiques choisies etc. pourraient également être explicitées. Cela viendrait compléter la vidéo actuellement disponible sur le site du Laboratoire d'innovation urbaine qui explique ce qu'est un gouvernement ouvert¹⁶. Disposer de sections pédagogiques expliquant le but des données ouvertes, la nature des formats disponibles et leur utilisation, avoir des espaces de formation tant pour les débutants que pour les experts, pourrait permettre d'augmenter le nombre d'utilisateurs du portail et de mieux orienter les messages qui lui sont destinés¹⁷. Cela permettrait de mieux répondre à la volonté du Portail de favoriser l'accessibilité aux données et leur valorisation.
- **Piste d'action n° 2** : À court terme, une page web sur le nouveau site du portail expliquant, les types de fichiers et formats fournis et leurs objectifs pourrait être créée. Des vidéos présentant des tutoriels pour exploiter ou explorer certains jeux de données pourraient par exemple être mis à la disposition. Cela permettrait – encore une fois – de mieux répondre à la volonté du Portail de favoriser l'accessibilité aux données et leur valorisation.
- **Piste d'action n°3** : À court terme, un sondage annuel permettant de connaître l'évolution des profils des utilisateurs, leurs projets, l'impact et la valeur créées grâce aux données pourrait être mis en œuvre. Une telle enquête offrirait la possibilité d'ajuster le service des données ouvertes aux différents types de réutilisateurs, que ce soit ceux qui consultent le site pour télécharger/demander un document spécifique par exemple ou ceux qui ont des projets de plus grande envergure.
- **Piste d'action n°4** : À moyen terme, la définition de la donnée, une explication de sa construction pourraient être ajoutées. Lorsqu'un jeu de données comporte une erreur, celle-ci pourrait être publiée afin d'assurer une traçabilité des jeux de données. Les jeux de données les plus anciens pourraient également figurer sur le site.
- **Piste d'action n°5** : À long terme, se doter d'une charte de qualité de données qui permettent à l'ensemble des arrondissements de la ville d'alimenter le portail des données ouvertes facilement. Les lignes directrices proposées par Données Québec pourraient constituer en ce sens un bon point de départ¹⁸.

¹⁶ <https://laburbain.montreal.ca/documentation>, Liens utiles :

<https://www.youtube.com/watch?v=bln9ZGd26KE>

¹⁷ Il est à noter que le nouveau portail des données ouvertes explique ce que sont les données et la gouvernance de celles-ci. Néanmoins, cette information reste peu mise en valeur sur le site.

¹⁸ Voir le site suivant : <https://www.donneesquebec.ca/lignes-directrices-sur-la-diffusion-de-donnees-ouvertes/>

- **Piste d'action n°6** : À long terme, pour favoriser l'enrichissement et le nombre de jeux de données disponibles, une uniformisation de la construction de la donnée mais également de la gouvernance qui l'entoure serait nécessaire au sein des arrondissements. Dans la mesure où la ville de Montréal dispose de 19 arrondissements où les modes de gestion des données peuvent être différents, cette piste d'action nécessitera un engagement politique clair et des ressources financières adéquates pour arriver à une uniformisation de construction de la donnée.

Ces pistes d'action permettraient de travailler la qualité de la donnée, mais également de mieux identifier ce qu'il est possible d'attendre d'un portail de données ouvertes municipal, afin de rééquilibrer les attentes et les réponses entre usagers et la plateforme. Cela permettrait de travailler l'aspect « transparence » des données ouvertes.

Pistes d'actions souhaitables

Les pistes d'actions souhaitables sont les pistes d'actions qu'il serait nécessaire de mettre en place – notamment pour répondre aux réutilisateurs qui souhaitent valoriser un produit réalisé à partir des données ouvertes, mais pas uniquement. Ces pistes d'actions exigent cependant l'implication d'acteurs et de soutiens logistiques autres que celui de l'équipe qui s'occupe du portail. Actuellement, le portail des données ouvertes de la ville de Montréal fait de l'utilisateur un acteur individuel « consommateur de données ». Cette vision ne permet pas un suivi des produits éventuellement créés par les usagers. Il s'agirait de créer davantage de possibilités d'innover en créant les bases d'un véritable écosystème :

- **Piste d'action n°7** : Mettre en réseau les différents usagers pour permettre le suivi et créer une interdépendance entre eux. La « plus-value » du portail consiste, outre la publication des données, à générer des relations entre acteurs. Certains réutilisateurs souhaitent que leurs produits soient mieux mis en valeur. Le portail actuel dispose d'une page où les différentes applications créées par des réutilisateurs sont présentées¹⁹. Les usagers interrogés n'en font pas mention et ont souligné la volonté que cette mise en valeur dépasse l'affichage actuel.
- **Piste d'action n°8** : Indiquer les autres fournisseurs de données qui pourraient être utiles aux différents usagers. Certains réutilisateurs ont indiqué qu'ils aimeraient disposer d'un guichet unique de l'ensemble des données ouvertes disponibles dans la province, plutôt que d'avoir différents sites offrant diverses données. L'indication pour certaines thématiques d'autres fournisseurs de données pourrait venir combler ce besoin.
- **Piste d'action n°9** : Offrir sur la plateforme des outils et des technologies dont les usagers pourraient avoir besoin pour innover.

En d'autres termes, cela nécessiterait d'augmenter le nombre d'acteurs pour s'occuper du site des données ouvertes afin de lui permettre de répondre à l'objectif d'innovation qui sous-tend ce type de démarche.

¹⁹ <https://donnees.montreal.ca/applications>

Pistes d'actions idéales

Les pistes d'actions idéales sont des pistes d'actions qui nécessitent une redéfinition du portail des données ouvertes et par conséquent exigent la participation d'un plus grand nombre d'acteurs et l'ajout de moyens humains et financiers conséquents pour permettre sa réalisation. Il s'agit principalement de faire du portail des données ouvertes de la ville de Montréal, un projet politique plus vaste – qui pourrait être indépendant de la municipalité – visant à favoriser la participation d'acteurs divers.

- **Piste d'action n° 10** : S'assurer que les acteurs s'engagent sur le long terme, afin de sortir de l'esprit de « consommation ponctuelle » de données, sans retour vers les autres acteurs. Il s'agit donc d'un engagement sur le long terme, permettant notamment aux nouveaux venus d'échanger avec les plus anciens.
- **Piste d'action n°11** : S'assurer de la représentativité d'une diversité d'acteurs gravitant autour du portail des données ouvertes. La charte des données numériques parle d'une « souveraineté numérique » en parlant de données venant de plusieurs acteurs cette notion pourrait être étendue aux acteurs impliqués.
- **Piste d'action n°12** : S'assurer que la plateforme soit une plateforme d'apprentissage continue en lien avec la disposition des outils et les interactions entre acteurs concernés.
- **Piste d'action n°13** : S'assurer que la ville dispose du leadership pour s'assurer de l'épanouissement et la pérennité des interactions créées.
- **Piste d'action n°14**: Créer un espace de production participative des données, réalisées par et pour les citoyens.

Ces différentes pistes d'actions invitent à repenser la définition et l'objectif du portail des données ouvertes de la ville de Montréal. Au sein des écrits scientifiques, il n'existe pas de définition unique et consensuelle d'un portail de données ouvertes. Le plus souvent, les écrits reprennent la définition des données ouvertes et précisent que leur mise en ligne par un gouvernement sur une plateforme constitue un portail de données ouvertes. D'autres mettent l'accent sur l'idée qu'un portail de données ouvertes constitue un ensemble de services (Danneels et al., 2017), tandis que d'autres réfèrent davantage aux acteurs concernés (Kassen, 2018). Il existe cependant un point commun quant à l'objectif poursuivi par un portail de données ouvertes, et ce, parallèlement aux promesses que nous évoquons dans la première partie de ce rapport : un portail de données ouvertes vise à favoriser la transparence, l'imputabilité, la participation et l'innovation. De tels objectifs ne peuvent être atteints sans une mise en relation des acteurs qui gravitent autour d'un tel dispositif. Cela ne peut se faire également sans une redéfinition de la notion de démocratisation, car l'accès seul aux données ne peut garantir l'exercice d'une citoyenneté numérique (Foth, Brynskov, et Ojala 2015). Néanmoins, il convient de remettre ces différents objectifs en lien avec le contexte spécifique de la municipalité et des compétences et moyens dont elle dispose. Une telle entreprise sous-entend également de clarifier les rôles de chacun : la municipalité doit-elle être responsable de ce projet ? Cela rentre-t-il dans sa mission ? Ou, à l'instar d'autres villes, une telle entreprise doit-elle être indépendante de la municipalité?

Références bibliographiques

- Barbrook, R. et Cameron, A. (1996). The Californian ideology. *Science as Culture*, 6(1), 44-72. doi: [10.1080/09505439609526455](https://doi.org/10.1080/09505439609526455)
- Batty, M. (2013). Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3), 274-279. doi: [10.1177/2043820613513390](https://doi.org/10.1177/2043820613513390)
- Boudreau, C. et Caron, D. (2016). *Les données ouvertes dans l'administration publique québécoise : Utilités, freins et pistes de solution*. Récupéré de http://www.enap.ca/cerberus/files/nouvelles/documents/La_recherche/NetGouv_Les_donneesouvertesdansadministrationquebecoise_2016.pdf
- Brandusescu, A. et Nwakanma, N. (2019). Gender equity. Dans B. Walker, T. Davies et F. Perini (dir.), *The State of Open Data: Histories and Horizons*. (Cape Town and Ottawa: African Minds and International Development Research Centre., p. 287-299).
- Breton, P. et Proulx, S. (2006). *L'explosion de la communication: à l'aube du XXIe siècle*. Montréal, Québec : Boréal : La Découverte.
- Capelle, Marthe, Jutand, Aline et Morandi, Franc (2018) Stratégies pédagogiques pour diffuser la culture des données ouvertes, *Revue Cossi*, n°4, en ligne : <https://www.revue-cossi.lerass-ceric.com/images/article-pdf/RevueCOSSI-n4-2018.pdf#page=20>
- Chignard, S. et Benyayer, L.-D. (2015). *Datanomics: les nouveaux business models des données*. Limoges : FYP éditions.
- Conradie, P. et Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, 31, S10-S17. doi: [10.1016/j.giq.2014.01.003](https://doi.org/10.1016/j.giq.2014.01.003)
- Church, James. (2017) « Civil Society and the Open Data Movement ». *A Quarterly Journal of Government Information Practice and Perspective* 45 (3).
- Danneels, Lieselot, Stijn Viaene, et Joachim Van den Bergh. 2017. « Open Data Platforms: Discussing Alternative Knowledge Epistemologies ». *Government Information Quarterly* 3 (3): 365-78. <https://doi.org/10.1016/j.giq.2017.08.007>.
- Dickner, Nicolas (2017) « Comprendre et manipuler les données ouvertes de l'administration publique. La situation au Gouvernement du Québec et à la Ville de Montréal ». Montréal: École de bibliothéconomie et des sciences de l'information.
- Dymytrova, Valentyna, et Françoise Paquienséguy (2017) « La réutilisation et les réutilisateurs des données ouvertes en France: une approche centrée sur les usagers ». *Revue Internationale des Gouvernements Ouverts*, 117-32.
- . 2019. « Des données aux services dans les Métropoles : enquête auprès des réutilisateurs professionnels de données ouvertes ». CODATA, France, DATA VALUE CHAIN.
- Dymytrova, Valentyna, Larroche, Valérie et Paquienséguy. 2018. « Cadres d'usages des données par des développeurs, des data scientists et des data journalistes », ANR OpenSensingCity 14-CE24-0029 – Elico – Livrable 3, Récupéré de : https://www.researchgate.net/publication/324976707_Livrable_3

- Dymytrova, Valentyna. (2018). Data journalisme, entre pratique créative innovante et nouvelle médiation experte ? Une analyse conjointe des discours et des productions journalistiques. Communication présentée au XXI Congrès de la SFSIC Création, créativité et médiations, Paris. Récupéré de <https://hal.archives-ouvertes.fr/hal-01817597/document>
- Ferro, E. et Orsella, M. (s. d.). *Eight Business Model Archetypes for PSI Re-Us*. Communication présentée au Paper presented at the “Open Data on the Web” Workshop, Google Campus, London. Récupéré de https://www.w3.org/2013/04/odw/odw13_submission_27.pdf
- Figer, J.-P. (2005). L'influence de l'évolution de la technologie sur les métiers des SSII ou 40 ans de surf sur la loi de Moore. *Entreprises et histoire*, 40(3), 15. doi: [10.3917/eh.040.0015](https://doi.org/10.3917/eh.040.0015)
- Foth, Marcus, Martin Brynskov, et Timo Ojala, éd. 2015. *Citizen's Right to the Digital City: Urban Interfaces, Activism, and Placemaking*. Singapore Heidelberg New York: Springer.
- Gascó-Hernández, M., E. G. Martin, L. Reggi, S. Pyo, et L. F. Luna-Reyes. 2017. « Citizen Co-Production through Open Data: Cases of Citizen Training and Engagement ». Dans *Proceedings of the 18th Annual International Conference on Digital Government Research - Dg.o '17*, 562-63. Staten Island, NY, USA: ACM Press. <https://doi.org/10.1145/3085228.3085252>.
- Goëta, Samuel, et Clément Mabi. 2014. « L'Open Data peut-il (encore) servir les citoyens ? » *Mouvements*, 2014, sect. 3, n°79.
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16(2). Récupéré de <https://firstmonday.org/ojs/index.php/fm/article/download/3316/2764>
- Hess, Charlotte, et Elinor Ostrom, éd. 2007. *Understanding knowledge as a commons: from theory to practice*. Cambridge, Mass: MIT Press.
- Janssen, Marijn, Yannis Charalabidis, et Anneke Zuiderwijk. 2012. « Benefits, Adoption Barriers and Myths of Open Data and Open Government ». *Information Systems Management*, 2012, sect. 29.
- Janssen, M. et Zuiderwijk, A. (2014). Infomediary Business Models for Connecting Open Data Providers and Users. *Social Science Computer Review*, 32(5), 694-711. doi: [10.1177/0894439314525902](https://doi.org/10.1177/0894439314525902)
- Johnson, J. A. (2014). From open data to information justice. *Ethics and Information Technology*, 16(4), 263-274. doi: [10.1007/s10676-014-9351-8](https://doi.org/10.1007/s10676-014-9351-8)
- Johnson, P. et Robinson, P. (2014). Civic Hackathons: Innovation, Procurement, or Civic Engagement?: Civic Hackathon: Procurement or Civic Engagement? *Review of Policy Research*, 31(4), 349-357. doi: [10.1111/ropr.12074](https://doi.org/10.1111/ropr.12074)
- Juanals, Brigitte et Pascual-Espuny, Cécile (2016). La santé environnementale à l'ère des données ouvertes: le cas de la qualité de l'air urbain en PACA. *Colloque international TICEMED, Xe édition*, laboratoires I3M (Universités Nice Sophia Antipolis, Toulon), CREM (Université de Lorraine, Haute-Alsace), IRSIC (Aix-Marseille Université), Oct 2016, Marseille, France, récupéré de. ([halshs-01569942](https://halshs.archives-ouvertes.fr/halshs-01569942))

- Kassen, Maxat. 2018. « Open Data and Its Institutional Ecosystems: A Comparative Cross-Jurisdictional Analysis of Open Data Platforms: OPEN DATA AND ITS INSTITUTIONAL ECOSYSTEMS ». *Canadian Public Administration* 61 (1): 109-29. <https://doi.org/10.1111/capa.12251>.
- Kobayashi, S., Kane, T. B. et Paton, C. (2018). The Privacy and Security Implications of Open Data in Healthcare. *Yearbook of Medical Informatics*, 27(1), 41-47. doi: [10.1055/s-0038-1641201](https://doi.org/10.1055/s-0038-1641201)
- Kubler, Sylvain, Jérémy Robert, Sebastian Neumaier, Jürgen Umbrich, et Yves Le Traon. 2018. « Comparison of Metadata Quality in Open Data Portals Using the Analytic Hierarchy Process ». *Government Information Quarterly* 35 (1): 13-29. <https://doi.org/10.1016/j.giq.2017.11.003>.
- Landry, J.-N. (2019). Urban development. Dans B. Walker, T. Davies et F. Perini (dir.), *The State of Open Data: Histories and Horizons*. (Cape Town and Ottawa: African Minds and International Development Research Centre., p. 226-236). (s. l. : n. é.).
- Marcon, Christian, Grosjean, Sylvie et Mallowan Monica (2018) Introduction : Méthodes et stratégies de gestion de l'information par les organisations : des "big data" aux "thick data", *Revue Cossi*, n°4, en ligne : <https://www.revue-cossi.lerass-eric.com/images/article-pdf/RevueCOSSI-n4-2018.pdf#page=20>
- Magalhaes, G. et Roseira, C. (2017). Open government data and the private sector: An empirical view on business models and value creation. *Government Information Quarterly*. doi: [10.1016/j.giq.2017.08.004](https://doi.org/10.1016/j.giq.2017.08.004)
- March, J. (1991), Exploration and Exploitation in Organizational Learning, *Organization science*, 2(1), 71-87.
- Martin, S., Foulonneau, M. et Turki, S. (2013). 1-5 Stars: Metadata on the Openness Level of Open Data Sets in Europe. Dans E. Garoufallou et J. Greenberg (dir.), *Metadata and Semantics Research* (vol. 390, p. 234-245). Cham : Springer International Publishing. doi: [10.1007/978-3-319-03437-9_24](https://doi.org/10.1007/978-3-319-03437-9_24)
- Meszaros, Branislav, Sitthida Samath, Sonia Guérin-Hamdi, et Céline Faure. 2015. « Livre blanc sur les données ouvertes ».
- Noveck, B. S. (2009). *Wiki government: how technology can make government better, democracy stronger, and citizens more powerful*. États-Unis: Brookings Institution Press.
- Noveck, B. S. (2012). Demand a more open source government [Conférence Webdiffusée]. Technology, Entertainment and Design. Récupéré de https://www.ted.com/talks/beth_noveck_demand_a_more_open_source_government#t-1025995
- Ouellet, M. (2016). *La révolution culturelle du capital: le capitalisme cybernétique dans la société globale de l'information*. Montréal : Éditions Écosociété.
- OCDE (2019) *Rapport sur les données ouvertes publiques : Encourager la maturité des politiques de données ouvertes pour un impact durable*, Paris : Éditions OCDE.
- Paquienséguy, Françoise, Larroche, Valérie, Peyrelong, Marie-France , Vila-Raimondi, Martine, Dymytrova. Valentyna (2016) *Synthèse des résultats de l'enquête auprès de ré-réutilisateurs de données ouvertes : livrable n°1*. [Rapport de recherche] SciencesPo Lyon; Enssib; Lyon3. 2016. hal-01432124

- Peeters, R. et Widlak, A. (2018). The digital cage: Administrative exclusion through information architecture – The case of the Dutch civil registry’s master data management system. *Government Information Quarterly*, 35(2), 175-183. doi: [10.1016/j.giq.2018.02.003](https://doi.org/10.1016/j.giq.2018.02.003)
- Pichot-Damon, Étienne. (2019a) « L’open Data, pas assez sexy pour la smart city ? » Dataactivist, récupéré de <https://medium.com/dataactivist/lopen-data-pas-assez-sexy-pour-la-smart-city-b7e0c5633905>
- Pichot-Damon, Étienne (2019b) « Des analytics à l’impact, pourquoi évaluer une politique open data ? », Dataactivist, récupéré de <https://medium.com/dataactivist/des-analytics-%C3%A0-limpact-pourquoi-%C3%A9valuer-une-politique-open-data-461504a34eca>
- Prido, Laëtitia (2017). « Données ouvertes, gouvernement ouvert : fondements épistémologiques et discours » Mémoire. Montréal, Québec, Université du Québec à Montréal, Maîtrise en communication.
- Proulx, S., Garcia, J. L. et Heaton, L. (dir.). (2014). *La contribution en ligne: pratiques participatives à l’ère du capitalisme informationnel*. Québec (Québec) : Presses de l’Université du Québec.
- Ruijter, Erna, Stephan Grimmelikhuijsen, Jochem van den Berg, et Albert Meijer. 2020. « Open Data Work: Understanding Open Data Usage from a Practice Lens ». *International Review of Administrative Sciences* 86 (1): 3-19. <https://doi.org/10.1177/0020852317753068>.
- Secrétariat du Conseil du trésor. 2019. « Défi A- Redonner les données publiques aux citoyens ». <https://www.tresor.gouv.qc.ca/information-resources/plan-daction-pour-laccessibilite-et-le-partage-des-donnees-ouvertes/defis/defi-a-redonner-les-donnees-publiques-aux-citoyens/?L=2>.
- Silberzahn, Léna, et Mathieu Saujot. 2018. « Stigmatisation ou stactivisme, quelle utilisation de l’ouverture des données dans la ville numérique réelle ? Pourquoi “ouvrir” les données ne suffit pas ». Audacities.
- The Economist*. 2010. « The data deluge », février 2010.
- Tim Berners-Lee. (s. d.). Linked Data - Design Issues. Dans *Linked Data*. Récupéré de <https://www.w3.org/DesignIssues/LinkedData.html>
- Turki, S., Martin, S. et Renault, S. (2017). How open data ecosystems are stimulated? Dans *Proceedings of the International Conference on Electronic Governance and Open Society Challenges in Eurasia - eGose '17* (p. 179-187). St. Petersburg, Russia : ACM Press. doi: [10.1145/3129757.3129787](https://doi.org/10.1145/3129757.3129787)
- Ubaldi, B. 2013. « Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives ». OECD Working Papers on Public Governance 22. Vol. 22. OECD Working Papers on Public Governance. <https://doi.org/10.1787/5k46bj4f03s7-en>.
- Verhulst, Stefaan, et Andrew Young. 2016. « Open Data Impact When Demand and Supply Meet. Key Findings of the Open Datat Impact Case Studies ». <https://www.thegovlab.org/static/files/publications/open-data-impact-key-findings.pdf>.

- Vetrò, Antonio, Lorenzo Canova, Marco Torchiano, Camilo Orozco Minotas, Raimondo Iemma, et Federico Morando. 2016. « Open Data Quality Measurement Framework: Definition and Application to Open Government Data ». *Government Information Quarterly* 33 (2): 325-37. <https://doi.org/10.1016/j.giq.2016.02.001>.
- Ville de Montréal. 2015. « Politique de données ouvertes ». <http://donnees.ville.montreal.qc.ca/portail/wp-content/uploads/2015/10/Politique-de-donn%C3%A9es-ouvertes.pdf>.
- . 2019. « Politique des données numériques, version préliminaire ». http://ville.montreal.qc.ca/pls/portal/docs/page/prt_vdm_fr/media/documents/politique_de_donnees_ville_de_montrereal.pdf.
- Weerakkody, Vishanth, Zahir Irani, Kawal Kapoor, Uthayasankar Sivarajah, et Yogesh K. Dwivedi. 2017. « Open Data and Its Usability: An Empirical View from the Citizen's Perspective ». *Information Systems Frontiers* 19 (2): 285-300. <https://doi.org/10.1007/s10796-016-9679-1>.
- Worthy, Ben. 2015. « The impact of Open Data in the UK : complex, unpredictable and Political ». *Public Administration*, 2015, sect. Vol.93, n°3.
- Zamith, R. (2019). Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism. *Digital Journalism*, 7(4), 470-489. doi: [10.1080/21670811.2018.1554409](https://doi.org/10.1080/21670811.2018.1554409)
- Zuiderwijk, A., Helbig, N., Gil-Garcia, J. R. et Janssen, M. (2014). Special Issue on Innovation through Open Data - A Review of the State-of-the-Art and an Emerging Research Agenda: Guest Editors' Introduction. *Journal of theoretical and applied electronic commerce research*, 9(2), 1-2. doi: [10.4067/S0718-18762014000200001](https://doi.org/10.4067/S0718-18762014000200001)

Annexe 1

Grille d'entrevue pour les usagers inscrits à la liste de diffusion du LIUM

| | Ce que l'on veut savoir | Exemple de questions |
|---|---|---|
| Type d'usagers | <p>Qui ? Usager ordinaire, entreprises, entrepreneurs, développeurs, milieux académiques, organismes, journalistes, autres</p> <p>Localisation</p> | <ul style="list-style-type: none"> - Quelle est votre fonction ? À quel titre utilisez-vous le portail de données ouvertes ? - Quels sont les trois premiers éléments de votre code postal ? |
| Type de demande | <p>Nature et fréquence de la demande 1) les demandes de données ou d'ouverture ; 2) les précisions sur des données spécifiques ; 3) des demandes d'information ; 4) des erreurs dans un jeu de données ; 5) des prises de contact ; 6) les demandes de produits personnalisés ; 7) les demandes qui n'ont rien à voir avec les données ouvertes ; 8) les plaintes (lien brisé, dossier qui ne télécharge pas) ; 9) Autres ?</p> <p>Objet de la demande Le jeu ou les jeux concernés</p> | <ul style="list-style-type: none"> - Quelle demande faites-vous en général ? (voir liste à gauche) - À quelle fréquence ? - Depuis quand ? - Quels jeux de données sont concernés ? - Utilisez-vous d'autres données que celles du portail ? - Comment avez-vous eu connaissance du portail de la ville ? - Quel est le format le plus adéquat pour vous ? |
| Chaîne de traitement des données | <p>Nature du traitement des données réalisée Description du chemin parcouru pour trouver et exploiter la donnée</p> | <ul style="list-style-type: none"> - Pouvez-vous nous décrire le traitement que vous faites des données (téléchargement, nettoyage, vérification, validation etc... ?) |
| Type d'utilisations | <p>Utilisation directe :</p> <ul style="list-style-type: none"> - Nombre et projets liés aux données - Nature du projet | <ul style="list-style-type: none"> - Lorsque vous avez les données demandées, qu'en faites-vous et dans quel but ? - Quels sont les projets qui en découlent, de quelle nature sont-ils ? - Développez-vous des applications, lesquelles ? Comment le projet est-il né ? Quel est l'objectif du projet ? - Vous manque-t-il parfois des données ? |
| | <p>Utilisation indirecte :</p> <ul style="list-style-type: none"> - Utilisation du projet issu des données ouvertes par d'autres personnes que vous (application, rapports de recherche etc) - Type et nature des publics touchés | <ul style="list-style-type: none"> - Quelles sont les personnes qui bénéficient ou qui sont visées par votre projet ? (décideurs politiques, résidents, organismes etc...) - Quel est leur profil ? Comment mesurez-vous les impacts de votre projet ? Êtes-vous en contact avec les usagers potentiels ? |
| Type d'impacts | Nature et nombre d'impacts éventuels | <ul style="list-style-type: none"> - Quels sont selon vous les impacts de votre projet issu des données ouvertes ? |

| | | |
|-----------------------------|--|---|
| | | <ul style="list-style-type: none"> - Exemple : nouveaux usages, utilisation d'une nouvelle technologie, modification des pratiques, amélioration des connaissances, découvertes scientifiques, innovation, réorientation de politiques publiques, révélation, réduction des coûts, création/perte d'emplois, meilleure diffusion de l'information etc. - Y a-t-il selon vous des impacts négatifs ? |
| Valeur d'usage créée | Intérêt, valeur utilité de la donnée pour l'utilisateur ? | <ul style="list-style-type: none"> - Dans quelle mesure le jeu de données téléchargé a-t-il répondu à votre demande ? - Comment qualifieriez-vous votre intérêt en tant qu'utilisateur pour ledit jeu de données ? - Comment qualifieriez-vous l'utilité dudit jeu de données pour vous ? - Vous arrive-t-il de télécharger ou de demander des jeux de données en vous disant que ce n'est pas utile maintenant mais que cela pourra l'être plus tard ? |

