



Les publics de données

Penser la datafication de la société

Sous la direction de

Florence Millerand, Alexandre Coutant,
Guillaume Latzko-Toth et Mélanie Millette



La place centrale des données dans notre monde met en lumière l'urgence de penser de manière critique le phénomène de *datafication* de la société. Répondant à la rareté de la recherche en français à ce propos, cet ouvrage rassemble des contributions québécoises, canadiennes, françaises et australienne avec des études de cas dans plusieurs domaines – des données ouvertes municipales aux applications de santé, en passant par l'information culturelle et l'activisme.

Qu'est-ce qu'un « public de données », et quelles sont les dimensions sociopolitiques de la mise en données, particulièrement en ce qui concerne les inégalités et l'inclusion de groupes minorisés ? Cet ouvrage explique ces enjeux importants en présentant les acteurs, les dispositifs sociotechniques et les collectifs façonnés par le processus de *datafication*.

Florence Millerand est professeure titulaire au Département de communication sociale et publique de l'Université du Québec à Montréal.

Alexandre Coutant est professeur agrégé au Département de communication sociale et publique de l'Université du Québec à Montréal.

Guillaume Latzko-Toth est professeur titulaire au Département d'information et de communication de l'Université Laval.

Mélanie Millette est professeure titulaire au Département de communication sociale et publique de l'Université du Québec à Montréal.

Avec les contributions de Kath Albury, Dominique Boullier, Jean Burgess, Nathalie Casemajor, Alexandre Coutant, Lucie Delias, Jérôme Denis, Maxime Harvey, Sarah Labelle, Guillaume Latzko-Toth, Mary Elizabeth Luka, Sarah Meunier, Anthony McCosker, Florence Millerand, Mélanie Millette, Nadia Seraiocco, Stéphane Vial et Rowan Wilken.

29,95 \$ • 24 €

Couverture : Image créée à l'aide d'outils d'intelligence artificielle.

Formats numériques en accès libre
www.pum.umontreal.ca

ISBN 978-2-7606-5224-8



LES PUBLICS DE DONNÉES

Sous la direction de
Florence Millerand, Alexandre Coutant,
Guillaume Latzko-Toth et Mélanie Millette

LES PUBLICS DE DONNÉES

Penser la *datafication* de la société

Les Presses de l'Université de Montréal

Catalogage avant publication de Bibliothèque et Archives nationales du Québec et Bibliothèque et Archives Canada

Titre: Les publics de données: penser la datafication de la société / sous la direction de Florence Millerand, Alexandre Coutant, Guillaume Latzko-Toth et Mélanie Millette.
Noms: Millerand, Florence, 1970- éditeur intellectuel. | Coutant, Alexandre, éditeur intellectuel. | Latzko-Toth, Guillaume, 1972- éditeur intellectuel. | Millette, Mélanie, 1980- éditeur intellectuel.

Collection: PUM.

Description: Mention de collection: PUM | Comprend des références bibliographiques.

Identifiants: Canadiana (livre imprimé) 20240028953 | Canadiana (livre numérique) 20240028961 | ISBN 9782760652248 | ISBN 9782760652255 (PDF) | ISBN 9782760652262 (EPUB)

Vedettes-matière: RVM: Société de l'information. | RVM: Médias numériques—Aspect social. | RVM: Fossé numérique.

Classification: LCC HM851.P83 2025 | CDD 303.48/33—dc23

Dépôt légal: 2^e trimestre 2025

Bibliothèque et Archives nationales du Québec

© Les Presses de l'Université de Montréal, 2025

Les Presses de l'Université de Montréal remercient de leur soutien financier le Fonds du livre du Canada, le Conseil des arts du Canada et la Société de développement des entreprises culturelles du Québec (SODEC).

Financé par le gouvernement du Canada
Funded by the Government of Canada

| **Canada**



Conseil des Arts
du Canada

Société
de développement
des entreprises
culturelles
Québec

IMPRIMÉ AU CANADA

PRÉFACE

Dominique Boullier

À toutes les époques de quantification de la modernité, il était entendu que les entités humaines que l'on recensait devaient se soumettre, sans rien dire, au savoir des experts qui prélevaient des données sur leur compte, leurs propriétés sociales, leurs préférences individuelles ou encore les traces de leurs comportements. Cette forme d'objectification semblait être une condition pour produire de l'objectivité, et donc de la science, en même temps qu'elle permettait de contrôler et de gouverner les populations, comme l'a si bien montré Alain Desrosières (1993).

Mais les conditions de production numérique des données ont changé avec la troisième ère de quantification, celle des traces et du calcul de tout, qui permet de suivre et quantifier leurs propagations (Boullier, 2023) après les recensements et les sondages d'opinion. La *datafication* a subi l'épreuve politique, technique et culturelle de la diffusion d'un savoir-faire de quantification dans tous les mondes sociaux. Les ressources de calcul et les dispositifs de collecte sont devenus accessibles à un coût marginal faible (Benkler, 2006) et la tendance à la quantification a quitté les sphères des institutions de contrôle, des gestionnaires d'entreprise obsessionnels ou des pirates informatiques compulsifs pour rendre possible la contribution de tous types de collectifs. Ce livre n'est pourtant pas une ode à l'extension infinie du monde du calcul, même s'il ne fait pas l'impasse sur cet effet de mode de l'*open data* (les données ouvertes), devenu un impératif parfois problématique. Il montre plutôt la mise en question de la *datafication* par plusieurs parties prenantes pour restituer aux données leur caractère d'« obtenues ». Dès lors qu'on dénature la donnée et qu'on dévoile la quantité de choix qu'il convient de faire pour l'obtenir, on rend enfin visibles les compétences de collectifs très variés pour produire d'autres données que

celles qu'on attend, pour remettre en cause les classifications et les procédures de collecte et pour rendre discutables ces supposées objectivations qui semblaient tombées du ciel.

Tous les secteurs sont concernés, comme la santé ou la culture, et dès lors, ce sont les statuts de ces parties prenantes qui changent d'aspect. Les entités humaines qui font l'objet de cette quantification sont-elles réductibles aux classes statistiques dans lesquelles on les avait placées jusqu'ici, sans discussion? Quelles facettes ont été occultées? Quelles catégories étaient biaisées? Qui peut le dire et qui peut le vérifier dans le processus même de production de la donnée? Ces quasi-publics que sont les «publics de données» – notion que propose ce livre – deviennent-ils alors plus largement collectifs et agissants? Les liens entre les individus connectés par ces corrélations validées par les concepteurs sont particulièrement ténus: peuvent-ils se rendre manifestes, puis produire une «action commune» (Livet et Thévenot, 1994), alors qu'on a la plupart du temps affaire à une «action à plusieurs» qui se limite à un «état partagé» ou à une facette partagée (tous ceux qui sont restés plus de dix secondes sur la page, tous ceux qui ont *liké* cette publication, tous ceux dont la localisation renvoie à une ville de moins de 15 000 habitants *et* qui commandent plus de trois fois par mois sur Amazon)? Peut-on reconnaître à ces agrégats une agentivité, pertinente dans ce processus, qui fasse perdre la distance que les experts voulaient mettre dans le calcul? L'engagement des *dataactivistes* est considérable – comme l'est celui des «patients» qui, on le découvrira dans ce livre, ignorent souvent qu'ils font l'objet d'un regroupement de données. Ce n'est pas parce qu'ils sont mis en données que ces patients/données deviennent des acteurs sociaux à part entière.

Les multiples cas présentés dans l'ouvrage restituent une forme de trouble dans l'objectivité, mais pour mieux en montrer la fécondité et les avantages pour les parties prenantes. Celles-ci sont à définir au cas par cas, car elles ne sont pas «concernées» au même titre, comme aurait dit Dewey (2003), qui inspire l'esprit de ce livre. Elles sont traditionnellement plutôt oubliées par la seule classe dotée du pouvoir de construire ce monde de fantômes, pour reprendre le «public fantôme» de Lippmann (1925). Cette «classe vectorialiste» (Wark et Degoutin, 2013) se distingue par sa capacité à manipuler le code, à penser des architectures de données et de calcul, à produire et à interpréter ses modèles, ce qui la différencie des citoyens ordinaires qui se débrouillent pour produire du code élémentaire. Les compétences de plus en plus évoluées et opaques pour le commun des mortels des

membres de cette classe vectorialiste poussent vers toujours plus d'autonomie dans leurs choix, leurs décisions, leurs paramétrages, leurs pondérations, au point de désencastrer le calcul de toutes ses conditions sociales de production, une évolution aussi critique que le désencastrement de l'économie au XIX^e siècle, comme l'a montré Polanyi (1944). Restituer la diversité de ces publics concernés, de leurs possibles émergences comme collectifs, et montrer le travail requis et les imaginaires mobilisés pour les faire exister, voilà les contributions majeures de ce livre pour une reprise de contrôle sur ces processus de quantification dès lors qu'ils deviennent *datafication*.

Références

- BENKLER, Yochai, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, Yale University Press, 2006.
- BOULLIER, Dominique, *Propagations. Un nouveau paradigme pour les sciences sociales*, Malakoff, Armand Colin, 2023.
- DESROSIÈRES, Alain, *La politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte, coll. « Textes à l'appui », série « Anthropologie des sciences et des techniques », 19-93.
- DEWEY, John, *Le public et ses problèmes. Œuvres philosophiques, vol. 2*, traduit par Joëlle Zask, Pau, Publ. de l'Université de Pau, Farrago L. Scheer, 2003.
- LIPPMANN, Walter, *The Phantom Public*, New Brunswick, Macmillian Co., 1925.
- LIVET, Pierre, et Laurent THÉVENOT, « Les catégories de l'action collective », *Analyse économique des conventions*, 1994, p. 139-167.
- POLANYI, Karl, *The Great Transformation*, New York, Farrar & Rinehart, 1944.
- WARK, McKenzie, et Christophe DEGOUTIN, « Nouvelles stratégies de la classe vectorialiste », *Multitudes* 54, n° 3, 2013, p. 191-198.

Introduction

*Florence Millerand, Alexandre Coutant,
Guillaume Latzko-Toth et Mélanie Millette*

Associées à la numérisation de la société, les données (*data*) occupent désormais une place prédominante dans toutes les sphères de la vie sociale. Qu'il s'agisse des gouvernements engagés dans des démarches d'ouverture de données publiques, des entreprises qui intègrent la science des données dans leurs stratégies d'affaires ou des internautes qui génèrent, souvent à leur insu, des quantités de traces monnayables par le simple fait d'utiliser le Web, les données sont désormais omniprésentes dans notre quotidien. Pourtant, pour le citoyen ordinaire, elles restent quelque chose d'abstrait. Alors que nous sommes exposés de partout à des discours sur le pouvoir des données et que nos activités sociales en sont imprégnées (Boullier, 2015), ce que *sont* les données, ce qu'elles nous *font* et l'influence qu'elles *ont* sur la société reste relativement obscur. Que produit cette circulation intense des données dans nos sociétés contemporaines ? Comment l'interpréter ?

Il importe d'interroger la place centrale dévolue aux données dans nos sociétés numérisées en réfléchissant sur les *publics* constitués autour d'ensembles (ou de jeux) de données divers et variés. Car il se forme autour des données des groupes qui sont constitués tantôt parce qu'un algorithme a repéré chez leurs membres un ensemble d'intérêts partagés, tantôt parce qu'une organisation ou une institution a anticipé des usages communs.

L'expression « publics de données » met en scène les groupes d'individus qui se trouvent liés en premier lieu par des données. Elle désigne les sujets sociaux qui interagissent ou sont censés interagir avec des données générées *sur eux, par eux* ou *pour eux*. Comme on le verra dans cet ouvrage, les utilisateurs anticipés d'une base de données développée pour le secteur culturel au Canada constituent un public de données au même titre que les usagers

d'un média social comme Facebook qui produisent des quantités de données dans le sillage de leur usage quotidien. De la même façon, les développeurs d'une application de santé mentale font partie des différents publics articulés autour de la production de données sur la santé, tout comme les résidents d'une ville composent l'un des publics considérés par une politique d'ouverture des données municipales. Notre objectif est de réfléchir sur les publics constitués autour de ces initiatives de données, terme que nous utilisons pour désigner des projets sociotechniques de production, de collecte, de traitement et de circulation de données dans des infrastructures numériques.

Quels sont les publics imaginés pour les initiatives de données ? Quels publics se saisissent réellement des données et avec quelles conséquences ? Comment les publics s'intéressent-ils aux données ? Voilà quelques-unes des questions auxquelles cet ouvrage entend apporter des éléments de réponse fondés sur des enquêtes empiriques et des études de cas. Ce faisant, nous proposons une approche *située* des processus de *datafication* en cours qui va au-delà d'une lecture globale pour s'intéresser à des contextes et à des dispositifs sociotechniques précis.

La *datafication* ou la mise en données de la société

La place centrale des données dans nos sociétés renvoie à un phénomène que le néologisme *datafication* cherche à saisir et que l'on peut traduire par « mise en données ». L'expression désigne le rôle de plus en plus grand que la production, la circulation, la collecte et le traitement des données par le truchement des technologies numériques jouent dans la vie quotidienne. La *datafication* fait référence, selon les contextes, à des phénomènes distincts : la transformation de l'activité sociale en données computationnelles, la valorisation marchande de celles-ci ou encore un nouveau mode de gouvernance. Mayer-Schoenberger et Cukier (2014) ont proposé le terme pour décrire la transformation de tous les aspects de l'activité sociale en données quantifiées. Étroitement associée à la vague des données massives (*big data*), la *datafication* renvoie alors aux possibilités de traitement permises par la collecte et l'analyse automatisées de données en continu.

Dans le monde marchand, le terme désigne une nouvelle logique de création de valeur basée sur l'exploitation des contenus et des données numériques générées par leur circulation (Couldry et Turow, 2014). Appliqué à l'échelle des organisations et des institutions, le phénomène est également associé à un mode de gouvernance « cybernétique », une prise de décision

fondée sur les données (Supiot, 2015). Parler de mise en données nécessite donc de s'intéresser aux processus qui contribuent à les produire, à les faire exister et à leur donner un sens. Autant le terme anglais que sa traduction française invitent à les appréhender comme un processus sociotechnique dynamique et non pas comme un état de fait.

Des chantiers d'envergure se fondant sur des logiques de grands ensembles – pensons aux milliards, voire aux milliards de milliards de données collectées par les infrastructures de la « ville intelligente » – à une variété d'initiatives à plus petite échelle – pensons cette fois à des collectifs d'activistes intéressés par des actions locales prônant la production de « contre-données » –, les projets de *datafication* ont en commun la tâche de générer une valeur économique et/ou sociale. Certes, les données existaient avant le numérique; les organisations et les États s'appuyaient déjà sur une série d'informations et de statistiques pour gouverner, avant l'adoption massive de l'Internet (on a longtemps fait les recensements sur papier). Mais le processus actuel de mise en données de la société procède à la fois d'un changement d'échelle et d'un renversement de perspective: *les données viennent désormais en premier*, avant, donc, tout questionnement qui motiverait leur collecte, et le simple fait d'en générer ou d'en publier semble constituer une finalité en soi.

Les données seraient d'ailleurs à l'origine d'une révolution (Kitchin, 2014) qui résulte du croisement de trois phénomènes concomitants ayant émergé autour de 2010: le mouvement d'ouverture des données (*open data*) né dans les administrations publiques, la conception et l'intégration d'une pléthore de dispositifs et d'infrastructures numériques (médias sociaux numériques, Internet des objets, infonuagique, etc.) dans toutes les sphères d'activité, et l'exploitation de données massives grâce à la puissance de calcul aujourd'hui disponible (*big data*). Plus récemment, la pandémie de COVID-19 a braqué les projecteurs sur le rôle crucial des données pour les politiques de santé publique. De même, les dernières innovations en matière d'intelligence artificielle sont devenues possibles grâce à l'accumulation de gigantesques quantités de données pour alimenter et entraîner les algorithmes d'apprentissage machine.

Les discours sur cette mise en données de la société se polarisent autour de deux visions qui opposent, d'un côté, le risque d'une surveillance « panoptique » par les États et le marché, et, de l'autre, de nouvelles manières de produire, de soigner et de vivre en société (Théviot, 2023; Lupton, 2019).

Certains dénoncent l'émergence d'un récit naïf, appelé « dataïsme » (van Dijck, 2014), ancré dans un imaginaire construit sur la puissance des

données, cet « or noir » du XXI^e siècle, et la croyance qu'une quantification généralisée de la société serait apte à retracer et à expliquer tous les types de comportements humains, une fois analysés au moyen de dispositifs numériques (Bowker, 2014). Ce récit milite en faveur d'une société « centrée sur les données », *datacentrique*, où les données sont considérées comme une ressource pour la croissance économique, l'innovation technologique, les percées scientifiques, etc. (van Dijck, 2014) – ce qui ne va pas sans résonner avec l'imaginaire du « progrès ».

L'ouverture des données gouvernementales rendrait possible la création de nouveaux services, généralement sous forme de nouvelles « applis » mobiles, tout en jouant le jeu démocratique et en conférant plus de transparence aux administrations publiques. Mais l'exploitation des données personnelles recueillies en ligne rend également possibles des stratégies publicitaires d'une efficacité redoutable, tout autant que des opérations de surveillance épidémiologique à une échelle inédite. Par ailleurs, les données sont présentées et perçues comme objectives et neutres, alors que la *datafication* tend à reproduire certaines inégalités, par exemple en privilégiant la perspective de groupes historiquement dominants au détriment de celle de groupes minorisés (Luka et Millette, 2018). Qu'ils soient pessimistes ou optimistes, les discours actuels semblent régis par une même *data doxa* (Smith, 2018), en vertu de laquelle la place centrale des données se présente comme un fait accompli indiscutable, ce qui contribue ainsi à sa banalisation.

Parmi les nombreuses questions que suscite cette conception des données comme ressources à exploiter, l'une, de taille, renvoie au fait que les personnes qui les produisent, consciemment ou non, ou à qui l'on pense en concevant des initiatives de données en ont rarement l'accès ou l'usage. En outre, dans la plupart des cas, les personnes et les groupes visés ignorent non seulement qu'ils font partie d'un public « cible » pour les données, mais aussi, le plus souvent, ignorent les critères à partir desquels on les a visés (Pink *et al.*, 2017). Et même lorsque les données sont mises à la disposition du public – pensons aux portails de données ouvertes des gouvernements et des municipalités –, le commun des mortels y perd son latin plus souvent qu'à son tour à essayer de les comprendre et de les interpréter (Ruppert, 2015).

En tant que phénomène d'intérêt pour les sciences humaines et sociales, la *datafication* est théorisée comme une composante inhérente du capitalisme de surveillance (Zuboff, 2018), une nouvelle forme de *data colonialism* dans laquelle les données sont la matière première extraite au détriment des populations concernées (Couldry et Mejias, 2019), mais aussi comme une

occasion de donner du pouvoir aux plus faibles en offrant des possibilités inédites de « surveillance par le bas » (*sousveillance*, Mann et Ferenbok, 2013) et d'activisme (Milan, 2017). En d'autres termes, les données tendent à donner du pouvoir aux acteurs sociaux qui ont les moyens de les rendre significatives et utiles pour eux (Powell, 2014). Pourtant, peu de recherches avant celles que nous venons de citer se sont intéressées aux collectifs, aux publics impliqués dans les processus de *datafication*.

Les publics de données

Public, contre-public (Fraser, 2001), non-public ou presque public (Dayan, 2000), et maintenant « publics de données », cette notion majeure du champ de recherche en communication que sont les *publics* se décline au fil des mutations sociales, politiques et technologiques. Dans la dernière décennie, les scientifiques ont multiplié les propositions pour rendre compte avec précision des transformations des publics dans le contexte des technologies numériques : publics connectés ou en réseau (Boyd, 2011), publics *ad hoc*, constitués autour de mots-clics (Bruns et Burgess, 2015), publics affectifs (Papacharissi, 2014), publics d'imitation (Zulli et Zulli, 2022), ou encore publics « datafiés » (Baack, 2015), pour n'en citer que quelques-unes. Cette profusion de termes est éloquent ; elle signale tout autant la centralité de la notion que l'inconfort conceptuel qui accompagne la réflexion sur ses dynamiques. Des groupes d'utilisateurs regroupés autour d'intérêts communs ou de contenus de niche aux masses rassemblées par les médias sociaux, ces collectifs mettent à mal les catégorisations habituelles. Plus généralement, les nuances apportées par chaque nouvelle expression pour tenter de les circonscrire reflètent l'ambivalence de longue date de la notion de public en sciences sociales, notion tiraillée entre diverses significations allant des audiences passives aux publics politisés (Livingstone, 2005a).

La notion de public a d'abord été théorisée dans ses liens avec les processus démocratiques. Le débat classique entre Lippmann et Dewey dans les années 1930 opposait deux visions du public : d'un côté, le public entendu comme une impossibilité ou, selon les mots de Lippman (2008), comme un « fantôme », terme qui dénote l'impuissance des individus à se faire une opinion devant la complexité du monde ; de l'autre, le public comme réel et nécessaire au bon fonctionnement des sociétés et des démocraties, qui doit prendre conscience de lui-même et politiser ses intérêts pour « sortir de son éclipse », selon l'expression de Dewey (2003). Plusieurs décennies plus tard,

Habermas (1978) proposait le concept de sphère publique et reconnaissait l'existence de publics multiples susceptibles de faire progresser les démocraties, à condition qu'ils s'engagent dans un débat rationnel. Les travaux de Fraser (2011) sur les contre-publics ont constitué un ajout important pour penser les publics oppositionnels, caractérisés par un comportement politique et des normes de prise de parole différentes, mais susceptibles de renforcer la démocratie participative par l'élargissement de l'espace discursif.

Ancré dans la perspective des études sur la communication et les médias, cet ouvrage s'inscrit dans la continuité des débats sur la notion de public, en cherchant plus précisément à interroger sa pertinence pour comprendre les formes actuelles des collectifs interpellés par les processus de mise en données. Les études en communication ont depuis longtemps distingué la notion de public de celle d'audience, cette dernière renvoyant à des construits quantitatifs sophistiqués produits par et pour les industries médiatiques, tandis que la notion de public est associée à l'idée d'un collectif en devenir, qui s'articule et se « publicise » dans le cadre de la réception d'un contenu commun (Cefaï et Pasquier, 2003). Pour autant, la porosité de la frontière séparant les deux notions ne disparaît jamais.

Les chercheuses et chercheurs en communication se sont aussi penchés sur l'articulation des notions de public et d'utilisateur (Assogba *et al.*, 2015). Car interroger ces deux notions conjointement ne va pas de soi : les travaux mobilisant les concepts de public, d'audience se sont longtemps cantonnés dans l'étude des médias et de leur réception, alors que les études sur les technologies d'information et de communication ont émergé dans le cadre d'une sociologie des usages et des utilisateurs qui considérait la notion de public comme peu adaptée à son objet. Toutes ces recherches ont eu tendance à sacraliser la figure de l'utilisateur en reléguant la notion de public aux études de réception médiatique (Morelli, Pignard-Cheynel et Baltazart, 2016). Cependant, si les notions d'utilisateur et de public ont évolué dans des cadres épistémologiques relativement distincts, leurs apports mutuels peuvent s'examiner à partir de deux axes de convergence (Morelli *et al.*, 2016).

Un premier axe renvoie au continuum passif-actif pour envisager le statut du public ou de l'utilisateur face à un contenu ou à un dispositif, autrement dit pour apprécier leur agentivité. Sur ce point, on a observé un mouvement similaire dans les recherches sur les médias et les technologies, recherches ayant évolué d'une conception plutôt passive à la reconnaissance d'une posture active, propice à négocier, voire à contester, l'interprétation d'un contenu par

le public ou la prescription d'un usage par l'utilisateur. Toutefois, la part à accorder aux logiques internes des publics et celle qui revient aux incitations des dispositifs dans leur constitution ne font l'objet d'aucun consensus. Et ce questionnement s'avère toujours aussi pertinent, sinon davantage, dans le contexte des processus de mise en données où des publics sont produits et « calculés » par l'entremise de boîtes noires algorithmiques.

Un second axe de convergence entre les notions de public et d'utilisateur renvoie à l'échelle d'analyse micro/macro et individuel/collectif. Aux conceptions en matière d'agrégats d'individus isolés ou d'individuation des pratiques est venue s'ajouter la notion de communauté comme point d'articulation entre les publics et les utilisateurs pour rendre compte de logiques plus collectives, par exemple avec les notions de communautés d'interprétation et de communautés d'intérêts. Ici encore, ce questionnement est d'autant plus pertinent que les logiques derrière la production de données numériques ont des propensions à la fois massifiantes et individualisantes.

Les publics de données peuvent être des « publics datafiés », puisqu'on peut les représenter sous forme de données, mais ils ne se réduisent pas à ce seul mode d'existence. Møller Hartley, Sørensen et Mathieu (2023, p. 13) proposent l'expression *datapublics*, que l'on pourrait traduire par des publics-données, pour explorer « les influences réciproques que l'un peut avoir sur l'autre sans présumer d'une hiérarchie ou d'une directionnalité » entre les *data*, d'un côté, et les publics, de l'autre. L'expression « publics datafiés » indiquerait à tort que seules les données ont un impact sur les publics, alors qu'un effet s'observe aussi en sens inverse.

La notion de publics de données gagne à être pensée comme un concept heuristique pour analyser des processus de *datafication* à travers le prisme des groupes de personnes qu'ils participent à produire. Ce faisant, on attire l'attention sur les implications sociopolitiques du phénomène de mise en données : quels publics sont ciblés ? Lesquels sont produits, exclus ou habilités, dans quels buts et avec quelles conséquences ?

La valeur heuristique de cette proposition réside essentiellement dans le déplacement du regard qu'elle impose : aborder la *datafication* non pas exclusivement à partir des données, mais en analysant les publics qui s'articulent autour de leur production, leur mise en circulation, leur utilisation ou non-utilisation. Par exemple, si nous nous intéressons aux données publiées par le portail de données ouvertes d'une métropole, nous pouvons nous demander ce que ces données nous apprennent sur les problèmes de logement dans cette ville. Nous pouvons nous interroger sur la manière dont

la question du logement est présentée « en données » et sur la façon dont elle a été traduite dans le cadre de processus de mise en données qu'il s'agit de déconstruire : quels aspects du logement sont documentés ou, au contraire, absents ? Comment les informations sont-elles organisées, étiquetées ? Lesquelles ont été laissées de côté, et avec quelles conséquences ? Et pour quelles catégories d'acteurs ? Enquêter sur les publics de données signifie de s'enquérir à la fois des publics investis dans les rôles de producteurs ou d'utilisateurs des données municipales, des publics qui sont *a priori* ou *de facto* exclus – par exemple pour des raisons d'accessibilité ou d'intelligibilité des jeux de données – et des publics ciblés par l'ouverture pour les confronter à ceux qui se saisissent réellement des données. Ce qui nous permet de comprendre, par exemple, pourquoi l'ouverture de jeux de données sur l'état du logement pourrait bénéficier davantage aux sociétés de développement immobilier qu'aux citoyens « ordinaires » (Millerand *et al.*, 2023).

Interroger les publics de données requiert d'examiner, d'une part, ce que la mise en données « compte », puis « rend public », c'est-à-dire les processus de publicisation impliqués et, d'autre part, en quoi la mise en données « fait public », c'est-à-dire quelles entités collectives sont constituées (ou non) par ces processus. La valeur heuristique de la notion de public apparaît alors dans toute sa richesse (Assogba *et al.*, 2015), d'abord dans sa dimension performative, puisque les choix de conception des dispositifs sont guidés par des représentations des publics – même si celles-ci peuvent se heurter en pratique à des réalités autres.

Il s'agit d'appréhender la question des publics, dans le contexte de processus de *datafication*, par celles des représentations et des discours, d'un côté, et des matérialisations empiriques, de l'autre. Les publics, même imaginaires, pèsent sur l'action. S'interroger sur ces construits, sur leur émergence et leur actualisation suppose de prendre en compte les discours, les pratiques et l'ensemble des médiations sociotechniques qui contribuent à les façonner.

Les contributions rassemblées dans cet ouvrage s'ancrent théoriquement dans le champ de la communication et des études médiatiques, mais elles s'enrichissent également de l'apport de concepts issus du champ des études des sciences et des techniques (*Science and Technology Studies*, STS) et des récentes études critiques sur les données (*Critical Data Studies*). Ainsi, plusieurs textes partagent des considérations théoriques, relevant notamment d'une approche sociotechnique, sur le statut à accorder à la donnée, sur l'étude des publics et des usagers, et sur la prise en compte des dispositifs techniques.

Ensemble structuré de traces retenues pour appréhender un phénomène, les données sont surtout des objets socialement, techniquement et culturellement construits (Gitelman, 2013). De là l'importance d'examiner les ensembles de données dans le cadre des processus, souvent invisibles, ayant mené à leur conception ainsi qu'à leur opérationnalisation (Denis, 2018), et avec eux, les différentes visions produites et véhiculées sur les collectifs représentés. En ce qui concerne l'appréhension des publics, tant les études sur la réception que celles sur la construction sociale des technologies et la sociologie des usages offrent des prises analytiques pour saisir les publics « modèles » (ou implicites) et les publics « empiriques », ainsi que les éventuels écarts entre les deux (Livingstone, 2005b). La mise à jour des publics de données imaginés cherchera ainsi à examiner les publics autorisés ou habilités (ou occultés), tout en prenant en compte ceux qui seront activés en pratique, ainsi que leurs interactions. Il s'agit par ailleurs de s'interroger sur les dispositifs sociotechniques, du point de vue de leur conception et des configurations d'acteurs et d'actions qu'ils incarnent. Ces configurations structurent des relations de pouvoir et matérialisent certaines conceptions de l'organisation sociale.

La notion de publics de données offre un outil pour penser les grandes tensions qui traversent celle de public et pour la réactualiser. Trois tensions fondamentales se trouvent effectivement mises en dialogue dans la notion de publics de données : une première tension entre une conception du public orientée vers le politique et une conception orientée vers le marché ; une deuxième tension entre une réduction des publics à leurs seuls échanges rationnels et une approche attentive à leurs multiples raisons d'attachement ; et une troisième tension entre les publics idéels et empiriques.

Ces axes constituent des pistes analytiques pour comprendre les points de vue parfois contradictoires des acteurs impliqués dans les processus de mise en données, qui peuvent s'appliquer à de multiples publics.

Le concept de publics de données propose un double programme pour la recherche sur le phénomène de la *datafication*. D'une part, il invite à examiner les publics imaginés par les producteurs de données et les concepteurs de services autour de ces données, et par eux, la manière dont certains enjeux s'articulent et se traduisent dans des initiatives de données. D'autre part, il ouvre un nouveau terrain pour remettre en question et attester la manière dont ces enjeux sont (ré)articulés, ignorés ou (ré)appropriés par des acteurs qui cherchent à « faire public » en se constituant en tant que collectifs autour d'ensembles de données. L'étude des publics de données pourrait

ainsi contribuer à décrire et à théoriser les nouvelles formes de relations qui se tissent entre les publics et les données et, finalement, à mieux comprendre le type de société que les processus de mise en données participent à produire ou à reproduire.

Références

- ASSOGBA, Henri *et al.*, « Les publics imaginés et réels des professionnels d'Internet », *Communication*, vol. 33, n° 2, 2015.
- BAACK, Stefan, « Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism », *Big Data & Society*, vol. 2, n° 2, 2015.
- BOULLIER, Dominique, *Sociologie du numérique*, Malakoff, Armand Colin, 2015.
- BOWKER, Geoffrey, « The Theory/Data Thing », *International Journal of Communication*, vol. 8, n° 2043, 2014, p. 1795-1799.
- BOYD, Danah, « Social network sites as networked publics: affordances, dynamics, and implications », dans PAPACHARISSI, Zizi, (dir.), *A Networked Self: Identity, Community, and Culture on Social Network Sites*, New York, Routledge, 2010, p. 39-58.
- BRUNS, Axel et BURGESS, Jean, « Twitter hashtags from ad hoc to calculated publics », dans Nathan RAMBUKANA (dir.), *Hashtag Publics: The Power and Politics of Discursive Networks*, New York, Peter Lang, 2015, p. 13-27.
- CEFAÏ, Daniel et PASQUIER, Dominique, *Les sens du public. Publics politiques, publics médiatiques*, Paris, Presses universitaires de France, 2003.
- COULDRY, Nick et MEJIAS, Ulises A., *The Costs of Connection. How Data Is Colonizing Human Life and Appropriating It for Capitalism*, Stanford, Stanford University Press, 2019.
- COULDRY, Nick et TUROW, Joseph, « Advertising, Big Data, and the Clearance of the Public Realm: Marketers' New Approaches to the Content Subsidy », *International Journal of Communication*, vol. 8, 2014, p. 1710-1726.
- DAYAN, Daniel, « Télévision: le presque-public », *Réseaux*, vol. 2, n° 100, 2000, p. 427-456.
- DENIS, Jérôme, *Le travail invisible des données. Éléments pour une sociologie des infrastructures scripturales*, Paris, Presses des Mines, 2018.
- DEWEY, John, *Le public et ses problèmes. Œuvres philosophiques*, vol. 2, traduit par Joëlle Zask, Pau, Publ. de l'Université de Pau, Farrago L. Scheer, 2003.
- FRASER, Nancy, « Repenser la sphère publique: une contribution à la critique de la démocratie telle qu'elle existe réellement », *Hermès*, vol. 3, n° 31, 2001, p. 125-156.
- , *Qu'est-ce que la justice sociale?*, Paris, La Découverte, 2011.
- GITELMAN, Lisa, « Raw Data » *Is an Oxymoron*, Cambridge, The MIT Press, 2013.
- HABERMAS, Jurgen, *L'espace public. Archéologie de la publicité comme dimension constitutive de la société bourgeoise*, Paris, Payot, 1978.

- KITCHIN, Rob, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Thousand Oaks, Sage Publications, 2014.
- LIPPMANN, Walter, *Le public fantôme*. Sous la direction de Bruno Latour, Paris, Demopolis, 2008.
- LIVINGSTONE, Sonia, « On the relation between audiences and publics », dans LIVINGSTONE, Sonia (dir.), *Audiences and Publics: When Cultural Engagement Matters for the Public Sphere*, Bristol, Intellect Books, 2005a, p. 17-41.
- , « Media audiences, interpreters, users », dans GILLESPIE, Marie (dir.), *Media Audiences*, Maidenhead, Open University Press, 2005b, p. 9-50.
- LUKA, Mary Elizabeth, et MILLETTE, Mélanie, « (Re)framing Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research », *Social Media + Society*, vol. 4, n° 2, 2018, p. 1-10.
- LUPTON, Deborah, *Data Selves: More-than-Human Perspectives*, Cambridge, Polity Press, 2019.
- MANN, Steve et FERENBOK, Joseph, « New Media and the Power Politics of Sousveillance in a Surveillance-Dominated World », *Surveillance & Society*, vol. 11, n° 1/2, 2013, p. 18-34.
- MAYER-SCHÖNBERGER, Viktor et CUKIER, Kenneth, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Boston, Houghton Mifflin Harcourt, 2014.
- MILAN, Stefania, « Data activism as the new frontier of media activism », dans *Media Activism in the Digital Age*, Milton Park, Routledge, 2017.
- MILLERAND, Florence *et al.*, « À la recherche du citoyen "ordinaire", les publics imaginés de l'ouverture des données publiques au niveau municipal », *Les Enjeux de l'information et de la communication*, vol. 23, n° 4, 2023, p. 109-129.
- MØLLER HARTLEY, Jannie, Kirk SØRENSEN, Jannick et MATHIEU, David, *Datapublics: The Construction of Publics in Datafied Democracies*, Bristol, University Press, 2023.
- MORELLI, Pierre, Nathalie PIGNARD-Cheynel et Didier BALTAZART (dir.), *Public et TIC: confrontations conceptuelles et recherches empiriques*, Nancy, Presses universitaires de Nancy, 2016.
- PAPACHARISSI, Zizi, *Affective Publics: Sentiment, Technology, and Politics*, Oxford, Oxford University Press, 2015.
- PINK, Sarah *et al.*, « Mundane Data: The Routines, Contingencies and Accomplishments of Digital Living », *Big Data & Society*, vol. 4, n° 1, 2017.
- POWELL, Alison, « 'Datafication', transparency, and good governance of the data city », dans O'HARA, Kieron, NGUYEN, Carolyn et HAYNES, Peter (dir.), *Digital Enlightenment Yearbook*, ISO Press Ebooks, 2014, p. 215-224.
- RUPPERT, Evelyn, « Doing the Transparent State: Open Government Data as Performance Indicators », dans Richard ROTTENBURG, *et al.* (dir.), *A World of Indicators: The Making of Governmental Knowledge through Quantification*, Cambridge, Cambridge University Press, 2015, p. 127-150.
- SMITH, Gavin J., « Data doxa: The affective consequences of data practices », *Big Data & Society*, vol. 5, n° 1, 2018.

SUPIOT, Alain, *La gouvernance par les nombres*, Paris, Fayard, 2015.

THEVIOT, Anaïs, *Gouverner par les données? Pour une sociologie politique du numérique*, Lyon, ENS, 2023.

VAN DIJCK, José, «Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology», *Surveillance & Society*, vol. 12, n° 2, 2014, p. 197-208.

ZUBOFF, Shoshana, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, New York, PublicAffairs, 2019.

ZULLI, Diana, et David James Zulli, «Extending the Internet meme: Conceptualizing technological mimesis and imitation publics on the TikTok platform», *New Media & Society* 24, n° 8, 2022, p. 1872-1890.

PREMIÈRE PARTIE

L'ouverture

CHAPITRE 1

Les publics « impensés » des données ouvertes

Quand les fournisseurs désignés
sont aussi les usagers

Florence Millerand et Sarah Meunier

Les décideurs politiques fondent de grands espoirs dans les données ouvertes. Cette ouverture laisse entrevoir l'amélioration de la relation entre les administrations et les citoyens, que ce soit par la stimulation de l'activité entrepreneuriale, l'amélioration de la qualité de vie des administrés ou le renforcement des institutions démocratiques (Bernardin et Jeannot, 2019; Kitchin, 2014). La mise à disposition libre et gratuite des données publiques permettrait l'implication plus directe des citoyens dans les processus décisionnels et susciterait ainsi leur collaboration, soit par le développement d'usages « citoyens » des données ouvertes, soit par l'action médiatrice d'acteurs intermédiaires (*startups*, groupes communautaires offrant des services de données ouvertes, activistes, journalisme citoyen, etc.).

Mais les promesses des données ouvertes s'appliquent également aux administrations publiques elles-mêmes, en ce qu'elles favoriseraient une plus grande efficacité des bureaucraties. D'une part, l'ouverture des données répondrait à une injonction de transparence : en donnant à voir leurs activités par les données, les administrations se feraient moins opaques aux yeux de la population (Broomfield et Ruetter, 2022). D'autre part, les possibilités d'analyse des données publiques faciliteraient l'évaluation et la mesure des programmes gouvernementaux et, de là, la possibilité d'une meilleure reddition de comptes auprès des administrés (Chevallier, 2018). À ces attentes

s'ajoute une volonté de décloisonner les services internes des administrations grâce à une circulation plus fluide des données dans le but de limiter les effets du travail en silo (Lehmans, 2017). En bref, les promesses de l'ouverture des données publiques visent l'amélioration de la communication à la fois interne et externe du gouvernement et des administrations publiques.

Ces promesses se traduisent par une prescription à développer une « culture de la donnée » au sein des organisations. Or cette prescription peine à se matérialiser dans les pratiques. C'est l'un des constats issus de notre étude sur le site des données ouvertes de la Ville de Montréal. Alors que nous enquêtons sur les différents publics du site, autant ceux que ciblait la Ville initialement que ceux que nous observions en pratique, nous avons été amenées à nous intéresser aux publics « internes », autrement dit aux différents services de la Ville et à leur relation au portail de données ouvertes.

Du mouvement des données ouvertes au développement d'une « culture de la donnée » dans les administrations

L'ouverture des données publiques s'est accélérée à partir des années 2010 avec la création de portails aux différents niveaux de l'administration publique, y compris les municipalités. La mise en œuvre des politiques d'ouverture s'est accompagnée de changements importants dans les administrations, en particulier du point de vue de la gestion de l'information. Ces changements se sont accentués à compter de 2015 avec l'émergence de la « ville intelligente » et l'instauration d'une nouvelle relation des villes avec leurs données, désormais produites et exploitées pour soutenir la gestion quotidienne (Courmont, 2021).

Un des objectifs les plus saillants de l'ouverture des données publiques est de « faire émerger une culture partagée des données, enjeu de leur appropriation effective par la société civile et dans les entreprises » (Lehmans, 2017, p. 17). L'élément central de cette culture de la donnée renvoie au passage d'une vision axée sur la production et la circulation de l'information à une autre, axée plutôt sur la production et la circulation de la *donnée*. L'acculturation des administrations publiques à la donnée passe par l'évolution des procédures et « routines » organisationnelles vers une recherche d'intégration des dynamiques de coopération et de participation autour des données. Les administrations doivent non seulement développer de nouvelles compétences techniques pour ouvrir leurs données, mais aussi mettre en place « un ensemble de pratiques informationnelles, de représen-

tations techniques et sociales et de savoirs partagés» autour des données (Lehmans, 2017, p. 19).

L'expression elle-même de « culture de la donnée », qui semble s'être imposée dans le monde des organisations, reste polysémique – notamment dans le discours public (Casemajor, cet ouvrage). La culture de la donnée engloberait « les différentes normes culturelles, les systèmes de valeurs et les croyances qui informent, encadrent et justifient les pratiques en matière de production, de traitement, de distribution ou d'utilisation des données, ainsi que les efforts consentis pour régir et façonner des formes particulières de pratiques en matière de données par divers moyens sociaux et techniques » (Bates, 2017, p. 5). La culture de la donnée renverrait donc à la fois à des normes et systèmes de valeurs susceptibles d'orienter les pratiques et à des cadres d'action, par exemple sur le plan de la gouvernance.

Les données sont produites dans des contextes spécifiques. Tant leur régime de propriété que les modalités techniques définissant les formats et conditions de stockage se rattachent directement à leurs contextes d'origine. Toute politique d'ouverture des données suppose donc en premier lieu de les « détacher » de leurs contextes initiaux (Courmont, 2021). Or ce « détachement » n'est pas sans provoquer de possibles frictions. Au sein d'une administration publique, certains services pourraient y voir des contraintes importantes, du fait par exemple de la publicité nouvelle faite à leurs données, des standardisations imposées par le processus d'ouverture, ou encore de la fréquence des mises à jour possibles des données. Ainsi, le déploiement d'une culture de la donnée pourrait forcer des services à revoir l'organisation globale de leurs procédures de gestion des données : redistribution des compétences, redéfinition des pratiques et modes d'organisation, mise en lumière de certains services et acteurs au sein des administrations, etc.

De l'instauration des publics par les dispositifs sociotechniques

Les dispositifs sociotechniques comme un portail de données ouvertes sont toujours conçus en lien avec les types d'usages que leurs concepteurs cherchent à susciter consciemment ou non (Morelli, Pignard-Cheynel et Baltazart, 2016). Leur design matérialise et projette leurs attentes et imaginaires sur les usages futurs du dispositif (Coutant et Vial, cet ouvrage) ; il participe ainsi à la configuration des types d'usages et des identités d'utilisateurs (Woolgar, 1990), par exemple, dans le cas de données publiques municipales, des usagers « citoyens », des entrepreneurs privés, des journalistes ou autres.

Plus largement, les dispositifs prévoient et incitent la formation de publics par un travail de « préconfiguration » (Grignon, 2015, p. 27) en trois dimensions analysables : l'incitation herméneutique à contempler un dispositif de ce genre comme remplissant certaines finalités, la formalisation pragmatique plus ou moins vague d'usages attendus et la précision normative de son bon usage. Certes, en pratique, les publics se conforment souvent en partie seulement à ce qu'on attend d'eux.

Il y a donc fabrication de publics par les pratiques et discours intégrés aux dispositifs sociotechniques, et cela dès leur conception, puis tout au long de leur déploiement et de leur usage. Une manière d'appréhender ces publics consiste à distinguer, d'un côté, les publics imaginés (ceux que souhaitent généralement les concepteurs) et, de l'autre, ceux qui émergent en pratique ; ces deux types de publics ne coïncident pas toujours. Dans le cas des portails de données ouvertes des administrations publiques, et à l'instar des dispositifs participatifs en général, de nombreuses études attestent d'un écart considérable entre les publics anticipés et les publics réels (par exemple Dymytrava et Paquienséguy, 2017 ; Goëta, 2016 ; Ruijter *et al.*, 2020). Ces initiatives restent plus souvent encore à la recherche de leur public « imaginé », à savoir le citoyen ordinaire, qu'elles peinent à faire advenir (Millerand *et al.*, 2023). En outre, et comme l'a bien montré Blondiaux (2003), les écarts par rapport aux publics imaginés par les acteurs politiques existent non seulement avec les publics observables empiriquement sur le terrain, mais avec ceux que s'attendent à rencontrer les chercheurs en sciences sociales à partir de leurs propres cadres d'analyse.

Ces constats incitent à l'analyse des dispositifs sociotechniques et de leurs configurations spécifiques, car celles-ci induisent des manières particulières de « faire public » ; simultanément, les discours et stratégies des acteurs contribuent à les façonner. Il s'agit alors de s'intéresser à la façon dont les acteurs problématisent la question de l'ouverture des données, à leur manière d'envisager les publics censés se saisir – ou non – des données ouvertes (publics souhaités, publics réels, publics inattendus), et aux conséquences ou retombées attendues, comme un meilleur service public, une plus grande efficacité organisationnelle, etc.

L'ouverture des données municipales : une étude de cas

À l'instar d'autres grandes métropoles du Canada, la Ville de Montréal propose un portail de données ouvertes depuis le début des années 2010.

Après plus de dix ans d'existence, force est de conclure que les retombées des réutilisations des données publiées sont mitigées. La réutilisation reste surtout le fait d'usagers professionnels et non pas de citoyens « ordinaires » – comme pour la grande majorité des administrations publiques, d'ailleurs (Diaz et Breux, 2021).

En cherchant à comprendre cet écart, nous nous sommes penchées sur les publics des données ouvertes de la Ville de Montréal tels qu'ils ont été anticipés à l'origine par les concepteurs du portail (les publics imaginés), pour les confronter aux publics effectifs (les publics réels). Concrètement, nous avons analysé la manière dont les publics sont définis, reconnus ou au contraire ignorés dans les discours et pratiques des acteurs de certains services de la Ville, dont celui qui est responsable du portail de données. Nous nous sommes aussi intéressées à l'évolution des conceptions des publics au fur et à mesure que les acteurs ont cherché à les faire advenir en pratique, que ce soit par des initiatives ponctuelles comme l'organisation d'hackathons ou par des changements apportés au portail pour rendre les jeux de données plus attractifs et plus faciles à comprendre (par l'ajout de fonctionnalités de visualisation des jeux de données, par exemple).

Deux événements ont retenu notre attention. D'une part, l'émergence d'un public inattendu d'usagers des données ouvertes, en l'occurrence les services internes à la Ville de Montréal, et d'autre part, le déploiement d'un outil de gestion visant à accélérer l'ouverture des données qui a cristallisé une vision du public interne principalement en tant que fournisseur de données et, par là même, érudé son statut d'utilisateur des données ouvertes. Notre analyse s'appuie sur trois sources : une dizaine d'entretiens menés de mars 2021 à juillet 2022 au sein de différents services de la Ville de Montréal, une analyse documentaire (publications officielles, documentation technique, etc.) couvrant la décennie d'existence du portail (2011 à 2021) et l'analyse du portail lui-même.

L'instauration d'un public de fournisseurs de données

Vers 2020, soit 9 ans après le déploiement du portail de données ouvertes, l'équipe responsable du portail met en place un outil de gestion pour répondre à deux exigences : une demande politique d'accélération de l'ouverture des données municipales, et l'amélioration des processus pour remédier à des difficultés opérationnelles. L'outil de gestion prend la forme d'une matrice des responsabilités qui formalise le processus d'ouverture des données en

assignant des responsabilités à chacun des intervenants. Pour l'équipe, ce nouveau mode d'organisation devra permettre non seulement l'accélération de l'ouverture des données publiques, mais aussi et surtout le développement d'une « culture de la donnée » au sein de la Ville :

Il y a 19 arrondissements qui ont des législations différentes, des façons de faire différentes. Il y a des systèmes corporatifs, il y a des arrondissements qui adhèrent, d'autres qui ont leurs plateformes à part... Donc pour [...] ramener toute l'information dans une base de données centralisée et automatisée, tout ça, ça complique la tâche. (Entretien 2, équipe responsable du portail de données ouvertes)

Au début des années 2010, la politique d'ouverture des données est avant tout pensée pour des publics externes, c'est-à-dire les citoyens. La Ville envisage donc l'ouverture des données au sein de ses services comme une nécessité. Elle attend de ses fonctionnaires qu'ils adhèrent à la politique d'ouverture en rendant disponibles les données de leurs services et arrondissements. Mais des difficultés surgissent dès les premières années de la mise en place du portail, liées à des craintes concernant le caractère sensible de certaines données ou à des ralentissements dans les procédures d'ouverture en raison d'une pénurie de personnel, notamment.

Le déploiement de l'outil de gestion se présente alors comme une solution pour optimiser l'organisation du travail. En tant qu'outil de management stratégique, il permet par exemple d'effectuer un tri efficace des données en amont de leur publication tout en améliorant le contrôle sur celles-ci. Différents rôles sont ainsi attribués aux employés de la Ville afin de faciliter et d'accélérer l'ouverture des données d'après leur contribution attendue. Concrètement, l'équipe responsable du portail se voit attribuer un rôle organisationnel clé dans le processus et acquiert ainsi une plus grande légitimité d'action. Les services et les arrondissements sont présentés comme les « contributeurs de données ». Chacun se dote d'un « coordinateur des données ouvertes » dûment identifié pour servir d'intermédiaire avec l'équipe responsable du portail. À ces coordinateurs s'ajoutent des « responsables techniques ». Les responsabilités légales et éthiques sont prises en charge par le Service du greffe, et la diffusion d'information, par le Service des ressources humaines et des communications.

Cette matrice de responsabilité s'appuie sur un organigramme et des responsabilités partagées entre tous les services municipaux, et renvoie les services internes avant tout à leur fonction de producteurs de données. À ce titre, elle se présente à la fois comme un outil apte à soutenir la décision politique d'accorder la priorité à l'ouverture des données tout en cherchant

à automatiser les processus à plus long terme et ainsi favoriser le développement d'une culture de la donnée.

Le développement d'une culture de la donnée en interne

Avec le déploiement de l'outil de gestion, l'équipe responsable du portail espère que la mise en place de ce dernier contribuera à élaborer une culture de la donnée interne en procédant à un travail de sensibilisation et de démythification autour des données ouvertes. Lorsqu'elle se déroule bien, la collaboration, par exemple entre la personne coordinatrice des données ouvertes dans un service et les analystes de l'équipe responsable du portail, incarne le processus d'ouverture formalisé par l'outil de gestion. Elle participe à l'élaboration progressive d'une culture de la donnée au sein dudit service: « C'est sûr qu'au point de vue organisationnel, l'objectif [est] de rendre les données disponibles en données ouvertes [...]. Il y a eu un temps où c'était peut-être moins clair, la donnée était un petit peu plus gardée au niveau des services pour les besoins internes. On n'avait pas cette culture-là. » (Entretien 6, Comptoir de la géomatique)

L'ouverture des données ne va pas de soi; la formalisation du processus clarifie et facilite le travail des services, en même temps qu'elle les contraint à y adhérer en tant que fournisseurs de données ouvertes « par défaut ». L'outil de gestion instaure ainsi un processus auquel les services sont soumis et qui est légitimé par la demande politique municipale. Cependant, des défis techniques de causes variées perdurent. Par exemple, la diversité des systèmes informatiques et des formats de données au sein de l'appareil municipal entrave régulièrement le travail: « [L'équipe du portail] n'est pas libre nécessairement d'aller faire toutes les extractions qu'ils veulent comme ils veulent, quand ils veulent, dans la multitude de logiciels de la Ville de Montréal parce qu'il y a beaucoup, beaucoup de logiciels, beaucoup d'entrepôts de données. » (Entretien 7, coordination des données ouvertes)

Dans ce contexte, l'équipe responsable du portail est fortement dépendante du Service des technologies de l'information, sans lequel elle ne peut extraire de données. La place de ce dernier devient centrale et les « TI » se constituent alors comme un intermédiaire incontournable de l'ouverture des données en interne: « Les unités [services] n'ont aucune idée où [les données] se trouvent. Et c'est là où on a besoin justement des partenaires TI, parce que c'est eux qui gèrent les données. Puis, nous [...] on n'a pas cette connaissance des systèmes. » (Entretien 3, Service des technologies de l'information)

Par ailleurs, il arrive que la révision des priorités d'ouverture de jeux de données vienne bouleverser l'agenda établi en amont. Les dynamiques de travail demeurent en effet soumises aux programmes politiques et, bien que les élus n'apparaissent pas clairement dans l'organisation du travail formalisée par l'outil de gestion, leurs demandes restent prioritaires afin de répondre à cette volonté de transparence qu'incarne l'ouverture des données. Par exemple, l'ouverture d'un jeu de données répertoriant les cas d'infestation d'insectes selon les arrondissements a reçu priorité à la suite des demandes externes répétées de comités de logements. Dans d'autres cas, il arrive que les données soient ouvertes, c'est-à-dire qu'elles passent par tout le processus d'ouverture, pour n'être finalement accessibles qu'en interne, par exemple lorsqu'elles engagent des responsabilités politiques qui concernent d'autres acteurs que la Ville. « Il y a un souci de sensibilité de la donnée [...]. Tout ce qui est financier, quand ça touche des fonds publics, c'est sensible. » (Entretien 7, coordination des données ouvertes)

D'un côté, en interne, on peut en venir à craindre une volonté citoyenne de demander des comptes à la Ville après la diffusion de données jugées sensibles. De l'autre, la demande politique exige une ouverture rapide de l'équipe responsable du portail et des services internes. On assiste alors à des modes d'action parfois contradictoires qui vont jusqu'à remettre en question le fonctionnement promu par l'outil de gestion et la culture de la donnée qu'il vise à déployer, en l'occurrence un processus d'ouverture très systématisé et balisé, avec des responsabilités ciblées par service et acteur impliqués.

Toutes les données municipales ne sont pas de même nature. Dans l'outil de gestion, celles-ci sont envisagées d'abord et avant tout comme des données quantitatives constituant de grands volumes. Les services qui produisent majoritairement ce type de données sont donc en mesure de répondre à la demande politique d'ouverture plus rapidement et plus complètement que ceux dont les pratiques professionnelles s'incarnent dans d'autres types de fonctionnements. En outre, l'hétérogénéité des données se reflète dans la disparité des logiciels, outils et systèmes informatiques employés par la Ville. Certains services utilisent des bases de données bien structurées convenant à de gros volumes de données standardisées (par exemple les données géomatiques), alors que d'autres fonctionnent sur la base d'outils maison adaptés à la gestion de données hétérogènes, foncièrement qualitatives :

Il y a beaucoup de données qui ne sont pas de la donnée automatisable [...] des rapports de direction, des notes techniques par des conseillers, des réponses à

des demandes médias. [...] Dans d'autres équipes, ça peut être plus axé sur des tableaux de bord. [Chez] nous, il y a plein de choses qui se mettent mal dans des tableaux de bord [...]. Donc ça ne se libère pas très bien en données ouvertes. (Entretien 7, coordination des données ouvertes)

En outre, les fonctionnaires peuvent être amenés à croiser des jeux de données, parfois manuellement, pour répondre à des besoins internes. Ils produisent ainsi des données tout à fait utiles et cohérentes pour leur service, mais dont la fiabilité et le sens peuvent soulever des questions, voire des craintes, dans un autre contexte. Dans d'autres cas, les fonctionnaires doivent mener un important travail sur leurs jeux de données pour les adapter au processus d'ouverture, ce qui peut engendrer des résistances au moment de les diffuser et conduire à des remises en cause ou même à l'abandon de l'ouverture.

Assurer la bonne circulation des données à l'intérieur comme à l'extérieur de l'administration, voilà bien l'un des principaux objectifs de la nouvelle culture de la donnée à développer au sein des administrations publiques (Lehmans, 2017). Mais les manières de faire propres à chacun des services (en fonction des compétences de métiers), le caractère local et situé des données, ainsi que la disparité des formats et systèmes de gestion peuvent se heurter à une logique d'uniformisation des manières de faire circuler les données. Ces frictions conduisent à la fois à interroger l'injonction d'une ouverture des données publiques à tout prix et celle d'une culture unique de la donnée plutôt que de cultures plurielles.

Dans le cas à l'étude, ni la culture de la donnée à développer ni l'outil de gestion qui cherche à l'opérationnaliser au sein de l'administration publique ne se déploient dans le vide. La lourdeur et le ralentissement du processus d'ouverture des données municipales n'incombent pas seulement aux services internes qui « résisteraient » à publier leurs données. Ces difficultés semblent surtout liées à la réorganisation du travail qui vient avec le processus d'ouverture et qui émane de la nouvelle valeur attribuée à la donnée (Lehmans, 2017). En pratique, l'injonction à l'ouverture des données publiques à tout prix au bénéfice d'un public pensé dans les termes génériques de « citoyens ordinaires » détermine (Grignon, 2015) le confinement des services internes dans un rôle unique de producteurs de données, sans égards à leur rôle d'usagers des données ouvertes.

Les services internes : un usager impensé

C'est malheureusement en interne qu'on a plus d'intérêt sur les données. Pourquoi? Parce qu'elles ne sont pas disponibles en interne... Beaucoup voient dans le portail des données ouvertes un moyen d'accéder aux données qu'ils n'ont pas en interne. [Le portail], il n'a pas été conçu pour ça; ce n'est pas normal que j'aie sur le portail des données ouvertes pour accéder à des données auxquelles j'ai droit, ça ne fonctionne pas. D'autant plus que ces données-là, même si on les publie avec le plus de transparence possible, elles sont quand même édulcorées puis parfois obfusquées ou anonymisées. [Or] en tant qu'employé interne, j'ai peut-être besoin d'un peu plus de précision... C'est un vrai enjeu que nos employés voient dans le portail des données ouvertes la principale source de leurs données. (Entretien 3, Service des technologies de l'information)

Les services internes à la Ville sont, dans les faits, à la fois fournisseurs et usagers des données ouvertes. Mais ce public interne, s'il est connu, n'en est pas moins inattendu. Comme l'exprime l'extrait ci-dessus, on déplore l'usage du portail par les services internes. Il s'agit d'un usage imprévu, déviant du « bon » usage du portail. Ainsi, le public interne du portail des données ouvertes de la Ville émerge dans les discours comme un public impensé. Pourtant, ses usages sont clairement reconnaissables, puisque dans les faits, le portail est la principale porte qui donne sur les données municipales en interne.

Les services l'utilisent d'abord quand ils n'ont pas accès aux données d'autres équipes. Régulièrement, des demandes de données faites en interne restent sans réponse, notamment par manque d'effectifs. Les employés se tournent alors vers le portail. Dans d'autres cas, la désuétude des systèmes informatiques pousse les services à privilégier le portail pour accéder plus facilement à leurs propres jeux de données. Ainsi, l'ouverture des données au public externe est venue faciliter de façon inattendue l'accès à ces mêmes données par un public interne.

Cet usage du portail par les services pose la question des types et de la qualité des données dont ils ont besoin en interne. Or, devant l'établissement de priorités politiques dont l'ouverture n'est pensée qu'en fonction des publics externes (la population), cette question n'est pas ou peu discutée. L'équipe responsable du portail n'ignore pas pour autant la question, car elle est consciente des risques d'inadéquation des données publiées sur le portail, en regard des besoins de précision par exemple, pour les services internes. Les jeux de données diffusés sur le portail peuvent en effet s'avérer incomplets pour les traitements nécessaires en interne.

Il arrive que le besoin d'accès aux données en interne accélère des ouvertures et même qu'il génère une occasion favorable de négociation pour l'équipe responsable du portail. En effet, du fait de son rôle pivot, celle-ci est amenée à jouer un rôle clé d'intermédiaire entre les différents services de la Ville, une fonction qui donne parfois lieu à des transactions insolites :

Souvent quand un service demande des données [...], nous, on joue un peu là-dessus : « si vous voulez notre aide pour obtenir cette donnée-là, on va vous aider, mais on le met en données ouvertes. Il n'y a pas d'entre-deux. Si vous voulez obtenir des données qui ne seront pas publiées, vous vous arrangerez vous-mêmes. » (Entretien 5, équipe responsable du portail de données ouvertes)

Pour accéder à certains jeux de données, les services doivent accepter leur rôle de fournisseurs de données ouvertes. Dans ce contexte, le public interne est renvoyé à sa fonction première de producteur de données tandis que l'équipe responsable du portail maintient en quelque sorte son propre public d'utilisateurs, en choisissant de mettre à disposition ou non certains jeux de données en interne avant leur publication. L'existence même de ce public interne n'est pas vraiment discutée et celui-ci n'est pas intégré dans le processus mis en place avec l'outil de gestion.

Un public négligé

L'équipe du portail de données ouvertes a concentré son travail sur le public externe après avoir constaté au fil des ans que la mise à disposition des données ne suffisait pas à les rendre compréhensibles à tout un chacun :

On s'est donné comme mission de voir croître cette culture de la donnée, c'est-à-dire de démocratiser la donnée et de faire en sorte qu'autant la communication que les services y voient un potentiel pour bonifier leur travail, améliorer leurs pratiques. Mais pour faire ce chemin-là, il faut se les approprier, les données, donc il faut avoir les outils nécessaires, comprendre les limites. C'est bien les données, mais... (Entretien, 4, équipe responsable du portail de données ouvertes)

En interne, on observe une contradiction : d'un côté, l'outil de gestion favorise l'émergence d'une culture de la donnée automatisée et centrée sur la production de données ouvertes, où les services internes sont reconnus comme des fournisseurs par défaut ; de l'autre, l'existence des services internes, en tant qu'utilisateurs des données ouvertes, et leurs éventuelles difficultés à consulter ou à réutiliser des données restent largement ignorées. Les services ne reçoivent pas d'accompagnement dans leurs réutilisations des

données ouvertes, ce qui pose problème, car les participants à l'enquête rapportent des compétences très inégales parmi les employés et entre les services.

Certains jeux de données étaient disponibles avant en format PDF sur l'ancien portail Web de la Ville. La différence, c'est qu'ils sont disponibles en CSV maintenant. [L]es gens sont comme dépassés, ils ne savent pas quoi faire avec ça. Pourtant, sur le portail, il y a des trucs interactifs: tu peux créer ton propre tableau, ta propre carte. Mais je ne suis pas sûre que tout le monde est rejoint au niveau de la littératie numérique. (Entretien 7, coordination des données ouvertes)

La littératie de données, c'est-à-dire la capacité à accéder aux données ouvertes, à les comprendre et à les employer pour ses propres besoins, n'est pas problématisée comme un enjeu organisationnel interne, à l'échelle de l'administration. Elle est appréhendée *de facto* comme un problème individuel, qui ne peut donc se régler que sur une base personnelle, comme le révèle des témoignages d'employés qui cherchent à s'autoformer. Cet impensé pose question, d'autant plus que certaines réutilisations de données ne se font que par le portail. Ce n'est donc pas seulement pour produire des données que des compétences numériques sont requises au sein des services, mais aussi pour être en mesure de les lire et de les interpréter.

En considérant que ce public interne du portail possède ou développera par lui-même les habiletés requises pour utiliser les données – puisqu'il les produit –, la diversité des profils et des compétences qui composent ce public reste largement ignorée (Delias, cet ouvrage). Cette vision contraste fortement avec la manière dont sont conçus les publics externes. Ceux-ci sont appréhendés sous la catégorie des « citoyens » et englobent des usagers aussi bien « spécialistes » – journalistes, entreprises, étudiants ou chercheurs, etc. – que profanes (Millerand *et al.*, 2023). Placés au cœur de la politique des données ouvertes depuis la création du portail, les « citoyens » constituent ainsi la première cible de l'ouverture de données municipales, et les villes ont consenti des efforts considérables depuis dix ans pour faire advenir ce public et lui en faciliter l'accès et la réutilisation, par exemple en introduisant des visualisations et des récits de données sur le portail (Millerand *et al.*, 2023).

Alors que les représentations de ces publics externes se sont construites sur l'idée que leurs niveaux de littératie numérique étaient variables, voire inexistants, et qu'il était donc nécessaire d'adapter le portail en conséquence, les publics internes n'ont pas eu la même considération. Or l'absence d'ac-

compagnement des publics internes se répercute sur les services eux-mêmes, pour lesquels la diffusion des données est aussi une question de visibilité : plus un service diffuse ses données, plus celles-ci ont des chances d'être utilisées, ce qui contribue ainsi à mettre en vitrine ou au contraire à invisibiliser les domaines et métiers derrière les données. Les publics internes apparaissent donc doublement invisibilisés : non seulement on ne les reconnaît pas comme des publics d'utilisateurs des données ouvertes, mais on ignore aussi leurs besoins de formation en littératie de données.

En tant qu'outil de management stratégique visant à favoriser le développement d'une culture de la donnée, le dispositif mis en place au sein de la Ville porte en lui un imaginaire qui emprunte à la fois aux données ouvertes – offrir un meilleur service aux citoyens en lien avec la mission de la Ville – et à une *data doxa* (Smith, 2018) où les données sont censées contribuer en elles-mêmes à améliorer les pratiques de travail, dans une logique de circulation et de valorisation de la donnée publique. Cependant, les frictions suscitées par l'émergence d'un public interne, préfiguré comme fournisseur de données (Grignon, 2015), mais qui apparaît *de facto* également comme utilisateur des données ouvertes, montrent les limites d'une politique tournée exclusivement vers les besoins des publics externes.

Reconnaître les services internes de la Ville en tant qu'utilisateurs des données ouvertes au même titre que les publics anticipés – citoyens, journalistes, entreprises, etc. – nécessite que l'on réinterroge la culture de la donnée et les dispositifs par lesquels elle s'incarne, de même que les formes d'organisation de l'administration publique municipale. Si chaque innovation sociotechnique suscite son lot d'attentes en matière d'amélioration des processus organisationnels, la donnée joue ici un rôle central de médiation dans ces attentes. La culture de la donnée répondrait non seulement à l'injonction d'ouverture des données publiques, mais aussi à l'amélioration du fonctionnement des administrations.

Par ailleurs, notre étude vient nuancer la critique selon laquelle les organisations se bureaucratisent à mesure qu'elles se technicisent, notamment par la mise en place de dispositifs et de processus d'ouverture de données. Contrairement à l'idée selon laquelle les services de la Ville se concentreraient surtout sur leur propre fonctionnement interne au détriment de leur mission de servir le public, la culture de la donnée qui se développe au sein de l'appareil

municipal vise d'abord et avant tout le service aux citoyens, jusqu'à en oublier ses propres publics internes. Faire croître une culture de la donnée dans les administrations publiques suppose de reconnaître la pluralité et l'hétérogénéité des services et des données elles-mêmes.

La Ville de Montréal a concentré ses efforts sur ses publics externes, allant jusqu'à formaliser des principes directeurs dans une Charte des données ouvertes qui s'adresse à toutes sortes de publics, y compris des publics minorisés, et qui reconnaisse l'inégale répartition des compétences numériques parmi eux (Delias, cet ouvrage). Le public interne s'est révélé être un utilisateur inattendu des données ouvertes pour lequel les dispositifs et principes élaborés en amont pour le citoyen « ordinaire » semblent peu transposables et mal adaptés. Des réflexions sur ces questions ont déjà cours, pourtant. Elles ne demandent qu'à être mises en œuvre.

Références

- BATES, JO, « Data cultures, power and the city », dans Rob KITCHIN, Tracey P. LAURIAULT, et Gavin MCARDLE (dir.) *Data and the City*, Londres, Routledge, 2017, p. 189-200.
- BERNARDIN, Stève et JEANNOT, Gilles, « La ville intelligente sans les villes? Interopérabilité, ouvertures et maîtrise des données publiques au sein des administrations municipales », *Réseaux*, vol. 218, n° 6, 2019, p. 9-37.
- BLONDIAUX, Loïc, « Publics imaginés et publics réels. La sollicitation des habitants dans une expérience de démocratie locale », dans Daniel CEFAÏ et Dominique PASQUIER (dir.), *Les Sens du public. Publics politiques, publics médiatiques*, Paris, Presses universitaires de France, 2003, p. 313-326.
- BROOMFIELD, Heather et Lisa REUTTER, « In search of the citizen in the datafication of public administration », *Big Data & Society*, vol. 9, n° 1, 2022.
- CHEVALLIER, Jacques, « Vers l'État-plateforme? », *Revue française d'administration publique*, vol. 167, n° 3, 2018, p. 627-637.
- COURMONT, Antoine, *Quand la donnée arrive en ville. Open data et gouvernance urbaine*, Fontaine, Presses universitaires de Grenoble, 2021.
- DIÁZ, Jérémy et BREUX, Sandra, « Les réutilisateurs et les réutilisatrices des données publiques ouvertes. Le cas de Montréal », *Institut national de la recherche scientifique*, Centre Urbanisation Culture Société, 2021.
- DYMYTROVA, Valentyna et PAQUIENSÉGUY, Françoise, « La réutilisation et les réutilisateurs des données ouvertes en France: une approche centrée sur les usagers », *Revue internationale des gouvernements ouverts*, vol. 5, 2017, p. 117-132.
- GOËTA, Samuel, « Instaurer des données, instaurer des publics: une enquête sociologique dans les coulisses de l'open data », thèse de doctorat en sociologie, Télécom Paris Tech, 2016.

- GRIGNON, Thomas, « L'expertise communicationnelle au prisme de ses instruments », *Communication & professionnalisation*, n° 3, 2015, p. 23-47.
- KITCHIN, Rob, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Thousand Oaks, Sage Publications, 2014.
- LEHMANS, Anne, « Données ouvertes et redéfinition de la culture de l'information dans les organisations », *Communication et organisation. Revue scientifique francophone en Communication organisationnelle*, n° 51, 2017, p. 15-26.
- MILLERAND, Florence *et al.*, « À la recherche du citoyen "ordinaire", les publics imaginés de l'ouverture des données publiques au niveau municipal », *Les Enjeux de l'information et de la communication*, vol. 23, n° 4, 2023, p. 109-129.
- MORELLI, Pierre, PIGNARD-CHEYNEL, Nathalie et BALTAZART, Didier (dir.), *Public et TIC: confrontations conceptuelles et recherches empiriques*, Nancy, Presses universitaires de Nancy, 2016.
- RUIJER, Erna *et al.*, « Open data work: Understanding open data usage from a practice lens », *International Review of Administrative Sciences*, vol. 86, n° 1, 2020, p. 3-19.
- SMITH, Gavin JD, « Data doxa: The affective consequences of data practices », *Big Data & Society*, vol. 5, n° 1, 2018.
- WOOLGAR, Steve, « Configuring the user: the case of usability trials », *The Sociological Review*, vol. 38, n° S1, 1990, p. 58-99.

CHAPITRE 2

Les figures du public dans les politiques d'ouverture des données

Sarah Labelle

C'est à partir de 2009 que les premières collectivités mettent en œuvre des actions pour publier des données issues de leurs services publics sur des sites Web dédiés, alors appelés « portails *open data* », désignés par le terme de « plateforme » aujourd'hui, et dont l'adresse URL comporte la plupart du temps le mot « data ». La production de cet instrument s'inscrit dans la tendance à l'uniformisation des modes d'action publique : la création de « portails » ou de « plateformes » constitue des médiations documentaires dont l'étude systématique des choix éditoriaux montre la grande similarité (Labelle et Le Corf, 2012).

Si les politiques publiques des données ont déjà donné lieu à des études sur leurs promesses démocratiques (Lehmans, 2018) et sur le potentiel d'émancipation par les données (Goëta et Mabi, 2014), l'appréhension des publics de données s'effectue par le profil de concepteurs et l'attente d'innovation associée (Le Corf, 2016). Les politiques des données sont conçues avec l'objectif de *re-définir* le cadre démocratique de la relation entre l'État et ses administrations, ses administrés, ses habitants, ses citoyens et ses acteurs économiques. Cela se traduit par la mise en œuvre de logiques d'accès aux informations et données produites par l'État dans le cadre de services publics et de services statistiques. Une telle publicisation a pour objectif affiché que ces données soient l'objet d'un traitement désigné le plus souvent par le terme de « réutilisation ». Ces données sont mises à disposition avec l'idée qu'il existe des publics qui souhaitent les mobiliser, que ce soit pour participer au débat public ou pour proposer un service.

De quelle manière les collectifs de l'ouverture, qu'ils soient des acteurs publics ou privés, envisagent-ils les publics des activités et dispositifs qu'ils produisent ? Quelles sont les figures du public que ces collectifs mobilisent et représentent ? Notre orientation de recherche s'inscrit dans une approche communicationnelle des conceptions des publics à partir d'une réflexion sur le politique. À partir des nombreuses médiations observées durant une enquête de terrain, nous avons pu distinguer plusieurs figures du public présentées et *re-présentées* par les collectifs.

L'enquête sur laquelle se fonde notre réflexion s'est déroulée de 2012 à 2018 auprès de deux associations d'envergure nationale, la Fondation Internet Nouvelle Génération et Open Data France, et d'une administration rattachée au service du premier ministre, Étalab. Durant cette période, nous avons intégré les équipes de ces organisations et participé à leurs activités. Nous avons désigné cette forme d'implication par le fait d'être une « chercheuse embarquée » (Labelle, 2020). Ce syntagme présente l'avantage non seulement de renvoyer à une pratique journalistique d'observation au plus près des personnes et des lieux, mais aussi de faire écho aux pratiques langagières des collectifs observés et de s'y inscrire facilement. Concrètement, cette pratique d'enquête nous offre l'occasion d'observer en situation des dispositifs et des pratiques (discursives ou extradiscursives) et de percevoir les normes et valeurs que mobilisent et conçoivent ceux qui jouent un rôle dans les politiques des données.

Cette démarche se situe dans une pratique de l'enquête qui vise à observer des actions situées. Elle encourage l'adoption d'un point de vue sensible : si l'enjeu n'a jamais été de démystifier ou de révéler certaines postures ou certaines pratiques, l'objectif était de décrire, de manière presque intime, un monde de l'intérieur qu'on peut, à certains égards, qualifier de tout petit monde, le tout petit monde des politiques d'ouverture des données.

Pour prendre pleinement en considération la dimension politique des réalités sociales qui nous intéressent, nous avons porté notre attention sur les processus et les médiations à l'œuvre qui relèvent de l'ouverture ou sont qualifiés comme lui appartenant. Nous ne les abordons pas à partir des modèles stratégiques ou des explications rationnelles des sujets impliqués, mais nous observons plutôt des « composites », c'est-à-dire des « unités socio-communicationnelles bornées et objectivables » (Le Marec, 2002, p. 57).

Ces composites sont des formes, des objets, des situations dont les collectifs observés participent à la production, à la valuation et à la circulation. Ces unités sont appréhendées comme un tout et non comme des fragments.

Ainsi, dans chaque situation, on observe la présence de processus et de médiations que les acteurs activent, mettent en œuvre ou prennent en compte, et qui contribuent à mettre en avant leur conception de l'ouverture. L'enjeu d'observer ces unités est d'étudier les conditions dans lesquelles les collectifs s'inscrivent dans leur environnement et entrent en relation avec des objets qu'ils fabriquent et manipulent. Nous analysons donc des discours et des dispositifs en tant qu'ils sont liés à des pratiques, à des logiques d'action et d'intervention et à des processus. Les six ans passés à être chercheuse embarquée nous ont permis de « [nous] situer le plus longtemps possible dans la zone de confrontation entre la réaction de l'observation au questionnement et la réaction du questionnement à l'observation » (Le Marec, 2002, p. 57) et ainsi de pouvoir poser une réflexion située sur les publics des politiques des données.

Dans les trois espaces d'enquête observés, il apparaît très vite que les collectifs revendiquent la capacité à définir un public des données, des politiques associées et comment ce public est non seulement associé à un certain type d'activité, mais aussi reconnu par ce qu'il produit avec les données. Ainsi, les deux associations et l'administration rattachée au service du premier ministre s'autorisent à définir leurs publics, tout comme le font les institutions culturelles où Le Marec observe le fait que le public des institutions n'est pas un ensemble préexistant, et que dans ces conditions, ce sont les institutions elles-mêmes qui travaillent à produire une définition lisible de leurs publics (Le Marec, 2007).

Nous cherchons à comprendre la manière dont le public est constitué en valeur, c'est-à-dire un fait établi en fonction des attentes et des intérêts (Dewey, 2011), pour les politiques de données. Pour ce faire, nous analysons différentes figures du public. Ces figures renvoient aux différentes conceptions du public révélées lors des enquêtes, dans les manières qu'ont les collectifs de désigner les activités et les modes d'implication des individus ou des groupes décrits et présentés comme jouant un rôle dans les politiques des données et conférant du sens aux activités et aux dispositifs produits par les collectifs. Nous avons relevé et qualifié trois figures dans l'optique de mettre en lumière le travail de valuation qu'effectuent les collectifs de l'ouverture, et notamment la désignation des publics en fonction de leur capacité à consulter et à traiter les données mises à disposition.

Les différences de conception sont perceptibles dans les manières de décrire les activités des publics et dans la nature du rapport symbolique à l'institution que les collectifs projettent sur les publics. Quels liens et quelles

attentes les collectifs pensent-ils que les publics manifestent ? Nous n'étudions donc pas à proprement parler les publics, ni leurs aspirations, ni même leurs activités, mais la manière dont ceux qui fabriquent les politiques publiques des données les pensent et les qualifient (Millerand et Meunier, cet ouvrage). Le besoin de désigner les publics est fort et conduit à proposer des dispositifs de nature hétérogène : il s'agit de faire advenir les publics qui sont constitués en « réserve ». La problématique peut alors se formuler en ces termes : comment les politiques publiques d'ouverture conçoivent-elles leurs publics ? Dans quelle mesure les politiques de données estiment-elles indispensable de produire et d'identifier des publics à même de pouvoir agir avec les données ? Comment ces publics sont-ils pensés pour garantir la qualité et la puissance des politiques de données ?

Le travail d'enquête a mis en évidence le fait que les publics sont une « réserve », c'est-à-dire un ensemble pluriel assemblé pour témoigner de la vitalité présente et future des politiques des données et leur conférer un pouvoir. À la suite de Louis Marin, Yves Jeanneret rappelle que le pouvoir est virtuel, autrement dit, qu'il constitue la possibilité d'agir sur un domaine (Jeanneret, 2019). L'enquête montre que les publics sont ainsi rendus présents par les discours et les dispositifs des collectifs : ils sont présentifiés et mis en « réserve » afin de renforcer la pertinence et la nécessité des politiques des données. Notre analyse pointe tout d'abord la manière dont les publics sont constitués en valeur par les collectifs de l'ouverture. Elle propose ensuite une qualification des différentes figures de publics qui mettent en exergue la dimension indispensable des publics en tant que « réserve » pour activer les politiques des données.

Le public comme valeur

La manière dont les publics sont pensés et travaillés dans l'espace public libéral est une question qu'Habermas (1993) et Fraser (2003) ont déjà traitée. Quant à la question des publics des politiques publiques, elle connaît un renouvellement depuis plusieurs années, notamment en sociologie de l'action publique. Gourgues et Mazeaud (2018) promeuvent une approche des publics qui vise en particulier à ne pas les penser uniquement en tant que destinataires des politiques publiques, mais aussi en tant que participants à la constitution de ces politiques. C'est ce sens qui informe la présente analyse, qui veut comprendre comment les publics sont une « réserve politique » des politiques des données. De prime abord, il apparaît évident que ces politiques revendiquent

leur inscription dans une perspective démocratique avec l'objectif de s'adresser à tout le monde et de fournir à des publics des dispositifs qui leur permettent d'agir. Ces publics sont alors une réserve dans le sens où ils sont considérés comme pouvant se mobiliser à tout moment.

Notre enquête nous a conduite à observer des situations extrêmement hétérogènes durant lesquelles les collectifs des politiques de données constituent les publics comme pouvant agir avec des données. Dans certains événements, les publics peuvent se trouver en situation de spectateurs à qui les politiques de données, et surtout les dispositifs produits, sont présentés et expliqués. Dans d'autres, ils sont placés en situation de participants, comme lors de la création d'une frise de neuf mètres sur les prénoms donnés à Rennes en 2012-2013. Cet atelier animé par Simon Chignard, alors conseiller stratégique de l'administration Étalab et consultant sur la campagne Infolab, visait à rendre visible « la longue traîne des prénoms » par la visualisation des données. D'autres dispositifs font appel au travail des publics comme la consultation mise en place dans le cadre de la loi pour une République numérique et qui invitait toute personne ou toute organisation à « co-écrire » la loi, à la commenter et à voter pour ou contre ses articles ou des amendements proposés (Alexis *et al.*, 2016). Ces différents ensembles sont désignés comme des publics par ceux qui produisent les médiations dispositives dans toute leur diversité. Cette hétérogénéité des situations souligne combien il est difficile d'étudier les publics et combien la qualification d'un groupe, constitué sur un temps bref par une activité plus ou moins précise, n'est pas aisée. Tout comme dans le champ culturel, la manière d'appréhender les publics reste complexe et suppose de dépasser quelque peu les cadres théoriques qu'on pourrait leur imposer.

Le public est considéré ici comme une valeur, au sens où Dewey l'entend, c'est-à-dire un fait désirable et semblable à ce qui s'inscrit dans l'action, avec une analyse des conséquences et des intérêts. Pour Dewey, « une valuation n'a lieu que lorsque quelque chose fait question : quand il y a des difficultés à écarter, un besoin, un manque ou une privation à combler, un conflit entre tendances à résoudre en changeant les conditions existantes » (Dewey, 2010, p. 120). Dans le cas des politiques publiques des données, le travail de valuation s'effectue par les collectifs parce qu'ils n'ont de cesse de se questionner sur ceux à qui ils s'adressent, d'une part, et qu'ils cherchent à modifier par leurs actions les modes d'existence et d'organisation de l'État, d'autre part.

Partant de cette proposition, notre démarche d'analyse vise donc à considérer l'importance, voire l'urgence associée à l'identification de publics

et à la production de dispositifs qui seraient attendus et qui répondent à des besoins de transformer l'action avec les données. S'il apparaît assez nettement que les politiques publiques des données ont constitué l'impératif de publics comme un horizon de leurs actions, cela se traduit par la multiplication des médiations et par l'hétérogénéité des dispositifs créés. S'observe alors la mise en œuvre d'une panoplie de médiations (Labelle, 2007), c'est-à-dire un ensemble de configurations matérielles disponibles et mobilisables qui s'ajustent en fonction des contextes et qui garantissent une opérativité sociale et symbolique aux politiques auxquelles elles participent. La production de dispositifs correspond à la manière d'agir des collectifs en fonction de leur perception et de leur attention à ce qui renvoie à des formes et des activités anticipées dans les politiques publiques des données. Cette panoplie de médiations correspond à la fois à l'actualisation d'un idéal politique et à la concrétisation d'un cadre d'action. Elle rend visible et active des conditions et des processus pour participer à un faire social qui convoque des publics.

En outre, ces nouveaux publics à produire correspondent à l'ambition de renouveau des formes de participation démocratique. Il s'agit de mettre en avant une forme idéale de la démocratie, celle qui « rend des comptes » comme le rappellent les membres de la mission Étalab lors de leurs interventions dans des événements, celle qui offre un cadre de participation active. Cette ambition favorise l'émergence d'un discours sur les publics, sur ses attentes et ses besoins. C'est un élément constant dans l'histoire de la démocratie que de chercher à mobiliser les publics. Il s'agit de pouvoir reconnaître le rôle du peuple, d'insister sur son implication dans les processus politiques.

Ce désir de participation est mis en avant dans les discours par l'attente de données par différents publics. Il n'y a qu'à entendre Tim Berners-Lee faisant scander « des données brutes maintenant!¹ » à l'assistance d'une conférence TED². Il s'agit bien de rendre audible et partagé le désir de données et ainsi, d'affirmer l'existence de publics larges et hétérogènes prêts à s'en saisir et à les traiter. Il est aussi possible de reconnaître cette valuation des publics dans le préambule de la charte du G8 pour l'ouverture des données publiques :

1. Traduction libre de *raw data now*.

2. La conférence donnée le 9 février 2009 à Long Beach vise à proposer à l'assistance la définition d'un nouveau cadre d'action qui s'appuie sur la transformation du Web du document en un Web des données.

S'ils recueillent une multitude de données, les gouvernements et les entreprises ne les partagent pas toujours selon des modalités permettant au public de les découvrir, de les utiliser et de les comprendre facilement. [...] C'est là une occasion manquée. [...] Les citoyens exigent aujourd'hui de pouvoir accéder à l'information et aux services par voie électronique³.

Deux motifs des discours sur l'ouverture sont présents dans ces quelques lignes : la nécessité de créer des dispositifs d'accès aux données, ce qui constitue la panoplie, et le désir de données des publics.

Pour explorer la manière dont les publics sont constitués comme une catégorie spécifique de l'action, voire comme instrument de légitimation de leur action, on peut distinguer trois figures de publics mobilisées par les collectifs de l'ouverture, à savoir le public spontané, le public désigné et le public désiré.

Le public spontané

Première figure de ce public, c'est un public spontané, qu'il serait peut-être préférable de qualifier de public postulé. Cependant, l'étymologie du terme spontané est intéressante à explorer. Cet adjectif vient du verbe latin *spondere* qui signifie « engager » ou « s'engager ». L'adjectif peut désigner un mouvement qui n'est pas stimulé, tout comme une action qui échappe à un cadre établi. Cette figure du public spontané sert à désigner une présence identifiable, mais indiscernable. L'existence de ce public est affirmée par les collectifs des politiques publiques qui cherchent sans cesse à souligner une activité qui serait à la fois extrêmement prégnante et identifiable dans le social, et extrêmement insaisissable, notamment parce que cette activité leur échappe. La valeur qu'ils attribuent à ce public est double : il s'agit, d'une part, de conférer du sens aux dispositifs par l'existence de destinataires, et, d'autre part, de faire acquérir une existence aux données par l'usage (leur consultation, leur traitement).

Cette figure du public émerge de deux manières au sein de notre enquête de terrain. D'un côté, il s'agit de personnes ou de collectifs qui voient dans les données des moyens d'agir. Ils s'organisent seuls, sans l'aide des institutions, et contribuent par leur engagement à inscrire les politiques des données dans une dynamique militante et sociale. De l'autre côté, les collectifs des données désignent une forme sociale, la « multitude » qui correspond à un ensemble hétérogène et reposant sur une absence d'organisation.

3. G8, *Charte du G8 pour l'ouverture des données publiques*, G8 UK, 2013.

Les collectifs de l'ouverture reconnaissent un premier public, que l'on peut qualifier de spontané, constitué de personnes aux profils militants et engagés qui créent des associations pour agir sur l'accès aux données publiques. Par exemple, à Nantes, autour de 2008-2009, Claire Gallon cherche d'abord des informations pour nourrir son activité militante écologiste. Elle découvre que certaines données sont accessibles, d'autres non, et commence alors à s'intéresser à la question de leur ouverture. Pour se documenter, elle lit en anglais, en espagnol et en français tout ce qui s'écrit sur le sujet. Elle s'approprie le vocabulaire et des savoir-faire techniques et cerne les enjeux politiques locaux et nationaux. Ce faisant, elle constitue les données comme des objets politiques qui offrent l'occasion de se réapproprier le débat public autour de sujets de société. Elle acquiert dès lors une connaissance fine qu'elle met au service de l'association Libertic, créée en 2009. Cette dernière conduit la mairie de Nantes et son maire Jean-Marc Ayrault à concevoir un portail d'accès aux données publiques de la Ville en 2011. Le second exemple est issu de l'association Regards citoyens, qui a créé et mis en place le site nosdeputes.fr (puis nossenateurs.fr) pour suivre quotidiennement l'activité des députés (et des sénateurs) en récupérant les données sur le site de l'Assemblée nationale (et du Sénat). Le site met en lumière l'activité de chaque élu de la Nation par des représentations des données : les jours de présence, le nombre de questions posées au gouvernement, de projets de loi présentés, etc.

Ces deux associations n'ont pas les mêmes objectifs, mais partagent des principes qu'elles définissent en ces termes : les données appartiennent à tous, et il est indispensable de pouvoir s'en servir. Dans cette optique, elles revendiquent la continuité de la pensée des *pirates* qui voient dans l'informatique et le Web une augmentation de la capacité d'agir. Ces collectifs se posent en « chiens de garde », c'est-à-dire en observateurs vigilants de la vie publique. La dimension spontanée de ce public tient au fait qu'il a la capacité de s'organiser seul, de déployer des dispositifs qui acquièrent une certaine pertinence dans l'espace public et d'être reconnu pour son expertise.

Le deuxième type de public spontané est, quant à lui, un public reconnu par les collectifs de l'ouverture comme appartenant à la population et donc comme large. Il est désigné par le mot « multitude ». L'équipe d'Étalab le mobilise constamment en réunion, en référence à la proposition d'Henri Verdier dans son livre, co-écrit avec Nicolas Colin, intitulé *L'âge de la multitude* (Colin et Verdier, 2015). Il s'agit de mettre en avant un public qui n'est pas organisé, qui n'a pas besoin d'être stimulé et qui ne s'organise pas sur la

base de l'appartenance à un groupe, mais qui participe à une activité souvent suscitée par des dispositifs informatisés. La multitude utilise des dispositifs peu contraignants du point de vue des savoir-faire techniques ou de l'engagement social nécessaires. Par exemple, les contributeurs de critiques sur des dispositifs culturels comme Sens critique ou les participants à une discussion publique médiatisée qui usent d'un mot-clic précédé d'un # au sein d'une « industrie médiatisante » (Jeanneret, 2014), c'est-à-dire un réseau social, sont constitutifs de cette multitude.

Le terme de multitude est fréquemment employé par les collectifs étudiés comme un clin d'œil à la figure tutélaire d'Henri Verdier, ambassadeur du numérique pour la France depuis 2018. Il permet de reconnaître le rôle déterminant de ce dernier dans la définition des politiques publiques de l'*open data* en France. Dans le livre *L'âge de la multitude*, la multitude est décrite à partir du lien établi entre des pratiques extrêmement hétérogènes de participation active à la circulation de formes culturelles et sociales. Cette uniformisation permet à l'auteur de constituer un ensemble de pratiques comme relevant de la même activité et d'essentialiser le public qui met en œuvre ces pratiques. La multitude se définit non seulement par l'idée de masse, de grand nombre, mais aussi par l'égalité des intelligences. La référence à la multitude offre ainsi la possibilité de souligner que l'État n'a pas besoin de stimuler les gens pour qu'ils participent à la politique des données, parce qu'ils sont déjà en train d'agir, quand bien même ils agissent ailleurs.

Le mot multitude intervient lorsque se pose de façon récurrente la question sur la manière dont les projets mis en œuvre par les membres de l'administration seront reçus et repris en dehors de l'administration. Il s'agit pour les membres de l'équipe de ne pas présumer des usages possibles, mais de prétendre que leurs dispositifs acquerront un usage social immédiatement. Au fil des réunions d'Étalab auxquelles nous avons assisté, la multitude a constitué à la fois un terme de reconnaissance mutuelle, une blague partagée au sein du groupe, et un slogan, une forme concise et expressive qui se place au cœur de la conduite de l'équipe par Henri Verdier. Cependant, cette multitude souvent convoquée n'est que peu rencontrée par les collectifs de l'ouverture. Le « capitalisme médiatique » est défini par Jeanneret comme le mode d'organisation économique contemporain autour d'industries médiatiques qui règlent les pratiques d'échanges, de publication et de circulation des productions sociales et culturelles, et qui en tirent des revenus. Si ce capitalisme parvient à convoquer cette multitude grâce à la création de

dispositifs capables d'être mis en usage, les politiques des données n'exercent pas le même attrait. La chose est perceptible dans les conversations informelles qui établissent le constat que les publics des données, cette fameuse multitude, ne sont pas visibles, car les dispositifs d'accès aux données ne sont pas encore tout à fait en adéquation avec des pratiques génériques.

La figure du public spontané souligne le fait que les collectifs de l'ouverture postulent l'existence d'un public actif, d'ores et déjà mobilisé sur les dispositifs informatisés, d'ores et déjà impliqué dans la production de formes culturelles de nature extrêmement variée, et la possibilité que ce public se mette à investir les données comme des formes qui autorisent l'action dans leur sphère ou dans l'espace public.

Le public désigné

La seconde figure de public est celle d'un public désigné. Ce public est celui auquel les collectifs de l'ouverture s'adressent, que ce soit lors d'événements ou sur les dispositifs et médiations qu'ils mettent en œuvre. L'adjectif désigné vient du verbe latin *designare*, qui signifie « marquer quelqu'un d'un signe ». Ce public se caractérise par la présence de qualités spécifiques, notamment de savoirs et savoir-faire techniques et, en particulier, informatiques, que ce soit en matière de logiciels ou de programmation. Ce public est reconnu par les collectifs de l'ouverture, qui lui attribuent des logiques d'action et un pouvoir particulier.

Ce public désigné, c'est globalement celui qui possède des compétences techniques en programmation, en conception ou en traitement des données. Ce sont ces traits qui définissent les fonctions et les réalisations qu'on attend de lui. Les collectifs de l'ouverture créent des dispositifs spécifiques pour mobiliser ce public, pour le repérer et pour entrer en relation avec lui. Ces dispositifs prennent principalement deux formes : les infrastructures d'accès aux données et les événements, par exemple de type hackathons, qui visent la création et le prototypage d'instruments ou d'artéfacts dans un temps restreint. Dans ces situations, le public est nommé à une fonction pour remplir une charge spécifique.

Ces événements ont deux caractéristiques principales. Premièrement, un événement repose sur la joie de participer que favorise une unité de lieu et de temps. Cet agencement vise en effet à créer une émulation qui facilite les rapprochements et les associations inédites entre participants. Les organisateurs cherchent à produire un état émotionnel qui s'appuie sur l'enthousiasme.

siasme où la joie que provoque la participation est censée renforcer l'engagement. Les mines sont réjouies, les postures décontractées. L'enthousiasme et la joie sont mobilisés pour partager et transmettre. Cela se traduit par le fait de se présenter comme une communauté agissante, dans une posture à la fois productive et positive. L'engagement se traduit aussi par la place faite à l'humour, que l'on peut mettre en relation avec l'esprit *lulz* des pirates (*lulz* est un terme dérivé de l'expression *lol – laughing out loud* – pour insister sur la dérision). Par exemple, lors de l'hackathon *Open Democracy Now*, un groupe a proposé de créer le « *candy crush* » de la démocratie, pour rendre les citoyens « *addicted* » au vote.

Deuxièmement, dans ces événements, il s'agit de produire des alliances entre des espaces différents, l'État et la société civile, les entreprises et de potentiels salariés. Ce qui est mis en avant, c'est le motif du « co- », de coconstruction, de la collaboration, de la coproduction, de la coopération. Ce motif « connote plus fortement l'action (avec l'idée de construction) et le faire ensemble (préfixe "co") à un niveau égal d'implication » (Belhadj, 2023, p. 136). L'événement met en place une certaine configuration sociale qui suppose l'implication d'un public qualifié. Le public est ici un adjuvant pour la production d'une politique des données.

L'euphorie et la collaboration sont deux registres communicationnels mobilisés non seulement durant les événements, mais aussi dans les publications. L'une et l'autre constituent des médiations qui permettent de faire de ce public désigné un public qui entraîne les autres. Ces manières d'agir et d'être ensemble confèrent en effet une place essentielle à la relation telle qu'elle peut s'établir et s'afficher. Ainsi, les photographies publiées lors des événements par les collectifs organisateurs – que ce soit sur les sites Web institutionnels ou sur les dispositifs promus par les industries médiatisantes – représentent cette euphorie et cette collaboration en acte et visent à donner envie de faire partie de ce public.

Les collectifs de l'ouverture voient dans ce public désigné la possibilité de pointer l'existence de personnes possédant les connaissances et les savoir-faire techniques nécessaires. Ces savoirs sont vérifiés et servent à marquer les personnes qui les possèdent. C'est par leur désignation que les collectifs constituent la valeur de ce public, telle qu'elle est reconnue et attendue pour son potentiel de transformation et d'innovation. En outre, les savoirs pratiques de la communication maîtrisés par les collectifs, c'est-à-dire la publication de photographies ou de comptes rendus et la publicisation des résultats de ces événements, jouent un rôle important dans la présentation et la valorisation

de ce public désigné. Ces savoirs communicationnels permettent de manifester au-delà de la sphère de l'ouverture l'existence de ce public et confèrent un certain pouvoir politique aux collectifs qui revendiquent de « bien communiquer » (Seurrat, 2019). Par ces procédés communicationnels, le public désigné est re-présenté comme la véritable réserve des politiques de données.

Le public désiré

Les collectifs des politiques publiques, dans les associations, les collectivités ou au sein de l'État, se posent de la même manière la problématique des publics absents. À l'instar des acteurs de la culture, ils cernent un non-public des données à l'adresse duquel ils ne vont cesser de proposer des dispositifs de médiation. Nous qualifions cette figure du public de désiré. L'adjectif désiré vient du verbe latin *desiderare* qui signifie « regretter l'absence ». Sur le plan étymologique, le terme se compose du préfixe *de*, qui marque l'absence, et du nom *sidus*, qui signifie « astre ».

Tout d'abord, la volonté d'élargir le cercle se traduit par la tentative de désigner le public qui n'est pas au rendez-vous. Dès notre premier contact avec Amandine Brugière, qui met en place la campagne Infolab en 2012, elle nous fait part de son constat, à savoir qu'elle décèle un problème : celui de « la hauteur de la marche » pour que les citoyens puissent s'approprier les données. Elle marque ainsi la distance, voire l'écart entre le public citoyen et les données pour lesquelles elle souhaite que l'Infolab s'engage dans la production de médiations. Lors de la préparation d'événements au sein du groupe Infolab, les discussions mettent en évidence une tension entre le fait que les données concernent « tout le monde » et le fait que ça ne parle qu'à un tout petit monde (et toujours le même). De cette tension dans la définition du public naît un ensemble hétérogène dont les caractéristiques sont difficiles à stabiliser, mais qui servent à déterminer à qui s'adresse l'Infolab. Ainsi, le public de l'Infolab, c'est l'étudiant, le jeune entrepreneur, le militant d'une cause, bref, ceux que plusieurs documents nomment les « M. et M^{me} Jourdain » des données. Cet ensemble peut être réuni sous le syntagme flou de « grand public », car l'objectif affiché est que les médiations aux données proposées s'adressent à tous. Ce public se caractérise par sa dimension ordinaire, sans véritable spécificité ni caractère distinctif. Les collectifs veulent rendre ce public capable d'user des données, le mettre en relation avec des données : cet objectif capacitaire justifie la création de multiples médiations.

Pour répondre à ces besoins hétérogènes, les équipes auxquelles nous avons participé ont créé des outils et des dispositifs pour favoriser l'appropriation de la culture des données (Casemajor, cet ouvrage). La création de modèles d'acculturation se réalise dans une perspective d'éducation citoyenne et d'idéal démocratique de participation. Les médiations sont pensées pour réduire l'écart. D'un côté, on réfléchit à la reconnaissance des compétences, et l'enjeu est alors de montrer que tout un chacun possède déjà des « briques » de culture des données qu'il est possible de mobiliser. De l'autre, la question récurrente, lors de la campagne Infolab notamment, est de savoir « pourquoi je rentre dans un Infolab ? » ou « pourquoi j'ouvre data.gouv ? ». Nombre de réunions ont visé à apporter une réponse à ces questions, qui révèlent qu'il n'y a pas de public *a priori*, mais que celui-ci est considéré comme peu éloigné, et qu'il suffit de susciter son intérêt en créant des médiations *ad hoc*. Des médiations qui lui révéleront que les données sont accessibles et manipulables ; celles qui dédramatiseront l'usage possible des données grâce à des outils dont on vante le côté « magique », par exemple récolter des données sans effort par le *scraping*, c'est-à-dire la récupération de données non structurées sur une page Web.

La manière dont les collectifs désirent ce public se manifeste par la permanence de sa présence dans les discours informels et par la volonté de pouvoir le nommer et lui attribuer des qualités qui peuvent constituer les prémisses d'une culture des données.

Les publics, un impératif pour les politiques de données publiques

Durant nos enquêtes, nous avons pu observer comment les sujets impliqués dans les politiques de données créent des dispositifs et mettent en œuvre des processus qui visent à constituer certaines personnes, et si possible le plus grand nombre, en publics. C'est parce que les collectifs de l'ouverture s'autorisent à parler au nom du public, et notamment à dire que les publics ont besoin des données pour agir, qu'ils créent autant de dispositifs pour pouvoir les capturer ou du moins le mettre en activité.

L'affirmation de l'existence de publics et l'idée que le public attend, voire désire toutes formes de médiations et de dispositifs pour accéder aux données comportent un caractère politique. Les publics reconnus servent aux collectifs de l'ouverture à délimiter leur champ d'action et à mettre en scène une certaine puissance de ce que la mise à disposition des données rend possible. On peut considérer les publics comme une valeur indispensable à la panoplie de l'ouverture, aux infrastructures de données comme aux événements, car la

manière dont les collectifs de l'ouverture pensent le public est intrinsèquement liée à leur mode d'action. Constituer des publics est un véritable impératif pour les politiques de données : le public est appréhendé sous l'angle de la nécessité de son existence. C'est une somme d'individus qui est reconnaissable et qui, surtout, est mobilisable au moment adéquat. Notre enquête nous a conduite à catégoriser trois figures de publics qui permettent de distinguer le travail de valuation qui est accompli par les collectifs de l'ouverture en fonction de différentes qualités et de plusieurs régimes d'action qui sont attribués aux publics. Le point commun à ces trois figures de publics est sa dimension de réserve, c'est-à-dire une existence latente qui peut être mobilisée à tout moment pour actualiser les politiques de données. Que ce soit les publics désiré, désigné ou spontané, les collectifs de l'ouverture ne cessent de penser leur existence et les convoquent régulièrement pour revendiquer le pouvoir de leur action de transformation.

Lors du sommet de l'Open Government Partnership, organisation internationale qui promeut les principes d'ouverture, qui se tient à Paris en 2016, Laure Lucchesi, alors directrice d'Etalab, nous désigne la salle, nous montre le nombre de participants et commente la conversation qui surgit en tous points de la salle : le public désigné a répondu à l'appel, il est présent, il se manifeste par son activité bruisante. Ainsi, les publics, qu'ils soient spontanés, désignés ou désirés, sont les résultats d'activités structurées et permettent de mettre en évidence la puissance de mobilisation des médiations.

Nous proposons une approche communicationnelle des manières dont les acteurs des politiques de données pensent et agissent pour faire advenir des figures de publics qui confèrent une épaisseur sociale et une teneur au déploiement de la panoplie de médiations de l'ouverture des données grâce au travail de valuation des publics. Cette approche génère une compréhension sociopolitique des dispositifs infocommunicationnels produits en vue de constituer le public en réserve politique nécessaire à l'actualisation des politiques de données. Créer une panoplie de dispositifs et affirmer l'existence de publics pour se les approprier apparaissent donc comme des processus de remédiation indispensables à l'agir « ingénieur » qui caractérise les politiques de données (Labelle, 2020).

Références

ALEXIS, Lucie *et al.*, « Vers le partage du savoir-écrire la loi ? Analyse sémio-discursive du dispositif republicque-numerique.fr », *Semen*, n° 42, 2016.

- BELHADJ, Amina, « Circulation d'énoncés et autorité en discours : la "réussite socio-discursive". Du vocabulaire de l'"économie numérique" (1998-2019) », thèse de doctorat, Université Paris Est Créteil, 2023.
- COLIN, Nicolas et Henri VERDIER, *L'âge de la multitude. Entreprendre et gouverner après la révolution numérique*, Malakoff, Armand Colin, 2^e éd., 2015.
- DEWEY, John, *La formation des valeurs*, Paris, La Découverte, 2011.
- , *Le public et ses problèmes*, traduit par Joëlle Zask, Paris, Gallimard, 2010.
- FRASER, Nancy, « Repenser l'espace public : une contribution à la critique de la démocratie réellement existante », dans Emmanuel RENAULT (dir.), *Où en est la théorie critique ?* Paris, La Découverte, 2003, p. 103-134.
- GOËTA, Samuel et Clément MABI, « L'open data peut-il (encore) servir les citoyens ? », *Mouvements*, vol. 79, n° 3, 2014, p. 81-91.
- GOURGUES, Guillaume et Alice MAZEAUD, *L'action publique saisie par ses « publics ». Gouvernement et (dés)ordre politique*, Québec, Septentrion, 2018.
- HABERMAS, Jürgen, *L'espace public*, Paris, Payot, 1993.
- JEANNERET, Yves, « L'urbanité, c'est là et ça va donner. Le motif de la donnée dans les représentations de la ville », *Questions de communication*, vol. 36, n° 2, 2019, p. 25-42.
- , *Critique de la trivialité. Les médiations de la communication, enjeu de pouvoir*, Le Havre, Éditions Non Standard, 2014.
- LABELLE, Sarah, « La ville inscrite dans la « société de l'information » : formes d'investissement d'un objet symbolique », thèse de doctorat en sciences de l'information et de la communication, Université Paris IV, 2007.
- , « Médiations et pouvoirs de l'agir ingénieur : mythes et infrastructures des politiques de données », HDR en sciences de l'information et de la communication, Université Paris Est, 2020.
- LABELLE, Sarah et Jean-Baptiste LE CORF, « Modalités de diffusion et processus documentaires, conditions du "détachement" des informations publiques. Analyse des discours législatifs et des portails open data territoriaux », *Les Enjeux de l'information et de la communication*, vol. 13/2, n° 2, 2012, p. 209-259.
- LE CORF, Jean-Baptiste, « Les pratiques d'innovation de services des développeurs Web dans les territoires : le cas des projets Open Data », *Communication et organisation*, n° 50, 2016, p. 123-136.
- LEHMANS, Anne, « Les réinventions de la démocratie à l'aune de l'ouverture des données : du discours de la participation aux contraintes de la gouvernance », *Les Enjeux de l'information et de la communication*, vol. 19/2, n° 2, 2018, p. 135-146.
- LE MAREC, Joëlle, « Situations de communication dans la pratique de recherche : du terrain aux composites », *Questions de terrains*, vol. 25, 2002, p. 15-40.
- , *Publics et musées. La confiance éprouvée*, Paris, L'Harmattan, 2007.
- SEURRAT, Aude, « Les savoirs sur la communication face à l'impératif d'efficacité : industrialisation, professionnalisation, médiation et évaluation dans la formation professionnelle courte à la communication », Mémoire pour l'HDR, Sorbonne Université, 2018.

DEUXIÈME PARTIE

La santé

CHAPITRE 3

La représentation des citoyens dans une plateforme de données en santé

Défis et ambiguïtés

Maxime Harvey et Guillaume Latzko-Toth

Traditionnellement, les données utilisées dans le domaine de la santé sont recueillies et analysées en définissant les individus comme des sujets de recherche ou des patients de soins cliniques (Bromley *et al.*, 2015). Les données visent alors à maintenir ou à améliorer les conditions de vie des individus. Certains chercheurs proposent maintenant un nouveau paradigme afin de prendre en compte ces personnes au sein de leur environnement et de les impliquer dans les activités scientifiques et thérapeutiques. Ce changement comporte non seulement de nouvelles données incluses comme indicateurs de santé populationnelle et environnementale, mais aussi une redéfinition des patients comme des « citoyens » (Alami, Gagon et Fortin, 2019), des « scientifiques citoyens » (Bietz, Patrick et Bloss, 2019), des « partenaires » (Pétre *et al.*, 2018) ou un « public » (L'Espérance *et al.*, 2021).

Pour étudier la manière dont émergent et sont configurés des publics de données dans ce contexte de transformation de la recherche en santé, nous nous sommes intéressés à une innovation sociotechnique nommée « Pulsar ». Ce projet hybride, à la fois plateforme technologique, réseau d'acteurs et infrastructure de données, a été amorcé dans le but d'« engager » les citoyens et de faire usage des données de manière plus « collaborative ». Pulsar vise le développement et la promotion de la « santé durable », une approche qui préconise la prise en compte de toutes les facettes de la santé et du bien-être des individus au sein de leur population et de leur environnement, ainsi que

sur le long terme (Alliance santé Québec, 2015). C'est à cette « vision » que doivent adhérer les acteurs du réseau Pulsar, qui comprend des membres du grand public ainsi que des acteurs issus des milieux de la recherche et des nouvelles technologies, du réseau de la santé et de plusieurs instances, notamment politiques. Tous ces acteurs sont rassemblés *autour* de et *par* une « plateforme collaborative de recherche et d'intervention en santé durable » (Alliance santé Québec, 2015). Celle-ci fournit des outils de recrutement, de gestion de la participation, de collecte, de stockage, d'analyse et de partage des données et de diffusion des résultats de recherche. Finalement, on trouve aussi liée à ce projet la Banque de données en santé durable, qui offre la possibilité de réutiliser les données recueillies avec la plateforme dans d'autres recherches.

Publics, représentation et imaginaires sociotechniques : une étude de cas

Le cadre conceptuel adopté pour l'étude du projet Pulsar se situe à la croisée des études en communication et des *Science & Technology Studies* (STS). Sur le plan communicationnel, c'est la notion de « public » qui constitue notre concept central ; à la fois comme *auditoire* (à qui s'adresse le dispositif) et *acteur* au sens de sujet politique concerné par un problème (Zask, 2008). Dans notre étude, les citoyens s'inscrivant sur la plateforme Pulsar, mais également les autres publics enrôlés par le dispositif, sont conceptualisés sous cette double dimension, à laquelle vient s'ajouter l'axe réel/imaginé dans la caractérisation des publics, c'est-à-dire les façons dont différents imaginaires guident la catégorisation des publics et les manières dont ces derniers s'approprient réellement les innovations (Assogba *et al.*, 2015). Se pose ensuite la question de leur « représentation » dans le procès d'innovation, articulée notamment par Hacking (1983), Lynch (1994) et surtout Latour (2004). Ce concept invite à porter attention non seulement à la construction sociale de ce qui est représenté (ici, la santé durable), mais aussi à l'organisation socio-matérielle qui se met en ordre par ce processus et qui assigne des positions à divers actants dont certains agissent comme porte-parole ou délégués pour d'autres – par exemple sous forme de structure de gouvernance et d'infrastructure numérique. Notre étude exige aussi de prendre en compte des « imaginaires sociotechniques » et la relation circulaire qui les lie aux innovations technoscientifiques. Portés collectivement, stabilisés institutionnellement et faisant exister publiquement des visions désirables du futur, ces imaginaires sont potentiellement réalisables par les innovations scientifiques

et technologiques, tout en jouant un rôle actif dans le développement de ces dernières (Jasanoff et Kim, 2015, p. 4). Notre analyse prend aussi en compte les approches critiques de la classification et des infrastructures de données, ainsi que les aspects sociopolitiques de la définition des catégories d'acteurs des données (Bowker, 2000).

La démarche d'enquête s'inscrit dans l'approche de l'étude de cas en STS, laquelle constitue un ensemble cohérent théorie-méthode¹ (Latzko-Toth, 2009). Développée dans le cadre d'une sociologie des controverses pour comprendre la genèse des dispositifs sociotechniques en tant que *réseaux* (Akrich, Callon et Latour, 2006), l'étude de cas consiste à suivre l'ensemble des acteurs impliqués dans le processus d'innovation, y compris ceux qui sont évoqués sans être parties prenantes du projet autrement que comme *représentés* – dans le discours ou par le truchement de porte-paroles. Dans le cadre de cette enquête, nous avons mené des entretiens qualitatifs avec 12 membres de l'équipe de Pulsar occupant ou ayant occupé divers rôles dans le développement et la gestion de la plateforme. Pour des raisons à la fois éthiques et organisationnelles qui se sont aussi avérées révélatrices de la représentation du public étudié, nous n'avons pas eu accès aux citoyens eux-mêmes. Cela étant dit, certains membres de l'équipe nous ont fait part de leurs interactions avec ces acteurs, notamment les deux responsables du comité de représentants du public et les deux responsables des communications qui ont interagi avec des citoyens et des usagers tout au long du développement de la plateforme. Nous avons également testé la plateforme en participant à deux projets en tant que sujets de recherche, en accord avec les responsables des équipes de recherche concernées. Enfin, l'analyse documentaire a porté sur un corpus d'environ 125 documents, spécifiquement des documents internes fournis par l'équipe de direction, différentes versions du site Web de Pulsar, le catalogue de données de la Banque de données en santé durable, les comptes de Pulsar sur les réseaux socionumériques, les infolettres envoyées aux membres de Pulsar et plusieurs articles de presse.

1. Clarke et Star (2008) ont introduit la notion de « paquet théorie/méthode » à propos de la sociologie des mondes sociaux, qui lie de manière inextricable théorisation ancrée et concepts sensibilisants. De la même manière, l'étude de cas en STS, qu'elle s'inscrive ou non dans une perspective interactionniste, mobilise conjointement et de manière récurrente les théories socioconstructivistes et l'ethnographie, l'analyse documentaire, la méthode « boule de neige » pour le recrutement d'informateurs, etc., afin de décrire un réseau et d'en suivre les actants.

Genèse du projet Pulsar

L'idée du projet Pulsar émerge en 2013, même s'il faut attendre 2018 pour que son nom soit officialisé au terme d'une consultation menée la même année. Une première phase de conceptualisation s'étire jusqu'en 2016, durant laquelle plusieurs forums regroupant divers acteurs du milieu de la santé de Québec forgent un consensus : la plateforme collaborative de recherche doit incarner un changement de paradigme vers la santé durable dans lequel les sujets de recherche ne sont plus des patients qui ont des maladies, mais des « personnes », c'est-à-dire des individus étudiés dans leur environnement (Alliance santé Québec, 2015). En 2017 commence la phase de développement, alors que le projet devient une plateforme dont la conception est confiée à la Direction des technologies de l'information (DTI) de l'Université Laval. Celle-ci met alors sur pied la structure de gouvernance du réseau et de la plateforme ainsi que l'infrastructure technique qui la soutient. Le site Web qui héberge cette plateforme voit le jour en 2019, alors que le projet entre dans sa phase de réalisation. Différentes fonctionnalités s'ajoutent par la suite à la demande des utilisateurs. Dès la première phase du projet, trois publics sont envisagés comme usagers de la plateforme ou bénéficiaires du projet : les « citoyens », les « chercheurs » et les « décideurs ».

D'abord, le public de citoyens est convoqué de plusieurs façons dans le développement de ce projet. Au moment des efforts de conceptualisation de la plateforme, il participe aux consultations, aux côtés d'acteurs du milieu de la recherche et d'institutions publiques en santé. Lors de la phase de développement, les responsables de Pulsar intègrent les citoyens dans la structure de gouvernance par une stratégie d'engagement qui vise à mettre sur pied un « comité de représentants du public ». Finalement, les citoyens sont imaginés en tant qu'utilisateurs de la plateforme lors de sa réalisation, c'est-à-dire qu'en testant l'infrastructure technique, on configure les limites de leurs actions dans l'espace de la plateforme et de leurs interactions avec les autres membres de Pulsar. C'est aussi à ce moment qu'apparaissent les usagers réels de la plateforme.

Au cours du développement de Pulsar, les citoyens sont imaginés aux côtés de deux autres publics : les chercheurs et les décideurs. Les chercheurs sont les responsables ou les collaborateurs des projets de recherche sur la santé qui utilisent la plateforme Pulsar et la Banque de données en santé durable adossée à celle-ci, et qui compose une partie intégrante du projet. Imaginons-les comme formant, grâce au partage de données, un réseau de

recherche interdisciplinaire rendu possible, d'une part, par l'infrastructure Pulsar et, d'autre part, par le modèle de recherche, développé en collaboration, qui recense les dimensions et les indicateurs de la santé durable, mais aussi les données, les outils et les mesures pour les exploiter. C'est le « comité scientifique » qui représente ce public de chercheurs au sein de la structure de gouvernance. Les décideurs, pour leur part, sont les différentes institutions partenaires de Pulsar, souvent gouvernementales ou paragonnementales, et parfois des acteurs du milieu de la santé. Le « comité des partenaires » les représente.

Les multiples représentations du citoyen

Les discours sur le projet Pulsar ont recours à plusieurs termes pour parler des « citoyens ». Trois catégorisations de ce public sont omniprésentes : les « participants », les « représentants du public » et les « utilisateurs ». Nous développons ici les différentes manières de les évoquer en tant que public en confrontant successivement les représentations qui leur sont associées.

Des participants aux projets de recherche

Les citoyens sont d'abord envisagés en relation avec Pulsar comme des « participants » aux projets de recherche. Au moment de formuler l'idée d'une plateforme collaborative de recherche en santé durable, les membres de l'Alliance santé Québec s'imaginaient transformer la recherche en santé en définissant le sujet de recherche non plus comme un patient ou un malade, mais comme une personne étudiée dans son environnement et un acteur dans le système de soin.

On distingue trois modes de participation de ce public dans les discours de Pulsar : en tant que « citoyens », les participants collaborent à l'avancement de la science et à l'amélioration de la santé de la population ; en tant que « patients », ils sont des acteurs aptes à modifier leur trajectoire de santé ; et en tant que « sujets de recherche », ils contribuent volontairement et de manière éclairée aux projets de recherche en fournissant des données, sans nécessairement recevoir de rétribution en retour.

Au début du projet, les concepteurs de la plateforme ont avancé qu'il serait stratégique d'offrir des « mécanismes structurants et systématiques » de participation citoyenne qui incluraient les citoyens « à toutes les étapes de la recherche » (communication interne, 2018). Il revient aux porteurs des

projets de recherche de manifester à l'équipe de réalisation de Pulsar leurs besoins en matière d'usage et de participation du public, que leurs développeurs traduiront alors en fonctionnalités sur la plateforme. Par exemple, pour avoir « une vision un peu plus holistique de la société » et aller chercher « des clientèles plus vulnérables, défavorisées, qui n'ont pas accès à l'ordinateur, qui n'ont pas accès à l'Internet », il est apparu à un chercheur qu'il serait avantageux d'avoir un consentement à la participation aux projets de recherche exprimé oralement ou par message texte (Simon, chercheur) ; il fallait à ce moment non seulement convaincre les informaticiens de la faisabilité de ces fonctionnalités, mais aussi rouvrir le cadre de gestion des données en partenariat avec le comité éthique de l'université partenaire du projet et les représentants du public. Comme le montre le commentaire de ce chercheur, la vision de la société incorporée dans l'infrastructure de Pulsar n'est pas limitée que par la quantité et la qualité des données recueillies avec la plateforme et conservées dans la Banque de données en santé durable, mais aussi par l'opacité de certaines fonctions – par exemple, le consentement écrit sur ordinateur – qui empêche de recruter certaines catégories de participants (ou de « clients » selon le vocabulaire). En tant qu'« infrastructure » de collaboration entre différents acteurs, notamment des chercheurs et des citoyens, la plateforme n'est « transparente » que pour certaines catégories de chercheurs et de citoyens. L'un des défis de Pulsar est d'inclure un maximum de citoyens sans être limité à l'expression des besoins de chaque projet de recherche.

En fait, la représentation des citoyens comme participants aux projets de recherche reste très générale au sein de Pulsar, parce que c'est en effet à chaque projet de sélectionner les participants selon ses besoins précis. Pulsar se présente plutôt comme un moyen d'accéder à un bassin de participants potentiels. Ainsi, l'équipe Pulsar valorise leurs outils de communication – Facebook, Twitter maintenant X, site Web, infolettre – pour former un « réseau » constitué de l'ensemble de ses membres inscrits et de ses « abonnés ». Ceux-ci forment un bassin de participants potentiels qui peut être exploité par de nouveaux projets de recherche grâce à la capacité de communication et de recrutement à grande échelle de Pulsar.

Cette représentation de l'équipe de réalisation de la plateforme est cependant contestée par certains chercheurs et leur réalité de terrain.

La portée de Pulsar, [...] son potentiel, c'est pas de générer des nouveaux participants pour des projets par-ci par-là. C'est pas ça qui va être bon pour Pulsar, c'est qu'il y ait le plus de projets possible dans Pulsar. Puis le recrutement va

FIGURE 3.1

Les citoyens participant à la recherche (pulsar.ca)**Moi, participant?**

Toute personne intéressée peut, sur une base libre et volontaire, participer à un projet réalisé avec PULSAR. Chaque projet est unique et son équipe sélectionne ses participants en fonction de critères prédéterminés. Vous souhaitez participer et savoir ce que cela impliquera dans votre vie? Visitez d'abord la page du projet sur notre site et posez ensuite des questions à l'équipe du projet. Cela vous permettra de faire un choix éclairé!

Pour plus d'information, visionnez notre [vidéo](#) et consultez la section «Participer à un projet» de notre [foire aux questions](#).

passer par les projets eux-mêmes, vous pouvez pas vous substituer à ça, ça marchera pas. Par expérience, ça fonctionnera pas. (Daniel, chercheur)

Ainsi, l'idée selon laquelle les participants seraient mis en contact avec des projets de recherche par le biais de Pulsar est remise en cause. Au lieu de recourir à un réseau, le recrutement par les projets se fait ou bien en parlant directement aux participants, un par un, ou bien par la présence médiatique d'un chercheur à la télévision ou à la radio, ce qui provoque un *boom* d'inscriptions. Cette proposition est donc un appel à différents moyens de communication avec les participants potentiels. Il y a aussi opposition en matière de valorisation de la participation des sujets de recherche entre le recrutement de nouveaux participants pour les recherches futures et la rétention des participants actuels pour réaliser les projets de recherche en cours. Le bassin de participants créé par Pulsar, présenté d'abord comme une innovation sociotechnique grâce au réseau Pulsar, est donc en réalité l'objet d'un débat où les uns veulent maximiser le potentiel de recrutement des membres de Pulsar grâce à son réseau, et les autres, minimiser le risque d'abandon des participants par un travail de recrutement effectué directement par les équipes de recherche.

Cette concurrence pour le recrutement trouve un écho dans la négociation de la valeur des données de santé collectées et archivées. Le recrutement d'une même personne par plusieurs projets de recherche facilite le croisement des données entre les projets. Cependant, si le recrutement de participants par plusieurs projets diminue effectivement la rétention des participants à un projet donné, alors il sera impossible de construire des jeux de données et de créer des trajectoires de santé sur le long terme. Il devient ainsi difficile de conjuguer collectivement – dans la communauté de chercheurs – et institutionnellement – dans l'équipe de Pulsar – les besoins actuels de participants

et de projets de recherche avec l'objectif de constituer des jeux de données qui permettront un jour de visualiser la santé durable.

Des représentants du public

En plus de les catégoriser comme des participants aux projets de recherche, Pulsar intègre aussi certains citoyens en tant que membres du « comité des représentants du public ». À l'origine du projet, le comité a été défini comme un groupe de dix citoyens engagés dans la coconstruction et la gestion de Pulsar. Il devait participer au choix des projets de recherche selon leur potentiel d'avoir un effet sur la santé de la population. Pour établir les priorités de recherche et évaluer leur pertinence, les membres du comité devaient fournir au comité scientifique chargé de la sélection des projets des informations permettant à celui-ci de tenir compte des besoins, des intérêts et des préoccupations du (grand) public. De plus, le comité devait consulter les travaux des comités « de direction », « de développement », « de partenaires » et « scientifique » afin de faire des recommandations à l'équipe de réalisation de Pulsar lors des différentes phases du projet, par exemple en validant le résultat des consultations publiques (communication interne, 2018).

Avant de créer le comité, l'équipe responsable du projet Pulsar a formé un atelier avec une quinzaine d'experts de différentes disciplines pour débattre de la participation citoyenne : « on essayait d'avoir des gens de tous horizons pour justement confronter les idées. On voulait pas juste des gens qui étudiaient la participation citoyenne, on voulait des gens qui l'utilisaient aussi là, donc de tous les styles. » (Sylvie, membre de l'équipe Pulsar) Les experts sont alors engagés dans des ateliers où la discussion autour du thème de la participation citoyenne les amènent à réfléchir « aux conditions gagnantes et aux bonnes pratiques à mettre en place afin de maximiser l'engagement du public dans toutes les étapes de la mise en œuvre et de l'opération de la plateforme, tant au niveau des instances décisionnelles qu'au niveau de sa participation en tant qu'utilisateur de cette plateforme » (rapport d'atelier). L'équipe de Pulsar met ensuite en action ces recommandations. L'une d'elles est de changer le terme de « participation citoyenne » initialement proposé pour la formule « engagement du public ». Selon ce groupe de travail, le terme « citoyen » renvoie de manière trop restrictive aux personnes ayant droit de vote, alors que le terme « public » inclut « toute personne de la province de Québec qui peut être un bénéficiaire direct ou indirect de

Pulsar » (communication interne, 2018). Dans cette conception, un bénéficiaire direct est une personne qui est en mesure de prendre des décisions, qui a le pouvoir d'influencer la construction de Pulsar, de participer à sa continuité et qui a un intérêt pour le concept de santé durable. Un bénéficiaire indirect, quant à lui, est une personne qui tire profit indirectement de l'influence que Pulsar aura sur la santé durable de la population et de son environnement (communication interne, 2018).

Selon cette définition, le « public » englobe l'ensemble des personnes directement ou indirectement bénéficiaires des projets de recherche en santé de Pulsar. Ce « grand public » existe *a priori* et n'est donc pas construit au sein du projet, mais il est « représenté » par lui (Laplanche-Servigne et Boas, 2019). En effet, lorsque les citoyens sont considérés comme des participants à la recherche, ils forment un public construit par les chercheurs étudiant une population donnée; de même, les utilisateurs de Pulsar constituent un public construit par les gestionnaires du projet recrutant des membres pour le réseau Pulsar et par la plateforme dont ils font usage. Le « grand public », pour sa part, est représenté. Les concepteurs de Pulsar ont déterminé deux attributs que doivent posséder les membres du « comité des représentants du public » pour rendre effective cette représentation. Premièrement, le comité doit être représentatif de la diversité de la population, c'est-à-dire de la diversité de « maladies », de « groupes vulnérables », de « niveaux de revenus et d'éducation », de « générations », de « secteurs d'emploi », de « familles », de « genres » et d'« intérêts » (communication interne, 2018). Deuxièmement, les membres du comité doivent être engagés dans la promotion de la santé et, plus particulièrement, partager l'intérêt de Pulsar envers la santé durable.

La principale stratégie de recrutement des promoteurs du projet pour former un comité de représentants du public dont les membres sont à la fois représentatifs de la population et engagés dans la santé durable consiste à passer par des intermédiaires, soit des organismes communautaires, des équipes de recherche et des professionnels du milieu de la santé. Cette stratégie soulève un premier défi lié au respect des deux caractéristiques citées ci-dessus. En effet, si l'on cherche à recruter des membres d'organismes communautaires – comme des groupes de soutien pour favoriser la représentation de diverses maladies –, ces représentants ne doivent pas défendre la mission de leur organisation en tant que membres du comité de représentant *du public*. On dit par exemple: « sera exclue une personne qui a un intérêt compétitif envers Pulsar, qui ne partage pas l'intérêt de santé durable, qui recherche à défendre seulement l'intérêt de son groupe ou qui a une cause

à défendre» (communication interne, 2018; c'est nous qui soulignons). En d'autres mots, d'un côté on tente de recruter des personnes représentatives de la population en cherchant des citoyens engagés dans une diversité d'organismes communautaires, mais de l'autre, on leur demande de ne pas défendre la cause de leur organisme et d'aligner leurs intérêts sur ceux de Pulsar. On relève ainsi une contradiction entre ce qu'on s'imagine être un *représentant de la diversité* et ce qu'on attend de cet individu en tant que *représentant du public* au sein du comité.

Le comité a été implanté à la suite des processus de consultation d'experts en engagement du public et de recrutement, puis il a participé à la conception de la plateforme. Il a notamment commenté le cadre de gestion de la Banque de données en santé durable. Cependant, le comité a rapidement été jugé « dysfonctionnel » et démantelé. Les principales conséquences de ce démantèlement sont de deux ordres.

Premièrement, l'implication des citoyens dans la construction et le fonctionnement de la plateforme a été repoussée *sine die*. Le comité, pour qui on avait imaginé une collaboration étroite au développement de Pulsar, a été remplacé par une « banque de partenaires » (Marie, membre de l'équipe Pulsar). Le rôle décisionnel du comité s'est mué en rôle consultatif, c'est-à-dire qu'il est possible grâce à la banque de solliciter certaines personnes pour des consultations *ad hoc*. Ce changement a des conséquences concrètes pour les citoyens utilisateurs de la plateforme. Par exemple, cette absence relative autour de la table des concepteurs est l'une des raisons pour lesquelles certaines fonctionnalités destinées aux utilisateurs n'ont pas été développées, notamment – et non la moindre – l'accès à leurs données.

Deuxièmement, la dysfonction du comité et son démantèlement ont donné lieu à la réévaluation de la représentation du public et à l'expression de visions opposées de l'engagement des représentants. En effet, le processus qui a remplacé le « comité de représentants du public » par une « banque de partenaires » a mis en jeu la nature de l'organisation de la participation citoyenne, sa place dans la structure de gouvernance de Pulsar et les rôles des acteurs engagés. L'« équipe de réalisation » de Pulsar, telle qu'on la présente publiquement, est alors apparue comme un bloc hétérogène. En effet, cette négociation a révélé les tensions entre les cultures des chercheurs, des gestionnaires de projet, des concepteurs des technologies – et en face, celle des comités éthique de l'université. De plus, une partie de l'équipe chargée de gérer la conception du projet est partie à la suite du lancement de la plateforme, remplacée par de nouveaux membres qui travaillent maintenant à

son fonctionnement à long terme ; le changement de régime amène avec lui une nouvelle vision de l'engagement du public. Celle-ci se rapproche du modèle connu dans le domaine médical sous les termes de « patient partenaire », que l'on trouve parfois dans les discours autour de Pulsar sous l'appellation hybride de « citoyen partenaire ». L'« approche patient partenaire » partage les prises de décision de soins entre les cliniciens et les patients, et favorise le développement de compétences pour l'intervention du patient dans le système de santé ; elle porte ainsi « à son apogée la reconnaissance de l'expertise du vécu de la maladie par le patient » (Pétre et al., 2018, p. 621). La redéfinition du *citoyen* entre (représentants du) « public » et « patient » et de sa *participation* entre « engagement » et « partenariat » est donc une remise en ordre d'un point de vue épistémologique – on passe du modèle interdisciplinaire résultant de la consultation d'une quinzaine de chercheurs à ce modèle médical et sociologique où l'on valorise l'intervention du patient au lieu de la représentation du public.

Des utilisateurs

Finalement, les citoyens constituent une catégorie d'« utilisateurs » de la plateforme Web aussi appelés « membres » du réseau Pulsar. Soulignons que les catégories de « participants », de « représentants du public » et d'« utilisateurs » sont interreliées. En effet, des fonctionnalités techniques laissent les utilisateurs de la plateforme fournir des données aux projets de recherche auxquels ils participent et, à l'inverse, laissent voir l'usage de leurs données par les équipes scientifiques. Ces fonctionnalités ont été réfléchies, proposées, développées et intégrées à la plateforme en collaboration avec le « comité de représentants du public ». À ce titre, les « utilisateurs » constituent à la fois des acteurs dans le dispositif et des publics imaginés pour les différentes fonctionnalités de la plateforme par l'équipe de réalisation de Pulsar et les chercheurs faisant usage de la plateforme. Ces fonctionnalités inscrivent ainsi les publics imaginés au sein de l'infrastructure de recherche de Pulsar. Nous avons déjà donné l'exemple du consentement à la participation aux projets de recherche, initialement prévu sur la plateforme Web ; en autorisant les consentements oraux et par messagerie texte, on inscrit de nouveaux usages au sein de l'infrastructure Pulsar et l'on permet le recrutement de nouvelles catégories d'usagers.

L'exemple de la fonctionnalité « tableau de bord personnalisé » est la plus emblématique de l'écart entre la représentation du citoyen, en tant

qu'« utilisateur » dans l'idéation du projet, et sa traduction concrète dans le dispositif. La version imaginée au départ par les concepteurs de la plateforme visait à offrir aux participants une fenêtre sur leurs propres données :

nous visons à mettre en place un outil très simple, sous forme de tableau de bord, qui permettra aux cliniciens et aux divers intervenants du milieu de mettre en lumière des portraits individuels et populationnels. Cela deviendra non seulement très aidant pour poser des diagnostics plus personnalisés, mais aussi pour cibler les besoins spécifiques d'un individu ou même d'une population donnée. (ULaval nouvelles, 2020)

[Il s'agit, pour] la population, d'avoir accès à son propre dossier. D'être en mesure [...] d'identifier un certain nombre d'indicateurs en fonction de son âge, de son sexe, de son genre [...] d'être en mesure de se comparer, de mieux se comprendre finalement. (Jonathan, membre de l'équipe Pulsar)

Certains intervenants ont envisagé le tableau de bord comme une sorte de contre-don aux participants qui ont fait don de leur temps (et de leurs données) à des projets de recherche, mais aussi comme une façon de les intéresser à participer de manière continue aux projets en les incitant à se connecter régulièrement sur la plateforme. Dans les faits, constamment retardé, le tableau de bord n'a jamais vu le jour. La vision décrite plus haut se heurte à une autre : pour l'un des responsables du projet, par exemple, il est risqué de donner trop de « *feedback* » aux participants à une recherche, surtout si elle s'étire dans le temps et vise un portrait populationnel longitudinal, car le retour d'information sur leurs données peut influencer le comportement des participants d'une cohorte et fausser les résultats d'une recherche. Il y a donc une tension entre la nécessité d'observer des changements naturels de comportements individuels et populationnels sur le long terme et, avec le tableau de bord, d'intervenir ponctuellement ou continuellement sur les comportements des participants en induisant par là même des changements.

Les citoyens ont fait valoir qu'ils attendaient une forme de retour pour s'engager sur la plateforme. Certains projets proposent donc des retours très sommaires des informations fournies, mais sans inclure de suivi continu comme le permettrait un tableau de bord personnalisé. D'autres envisagent la possibilité de tableaux de bord multiples, mais communs :

Les données individuelles versus mettre en valeur ce que Pulsar peut faire pour la société. [...] J'étais plus du côté de mettre en valeur ce qu'on fait et peut-être avoir des tableaux de bord, mais plus généraux [...] pas sur des données individuelles pour chaque individu. On a toujours été très prudents par rapport à ça, [...] mais

en même temps les sujets nous disaient « Faut qu'on nous en donne », donc ça a été le meilleur compromis qu'on a trouvé. (Daniel, membre de l'équipe Pulsar)

En somme, les perceptions contradictoires des finalités du tableau de bord personnalisé révèlent des publics imaginés de citoyens de nature divergente parmi les concepteurs de la plateforme. D'un côté, des individus-patients outillés pour une intervention sur leurs comportements individuels, ce qui aurait des retombées positives pour leur santé, mais potentiellement négatives pour les projets de recherche qui veulent étudier des changements comportementaux naturels. De l'autre, des individus-sujets de recherche « mis en données » exposant des trajectoires de santé sur le long terme et susceptibles d'influencer les décideurs politiques pour qu'ils élaborent et mettent en œuvre des politiques de santé durable pour le bien du grand public dans son ensemble.

Représentations des citoyens : défis et ambiguïtés ou conflit et confusion ?

Le cas de la plateforme de recherche Pulsar met en lumière la pluralité des imaginaires se rapportant aux citoyens lorsqu'on porte attention aux différents acteurs qui les représentent et aux transformations dans l'organisation sociale au sein de laquelle s'expriment et se matérialisent ces représentations. Lors de la phase d'idéation de Pulsar, trois publics se sont constitués : les citoyens, les chercheurs et les partenaires. Dans la phase de conception, ces publics ont été intégrés au réseau, à la plateforme Web, ainsi que dans la structure de gouvernance de Pulsar. Puis à l'étape de la réalisation du dispositif, le public de chercheurs a pris le dessus sur les deux autres, ce que nous avons pu observer d'au moins trois manières. À l'échelle du réseau, d'abord, les communications avec les citoyens se sont raréfiées. Ainsi, l'info-lettre, d'abord mensuelle, n'a pas connu de nouvelle édition en près d'un an ; les invitations à des activités dans la section « Actualités » du site (pulsar.ca) sont passées d'une demi-douzaine environ en 2019 et en 2020, à trois en 2021 et à une seule en 2022 ; une même diminution des publications s'observe sur Twitter et Facebook. À l'échelle de la plateforme, ensuite, la conception d'un tableau de bord permettant une forme de restitution des résultats de recherche aux participants s'est vue constamment repoussée. Et à l'échelle de la structure de gouvernance, enfin, le comité de représentants du public qui devait remplir un rôle décisionnel dans le développement de Pulsar a été dissous et remplacé par une « banque de partenaires » pouvant être

sélectionnés pour constituer des groupes de discussion et ainsi participer de manière *ad hoc* et à titre consultatif seulement aux décisions de l'équipe de gestion. S'il y a une représentation durable de la participation citoyenne dans une innovation sociotechnique telle que Pulsar, c'est entre autres parce que la valeur des jeux de données qu'elle recueille dépend des catégories de public incluses comme participants aux projets de recherche, qui relèvent elles-mêmes des fonctionnalités intégrées à l'infrastructure numérique et offertes à ses utilisateurs. La conception de ces fonctionnalités est elle-même soutenue par l'imaginaire relié à la santé durable qui invite à l'engagement de représentants du public au sein de la structure de gouvernance du projet. Le défi est de faire en sorte qu'au-delà de la pérennité de la représentation du citoyen, il n'y ait pas d'ambiguïtés entre les catégories de « participants » au projet de recherche, d'« utilisateurs » et de « représentants du public » et de leurs multiples rôles dans la représentation de la santé (au sens symbolique) et de la société (au sens politique).

Ces enjeux illustrent bien les défis et les ambiguïtés de l'inclusion des citoyens dans les initiatives de données en santé. Il semble en effet plus facile, pour une organisation à caractère scientifique, de s'en tenir à des actions qui ne sortent pas du cadre des pratiques établies et reconnues au sein de l'organisation de recherche universitaire qui l'encadre et de les adapter au contexte numérique. En revanche, intervenir auprès des différents publics auxquels elle s'adresse exige de développer un réseau élargi de partenaires et d'innover en matière d'infrastructure tant sur le plan technologique qu'organisationnel. C'est ainsi qu'il est possible de mobiliser le public de citoyens à la fois comme auditoire des communications partagées par ce réseau et acteur de celui-ci. Or la négociation entre les différentes visions de ce qu'est une intervention adéquate – la visualisation de données pour les citoyens, les chercheurs ou les décideurs par exemple – freine la réalisation de nouvelles fonctionnalités à l'intention de ces publics. Lorsque les défis et les ambiguïtés de la représentation des citoyens deviennent des conflits et génèrent de la confusion dans l'équipe de réalisation de la plateforme, c'est alors qu'il devient difficile d'imaginer un futur durable pour cette représentation.

Références

- AKRICH, Madeleine, CALLON, Michel et LATOUR, Bruno, *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'École des Mines de Paris, 2006.
- ALAMI, Hassane, GAGNON, Marie-Pierre et FORTIN, Jean-Paul, « Conditions organisationnelles et systémiques à l'implication des citoyens-patients dans le développement de la télésanté au Québec », *Santé publique*, vol. 31, n° 1, 2019, p. 125-135.

- ALLIANCE SANTÉ QUÉBEC, *Qu'est-ce que la santé durable?* 2015.
- ASSOGBA, Henri *et al.*, « Les publics imaginés et réels des professionnels d'Internet », *Communication*, vol. 33, n° 2, 2015, p. 1-13.
- BIETZ, Matthew, PATRICK, Kevin, et BLOSS, Cinnamon, « Data Donation as a Model for Citizen Science Health Research », *Citizen Science: Theory and Practice*, vol. 4, n° 1, 2019, p. 1-11.
- BOWKER, Geoffrey C., « Biodiversity, Datadiversity », *Social Studies of Science*, vol. 30, n° 5, p. 643-683.
- BROMLEY, Elizabeth *et al.*, « From Subject to Participant: Ethics and the Evolving Role of Community in Health Research », *American journal of Public Health*, vol. 105, n° 5, 2015, p. 900-908.
- CLARKE, Adele E. et STAR, Susan Leigh, « The Social Worlds Framework: A Theory/ Method Package », dans HACKETT, Edward J. *et al.* (dir.), *The Handbook of Science and Technology Studies*, 3^e éd., Cambridge, The MIT Press, 2008, p. 113-137.
- HACKING, Ian, *Representing and Intervening: Introductory Topics in the Philosophy of Natural Science*, Cambridge, Cambridge University Press, 1983.
- JASANOFF, Sheila et KIM, Sang-Hyun, *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power*, Chicago, University of Chicago Press, 2015.
- LAPLANCHE-SERVIGNE, Soline et SA VILAS BOAS, Marie-Hélène, « Introduction. Les catégorisations des publics minorisés en questions », *Participations*, vol. 25, n° 3, 2019, p. 5-31.
- LATOUR, Bruno, *Politiques de la nature. Comment faire entrer les sciences en démocratie*, Paris, La Découverte, 2004.
- LATZKO-TOTH, Guillaume, « L'étude de cas en sociologie des sciences et des techniques », *Note de recherche CIRST*, n° 3, 2009.
- L'ESPÉRANCE, Audrey *et al.*, « Developing a Canadian evaluation framework for patient and public engagement in research: study protocol », *Research Involvement and Engagement*, vol. 7, n° 10, 2021.
- LYNCH, Michael, « Representation Is Overrated: Some Critical Remarks about the Use of the Concept of Representation in Science Studies », *Configurations* 2, n° 1, 1994, p. 137-149.
- PÉTRÉ, Benoît *et al.*, « L'approche patient partenaire de soins en question », *Revue médicale de Liège*, vol. 73, n° 12, 2018, p. 620-625.
- STAR, Susan Leigh, et RUHLER, Karen, « Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces », *Information Systems Research*, vol. 7, n° 1, 1996, p. 111-134.
- ZASK, Joëlle, « Le public chez Dewey: une union sociale plurielle », *Tracés. Revue de sciences humaines*, vol. 15, 2008, p. 169-189.

CHAPITRE 4

Le codesign d'un service numérique de santé mentale

Concilier les imaginaires et les attentes
autour des données

Alexandre Coutant et Stéphane Vial

Les données numériques ont infiltré toutes les dimensions de nos quotidiens. Leur omniprésence nous pousse à les considérer comme centrales pour penser et agir, au point où il est devenu primordial pour beaucoup d'accompagner le développement d'une culture de la donnée chez les citoyens et les travailleurs (Casemajor, cet ouvrage). Le domaine de l'innovation est particulièrement sensible à cette représentation, qui renouvelle un imaginaire techniciste, où la définition précise de ce que sont ces données et les techniques permettant de les exploiter restent évoquées de manière imprécise (Crawford, Gray, et Miltner, 2014). Il en résulte que la réalité concrète de ce que l'on peut faire avec ces données demeure complexe à appréhender et ambivalente dans ses conséquences. Pourtant, la littérature s'intéressant à la mise en données de différents projets souligne sans équivoque à quel point les dispositifs ainsi conçus incarnent des visions précises de leurs finalités et des rôles attendus de chaque acteur impliqué, qui généreront une attention accrue envers certaines données plutôt que d'autres. Une forme d'impensé social entourant la mobilisation de ce « matériau » central résulte donc de l'effet combiné de l'omniprésence de discours d'accompagnement plus promotionnels qu'informatifs, d'un engouement peu réfléchi des organisations et des innovateurs pour des solutions exploitant « la donnée », et des faibles prises en compte des usagers finaux dans le processus de conception. Elle

nourrit une littérature critique constatant les décalages entre les promesses d'un monde meilleur par les données et les effets concrets des dispositifs mis en place (Cheung, 2020). Au sein de ces ambivalences, notre chapitre se penche sur la pertinence de la recherche de l'exploitation systématique de grandes masses de données dans le domaine de la cybersanté mentale, qu'il s'agisse de les mobiliser pour tenter de faire sens des activités en ligne liées à ces enjeux, de cadrer des processus d'innovation ou d'imaginer des fonctionnalités à offrir aux futurs usagers. Le domaine de la santé en général et spécifiquement celui de la santé mentale (Sonja *et al.*, 2018) sont effectivement très représentatifs de l'engouement envers les données comme nouveau levier d'amélioration des systèmes de santé et de l'accès aux soins. Mais cet engouement se heurte aux difficultés à exploiter les nombreuses données reliées à la santé, à mettre en place une coordination acceptée entre les acteurs concernés et à tenir compte d'un ensemble de contraintes éthiques particulièrement sensibles (Househ *et al.*, 2017; Casemajor, cet ouvrage).

Précisons qu'il ne s'agit pas de déplorer le fait que les technologies s'accompagnent d'imaginaires, dont il a été démontré qu'ils sont consubstantiels à notre rapport à la technique et qu'ils jouent même un rôle mobilisateur auprès des innovateurs comme des premiers utilisateurs, malgré les représentations discutables qu'ils génèrent. Notre perspective critique n'encourage pas non plus une méfiance *a priori* envers tout nouveau développement technique permettant de définir, formaliser, stocker et exploiter des données (Theviot, 2023). L'objectif consiste plutôt à envisager comment mettre en place un processus d'innovation où une variété d'imaginaires demeure audible, malgré des visions les plus technicistes tonitruantes. Nous rejoignons ici la perspective d'Alain Gras, pour qui « [l]'essence de la technique n'est pas technique : elle doit se chercher dans l'environnement éthique et la vision du monde dans laquelle elle fait sens » (Gras, 2013, p. 19). L'enjeu est double. Il s'agit fondamentalement de s'assurer que tous les acteurs impliqués par les innovations demeurent audibles dans un contexte où ce sont souvent les usagers les plus vulnérables qui se trouvent réduits au silence (Desbois, 2018) et où les logiques organisationnelles tendent à supplanter l'orientation vers le service aux usagers (Mahnke et Nielsen, 2021). En somme, de s'assurer que le processus de conception ne prive pas les premiers intéressés de leur capacité à « faire public », c'est-à-dire, dans le cadre d'innovations autour des données, à teinter les solutions sociotechniques de leurs propres attentes et représentations de manière qu'elles répondent à leurs problèmes et non uniquement à ceux imaginés par les concepteurs ou les

acteurs institués. Il s'agit aussi d'une question pratique. En effet, dans un contexte de ressources limitées, l'attention, légitimement accordée, au potentiel des technologies fortement portées par un imaginaire de la donnée peut masquer des possibilités moins fascinantes, mais ayant démontré leur faisabilité, collant davantage aux attentes exprimées des principaux bénéficiaires ou garantissant la compatibilité avec les modes de fonctionnement existants des acteurs (Gaglio et Mathieu-Fritz, 2018). Si la valeur d'une innovation s'évalue à son appropriation sociale, alors l'histoire de l'innovation suggère que ces critères moins sensationnels constituent pourtant les meilleures garanties de réussite d'un projet.

Le contexte de la cybersanté mentale

Également appelée « santé mentale numérique », la cybersanté mentale (*e-mental health*) est « un domaine de recherche et d'intervention qui s'appuie sur les technologies numériques pour fournir des soins et de l'information en matière de santé mentale, en complément aux soins traditionnels¹ ». Son essor repose sur la prolifération des technologies numériques, dans un contexte de crise de l'accès aux services de soins en santé mentale (McGrath *et al.*, 2018). Chaque année au Canada, une personne sur cinq aura un problème de santé mentale ou une maladie mentale². Pour autant, l'accès aux soins est extrêmement difficile, décourageant et inégalitaire. Face à ce constat, les études évaluant les apports de la cybersanté mentale se multiplient. Outre leur accessibilité en ligne, contribuant à réduire les temps d'attente et à éliminer partiellement les contraintes géographiques³, les innovations dans ce domaine tendent à faciliter l'autogestion ou la télépsychothérapie de certains troubles, comme la dépression ou le stress post-traumatique.

L'engouement pour la cybersanté mentale se traduit sur plusieurs fronts. Sur le front des services offerts, les applications et autres interventions numériques connaissent un vaste essor. Sur celui du système de santé, ce domaine figure comme un moyen d'accès privilégié à de nouvelles connaissances, comme un environnement propice à la surveillance de comportements et

1. Commission de la santé mentale du Canada (2019), « Cybersanté mentale » : <https://commissionsantementale.ca>

2. Association canadienne pour la santé mentale (n.d.), « Faits saillants » : <https://cmha.ca>

3. Commission de la santé mentale du Canada (2019), « Cybersanté mentale » : <https://commissionsantementale.ca>

comme un levier pour la numérisation des services (Drissi *et al.*, 2020). Sur le front commercial, enfin, il est à la source d'une effervescence problématique de création d'applications, qui entretient notamment la confusion entre « bien-être » et santé (Lindenmeyer et d'Ortho, 2020).

Sur ce dernier point, soulignons que le marché des applications s'intéressant à la santé mentale s'avère préoccupant. La littérature souligne en effet un manque de validation scientifique et d'évaluation rigoureuse de celles-ci. Elle soulève aussi de profondes inquiétudes sur la confidentialité et la sécurité des données collectées. Elle révèle les barrières commerciales à l'engagement des utilisateurs et la pérennité variable des applications. Enfin, elle déplore leur médiocre utilisabilité et le faible engagement des utilisateurs qu'elles génèrent (Torous *et al.*, 2018; Larsen *et al.*, 2019). Ainsi, une étude basée sur 93 applications en santé mentale a montré que la rétention générale des utilisateurs est très faible, avec un taux de rétention médian à 15 jours de 3,9 % et à 30 jours de 3,3 % (Baumel *et al.*, 2019). Quarante-six pour cent de ces applications ont moins d'un utilisateur actif mensuel, alors que le coût moyen de développement de telles applications est de 425 000 \$ US, selon une enquête menée par Research 2 Guidance⁴.

Faciliter l'accès aux soins : le projet Mentallys

C'est dans ce contexte peu encourageant que s'inscrit l'élaboration du projet Mentallys⁵. Initialement soutenu par les Fonds de recherche du Québec dans le cadre du programme intersectoriel Audace (2020-2021) et par la Fondation de l'Institut universitaire en santé mentale de Montréal, Mentallys a bénéficié de l'accompagnement de l'Esplanade dans le cadre de l'intégration à sa septième cohorte Collision⁶. Le Conseil de recherche en sciences humaines du Canada l'a ensuite appuyé dans le cadre de ses programmes Engagement partenarial (2022-2023) et Savoir (2023-2026). La diversité de ces soutiens et de leurs finalités – recherche-action, aide à la conception, recherche partenariale, recherche fondamentale – illustre bien la raison d'être de ce projet,

4. *mHealth Developer Economics: How mHealth app publishers are monetizing their apps*. Berlin, Research 2 Guidance, 2018.

5. Pour en savoir plus sur Mentallys : <https://news.mentallys.com>

6. L'Esplanade Québec est un accélérateur de projets québécois spécialisé dans les dimensions sociales et environnementales. Le programme Collision accompagne les lauréats de ce concours pendant trois mois afin de les aider à lancer leur projet. Pour plus d'informations : <https://esplanade.quebec>

qui cherche non seulement à développer une application utile pour deux publics (les patients et les services de santé), mais aussi à en assurer la pérennité par la constitution d'un réseau d'acteurs intéressés à son bon fonctionnement. Saluée par le Prix Partenariat 2022 de l'Université du Québec à Montréal (concours « De l'idée à l'innovation »), cette démarche fondamentalement partenariale implique plusieurs établissements publics de santé du Québec, notamment le Centre intégré universitaire de santé et de services sociaux (CIUSSS) de l'Est-de-l'île-de-Montréal et le CIUSSS de l'Estrie CHUS.

Les ambitions initiales étaient énormes. Il s'agissait de « concevoir, développer, tester, promouvoir et évaluer un modèle d'application mobile unique couvrant à elle seule la prise en charge de première ligne et à grande échelle des troubles dépressifs et anxieux les plus courants » (extrait du dossier de soumission au programme Audace). De manière plus détaillée, on distingue six dimensions innovantes, parmi lesquelles l'accumulation et le traitement à des fins de connaissance de vastes volumes de données : précisément, il est prévu d'« expérimenter le potentiel des technologies d'intelligence artificielle [...] pour la détection précoce de symptômes et troubles mentaux par analyse automatique de grands volumes de messages » (extrait du dossier de soumission au programme Audace). Il peut paraître naïf, avec quelques années de recul, de rêver à de l'exploitation de volumes importants de textes portant sur la maladie mentale quand on considère toutes les difficultés d'accès et les enjeux d'utilisation des applications constatés dans la littérature scientifique pertinente. Ajoutons que le dossier n'informe pas davantage ses lecteurs sur la manière dont cet objectif serait rempli, ce qui tranche avec les exigences extrêmement élevées de démonstration méthodologique usuellement appliquées aux demandes de subventions en sciences humaines et sociales. L'incohérence budgétaire peut enfin retenir l'attention, tant cette dimension à elle seule mobiliserait l'essentiel des fonds obtenus. Mais ce point illustre justement la force des imaginaires de la donnée, puisque ni les chercheurs ni les évaluateurs de la demande n'ont émis de réserves majeures face à l'ambition démesurée de cet objectif initial (il faut dire que ce type de programme subventionnaire vise explicitement à « soutenir des projets audacieux, voire à risque⁷ »). Ce qui nous intéresse va donc concerner les garde-fous qui ont mené à l'abandon rapide de cette option.

En effet, orientés par les commentaires interdisciplinaires sur les risques qu'un tel projet pourrait courir, et notamment par les analyses en sciences

7. Fonds de recherche du Québec, Programme AUDACE : <https://frq.gouv.qc.ca>

sociales décrivant les effets de l'engouement irréfléchi pour le numérique, nous avons veillé à mettre en place une démarche d'innovation capable d'identifier la survenue de tels problèmes. La solution a consisté à organiser le travail selon une logique de codesign pour la cybersanté mentale. Le design pour la cybersanté mentale renvoie à l'ensemble des stratégies de conception et de création centrées sur l'être humain qui permettent de définir la structure, le fonctionnement et la forme d'un service numérique en santé mentale dans une perspective de haute qualité éthique et expérientielle (Vial, Boudhraâ, et Dumont, 2022). Le codesign en constitue une application particulière, davantage attentive à conserver l'implication de toutes les parties concernées par l'innovation tout au long du processus de conception. L'objectif poursuivi est de faire de ce processus une « conversation sociale » (Manzini, 2015) où des « designers experts » (chercheurs, praticiens et étudiants en design) coordonnent *tout au long du projet* les échanges avec des personnes non formées au design (« design diffus » porté par des patients, familles de patients, pairs aidants, psychologues, psychiatres, professionnels de la santé et chercheurs de diverses disciplines). Le processus itératif en découlant fait du projet un objet collectif très évolutif et intégrant le « design diffus » qui émane des propositions et remarques des participants.

L'identification des participants s'est appuyée sur la littérature qui documente les processus de codesign et recommande d'inclure systématiquement les utilisateurs finaux (le projet se destinait d'emblée aux patients et aux professionnels). Leur recrutement a été facilité par l'appartenance du chercheur principal au Centre de recherche de l'Institut universitaire en santé mentale de Montréal, au sein du CIUSSS de l'Est-de-l'Île-de-Montréal. Enfin, la mobilisation d'expertises disciplinaires a aussi bénéficié de la revue de littérature interdisciplinaire et des occasions de collaboration permises par l'Université du Québec à Montréal grâce au soutien d'une chaire de recherche stratégique en design pour la cybersanté mentale et à celui de son Institut Santé et société regroupant toutes ses facultés.

Détails sur le processus de codesign

Le processus de codesign de Mentallys a connu quatre grandes phases. Premièrement, de juin à décembre 2020, les chercheurs et partenaires ont formalisé les thèmes et les pistes sur lesquels mettre en œuvre le codesign. Deuxièmement, de janvier à juin 2021, six cyberateliers de codesign ont eu lieu : deux avec des cliniciens uniquement (n=7), deux avec des pairs

aidants (n=5) et des personnes ayant une expérience vécue des problèmes de santé mentale (n=2), et deux avec l'ensemble de ces personnes (n=14). Six facilitateurs y ont participé. Cette phase de codesign reposait sur la pensée divergente (Cross, 2008) dans le but de générer des idées et de répertorier les caractéristiques souhaitables pour l'application Mentallys par l'entremise de prototypes imaginaires et de conversations sociales avec différents types d'utilisateurs finaux (Vial *et al.*, 2023). Troisièmement, un travail préliminaire axé sur le modèle d'affaires et la pérennisation de la solution s'est déroulé d'octobre à décembre 2021 dans le cadre de l'intégration à la septième cohorte Collision de l'Esplanade, un accélérateur montréalais de projets entrepreneuriaux à portée sociale. Cette phase intermédiaire a dévoilé d'importants enjeux de viabilité permettant d'orienter la suite du codesign. Et quatrièmement, une seconde série d'ateliers de codesign a eu lieu, mais cette fois sous la forme particulière de dix séances de codesign en dyade. Cette phase de codesign reposait cette fois-ci sur la pensée convergente (Cross, 2008) dans le but de valider les idées qui peuvent satisfaire les utilisateurs finaux dans le monde réel grâce à des prototypes réels. Grâce au format dyadique, il a été possible d'explorer des sujets sensibles et confidentiels en instaurant un climat de confiance plus aisément qu'en groupe. Pour cette raison, l'accent a porté sur l'analyse qualitative détaillée, d'où l'implication de seuls trois participants (n=3) et un facilitateur (Vial *et al.*, 2023).

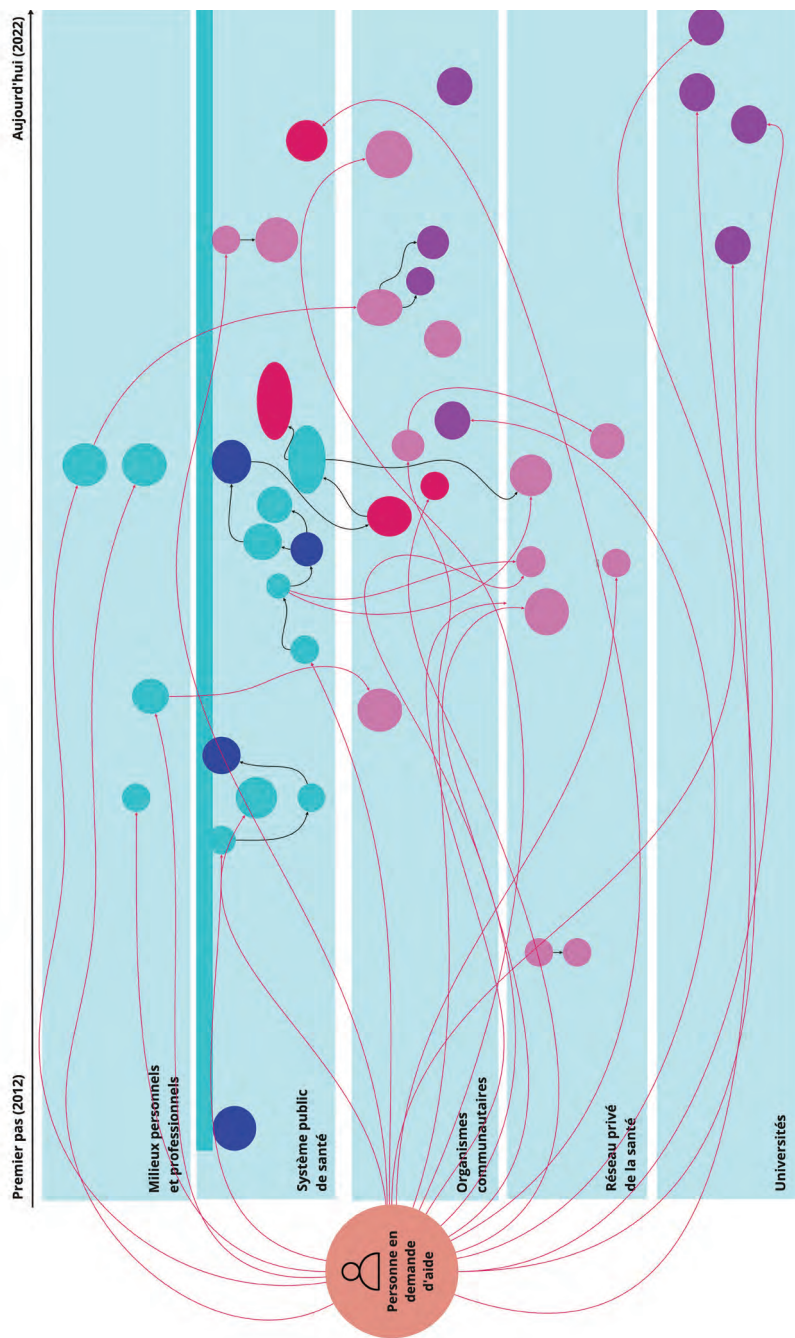
La conversation sociale que favorise le codesign a permis d'explorer en profondeur la désirabilité de la solution Mentallys en combinant une phase de codesign en groupes axée sur la pensée divergente (générer des idées multiples) avec une phase de codesign en dyades axée sur la pensée convergente (choisir et raffiner les idées clés). Les fonctionnalités retenues pour le concept final de l'application dérivent directement de ce processus de codesign et elles ont été dévoilées publiquement en novembre 2023 lors d'une activité avec les parties prenantes. Soumises à l'analyse et au commentaire des cochercheurs, elles ont ensuite fait l'objet d'évaluations de faisabilité par le partenaire technologique responsable du développement logiciel, et elles seront prochainement mises à l'épreuve lors d'une expérimentation réelle avec les milieux preneurs partenaires (services publics chargés de l'accès aux soins de santé mentale).

Les évolutions du projet guidées par les usagers

« Je m'attends à être mis directement en contact avec un intervenant. Je m'attends donc qu'on me demande mes besoins pour me mettre en relation avec le bon professionnel. » Ce verbatim d'un patient participant aux ateliers de codesign résume l'utilité de la démarche. Sa force consiste effectivement à redonner de la vigueur aux autres discours que ceux portés par les promoteurs de solutions « guidées par les données ». Dans le cas de Mentallys, la chose a servi à raviver systématiquement la conscience de l'urgence de fournir avant tout des services pour orienter les patients et à consigner le parcours de soin de manière que les différents professionnels de la santé puissent en prendre connaissance. La définition des données s'en trouve changée : à un ensemble imprécis de traces peu discriminées, nécessairement massives s'il s'agit de nourrir une IA, s'oppose une autre forme de richesse, celle des expériences constituant le parcours de soin et de difficulté d'en faire un tout cohérent et partageable de manière détaillée avec les différents professionnels rencontrés. En effet, les échanges ont permis de reconstituer des trajectoires d'une complexité incroyable, que les professionnels de santé ont peu de moyens de consigner et de centraliser, sauf à consacrer un temps de consultation, déjà rare, à le reconstituer. Un exemple de parcours décrit par l'une des participantes, Geneviève, reproduit ci-dessous, donne la mesure de cette complexité (figure 4.1). Il s'agit d'un parcours de soin qui s'étend de 2014 à 2021 et qui inclut plus d'une trentaine d'événements de soin dans les réseaux public, communautaire et privé. Les points bleus représentent les hospitalisations et les suivis psychiatriques. Les points turquoise désignent les initiatives de Geneviève, personnelles ou suivant une suggestion, mais ne découlant pas d'un transfert obligatoire ou d'une requête de médecin. Les points roses marquent des initiatives personnelles que des médecins traitants auraient dû proposer, mais que Geneviève a cherchées d'elle-même comme solutions de rechange pour se rétablir. Les points rouges illustrent des prises d'actions et les investissements plus approfondis de Geneviève pour se rétablir. Enfin, les points mauves indiquent les acquisitions et partages de connaissances qui aident Geneviève à se rétablir. Par souci de confidentialité, l'image omet le contenu de chaque étape du parcours de soin, son but étant essentiellement de donner à voir le caractère labyrinthique, long et décourageant d'un tel parcours. Chaque ligne reliant une étape à l'icône de profil de la participante représente une démarche entreprise pour trouver de l'aide.

FIGURE 4.1

Exemple de trajectoire d'une patiente cherchant à obtenir des soins en santé mentale (image produite par la patiente lors d'un atelier de codesign à partir d'un gabarit fourni)



Ce sont aussi des difficultés très concrètes qui émergent prioritairement du processus que nous avons mené et qui viennent rappeler qu'au-delà du traitement de données, il s'agit déjà de rendre des services. Par exemple : la mise en relation avec un professionnel adapté à ses besoins, des conseils et outils d'autodiagnostic, ainsi que des fonctionnalités simplifiant le suivi administratif de son parcours. Les échanges lors des ateliers ont révélé la convergence des attentes et les accords entre acteurs concernés autour des enjeux prioritaires, même si quelques divergences mineures ont pu apparaître ici et là.

Insistons sur l'utilité de la démarche donnant aux publics une voix pour exprimer les problèmes qui les préoccupent tout au long du processus, car les multiples épreuves économiques, techniques, organisationnelles et juridiques qui ont jalonné le projet Mentallys auraient aisément pu détourner ses responsables de la nécessité de développer ces services.

Dès la première phase d'ateliers de codesign axés sur la désirabilité, les fonctionnalités envisagées ont beaucoup évolué, pour se recentrer sur trois grandes dimensions : trouver de l'aide (trouver facilement des professionnels et des ressources, clavarder avec des pairs aidants, etc.) ; prendre soin de soi (autosoins, outils de soutien à l'autogestion, bibliothérapie⁸, vidéos, etc.) ; documenter son parcours de soin (gestion et partage du parcours de soin, ses intervenants, histoire et ses notes personnelles).

À nouveau, ce qu'on entend par données est clair, et très divergent de ce vers quoi mènent les imaginaires de la donnée. *Trouver de l'aide* se distingue de la documentation toujours plus poussée des patients pour rappeler la nécessité de jeux de données précis concernant les services, les ressources et les soutiens disponibles ainsi que les façons d'y accéder. *Prendre soin de soi* rappelle la masse de contenus accessibles pour accompagner le patient, qui soulève des enjeux de fiabilité des informations et du classement selon la pertinence pour ce dernier. *Documenter mon parcours de soin* concerne la nécessité déjà évoquée de pouvoir rendre compte d'un cheminement complexe et opaque.

Les enjeux sociotechniques demeurent majeurs : comment assurer la pertinence des mises en relation suggérées sur la base des informations disponibles sur l'utilisateur ? Comment tenir à jour la liste des sources et outils rendus accessibles en s'assurant de leur validation scientifique ? Comment assurer la collaboration autour de l'historique enregistré du patient en respectant les enjeux de vie privée et de sécurité des informations ? Comment

8. La bibliothérapie concerne tous les emplois thérapeutiques de la lecture en santé mentale.

s'assurer d'une prise en main facile de l'application en envisageant des contextes d'usage parfois fortement perturbés? Cependant, l'exploration et la sélection des solutions à ces enjeux au cours du codesign se trouvent simplifiées par la précision des attentes des futurs usagers, qui permet d'évaluer les intérêts et les coûts de ces solutions.

Tirer parti des méthodes de conception s'appuyant sur les usagers

Il est essentiel de prendre du recul et de réfléchir à la pertinence du développement de certaines solutions technologiques. La collaboration interdisciplinaire est particulièrement fructueuse pour répondre à la question précise de la manière d'innover en composant avec les imaginaires accompagnant notre domaine d'intervention. Elle concerne la complémentarité entre les sciences du design et les sciences de l'information et de la communication.

La perspective qui sous-tend le projet relève principalement de la première discipline. Au sein de ce grand ensemble de démarches s'intéressant aux manières de concevoir, elle s'inscrit dans celles centrées sur les utilisateurs, revendiquant une « humanisation innovante des technologies » relevant de la grande famille des pratiques de design centré sur l'humain (Vial, Boudhraâ, et Dumont, 2022). Cette humanisation renvoie à la capacité revendiquée et éprouvée du design à créer des technologies faciles à utiliser sur le plan cognitif et attractives sur le plan émotif, ainsi qu'à sa capacité à matérialiser en elles des valeurs morales. En assumant cette ambition, la discipline s'éloigne d'une vision instrumentale, neutre et désincarnée de l'innovation pour assumer son rôle dans le changement social. L'approche du codesign incarne cet engagement par des méthodes cherchant à générer une « créativité collective appliquée à l'ensemble du processus de design », découlant de la « créativité des designers et des personnes n'ayant pas reçu de formation en design qui travaillent ensemble dans le processus de conception » (Sanders et Stappers, 2008). En somme, selon cette approche, l'utilisateur n'est pas un *sujet de conception*, mais un *partenaire de conception*, comme le disent Sanders et Stappers. Un vocabulaire qui fait écho aux conceptions participatives des publics évoquées précédemment.

Mais comment s'assurer de bien comprendre ce partenaire? C'est ici que les sciences de l'information et de la communication apportent leur complémentarité, grâce à leur robuste lecture critique des phénomènes sociotechniques (Proulx, 2005), documentant notamment les enjeux spécifiques aux données en santé (Clavier et Paganelli, 2019). Le défi de convertir

cette compréhension en intervention demeure cependant central pour cette discipline plus orientée vers la recherche fondamentale, dès lors qu'il s'agit de participation à des processus d'innovation. La convergence avec les sciences du design est devenue possible par l'emploi de la méthode CAUTIC⁹, développée en sociologie des usages dès les années 1990 (Mallein et Toussaint, 1994) et formalisée pour le grand public dans un manuel en 2014 (Pizelle *et al.*, 2014). Cette méthode, structurée pour s'inscrire dans les temporalités des projets¹⁰, s'inspire d'ailleurs en partie de travaux en design. À ce titre, son intérêt consiste moins dans la concrétisation des solutions, que les expertises en design ont bien davantage approfondie, que dans le recensement des dimensions pertinentes à observer lors de la consultation des futurs usagers et de leurs contextes d'usage, qui se trouvent formalisées jusque dans les grilles de codage des entretiens et des ateliers menés. Outre toutes les petites routines, craintes, frustrations ou satisfactions qui nous renseignent sur l'épaisseur des pratiques sociales au sein desquelles s'insérera l'innovation, c'est aussi un changement de perspective face à la vision « destructrice » de l'innovation. Ainsi, « il s'agit donc de s'inscrire dans une logique d'hybridation de la technique nouvelle aux techniques existantes (via des ajustements et des négociations avec les usages existants), plutôt que de penser que la technique nouvelle va s'imposer de fait et "révolutionner" les usages en se substituant totalement aux pratiques anciennes (logique d'imposition) » (Pizelle *et al.*, 2014, p. 40).

Reconnaître les données pertinentes

Notre expérience permet de soutenir que l'approche découlant de la mobilisation croisée des sciences du design et des sciences de l'information et de la communication constitue un efficace garde-fou contre un ensemble d'imaginaires technicistes dominants. Elle ne saurait pour autant garantir le succès d'une innovation. Son utilité, bien plus modeste, réside plutôt dans sa contribution à la reconnaissance des écueils les plus évidents. Un apport circonscrit, mais bien utile si l'on se réfère aux nombreux cas d'échecs d'innovations ou de dispositifs venant ajouter aux contraintes des usagers, que la littérature évoquée plus haut documente.

9. Pour « conception assistée par les usages pour les technologies, l'innovation et le changement ».

10. Les auteurs affirment ne pas avoir besoin de plus de deux mois pour mettre en œuvre leur démarche de vérification de la désirabilité d'une idée d'innovation.

L'expérience de Mentallys met en évidence deux de ces écueils. Le premier concerne le périmètre de ce qui est entendu par « données », que le codesign ouvre à nouveau au-delà des automatismes de pensée pouvant découler des imaginaires de la donnée. C'est ainsi un ensemble plus varié d'éléments potentiellement importants qui a été colligé lors des ateliers. Le second concerne la sélection des solutions en fonction des données considérées comme importantes. Ici, garantir la parole de toutes les parties prenantes leur permet de jouer leur rôle de publics sonnante l'alerte au sujet des problèmes qu'ils rencontrent et proposant les solutions qu'ils imaginent ou attendent. Les avantages sont doubles : guider les choix d'innovation selon des critères en phase avec le problème tel qu'il se pose socialement et soumettre à nouveau les solutions toutes faites, charriées par les imaginaires, à l'épreuve de leur pertinence pour les publics (Millerand, Millette et Seraiocco, cet ouvrage ; Delias, cet ouvrage).

Ce type de projet encourage ainsi une évolution des méthodologies de conception de manière à intégrer une phase explicite d'interrogation des données sur lesquelles fonder les choix de conception à l'aune des problèmes sociaux identifiés par les publics. En effet, d'aucuns soulèvent depuis longtemps la nécessité de comprendre et d'intégrer les usagers, comme leurs contextes d'usage, dans les projets d'innovation. Dans les sciences du design, cela fait partie des bases de tout projet de conception (Norman, 1988). Cette conscience a même mené à l'idée d'une maîtrise d'usages, qui viendrait éclairer les traditionnelles fonctions liées aux maîtrises d'œuvre et d'ouvrage en se faisant la porte-parole des attentes et conditions d'appropriation des usagers tout au long du processus d'innovation (Coutant, 2013). Les recherches contemporaines témoignent de la permanence de leur invisibilisation et des conséquences négatives en découlant, sauf dans le milieu des *startups* où la culture design est fortement présente. Ce paradoxe nous paraît s'expliquer au moins en partie par la reconnaissance encore inachevée des approches orientées vers les usagers et des « cultures de conception » qu'elles véhiculent (Bonnet, 2021). En effet, une logique d'innovation ainsi que des méthodes ou des outils existent et ont fait leurs preuves. Le vocabulaire associé aux usagers et au design a aussi fait son chemin dans les discours portant sur l'innovation. Le rôle de l'innovation, sociale comme technique, dans la mise à jour de solutions à des problèmes formalisés par les publics est régulièrement proposé. Cependant, beaucoup d'usages superficiels demeurent (par exemple, employer les outils de tests sans avoir adopté la logique de dialogue avec les usagers, évoquer les usagers selon l'idée que l'on s'en fait sans vérifier sa réalité, etc.)

ou l'écoute qui leur est accordée reste limitée (reconnaissable cette fois par l'intervention tardive des designers, par le pouvoir décisionnel limité des usagers ou par une meilleure écoute obtenue de la part d'autres professions, qu'elles relèvent de la finance, du marketing, de la gestion ou bien entendu de disciplines techniques). C'est le cas notamment d'organisations fortement structurées ou au degré d'expertise très élevé, où les cultures du génie et de la planification sont historiquement très ancrées; c'est aussi celui de secteurs reposant beaucoup sur l'expertise numérique, qui demeurent soumis aux discours de promotion et où les professionnels pouvant réaliser techniquement les dispositifs conservent un pouvoir d'orientation important (Coutant et Domenget, 2015). Dans ces contextes encore très nombreux, faute d'avoir intégré des manières concrètes de mettre en œuvre les approches orientées vers les usagers dans les projets, il apparaît logique que la sensibilité à ces dernières en reste au stade de la bonne intention. Face à ce constat, le type de démarche de design et de codesign développé avec le projet Mentallys nous paraît une bonne occasion, dans les sciences de l'information et de la communication, d'instrumenter la maîtrise d'usages pour favoriser sa plus grande prégnance dans les projets innovants. Réciproquement, au sein d'un projet ancré en sciences du design comme Mentallys, la collaboration avec les sciences de l'information et de la communication est une opportunité de déconstruire les dogmes technicistes (fut-ce celui de l'intelligence artificielle) afin de faire de la situation d'innovation qu'est Mentallys l'occasion de définir « une nouvelle trajectoire imaginaire » (Gras, 2013, p. 44) pour la santé mentale numérique.

Références

- BAUMEL, Amit *et al.*, « Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis », *Journal of Medical Internet Research*, vol. 21, n° 9, 2019.
- BONNET, Fabien, « Cultures de conception – Entre “progrès”, “innovation” et “stratégie”, quels signes, quels dispositifs et quels modèles d'organisation pour se projeter aujourd'hui? », *Revue française des sciences de l'information et de la communication*, vol. 23, 2021.
- CHEUNG, Sarah, « Disambiguating the benefits and risks from public health data in the digital economy », *Big Data & Society*, vol. 7, n° 1, 2020.
- CLAVIER, Viviane et PAGANELLI, Céline, « L'ouverture des données de la recherche dans le cadre d'un projet pluridisciplinaire entre SIC et informatique: le cas des médias sociaux de santé », *Études de communication. Langages, information, médiations*, n° 52, 2019, p. 117-136.

- COUTANT, Alexandre, « Accompagner le développement de services sociotechniques », dans Alain KIYINDOU, Christian LE MOËNNE, et Béatrice VACHER (dir.), *Communication et débat public : les réseaux numériques au service de la démocratie ?* Paris, L'Harmattan, 2013, p. 155-164.
- COUTANT, Alexandre et DOMENGET, Jean-Claude, « Le communicateur bousculé par le numérique : quelles compétences à transmettre ? », *Revue Communication & professionnalisation*, n° 3, 2015, p. 5-20.
- CRAWFORD, Kate, GRAY, Mary L. et MILTNER, Kate, « Critiquing Big Data: Politics, Ethics, Epistemology », *International Journal of Communication*, n° 8, 2014, p. 1663-1672.
- CROSS, Nigel. *Engineering Design Methods Strategies for Product Design*. 4^e édition. Hoboken, New Jersey, U.S. John Wiley & Sons ; 2008.
- DESBOIS, Dominique, « La télémédecine, un instrument susceptible de corriger les inégalités de santé ? », *Terminal. Technologie de l'information, culture & société*, n° 122, 2018.
- DRISSI, Nidal *et al.*, « Connected Mental Health: Systematic Mapping Study », *Journal of Medical Internet Research*, vol. 22, n° 8, 2020.
- GAGLIO, Gérald et MATHIEU-FRITZ, Alexandre, « Les pratiques médicales et soignantes à distance », *Réseaux*, n° 207, 2018, p. 9-24.
- GRAS, Alain, *Les imaginaires de l'innovation technique : regard anthropologique sur le passé dans la perspective d'un avenir incertain*, Paris, Éd. Manucius, coll. « Modélisations des imaginaires », 2013.
- HOUSEH, Mowafa S. *et al.*, « Big Data, Big Problems: A Healthcare Perspective », *Studies in Health Technology and Informatics*, vol. 238, 2017, p. 36-39.
- LARSEN, Mark Erik *et al.*, « Using science to sell apps: Evaluation of mental health app store quality claims », *Npj Digital Medicine*, vol. 2, n° 1, 2019, p. 1-6.
- LINDENMEYER, Cristina, et D'ORTHO, Marie-Pia, *Santé connectée*, éditions CNRS, 2020.
- MAHNKE SKRUBBELTRANG, Martina et NIELSEN, Mikka, « EHealth platforms as user-data communication: Examining patients' struggles with digital health data », *Nordicom Review* vol. 42, n° 4, 2021, p. 45-58.
- MALLEIN, Philippe et TOUSSAINT, Yves, « Technologies de l'information et de la communication : sociologie pour la Conception Assistée par l'Usage », *Communications & stratégies, Cahiers de l'IDATE* 5, vol. 15, 1994, p. 77-99.
- MANZINI, Ezio, *When everybody designs: An introduction to design for social innovation*, Cambridge, The MIT Press, 2015.
- MCGRATH, Patrick *et al.*, *Trousse d'outils pour l'adoption de services de cybersanté mentale*, Ottawa, Commission de la santé mentale du Canada, 2018.
- NORMAN, Don, *The Design of Everyday Things*, New York, Basic Books, 2013.
- PIZELLE, Pascal *et al.*, *Innover par les usages : I(nnovation) = [T(echno) + H(umain)] x M(éthodes)*, Grenoble, Éditions d'Innovation, 2014.
- PROULX, Serge, « Penser les usages des technologies de l'information et de la communication aujourd'hui : enjeux-modèles-tendances », dans VEIRIA, Lise et PINÈDE,

- Nathalie (dir.), *Enjeux et usages des TIC: aspects sociaux et culturels*, vol. 1, Bordeaux, Presses universitaires de Bordeaux, 2005, p. 7-20.
- REY, Olivier, *Quand le monde s'est fait nombre*, Paris, Stock, 2017.
- SANDERS, Elizabeth B.-N. et JAN STAPPERS, Pieter, « Co-creation and the new landscapes of design », *CoDesign*, vol. 4, n° 1, 2008, p. 5-18.
- SONJA, Marjanovic *et al.*, « Understanding value in health data ecosystems: A review of current evidence and ways forward », *Rand Health Quarterly*, vol. 7, n° 2, 2018.
- THEVIOT, Anaïs, *Gouverner par les données? Pour une sociologie politique du numérique*, Lyon, Éditions ENS, 2023.
- TOROUS, John *et al.*, « Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements », *BMJ Mental Health*, vol. 21, n° 3, 2018, p. 116-119.
- VIAL, Stéphane, BOUDHRAÂ, Sana et DUMONT, Mathieu, « Human-Centered Design Approaches in Digital Mental Health Interventions: Exploratory Mapping Review », *JMIR Mental Health*, vol. 9, n° 6, 2022.
- VIAL, Stéphane *et al.*, « Developing A Mobile App With a Human-Centered Design Lens to Improve Access to Mental Health Care (Mentallys Project): Protocol for an Initial Co-Design Process ». *JMIR Research Protocols* 12, n° 1, 2023.

TROISIÈME PARTIE

La culture

CHAPITRE 5

Les cultures de données

Genèse d'une notion, enjeux épistémologiques et implications pratiques

Nathalie Casemajor

En 2014, Microsoft a lancé un cycle de formation à « la culture des données » (*data culture*) destinée aux analystes de données, aux architectes de données et aux autres spécialistes de la science des données. La « culture des données » y est présentée comme un champ de compétence et une mentalité à faire partager par l'ensemble des membres de l'organisation. « Aujourd'hui, les organisations doivent adopter une culture axée sur les données, sans quoi elles risquent de finir à la traîne¹ », affirme Microsoft dans la promotion de sa plateforme infonuagique Azure. L'expression de « culture des données » (ou « culture de la donnée ») a ainsi connu une diffusion accélérée depuis les années 2010. Elle désigne au sens large un ensemble de dispositions permettant de s'orienter et de travailler dans un monde de plus en plus médiatisé par les données numériques.

Malgré un usage croissant, seule une poignée de travaux se sont véritablement attelés à définir et à conceptualiser cette notion. En quoi l'ensemble des relations aux données numériques peut-il constituer une culture ? L'objectif de ce chapitre est d'interroger la notion de culture de données pour en éclairer la genèse, les enjeux épistémologiques et les implications pratiques. Dans ce cadre, analyser la relation entre données et culture permet

1. « Organizations today must embrace a data-driven culture or risk being left behind » (<https://azure.microsoft.com/>).

d'éclairer les modes de composition des « publics de données » à l'intérieur d'une formation discursive plus large qui est celle des cultures de données. J'adopte ici la forme doublement plurielle de *cultures de données* pour souligner la variété des cultures et des données concernées.

Cette réflexion s'appuie principalement sur une revue critique de la littérature francophone et anglophone. Elle est enrichie par des exemples tirés du secteur professionnel des arts et de la culture au Québec. Ces exemples proviennent d'entretiens de groupe exploratoires menés en 2021 avec l'objectif d'identifier les acteurs clés et les enjeux liés à la culture de données dans ce secteur². Dans ce cadre, nous avons mené trois entretiens collectifs auprès de 16 personnes exerçant toutes une activité professionnelle dans les domaines des bibliothèques, des archives, du spectacle vivant, du livre, du cinéma et de la musique (Sirois *et al.*, 2021). En croisant ces données empiriques exploratoires avec une analyse approfondie de la littérature, je propose des orientations pour analyser les pratiques professionnelles relatives à la culture de données, orientations approfondies dans un travail empirique subséquent (Casemajor, à paraître), où elles ont été appliquées à l'étude du secteur des bibliothèques et des archives. Par l'interrogation critique de la notion de culture de données et de ses présuppositions, ce chapitre vise à contribuer à sa conceptualisation dans le champ des études critiques de données et des études culturelles.

Enjeux : interroger les présupposés des cultures de données

La notion de cultures de données est principalement associée à deux types de données : les données ouvertes et les données massives. Ainsi, la Ville de Montréal a lancé en 2020 sa nouvelle plateforme de données ouvertes avec l'ambition de « développ[er] une culture de la donnée à Montréal³ » (Millerand et Meunier, cet ouvrage). La création en 2018 à Montréal d'un *Pôle sur les données massives* en culture visait quant à elle à « développer une culture de la donnée » dans les organisations culturelles⁴. Situés dans différentes logiques d'exploitation des données (l'une à vocation citoyenne et l'autre

2. Dans le cadre du projet de recherche *La mise en données des publics de la culture. Portrait des pratiques actuelles* (2018-2021), dirigé par Guillaume Sirois, avec Nathalie Casemajor et Guy Bellavance, financé par InventT (Université de Montréal).

3. Ville de Montréal (2020), « Dévoilement du nouveau site web des données ouvertes » (<https://donnees.montreal.ca>).

4. Synapse C, « Outils » (<https://synapsec.ca/>).

davantage commerciale), ces deux discours ont en commun l'idée que « la culture des données » doit être développée.

De fait, l'acquisition d'une culture de données figure dans les discours qui en font la promotion comme une nécessité pour s'adapter aux environnements numériques émergents. L'attrait pour cette nouvelle « culture », voire l'injonction à l'adopter, est alimenté par une croyance en l'efficacité technique et économique d'un modèle d'activité gouverné par l'exploitation des données numériques. Les données sont en effet devenues le moteur d'un nouveau paradigme de modernisation et d'innovation socioéconomique. Ce paradigme est modelé par le succès phénoménal des grandes plateformes du Web dont l'activité se fonde sur une triple logique d'extraction, d'accumulation et de valorisation des données. Un tel succès rend désirable l'adoption d'une culture de données similaire par laquelle toute organisation pourrait reproduire ce modèle et accroître ainsi sa valeur économique et sociale.

Dans le monde du travail, l'adoption d'une culture de données pose donc la question d'une transformation des pratiques professionnelles pour répondre à ce nouveau paradigme organisationnel de l'innovation par les données. Cette transformation se traduit par l'émergence de nouveaux métiers et de nouvelles compétences qui permettraient d'intégrer les données numériques dans les orientations stratégiques et dans les processus opérationnels des organisations. Poussé par la vague ascendante des données massives et de l'intelligence artificielle, le domaine de compétence qui s'est développé au croisement de l'analyse des données et de l'informatique fait office d'incontournable dans ces discours. Pour des entreprises comme Microsoft, l'injonction à développer une culture de données revêt aussi un intérêt commercial : diffuser cette vision de l'innovation et les compétences qui lui sont liées nourrit les ventes des produits et services mis en marché par les industries de la donnée (bases de données, serveurs, applications d'apprentissage machine, tableaux de bord de visualisation des données).

Or ces discours sont chargés de présupposés normatifs rarement interrogés. Ils reposent sur les prémisses suivantes : la *tabula rasa*, autrement dit le fait que les professionnels seraient exempts de toute culture de données préalable ; l'*urgence* à acquérir « une culture » des données face à la menace de perdre sa pertinence et de disparaître ; et l'*efficacité* de cette culture qui, en optimisant la prise de décision, garantirait la croissance et le succès des organisations. Étant donné la diffusion croissante de ce discours dans les secteurs privé et public, voire associatif, il est important d'interroger les valeurs qui le sous-tendent.

À travers la notion de *cultures* de données se noue en effet l'enjeu politique du rapport entre savoir (les informations et connaissances fournies par l'analyse des données numériques) et pouvoir (la prise de décision dans les organisations, l'injonction à se conformer à un nouveau modèle d'activité, les potentielles atteintes à la vie privée des usagers). Ces enjeux concernent non seulement les professionnels sommés d'adapter leurs pratiques de travail, mais aussi leur relation avec les récepteurs des biens et services produits par cette nouvelle organisation du travail : clients, usagers ou publics appréhendés par leur « mise en donnée ». Pour mieux comprendre les enjeux de cette notion et ses implications pratiques dans la mise en données de la société, il est nécessaire d'en creuser la conceptualisation. Dans quelle mesure les différents contextes d'émergence de la notion influencent-ils les approches conceptuelles des cultures de données ? Et en retour, quels sont les différents effets heuristiques et pratiques de ces conceptualisations ?

Problématiser les cultures de données

L'expression anglaise *data culture* a émergé dès les années 1960 avec le boom de l'industrie informatique. Tombée en désuétude dans les années 1980 et 1990, elle est alors supplantée par celle d'« *information culture* », comme le montre une comparaison des trajectoires des deux expressions dans Google Ngram Viewer. L'expression a suscité un nouvel intérêt dans les années 2010 en même temps que la croissance exponentielle du phénomène de « mise en données » (*datafication*) de la société, avec en corollaire l'émergence de nouveaux enjeux liés à la « littératie des données ». L'étude des cultures de données avoisine ces autres notions tout en s'en distinguant.

S'ils sont connexes, les termes de donnée et d'information n'ont pas le même sens pour autant. Une donnée est un élément de description basique de la réalité qui peut être codé sous forme numérique et traité par des systèmes automatisés. C'est seulement après avoir été interprétée et mise en contexte qu'une donnée prend la valeur d'information. Quant à la littératie des données, elle renvoie, au sens large, aux connaissances et aux compétences nécessaires pour comprendre et utiliser les données, et pour réfléchir à leurs enjeux dans l'environnement numérique. Alors que la notion de littératie des données concerne davantage les apprentissages engagés sur le plan individuel, la notion de cultures de données est plus large : par culture, on entend ici *un ensemble de répertoires de sensibilités et de significations collectives partagés au sein d'un groupe*.

C'est à l'intérieur de ce cadre culturel plus large qu'on peut situer la notion de *public de données* : les manières d'imaginer les publics comme entités collectives au sein d'environnements numériques (Litt et Hargittai, 2016), les manières de les constituer en pratique (par la collecte de données, la médiation algorithmique et les mesures d'audience ; Gillespie, 2014), les manières dont ils s'assemblent effectivement dans les environnements numériques (Marres, 2005), et les manières de se confronter à leur agentivité (Møller Hartley *et al.*, 2021) – toutes dépendent en bonne partie de schémas d'action et d'interprétation d'ordre culturel.

Parmi les précurseurs de l'étude des cultures de données, le sociologue des sciences et technologies Geoffrey Bowker (2000) s'est intéressé dès la fin des années 1990 à l'existence de « cultures locales de données » (*local coding culture*) dans les infrastructures d'information. Il entendait par là des schémas locaux de classification de l'information propres à différentes disciplines scientifiques qui collaborent dans des projets de recherche où une multitude de données de différentes provenances (une « diversité des données » ou *datadiversity*) doivent être rassemblées dans un même système d'information.

Dans les travaux contemporains, on peut identifier deux principaux courants ayant creusé le sujet des cultures de données : d'une part, les études critiques des données initiées dans le champ anglophone (*critical data studies*) et, d'autre part, les études françaises en sciences de l'information et de la communication. S'y ajoutent de manière connexe les travaux anglophones en science de l'information (*information studies*, sous l'angle de la gestion des systèmes d'information), en science de la gestion et en économie de la culture.

Il reste cependant plusieurs lacunes à combler dans la conceptualisation de cette notion. Les définitions existantes tendent à se limiter à des objets restreints et à ignorer les travaux menés dans d'autres champs de recherche. En outre, peu de croisements existent entre les conceptualisations anglophones et francophones, qui se développent en parallèle.

Analyse croisée des définitions existantes

Si de nombreux travaux universitaires utilisent la notion de culture(s) de données, très peu la problématisent réellement pour en proposer une définition explicite et argumentée. L'examen des travaux existants a permis d'identifier quatre principales définitions de la notion de cultures de données

(Tableau 5.1). Ces résultats sont le fruit d'une interrogation systématique des bases de données Scopus, Isidore et Cairn en novembre 2022, couplée à une recherche exploratoire sur Google Scholar et à un processus de « boule de neige ». Ces définitions ont été sélectionnées pour la profondeur de leur élaboration épistémologique, au-delà d'un simple usage de l'expression non problématisé. Celles des auteurs anglophones (Bates, Burgess) s'ancrent dans les études critiques des données (Kitchin 2014), tandis que celles des auteurs francophones (Labelle, Lehmans) s'inscrivent dans le prolongement d'une perspective critique sur la culture de l'information, issue des sciences de l'information et de la communication.

Les quatre définitions se rejoignent autour de quelques éléments centraux, mais elles se distinguent aussi par leurs objets, qui viennent influencer le cadrage de la notion. Toutes concordent sur les points suivants : il existe une *diversité* des cultures de données ; elles sont *situées* dans des contextes sociaux, matériels et culturels particuliers qui contribuent à leur donner forme ; et elles sous-tendent un engagement dans l'*action* et dans la production de *signification* par une multiplicité d'acteurs (entreprises, pouvoirs publics, société civile, travailleurs, usagers ordinaires).

Dans le champ des sciences de l'information et de la communication, les travaux qui ont abordé de front cette notion croisent les domaines de la documentation, de l'éducation et de la communication organisationnelle. Lehmans (2017, 2018) et Labelle *et al.* (2017) ont en commun une perspective centrée sur les transformations des modèles organisationnels et sur les nouvelles compétences professionnelles liées aux données : leur définition insiste sur le partage d'une vision commune par les membres de l'organisation et sur la compréhension critique du sens de l'action.

Labelle *et al.* ont élaboré leur définition dans un cahier visant la formation professionnelle produit par la Fondation Internet Nouvelle Génération (la FING, une fondation française spécialisée dans l'accompagnement des transformations numériques des organisations) ; elle est donc davantage orientée vers le déploiement et l'appropriation d'une « nouvelle » culture des données dans les organisations. Lehmans ancre plutôt ses réflexions dans l'étude des initiatives de données ouvertes au sein des collectivités territoriales pour montrer que la transformation des pratiques organisationnelles n'est pas linéaire : elle génère des conflits de valeur, des contradictions, des bricolages et des détournements de la part des membres des organisations.

Du côté des études critiques des données, les travaux qui se sont attelés à définir la notion s'inscrivent dans l'étude des phénomènes de la ville intel-

TABLEAU 5.1

Quatre définitions de la notion de cultures de données

Auteur	Objet et approche	Définition
Labelle <i>et al.</i> , FING (2017)	Stratégie organisationnelle. Sciences de l'information et de la communication.	« [...] la connaissance de gestes et de processus techniques, intellectuels et sociaux de traitement des données dans des situations professionnelles ou quotidiennes. La culture des données est composée de savoirs et savoir-faire issus des statistiques et de l'informatique. La culture des données suppose une compréhension des médias et de la circulation de l'information. »
Lehmans (2017)	Données ouvertes (organisations professionnelles). Sciences de l'information et de la communication.	« La culture de la donnée est composée de savoirs, de valeurs et de compétences statistiques, informatiques, médiatiques, et informationnelles qui donnent les moyens aux professionnels de comprendre le sens de leur activité de collecte et de mise en circulation des données, de la situer dans son contexte social d'usage, et de coordonner les gestes et les procédures professionnels. Cette culture est nécessairement propre à chaque organisation qui construit ses pratiques. »
Bates (2017)	Ville intelligente. Étude critique des données.	« les différentes normes culturelles, systèmes de valeurs et croyances qui informent, encadrent et justifient les pratiques des gens en matière de production, de traitement, de distribution ou d'utilisation des données (pratiques de données), ainsi que leurs efforts pour gouverner et façonner des formes particulières de pratiques de données par divers moyens sociaux et techniques ¹ . » (traduction libre)
Burgess (2021)	Apps numériques et médias sociaux. Étude critique des données et études culturelles.	« [...] les pratiques qui se forment en rapport avec et en réponse aux médias sociaux et autres données (et traces de données) que les gens génèrent dans leur vie quotidienne. Ces pratiques se forment à partir de nos divers engagements, expériences et approches de la compréhension et de la négociation de ces données ² . » (traduction libre)

1. « [...] the different cultural norms, value systems and beliefs, that inform, frame and justify peoples' practices of data production, processing, distribution, or use (data practices), as well as their efforts to govern and shape particular forms of data practices through a variety of social and technical means. »

2. « [...] the practices that form around and in response to the social media and other data (and data trails) that people generate as we go about our daily lives. These practices form from our diverse engagements with, experiences of, and approaches to understanding and negotiating these data. »

ligente, de la participation civique et de leurs enjeux sociopolitiques. C'est le cas de Bates (2017) qui met en rapport les notions de pratique de données, de culture de données et de gouvernance des données. Sa définition insiste sur une double dimension d'analyse spatiale : elle prend en compte à la fois la spécificité locale des différents sites de pratique et leur interconnexion dans un réseau plus large d'environnements techniques, de politiques publiques, de législations et de systèmes économiques.

D'autre part, les travaux de Burgess (2017, 2022) s'inscrivent dans l'étude sur le genre et la sexualité pour étudier la transformation du rapport à l'intimité dans le contexte des médias numériques. Cette autrice développe une perspective inspirée des études culturelles pour étudier des applications de rencontre. Sa définition propose une approche vernaculaire des cultures de données, centrée sur l'expérience ordinaire des médias siconumériques par leurs usagers. Cette conceptualisation met en lumière l'agentivité et la subjectivité des usagers ordinaires qui contribuent eux aussi à forger les cultures de données dans les activités de la vie quotidienne (Burgess *et al.*, cet ouvrage).

Proposition de conceptualisation

Cette recension montre que l'étude des cultures de données se décline sur différents plans de la notion de culture, une des plus polysémiques du vocabulaire des sciences humaines et sociales. Je propose de distinguer trois plans : celui de la *culture holistique*, celui de la *culture de production* et celui de la *culture d'usage* (chacun à envisager dans sa pluralité). Ces trois plans se superposent et ne sont pas mutuellement exclusifs.

Le plan le plus large et le plus général est celui des cultures de données au sens holistique : il englobe les effets croisés des médiations culturelles et techniques sur les dimensions symboliques de la vie. Dans une perspective ethnoculturelle ou identitaire, la culture désigne un ensemble de traits et de dispositions sociosymboliques incorporés, largement partagés au sein d'un groupe social élargi à l'échelle d'une société, d'une nation ou d'un collectif d'appartenance, et constitutifs de son identité et de son rapport au monde. S'il peut être pertinent d'envisager les cultures de données sous l'angle de leurs spécificités nationales et locales, l'approche suggérée ici consiste plutôt à adopter la perspective d'une anthropologie culturelle des médias au sens d'une analyse des médiations technosymboliques propres aux données numériques. Il s'agit alors d'étudier les environnements techniques de mise en données comme médiateurs d'un rapport expérientiel au monde, dans ses dimensions

tant pratiques que sensibles et affectives. En effet, plusieurs dimensions fondamentales de la vie, comme le rapport à la mobilité, au savoir ou à la santé, sont de plus en plus expérimentés par l'intermédiaire de dispositifs numériques (applications, services) fondés sur la collecte et l'exploitation de données. Les *cadres d'expérience d'un monde mis en données* apparaissent ainsi comme un trait culturel de la modernité technique.

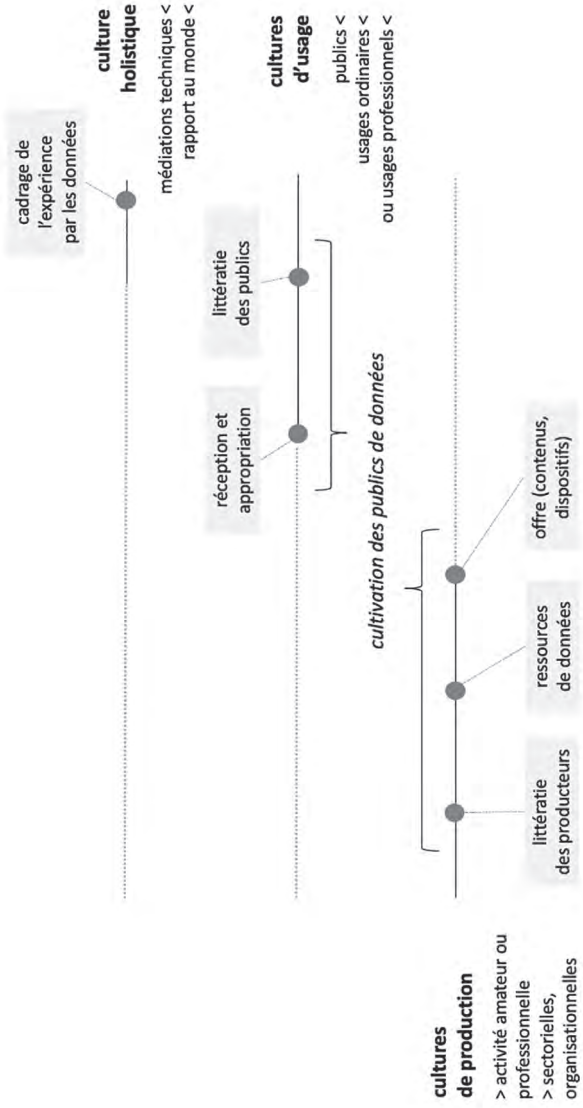
Le plan des cultures de production de données – et de dispositifs fondés sur des ensembles de données – est resserré sur l'activité professionnelle ou amateur visant la création ou la transformation de données. Ces cultures de production possèdent des caractéristiques, des normes et des valeurs propres à un groupe plus ou moins circonscrit (secteur socioéconomique, métier, organisation). Ce plan permet d'entrer davantage en profondeur dans l'analyse de contextes situés. En prenant le secteur professionnel des arts et de la culture comme terrain de référence, on peut distinguer des catégories d'acteurs situés dans différents pôles d'activité.

Du côté des producteurs, les pratiques relevant des cultures de données impliquent une activité de *constitution de ressources de données* (Luka, cet ouvrage). La nature de ces ressources est propre à chaque secteur, voire à chaque organisation; par exemple dans le secteur du spectacle vivant, les données transactionnelles issues des plateformes de billetterie sont mises en rapport avec les métadonnées descriptives des œuvres diffusées. Ces ressources contribuent à la *production d'une offre orientée par l'analyse des données (data-driven)*, telle la recommandation personnalisée de spectacles en fonction des choix précédents de groupes de spectateurs semblables. Le développement de cette offre s'appuie sur l'acquisition d'une *littératie professionnelle* des données nécessaire pour utiliser des outils spécialisés de collecte, de gestion et d'analyse des données, par exemple les bases de données ou Google Analytics. Un ensemble d'intermédiaires (plateformes, acteurs de la société civile tels Wikidata, Synapse C au Québec ou la FING en France) contribuent également à façonner les cultures de données au sens où ils participent à la médiation des données et à la formation en littératie des données.

Le plan des cultures d'usage se situe du côté des usagers et des publics⁵ des dispositifs de données. Ils développent des pratiques liées à la *réception de l'offre* et aux *usages quotidiens (ordinaires)* des dispositifs connectés. Les

5. Les termes d'usager et de public sont à distinguer: le premier renvoie aux formes d'utilisation et d'appropriation des dispositifs, tandis que le second suppose en outre une communauté d'intérêts et l'expression d'une position (Møller Hartley *et al.*, 2021; Millerand *et al.*, cet ouvrage).

FIGURE 5.1
Les plans culturels du rapport aux données



cultures d'usage étant influencées par une multiplicité de facteurs sociaux, culturels et techniques, leur compréhension implique l'analyse de contextes – les études de cas présentés dans le présent ouvrage en sont l'exemple. L'activité des publics de données engage elle aussi une *littératie*; toutefois dans le cas des usagers dits ordinaires, cette littératie ne nécessite pas de compétences de collecte et d'analyse de données, mais plutôt des savoir-faire qui intègrent les dispositifs de données dans les pratiques quotidiennes.

C'est à l'intersection de ces deux derniers plans des cultures de données, celui de l'activité de production de l'offre et celui des usages, que l'on peut situer la rencontre, l'écart ou le décalage entre les représentations que les professionnels se font des publics (les publics imaginés) et leurs pratiques concrètes d'appropriation des dispositifs (les publics effectifs) (voir Harvey et Latzko-Toth, cet ouvrage; Millerand et Meunier, cet ouvrage). Pour creuser l'analyse de ces points de contact entre publics imaginés et effectifs, la notion de *cultivation*⁶ des publics de données (Møller Hartley *et al.*, 2021) apparaît féconde, car elle permet d'associer plusieurs dynamiques en jeu : les intentions et les représentations des producteurs, les affordances des dispositifs de données et des algorithmes dans leur potentiel de « calcul » et de « programmabilité » des publics (Gillespie, 2014), ainsi que l'irréductible incertitude de leurs résultats en ce qui a trait à la formation et à l'agentivité des publics.

Le découpage proposé ici entre différents plans culturels et pôles d'activité résonne avec le découpage d'Albury *et al.* (2021) qui proposent quatre acceptions du terme de cultures de donnée : « les modes de cultivation des données », « les cultures de production », « les cultures d'usage » et « la mise en données de la culture ». Le terme de *cultivation* des données que l'on retrouve ici désigne les manières dont « diverses formes de données sont créées, nettoyées, ordonnées, récoltées et entrecroisées – par des acteurs multiples et distribués, mais connectés, notamment des entreprises, des gouvernements, des développeurs, des annonceurs et des utilisateurs ». Cette acception nous rappelle que si la reconnaissance d'un plan de production *professionnelle* de dispositifs de données permet de creuser les logiques propres au monde du travail et aux organisations d'un secteur, elle ne doit pas obscurcir le fait que les données sont coproduites par une multiplicité d'acteurs, incluant les publics. Ainsi un « public de données » n'est pas

6. Dans les études des médias, le terme de « cultivation » a été popularisé par Gerbner (1998) qui l'a mobilisé pour analyser la réception des contenus télévisuels et ses effets sur le partage de mêmes conceptions du monde chez des publics différenciés. L'usage de ce terme dans la littérature récente sur les cultures de données est toutefois sensiblement différent.

réductible au produit d'une culture de données particulière, mais il contribue tout autant à la coproduire par ses pratiques de consommation et de contribution sur les plateformes, par ses appropriations des dispositifs de données et par ses tactiques de détournement.

Éléments de définition et dimensions d'analyse

À la lumière de ces réflexions, je propose ci-dessous une synthèse des principaux éléments de définition de la notion de cultures de données ainsi qu'une série de dimensions d'analyse. Ce cadre a été construit pour être mobilisé dans un chantier de recherche sur l'activité professionnelle du secteur des arts et de la culture. L'objectif n'est pas de proposer une définition généralisable à tous les objets de recherche, ce qui escamoterait les spécificités de chacun et limiterait la profondeur des analyses possibles. Il s'agit plutôt de dégager des constantes dans les définitions proposées à ce jour, et de recenser des dimensions d'analyse qu'il faudrait nécessairement adapter en fonction des objets considérés.

Les éléments de définition proposés sont les suivants :

- Une culture de données est un assemblage de répertoires collectifs de représentations, d'actions et de justifications situé dans un contexte spécifique et caractérisé par une sensibilité, une rationalité et un univers de sens construits en relation avec des ensembles de données.
- Ces répertoires contribuent à façonner les données, leur statut et leurs effets sur le monde. Ils reposent sur des valeurs, des normes (explicites ou implicites), des littératies, des pratiques, des affects et des dispositifs techniques.
- Ces éléments forment un système possédant une cohérence interne, mais non uniforme, tout en étant traversé par des lignes d'ambivalence, de tension et de contradiction. Ils s'enchevêtrent avec d'autres cultures de données, toutes animées par des dynamiques à la fois locales et globales.

Le tableau de la page suivante détaille les six principales dimensions retenues pour étudier ces répertoires : les *valeurs* associées à une culture de données spécifique, les *normes* qui lui sont propres, la *littératie* des données qu'elle présuppose, les *pratiques* par lesquelles elle se déploie, les *affects* qu'elle véhicule et les *dispositifs matériels* qui sous-tendent sa formation. Ce tableau

développe ces dimensions en les illustrant par des exemples issus du secteur des arts et de la culture au Québec.

Dans l'optique d'étudier la formation et l'évolution des cultures de données d'un secteur professionnel ou d'une organisation en particulier, il est utile d'interroger la relation entre la notion de culture de données et celles de culture organisationnelle et de culture professionnelle. La première prend l'organisation comme milieu structurant pour la formation de cultures spécifiques et localisées, relativement partagées au sein d'un groupe délimité par une activité de coopération visant un but commun (dans une entreprise, une administration, une association). La seconde concerne les cultures spécifiques à chaque groupe professionnel sous l'angle des métiers. Ces cultures sont forgées par des processus de scolarisation (formation, diplôme) et de socialisation au sein d'organisations et d'associations professionnelles.

On peut tirer de la sociologie des professions d'importants éléments de conceptualisation des cultures de données. Les travaux de Monjardet en particulier proposent une critique de l'approche dite «culturaliste» de la culture professionnelle, jugée réductrice au sens où elle chercherait à répertorier des attributs différentiels assignés à l'ensemble du groupe, c'est-à-dire «un catalogue de traits (ou propriétés) partagés (à des degrés divers) par tous les professionnels du même secteur» (p. 408). Montjardet met plutôt en évidence «à la fois un pluralisme considérable des attitudes, valeurs et pratiques et un langage commun, une identique mobilisation autour des enjeux sous-jacents aux acceptions différentes, souvent conflictuelles, des mêmes missions, du même métier» (p. 409). Ainsi, bien que l'étude des groupes professionnels délimités par le partage d'un même métier ne soit qu'un des aspects de la composition des cultures de données, les travaux de Montjardet invitent judicieusement à les envisager comme des objets non réductibles à une liste de propriétés. Ainsi, il serait réducteur de parler d'une culture de données propre aux bibliothécaires, qui se différencierait de manière homogène par rapport à une culture des données du secteur muséal ou musical.

Problématiser la notion de cultures de données engage à faire la distinction entre ses propriétés heuristiques et les présupposés véhiculés par certains discours actuels sur l'innovation organisationnelle. L'usage croissant de cette notion dans les organisations reflète la nécessité pratique de s'adapter aux

TABLEAU 5.2

Dimensions d'analyse des cultures de données

Dimension	Description	Exemples
Valeurs	Croyances fortes au sujet des données, fondées sur des principes d'action supérieurs.	<i>L'objectivité</i> et la <i>complétude</i> des données comportementales (tel l'historique d'emprunt de documents recueilli par les comptes des abonnés des bibliothèques) permettraient de comprendre les pratiques des publics dans leur globalité, voire de « laisser parler les publics » en toute transparence par la mesure de leurs comportements. <i>L'efficacité technique</i> des dispositifs d'intelligence artificielle et de recommandation algorithmique permettrait de mettre en relation « la bonne offre » avec « le bon public ».
Normes	Cadres sociaux qui régulent les comportements liés aux données; normes d'usage; normes professionnelles explicites (p. ex. code déontologique) ou implicites (normes pratiques).	La <i>protection des données personnelles</i> des publics fait partie des normes professionnelles des bibliothécaires et des archivistes. Cette norme est transmise durant la formation scolaire et enchâssée dans des cadres légaux (comme la <i>Loi sur les archives</i> du Québec qui fixe les conditions de protection des renseignements personnels dans les archives privées) et dans des politiques institutionnelles (telle la <i>Politique de confidentialité</i> de Bibliothèque et Archives nationales du Québec). Ces normes peuvent modeler ou freiner les pratiques d'exploitation des données comportementales.
Littératies	Savoirs et savoir-faire liés au cycle de vie des données (production/extraction, stockage, traitement, gestion, exploitation, partage et diffusion des données) et aux usages des dispositifs de données. Les connaissances et les compétences liées aux données touchent notamment aux domaines de la statistique, de l'informatique, de l'information-documentation, du droit et des médias.	Dans le domaine des bibliothèques et des archives, la littératie des données s'est d'abord développée via <i>l'informatisation des bases de données documentaires</i> (catalogage et indexation des documents) puis dans le domaine des <i>données ouvertes</i> qui concerne essentiellement les métadonnées descriptives des documents. La littératie des <i>données comportementales</i> est encore peu répandue chez les bibliothécaires et les archivistes, contrairement aux analystes de données et aux spécialistes du marketing dans les services de communication des institutions.

Dimension	Description	Exemples
Pratiques	Schémas observables d'activité, d'usage et de relations sociales stabilisés dans l'organisation de l'activité. Ces schémas se déploient dans un langage (vocabulaire), dans des processus opérationnels (procédures, protocoles, routines, modes de prise de décision, usages), dans des attitudes ou des comportements et dans des modes d'interaction.	La <i>prise de décision concernant l'acquisition de nouveaux ouvrages</i> à intégrer dans les collections des bibliothèques est une pratique qui peut être influencée par les données comportementales : l'analyse de l'historique d'emprunt des abonnés peut orienter les acquisitions vers des types de documents populaires auprès des publics.
Affects	Formes de sensibilité relatives à la mise en données et à l'expérience des dispositifs de données ; aspects émotionnels des pratiques de données ; formes symboliques et esthétiques de mise en représentation des données (p. ex. visualisation de données).	L'analyse des données comportementales soulève chez certains professionnels des bibliothèques <i>l'espoir</i> d'exploiter une « mine d'or sur laquelle on dort », mais aussi la <i>résignation</i> que cela ne soit qu'« un beau rêve qui ne se réalisera qu'aux États-Unis, jamais au Québec » en raison des limites technologiques des systèmes actuellement utilisés dans la province.
Dispositifs	Supports matériels, outils, systèmes techniques et plateformes de données dont les potentialités pratiques, normatives et axiologiques participent au cadrage des cultures de données.	Dans le secteur des bibliothèques, <i>les plateformes de prêt de livres numériques</i> sont les plus avancées en termes d'expérimentation avec les données comportementales : elles « nagent dans les données », tandis que la plupart des bibliothèques municipales opèrent avec des <i>systèmes au bord de l'obsolescence</i> qui ne leur permettent pas d'envisager de telles pratiques.

mutations numériques des environnements de travail. Mais elle s'inscrit aussi plus largement dans une formation discursive, au sens foucauldien, caractérisée par la « mise en données » de la société, autrement dit une configuration contemporaine du rapport entre dispositifs de données, savoirs et pouvoir. Dans ce contexte, l'injonction à développer de nouvelles cultures de données est alimentée par les effets idéologiques du *dataïsme*, c'est-à-dire « la croyance répandue dans la quantification objective et le suivi potentiel de tous les types de comportement humain et de la socialité grâce aux technologies numériques connectées » (van Dijck, 2014, p. 198). Faire des cultures de données un objet de recherche implique donc de les historiciser de manière à saisir comment elles s'inscrivent dans des formes discursives plus larges, mais aussi d'en contextualiser les formes matérielles et les logiques d'action propres à divers milieux de pratique.

Conceptualisée de manière critique, la notion de cultures de données offre un éclairage sur les publics de données. Elle invite à prêter attention aux répertoires de sensibilité et de signification collective qui contribuent à forger le rapport entre production de dispositifs de données et formation des publics. Ce rapport engage une dynamique de coproduction entre cultures et publics de données. D'un côté, les cultures de données engagent des schémas d'action et d'interprétation (pratiques, normes, valeurs, affects, littératies) qui contribuent à forger des visions (publics imaginés), des architectures algorithmiques (publics calculés) et des mesures des publics (audience). Mais en corollaire, la formation effective de publics et leur agentivité (usages, appropriations, expressions) agit sur les répertoires culturels de la mise en données. Cette agentivité des publics peut aussi bien contribuer à reproduire des répertoires en place qu'à engendrer des décalages transformatifs. Ainsi, les cultures de données ne sont pas des corpus de répertoires stables, mais plutôt des processus actifs de *cultivation*, travaillés par une multiplicité d'acteurs hétérogènes, dont les publics.

Références

- ALBURY, Kath et BURGESS, Jean *et al.*, « Everyday Data Cultures » dans *AoIR Selected Papers of Internet Research*, 2021, p. 1-3.
- BATES, Jo, « Data cultures, power and the city », dans Rob KITCHIN, Tracey LAURIAULT et Gavin McARDLE (dir.) *Data and the City*, Londres, Routledge, 2017, p. 189-200.
- BOWKER, Geoffrey, « Biodiversity Datadiversity », *Social Studies of Science*, vol. 30, n° 5, 2000, p. 643-683.

- BURGESS, Jean, « Hookup apps' vernacular data cultures », communication présentée à *The Social Life of Data*, Melbourne, RMIT University, 2017.
- BURGESS, Jean *et al.*, *Everyday Data Cultures*, Hoboken, John Wiley & Sons, 2022.
- CASEMAJOR, Nathalie, « Développer la culture de données ? Au-delà de la découvrabilité, des significations plurielles », dans Anne JONCHÉRY, Amandine LOUGUET et Vincent BERRY (dir.) *Culture en régime numérique. Questionner les pratiques, catégories et méthodes*, Paris, Presses de Sciences-Po, à paraître.
- GERBNER, George, « Cultivation analysis: An overview », *Mass Communication and Society*, vol. 1, n° 3-4, 1998, p. 175-194
- GILLESPIE, Tarleton, « The relevance of algorithms », dans Tarleton GILLESPIE, Pablo BOCZKOWSKI et Kirsten FOOT (dir.), *Media technologies: Essays on communication, materiality, and society*, Cambridge, The MIT Press, 2014, p. 167-194.
- KITCHIN, Rob, « Big Data, new epistemologies and paradigm shifts », *Big Data & Society*, vol. 1, n° 1, 2014.
- LABELLE, Sarah *et al.*, *Vers une organisation data driven : mettre la culture des données au cœur de l'organisation*, Infolab – FING, 2017.
- LEHMANS, Anne, « Données ouvertes et redéfinition de la culture de l'information dans les organisations. Vers une culture de la donnée », *Communication et organisation*, vol. 51, 2017, p. 15-26.
- LEHMANS, Anne, « L'horizon d'une culture de la donnée ouverte: de l'utopie aux pratiques de gouvernance des données », *Revue COSSI: communication, organisation, société du savoir et information*, n° 4, 2018.
- LITT, Eden et HARGITTAI, Eszter, « The Imagined Audience on Social Network Sites », *Social Media+ Society*, vol. 2., n° 1, 2016.
- MARRES, Noortje S., *No issue, no public: Democratic deficits after the displacement of politics*, Enschede, Ipskamp Printpartners, 2005.
- MØLLER, HARTLEY, J. *et al.*, « Researching publics in datafied societies: Insights from four approaches to the concept of 'publics' and a (hybrid) research agenda », *New Media & Society* (online first), 2021.
- MONJARDET, Dominique, « La culture professionnelle des policiers », *Revue française de sociologie*, vol. 35, n° 3, 1994, p. 393-411.
- SIROIS, Guillaume *et al.*, « Cultural participation through digital technology: A puzzling issue for cultural governance », dans Devin BEAUREGARD et Jonathan PAQUETTE (dir.), *Canadian Cultural Policy in Transition*, Londres, Routledge, 2021, p. 129-143.
- VAN DIJCK, José, « Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology », *Surveillance & society*, vol. 12, n° 2, p. 197-208.

CHAPITRE 6

La mobilisation de publics de niche

Création d'une base de données
dans le secteur culturel

Mary Elizabeth Luka

Le secteur culturel se caractérise par la grande diversité des organisations et acteurs impliqués dans le travail de création, mais aussi, paradoxalement, par la rareté des ressources financières et informationnelles disponibles¹. Le partage d'informations et de connaissances se heurte en effet au manque de ressources à l'échelle du secteur, ce qui limite sa capacité à s'adapter aux exigences de la transition numérique. Artifex² est une base de données développée pour le secteur culturel et dont nous avons suivi l'évolution depuis sa création en 2019 jusqu'en 2023³.

Artifex est la création conjointe d'une équipe de recherche de l'Université de Toronto à Scarborough (UTSC) et de Mobilisation culturelle, un organisme de soutien aux arts spécialisé dans le partage de la recherche et des connaissances dans le secteur culturel. La base de données regroupe un vaste ensemble de ressources offertes en libre accès: trousse d'outils, jeux

1. Texte traduit de l'anglais par Maude Mainguy, avec la collaboration de Florence Millerand et de Mélanie Millette.

2. <https://criticaldigitalmethods.ca/artifex/>

3. L'autrice a fait appel aux sources de financement suivantes pour soutenir le développement d'Artifex: financement du groupe de travail de l'UTSC; bourse Connaught Early Scholar de l'Université de Toronto; projet UTSC Cluster of Scholarly Prominence, Urban Just Transitions; bourse postdoctorale Mitacs de Mobilisation culturelle. L'organisme Mobilisation culturelle a obtenu des fonds du Conseil des arts du Canada pour soutenir Artifex, en particulier l'organisation de datathons.

de données, cadres réglementaires, programmes de financement, etc. La particularité du projet tient à sa dimension collaborative puisqu'Artifex a été développée par la mobilisation de nombreux acteurs, organisations et partenaires du milieu culturel. Ainsi, des décideurs, des artistes, des activistes, des bailleurs de fonds et des dirigeants d'organisations, entre autres, ont contribué au projet aux côtés de l'équipe universitaire et de Mobilisation culturelle. Au cours des quatre années qu'a duré le projet à ce jour, Artifex a connu quatre changements majeurs correspondant à autant de phases, chacune étant le résultat – souvent inattendu – des activités de la précédente et, surtout, de la diversification progressive des groupes d'utilisateurs impliqués (Coutant et Vial, cet ouvrage).

Sur le plan méthodologique, nous avons déployé une stratégie autoethnographique critique (Holman Jones, 2017), complétée par une observation participante (Cresswell, 2014), par l'analyse d'archives, de notes de terrain et de notes rédigées par l'équipe de recherche et, finalement par l'étude de la base de données elle-même (Luka, 2014, 2022).

Bien que des décalages entre les utilisations imaginées *et* réelles d'Artifex ont suscité certaines frictions, les processus décisionnels qui en ont résulté ont abouti à un ensemble de pratiques robustes et éthiques pour la collecte, l'organisation et le partage de connaissances utiles au secteur culturel. En cela, cette expérience de développement collaboratif d'une ressource partagée vise à dynamiser les collaborations avec et au sein du secteur culturel, afin que les partenariats puissent s'articuler et se développer de manière utile à la fois au monde de la recherche universitaire et au monde des arts et de la culture.

Planter le décor : le secteur culturel, ses « publics de niche » et ses ressources limitées

Le secteur culturel comprend des organisations et des activités extrêmement variées – qui peuvent être sans but lucratif, à but lucratif ou un peu des deux – dédiées au travail professionnel de création ; ainsi, le secteur regroupe autant les artistes du spectacle, les concepteurs graphiques que les administrateurs et entrepreneurs d'art, etc. (Statistique Canada, 2022 ; Conseil des arts du Canada, s. d.). Une manière d'appréhender cette diversité d'organisations et d'acteurs est de les considérer comme une série de « publics de niche⁴ » (Luka,

4. Le terme « *narrowcast audiences* » n'a pas d'équivalent en français. La traduction proposée, les « publics de niche », cherche à rendre l'idée d'un auditoire segmenté ou de publics spécifiques.

2022; Bazalgette, 2009). Dans ce chapitre, nous mobilisons la notion de publics de niche pour souligner la manière dont des groupes de créateurs et d'utilisateurs de contenus peuvent se distinguer sur le plan conceptuel en fonction de leurs intérêts, de leurs expertises et de leurs niveaux d'engagement, mais aussi, sur le plan pragmatique, pour comprendre comment, ensemble, ils peuvent influencer le développement d'une ressource comme Artifex.

L'utilisation que nous faisons du terme s'inspire des propos de Peter Bazalgette (2009) sur les « *narrowcast audiences* » de la radiodiffusion publique et des Conseils des arts au Royaume-Uni. Bazalgette (2009) suggère d'aborder les publics de niche de manière différente afin que les bailleurs de fonds et les décideurs du secteur culturel puissent développer une gamme de contenus qui soit suffisamment attrayante pour la grande variété des publics. Cette pratique, qui consiste à déterminer les caractéristiques particulières des différents auditoires ou publics intéressés par le partage et l'accès à un même contenu, a connu une croissance exponentielle et a fait l'objet d'une définition plus précise à l'ère d'Internet. Cela dit, la notion existe depuis longtemps dans les milieux de la production médiatique et des arts de la scène (Luka, 2015, 2022b), où les publics de niche sont ciblés par des campagnes de marketing et de promotion.

En ce qui concerne le financement des arts, les décideurs et les bailleurs de fonds expriment souvent le désir de soutenir et de financer les arts de manière que le secteur puisse rendre des comptes aux contribuables, mais certains choisissent de mettre l'accent sur l'avant-gardisme dans le développement d'une pratique, alors que d'autres favorisent la popularité en tant que mesure principale du succès d'une production culturelle. Du côté des artistes, à un moment donné de leur carrière, ils peuvent avoir besoin d'apparaître dans des expositions de grande envergure pour élargir leur public, tandis que d'autres peuvent vouloir se concentrer sur des activités qui les amènent à discuter avec leurs pairs ou avec des groupes particuliers.

Dans le secteur des arts et de la culture, la concurrence forte pour l'accès aux ressources et leur partage vient complexifier la donne. Selon notre propre expérience en tant que productrice, décideuse politique et chercheuse dans le domaine médiatique et culturel, il est de notoriété publique que les priorités des gouvernements en place limitent le nombre de programmes de subvention ; que les bailleurs de fonds et les organismes artistiques ne comprennent pas bien les particularités de l'investissement des entreprises dans les organismes créatifs ; et que, par conséquent, il existe peu de ressources ou de structures qui soutiennent le partage des connaissances (Wilhelm, 2019 ;

Gattinger et Saint-Pierre, 2021). De plus, comme le montrent les études sur la production culturelle et médiatique, telles que celles menées par Bannerman (2020), Hesmondhalgh (2012) et Mayer (2011), de nombreux sous-secteurs d'activités sont cloisonnés artificiellement en raison de leurs régimes budgétaires, par exemple les galeries et les musées, les entreprises de production médiatique indépendantes, les salles de spectacle, etc. À cette complexité s'ajoutent les différences de compétences entre les niveaux municipal, provincial, national et mondial, alors que les travailleurs créatifs peuvent théoriquement se mouvoir aisément entre les différents domaines d'activité en raison de la transférabilité de leurs savoirs et de leurs savoir-faire. Par exemple, un scénographe pourrait travailler en principe à la fois pour un théâtre en Saskatchewan, une émission de télévision en Ontario ou un projet cinématographique en France (en fonction des règlements en matière de résidence).

Ce type de concurrence et le sentiment qui en découle, selon lequel le temps et les ressources sont limités, ont toujours empêché le partage des connaissances et le renforcement des capacités dans le secteur culturel. Ainsi, une évaluation récente menée pour le Conseil des arts du Canada de son fonds Stratégie numérique (2018-2021), destiné à stimuler la transformation numérique du secteur des arts, a abouti à trois constats : les besoins en matière de culture numérique sont grands, la capacité d'établir des partenariats est faible et il existe peu d'organisations de développement des compétences à la disposition du secteur (Nordicity, 2022). Ces facteurs ont limité le succès du fonds Stratégie numérique lui-même, en particulier au cours de sa première année, d'abord parce que les organisations artistiques ne savaient même pas ce qu'elles pouvaient déposer comme demandes de financement.

Réduire la rareté des ressources et la concurrence qui en découle est crucial, parce que le secteur culturel contribue vivement à l'économie, avec des répercussions positives sur les plans social, civique, éducatif et de la santé (Crossick et Kaszynska, 2019 ; Essig, 2018 ; Thompson et Campbell, 2022). Alors que l'emploi connaissait un ralentissement important causé par la pandémie de COVID-19, le secteur culturel a compté pour près de 2,7 % du produit intérieur brut (PIB) du Canada et pour 3,4 % de ses emplois (Statistique Canada, 2022). À l'heure où le secteur culturel commence à se remettre des effets de la pandémie, il se tourne vers le monde numérique où le partage des connaissances entre les multiples publics de niche promet de contribuer de manière significative au développement social. Tirer parti des

ressources et des connaissances d'un secteur rempli de petites et moyennes organisations et de dizaines de milliers de travailleurs indépendants apparaît de plus en plus vital.

Créer les conditions d'une ressource partagée pour le secteur culturel

Avant qu'Artifex ne soit imaginé, nous avons élaboré une ressource centrée sur les environnements collectifs de travail créatif, c'est-à-dire les centres de création, les organismes de services en art et les réseaux professionnels. Cette base de données, nommée *Creative Hubs and Networks Database* (CHND) s'est avérée être une ressource à la fois précurseuse et complémentaire d'Artifex. Alors que les éléments contenus dans la CHND pouvaient directement servir au secteur culturel (par exemple, des informations sur des programmes, des services et des ressources de renforcement des capacités), la base de données elle-même a constitué un référentiel de ressources de premier ordre à l'échelle du secteur.

Mais il s'agissait avant tout d'une ressource scientifique, et les membres du secteur, dont les artistes et les directions des centres de création, trouvaient la base de données trop dense et trop érudite pour servir au quotidien. Ce constat nous a conduit à mettre en place un partenariat avec Mobilisation culturelle pour développer ce qui semblait être à l'origine un projet secondaire, mais qui est devenu une passerelle de collaboration évolutive entre le secteur de la culture et le monde universitaire.

Développer Artifex : articuler de multiples publics de niche

L'expérience de la conception d'Artifex donne à comprendre comment une diversité de publics de niche du secteur culturel peut créer conjointement une ressource partagée qui soit directement reliée aux besoins des membres du secteur culturel. À l'origine, nous nous attendions à ce que les organisations répertoriées dans la CHND, soit les centres de création, les organismes de services en art et les réseaux professionnels, constituent les principaux publics d'Artifex, aux côtés des chercheurs, des décideurs et des bailleurs de fonds. Cependant, d'autres publics ont émergé, dont les bénévoles de Mobilisation culturelle et les titulaires des droits d'auteur.

Phase 1 : Passer de dossier de fichiers à ressource partagée

La première phase a commencé très simplement au printemps 2019. L'une des cofondatrices de Mobilisation culturelle nous a présenté un dossier contenant une liste de 30 ressources en format numérique ou papier, que d'autres membres de l'organisme (qui était naissant à l'époque) jugeaient utile de partager avec l'ensemble du secteur culturel. Grâce à l'obtention de fonds pour l'embauche d'une assistante de recherche, nous avons organisé le matériel et vérifié auprès des titulaires des droits d'auteur si nous avions le loisir de diffuser publiquement les ressources. Concrètement, cette ressource a été mise en place au cours de l'été 2019 sous la forme d'une feuille de calcul téléversée dans Google Drive et réservée aux membres de Mobilisation culturelle.

Cette première étape s'est révélée être le fondement d'un partenariat de recherche qui continue de promouvoir le partage de connaissances sur les occasions favorables et les stratégies d'analyse d'impact dans le secteur, que nous avons explorées ailleurs (Luka, 2022a, 2022b). Trois autres décisions clés, impliquant différents publics de niche, ont contribué à façonner Artifex, chaque étape en augmentant la taille et modifiant le type de ressource développée, de même que le processus de collaboration qui le sous-tend.

Phase 2 : Constituer une ressource publique

La deuxième phase a commencé à la fin de l'année 2019 à partir du moment où le projet s'est mué en une ressource publique et qu'il a pris le nom d'Artifex. L'équipe de recherche de l'UTSC a considéré à l'époque que les huit groupes d'utilisateurs potentiels identifiés pour la CHND pouvaient aussi avoir leur pertinence pour Artifex. Ces groupes d'utilisateurs comprenaient les administrateurs des arts, les « artrepreneurs » ou entrepreneurs, les travailleurs créatifs, les bailleurs de fonds, les médias, les décideurs et les chercheurs, en plus d'une catégorie fourre-tout d'intérêt général. Puis, à la suite de discussions menées avec Mobilisation culturelle, l'équipe de recherche de l'UTSC s'est aperçue que d'autres groupes, à savoir les artistes, les petites et moyennes organisations artistiques ainsi que les personnes extérieures au secteur culturel intéressées par l'utilisation de méthodes artistiques dans un cadre communautaire, seraient plus susceptibles d'utiliser Artifex que la CHND. Artifex a alors vu le jour comme instrument public regroupant 110 ressources au total, dont 80 liées à la lutte contre la pandémie de COVID-19 dans le secteur culturel.

La deuxième phase a débouché sur la mise en place de pratiques de gestion de la propriété intellectuelle fondées sur une vérification du contenu et l'intégration de considérations éthiques. Pour cela, l'équipe de recherche de l'UTSC a communiqué directement avec les propriétaires de chacune des ressources par courriel. Cette approche a permis d'atteindre deux objectifs. Premièrement, elle a assuré une utilisation respectueuse et légale des ressources. Deuxièmement, elle a introduit, au sein de Mobilisation culturelle, le principe de demande d'autorisation systématique auprès des propriétaires, avant d'inclure tous liens et documents dans une ressource publique comme Artifex. L'idée de demander un consentement, de procéder à une vérification des faits et de solliciter l'autorisation de partager des ressources est étonnamment méconnue dans le secteur culturel. Il faut dire que ce processus nécessite un travail non négligeable de communication et de documentation au sein de milieux de travail où le personnel reste rare. Pourtant, bien des intervenants du secteur décrient depuis longtemps la nature souvent extractive de certaines recherches ou l'utilisation de productions artistiques sans autorisation (Sicondolfo *et al.*, 2022). Malgré tout, de nombreuses personnes pensent que si une ressource a fait l'objet d'un partage public par un tiers, par exemple, sur un site Web, il n'est pas nécessaire d'obtenir une autorisation pour la reprendre.

Durant cette deuxième phase, l'équipe de recherche de l'UTSC s'est attachée à former les bénévoles et le personnel de Mobilisation culturelle en matière éthique pour la collecte, l'organisation et la présentation des données, à partir des questions développées dans Luka et Millette (2018) pour le développement de pratiques éthiques en recherche féministe. À cette époque, l'information sur la façon dont les ressources étaient recueillies et organisées au sein d'Artifex était consignée sur la page d'Artifex, y compris la manière dont les utilisateurs pourraient les réutiliser. Cependant, il n'existait encore aucun protocole formel pour l'obtention d'un consentement et l'autorisation d'utiliser le matériel. En outre, il n'existait pas davantage d'accord officiel entre l'équipe de recherche de l'UTSC et l'organisme Mobilisation culturelle sur la manière d'alimenter Artifex, à l'exception d'une entente selon laquelle l'organisme recevrait les soumissions par le truchement d'une adresse électronique dédiée à cette fonction. Si ces manques, à l'origine de l'écueil survenu à la troisième phase, apparaissent évidents avec le recul, ce n'était pas le cas à l'époque.

Phase 3 : Organiser des datathons et établir des collaborations

La troisième phase a débuté à la fin de 2021 avec la création de datathons, au moment même où une nouvelle collaboration entre l'équipe de recherche de l'UTSC, Mobilisation culturelle et un groupe élargi d'universités, de bailleurs de fonds et de décideurs politiques se mettait en place. La moitié des 400 ressources que comprend (à ce jour) Artifex sont le fruit de trois datathons qui se sont déroulés d'octobre 2021 à novembre 2022. Les datathons ont marqué un tournant dans l'histoire d'Artifex, puisqu'ils ont représenté le moment où Mobilisation culturelle a pris presque entièrement la responsabilité du mode d'alimentation de la ressource.

Les trois datathons se sont tenus en ligne, sur trois jours, en collaboration avec des partenaires des industries culturelles. Ils ont abordé les thèmes de la santé mentale, de l'analyse d'impact et des changements climatiques. Inspiré du terme « hackathon », le datathon relève d'une approche inclusive qui vise à favoriser la participation. Les participants à l'événement étaient invités à fournir, par courriel ou via Zoom, des suggestions de ressources pour alimenter Artifex. Chaque événement avait pour but d'améliorer Artifex en tant que ressource sectorielle par l'ajout d'outils, d'ensembles de données et d'autres référentiels pertinents. Les activités comprenaient des conférences et des ateliers sur Zoom organisés par des experts en santé mentale, des bailleurs de fonds et des décideurs politiques intéressés par les mesures d'impact, des artistes-activistes ou des universitaires impliqués dans la recherche de moyens pour une transition vers la carboneutralité. Dans les jours qui ont suivi, les bénévoles de Mobilisation culturelle se sont occupés de saisir l'ensemble des données recueillies dans Artifex.

Au cours du troisième datathon (appelé « climatathon »), il est apparu très clairement que les bénévoles de Mobilisation culturelle n'avaient pas été suffisamment formés pour saisir les données dans Artifex ni pour vérifier si les responsables d'Artifex avaient obtenu l'autorisation d'utiliser les ressources par écrit auprès des titulaires des droits d'auteur. Du même coup, l'équipe de recherche de l'UTSC a pu constater qu'elle avait inadéquatement pris en compte deux autres publics restreints impliqués dans Artifex, à savoir l'équipe de bénévoles de l'organisme et les titulaires des droits d'auteur. En outre, ce constat illustre à quel point les décisions sont cruciales dans les processus de mise en données (*datafication*) impliqués dans des pratiques de collecte et de traitement de l'information ; ces pratiques peuvent subir

l'influence des dynamiques de pouvoir en place au sein du milieu de travail et même, parfois, entrer en conflit avec d'autres pratiques.

Les bénévoles ayant participé aux datathons, issus d'autres milieux de pratiques dans le secteur, ne s'attendaient pas à devoir solliciter le consentement et la permission d'utiliser les ressources auprès des titulaires des droits d'auteurs. En outre, ils n'étaient pas familiers avec l'utilisation des mots-clés, des catégories et des métadonnées ayant servi à organiser l'information dans Artifex. Par conséquent, non seulement la saisie des renseignements n'était pas uniforme de bout en bout, mais aucune autorisation écrite n'existait, ce qui signifiait que la base de données n'était ni exacte, ni fonctionnelle. Une fois cette lacune mise à jour, les ressources enregistrées au cours des deux premiers datathons ont été temporairement supprimées afin que l'équipe de l'UTSC puisse apporter les correctifs nécessaires. Cela a mené à la quatrième et dernière phase du développement d'Artifex, qui a consisté à mettre en place un protocole de traitement des données et à le publier sur le site Web pour que les bénévoles puissent le suivre sans difficulté.

Phase 4 : Mettre en place les protocoles éthiques

La réalisation, à l'automne 2022, que les protocoles éthiques nécessaires n'avaient pas suivi le rythme de soumissions de nouvelles ressources grâce aux datathons a précipité la quatrième phase du développement d'Artifex. L'arrivée d'une chercheuse postdoctorale a permis de remédier en bonne partie à ce problème. En outre, le nettoyage de la base de données, la vérification des mots-clés et des métadonnées notamment, mais aussi la réorganisation des catégories sur le site Web, ont nécessité des efforts considérables pour donner à la base de données sa forme actuelle. À chacune de ces étapes, c'est l'évolution des priorités des partenaires et des groupes d'utilisateurs potentiels qui a poussé les responsables d'Artifex à faire des changements dans la base de données.

Concrètement, pour cette phase, nous avons remanié la présentation d'Artifex sur le site Web, revu les crédits et les licences, puis précisé la manière d'obtenir un consentement. Nous avons opté pour cette approche en sachant que de nombreux bénévoles de Mobilisation culturelle contribueraient à Artifex dans les années à venir. La meilleure façon d'assurer le respect des protocoles de consentement et de gestion éthique des droits d'auteur était de publier le processus en détail, dans une langue simple,

directement sur le site Web. De plus, la base de données s'est enrichie d'une colonne intitulée « Contacté pour obtenir le consentement ». Aujourd'hui, Artifex compte près de 400 ressources à jour, dont des jeux de données, des documents et des cadres réglementaires, des politiques, des programmes de financement, des outils, un référentiel et plus. Un glossaire de 175 mots-clés, similaire à celui utilisé dans la CHND, facilite la recherche de ressources pour les chercheurs et les membres du secteur culturel en intégrant des mots-clés suggérés par les artistes, les décideurs politiques et les bailleurs de fonds. L'avenir nous dira si le protocole en place continuera d'assurer l'utilisation et le partage éthique des ressources.

L'histoire d'Artifex est celle de la manière dont une base de données évolue en réponse aux publics de niche qui y collaborent ; c'est aussi celle d'un projet qui aboutit malgré les embûches. Qu'il s'agisse de composer avec des événements inattendus comme une pandémie ou d'apprivoiser différentes façons de travailler et de revoir les hypothèses de départ d'un projet sans perdre de vue l'objectif de partage des connaissances, les éléments importants de la gestion de données ont été nombreux.

Au début du projet, la trentaine de ressources qui composaient Artifex provenaient d'un public restreint : les membres de Mobilisation culturelle. L'équipe de recherche de l'UTSC entendait alors nettoyer et organiser ce matériel pour qu'il puisse servir aux membres de l'organisme. Il est rapidement devenu évident que cette collaboration avait le potentiel de créer une ressource publique plus large, destiné à plus de publics restreints, notamment dans le contexte de la pandémie de COVID-19, lorsque le secteur culturel a commencé à investir massivement l'environnement numérique. L'équipe de l'UTSC s'est alors mise à la recherche de nouvelles ressources à ajouter pour ces publics. Elle a également établi une série de considérations relatives au consentement et à la réutilisation du matériel. Puis, alors que Mobilisation culturelle trouvait sa place en tant qu'organisme de soutien aux arts de la communauté pendant la pandémie, Artifex a constitué une ressource inédite, soutenue par des chercheurs, permettant de déposer de l'information pertinente que divers publics restreints du secteur culturel pouvaient s'échanger.

Un nombre important de ressources sur la santé mentale, l'évaluation des conséquences et les changements climatiques se sont ajoutées dans le cadre de datathons par les publics restreints concernés. La mobilisation de

bénévoles et de stagiaires (un autre public restreint) pour organiser et gérer les datathons, mais aussi pour alimenter Artifex, a entraîné l'échec temporaire des protocoles établis pour assurer les demandes de consentement et la saine gestion des droits d'auteur. Il a donc fallu, par la suite, consigner avec minutie les protocoles partagés entre l'organisme et l'équipe de l'UTSC. En fin de compte, cette collaboration a donné lieu à une approche véritablement partagée de la recherche éthique dans, sur et pour le secteur culturel, qui a évolué en fonction de la disponibilité des fonds et des ressources au fil du temps. Bien que le projet soit modeste, la base de données Artifex et les processus de collaboration qui la sous-tendent ont le potentiel d'évoluer et de s'étendre à d'autres secteurs et à l'université.

Références

- BANNERMAN, Sara, *Canadian Communication Policy & Law*, Toronto, Canadian Scholars, 2020.
- BAZALGETTE, Peter, « Public Service Narrowcasting », *Prospect Magazine*, <https://www.prospectmagazine.co.uk>, 2009.
- CONSEIL DES ARTS DU CANADA, Glossaire, <https://conseildesarts.ca/glossaire>, s.d.
- CRESSWELL, John W., *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, Thousand Oaks, Sage Publications, 2014.
- CROSSICK, Geoffrey et KASZYNSKA, Patrycja, *The Social Impact of Cultural Districts. Global Cultural Districts Network*, Global Cultural Districts Network, 2019.
- ESSIG, Linda, « Value Creation and Evaluation in Arts Incubators », *International Journal of Arts Management*, vol. 20, n° 2, 2018, p. 32-45.
- GATTINGER, Monica et SAINT-PIERRE, Diane, *Cultural Policy: Origins, Evolution, and Implementation in Canada's Provinces and Territories*, Ottawa, Presses de l'Université d'Ottawa, 2021.
- HESMONDHALGH, David, *The Cultural Industries*, 3^e éd., Londres, Sage Publications, 2012.
- HOLMAN JONES, Stacy, « Living Bodies of Thought: The "Critical" in Critical Auto-ethnography », *Qualitative Inquiry*, vol. 22, n° 4, 2016, p. 228-237.
- LUKA, Mary Elizabeth, *Towards creative citizenship: Collaborative cultural production at CBC ArtSpots (1997-2008)*, Thèse de doctorat, Université de Concordia, 2014.
- , « CBC ArtSpots and the activation of creative citizenship » dans MAYER, Vicki, BANKS, Miranda et CONOR, Bridget (dir.), *Production Studies*, vol. II, Londres, Routledge, 2015, p. 164-172.
- , « The New Main Street: Reshaping the Canadian Creative Ecosystem », dans BEAUREGARD, Devin et PAQUETTE, Jonathan (dir.), *Canadian Cultural Policy in Transition*, Londres, Routledge, 2022a, p. 210-221.

- , « Creative hubs : Sites of community and creative work », dans THOMPSON, Cheryl et CAMPBELL, Miranda, (dir.), *Creative Industries in Canada*, Toronto, Canadian Scholars, 2022b, p. 157-179.
- LUKA, Mary Elizabeth et MILLETTE, Mélanie, « Re(framing) Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research », *Social Media & Society*, numéro spécial, 2018, p. 1-10.
- MAYER, Vicky, *Below the Line: Producers and Production Studies in the New Television Economy*, Durham, Duke University Press, 2011.
- NORDICITY, *Evaluation of the Canada Council for the Arts' Digital Strategy Fund and Sectoral Needs Assessment: Executive Summary*, 2022.
- SICONDOLFO, Claudia *et al.*, « Archive/Counter-Archive: Activating Principles of Respect in Archival Policy Development », *ESSACHESS: Journal for Communication Studies*, vol. 15, n° 1(29), 2022, p. 153-175.
- STATISTIQUE CANADA, *Statistiques sur la culture*, <https://www.statcan.gc.ca/fr>.
- , *Indicateurs provinciaux et territoriaux de la culture*, <https://www150.statcan.gc.ca>, 2022.
- THOMPSON, Cheryl et CAMPBELL, Miranda, *Creative Industries in Canada*, Toronto, Canadian Scholars, 2022.
- WILHELM, Kelly, « A balancing act: Supporting the arts in Canada », *The Philanthropist Journal*, <https://thephilanthropist.ca>, 2019.

QUATRIÈME PARTIE

L'inclusion

CHAPITRE 7

Data for Good

La science des données au service des publics minorisés et des milieux communautaires

Mélanie Millette, Florence Millerand et Nadia Seraiocco

Data for Good (DFG) est un organisme sans but lucratif canadien dont la mission consiste à mettre l'analyse de données au service du changement social: « Nous sommes un collectif de personnes passionnées du social qui rendent possible le changement social en faisant parler les données¹. » Fondé en 2013², l'organisme est issu de la mouvance nord-américaine *data kind* qui vise le bien commun (Ermoshina, 2018). En 2021, DFG comptait plus de 8 000 bénévoles répartis en différentes sections sur le territoire canadien: à la section fondatrice créée à Toronto s'en sont ajoutées huit autres au fil des ans à Calgary, à Vancouver, à Edmonton, à Régina, à Waterloo, à Ottawa, à Montréal et dans les Maritimes. L'organisme partage des caractéristiques des mouvements *grass roots*, dans le sens où toutes les activités sont organisées par des bénévoles (Ermoshina, 2018), qui sont en l'occurrence des personnes formées en science des données, en informatique, en marketing, en intelligence artificielle, etc. Comme son nom l'indique, DFG cherche à exploiter la science des données « pour le bien », c'est-à-dire que les bénévoles qui s'y

1. « *We are a collective of passionate socially minded people who are empowering social change makers to be better by allowing their data to speak to them.* » (traduction libre, dataforgood.ca)

2. Joy Robson, une gestionnaire stratégie et marketing, et Victor Anjos, un scientifique de données et chef des technologies, ont fondé Data for Good. Ils sont toujours actifs dans l'organisation (<https://dataforgood.ca/>).

investissent mettent leurs expertises au service d'organismes communautaires ou humanitaires. Ceux-ci bénéficient donc gracieusement des services de DFG; l'expertise et l'intégralité du travail d'analyse des données, y compris la production de visualisations et la rédaction de rapports, sont gratuites.

Nous nous sommes intéressées à la section montréalaise de DFG³ avec un questionnement centré initialement sur le type d'activisme de données incarné par le travail des bénévoles. Nous avons mené une enquête ethnographique de plus de deux ans, de 2019 à 2021, alors que DFG Montréal travaillait principalement avec un organisme à but non lucratif, le Réseau des maisons pour femmes victimes de violence conjugale⁴. Sa mission était d'analyser les données que lui fournissait le Réseau sur ses activités régulières pour produire des rapports que ses responsables pourraient employer notamment pour améliorer leurs opérations ainsi que dans leurs activités de démarchage auprès des bailleurs de fonds gouvernementaux. Nous nous sommes également intéressées à d'autres sections de DFG au Canada pour mieux comprendre la nature du travail de leurs bénévoles et leur ancrage local, elles qui œuvrent au plus près des besoins des organismes communautaires qu'elles desservent.

Les publics : repères théoriques

Data for Good met en relation différents types d'acteurs : les bénévoles en science des données qui travaillent sur les données des organismes, les organismes communautaires (y compris leurs cadres, leurs membres, leurs employés et leurs bénévoles), mais aussi, dans le cas étudié, les femmes victimes de violence et finalement les instances politiques auxquelles sont présentés les rapports basés sur les données. Nous considérons l'ensemble de ces acteurs comme des « publics » constitués autour du problème des violences faites aux femmes et, plus spécifiquement, autour des besoins de soutien sur le terrain pour ces femmes.

Les publics et les problèmes

Les chercheuses et les chercheurs en communication ont théorisé la notion de public en parallèle à celle d'audience pour définir les collectifs articulés autour de pratiques sociales (Livingstone, 2004; Katz et Dayan, 2012), comme

3. La section montréalaise a été créée en 2016.

4. Le nom de l'organisme a été anonymisé.

la réception d'une production culturelle ou la constitution d'un collectif autour d'une cause politique. Typiquement, l'audience renvoie à une acception commerciale (un regroupement de personnes qui consomment un contenu), alors que le public renvoie à une acception politique (un acteur collectif de la vie politique).

La conception politique de la notion a été développée notamment par Dewey (2003). Pour ce dernier, un public doit être en mesure d'estimer la nature des questions d'intérêt qui le concernent et la manière de les traiter. Cette vision renvoie à un public «compétent». Autrement dit, un collectif constitué en tant que public aura acquis les compétences pour reconnaître ses intérêts communs et les enjeux qu'ils soulèvent. S'inscrivant dans cette approche, Marres (2005) rappelle à la suite de Dewey la condition préalable à l'émergence d'un public : celui-ci ne peut se constituer que dans un processus de définition d'un enjeu ou d'un problème d'intérêt commun, par exemple les violences faites aux femmes. Dans cette perspective, un groupe se constitue comme public *en même temps* qu'un problème est défini. En l'absence de problème, il n'y a donc pas de public et vice versa⁵. Cette manière de comprendre le public l'inscrit en public concerné par le problème et apte à trouver des solutions. Le concept de «contre-public» proposé par Fraser (2001) s'inscrit dans une démarche différente, mais complémentaire, en permettant de penser la dialectique entre des publics «dominants», qui parviennent à définir les problèmes collectifs prioritaires, et «subalternes», qui luttent pour la reconnaissance de problèmes négligés dans l'espace public hégémonique.

Dans la conception de Dewey, le public renvoie à l'ensemble des personnes touchées par le problème en question. Il est donc nécessairement pluriel et rassemble des individus hétérogènes, liés par les conséquences d'une activité sociale (Zask, 2008). Or, pour que ce public sorte de sa passivité – car il peut être touché, mais ne rien y changer –, il doit pouvoir emprunter une démarche d'enquête qui le conduira à développer sa compétence – en s'instruisant sur ses problèmes et leurs causes – et, ainsi, à devenir politiquement actif. Cette vision suppose que les personnes concernées ont accès à l'information pertinente, y compris les «données», et qu'elles sachent les interpréter pour formuler le problème, puis agir.

Les organismes communautaires comme le Réseau des maisons des femmes (RMF) ont des missions sociales et sont amenés à porter des revendications sociopolitiques auprès d'instances de tous les ordres de gouvernement.

5. Selon Marres : «*no public, no issue*».

La raison d'être de DFG est précisément d'outiller ces organismes et, ce faisant, de prolonger leur mission par le travail concret de ses bénévoles sur leurs données. Il est ainsi possible de prendre en compte à la fois un organisme – dans notre cas le Réseau des maisons des femmes – et DFG comme des publics constitués pour s'attaquer aux problèmes de la violence envers les femmes, de la précarité et de la sécurité de ces dernières. Or qu'en est-il des femmes elles-mêmes ?

Les femmes usagères du Réseau comme public minorisé

Les femmes victimes de violence conjugale représentent ce qui est décrit sur le terrain comme une « clientèle vulnérable », parce qu'extrêmement précaire. Selon les cas, les femmes soutenues par le RMF ont fui un conjoint violent, elles vivent sous la menace, elles portent des séquelles de violences physiques et psychologiques, elles tentent de retrouver leur autonomie financière, elles cherchent à assurer leur sécurité et celle de leurs enfants, etc. Bien que directement concernées par le problème au cœur de la constitution des publics étudiés, elles ont un statut particulier. En effet, ces femmes sont les personnes dont l'expérience est dataifiée, mais aussi les personnes les plus directement concernées par le problème. La précarité dans laquelle la violence les place entrave leur capacité à se mobiliser sous forme de publics aptes à se saisir des données, au contraire des autres publics évoqués.

La notion de « public minorisé » (Delias, cet ouvrage) renvoie aux travaux de Guillaumin (1985), ainsi qu'à ceux de Laplanche-Servigne et Sa Vilas Boas (2019). L'apport de Guillaumin à la notion de minorité se distingue d'autres définitions où l'asymétrie du nombre entraînerait une asymétrie du pouvoir. Guillaumin aborde plutôt la minorité dans une perspective politique et critique. La signification accordée à « minorité » doit dépasser l'asymétrie du nombre pour considérer que des personnes sont rendues minoritaires, donc minorisées. C'est-à-dire que le moindre pouvoir des personnes minorisées est le résultat de rapports créés historiquement, culturellement et politiquement – ceux qui s'articulent autour des genres, des races et des classes sociales, par exemple. Selon Guillaumin, des personnes se retrouvent minorisées dans leur capacité à exercer un pouvoir d'agir et à contribuer à la société à cause des rapports de domination qui bénéficient à d'autres personnes, dites « majoritaires », dans cette même société.

Ainsi, parler de minorité naturalise des rapports sociaux construits et ancrés historiquement. Les publics minorisés sont « faits minoritaires » par

la reproduction de rapports de domination qui les limitent, sans nécessairement qu'il y ait asymétrie du nombre (Laplanche-Servigne et Sa Vilas Boas, 2019), comme dans le cas du patriarcat et des violences de genre qui limitent et touchent les femmes (Descarries, 2005).

Ainsi, les femmes directement concernées par le problème ont une capacité d'agir limitée par leur minorisation, par les violences genrées subies, et constituent un public spécifique dans le cas de notre ethnographie. Le problème les concerne en premier lieu et les a amenées à utiliser les ressources du RMF, ce qui met en données une partie de leur expérience dans la quête, précisément, de trouver une solution au problème de la violence.

L'activisme de données pour penser le public qu'est Data for Good

En tant que public, DFG se constitue autour des données apportées par le Réseau, afin de les interpréter et de mieux montrer les besoins de l'organisme auprès des gouvernements. Le public que constitue DFG par la mobilisation de ses bénévoles s'inscrit dans le contexte de l'activisme de données. Dans sa définition la plus large, l'activisme de données est une pratique médiatique qui découle de l'activisme en ligne (Milan, 2017).

Alors que certaines formes d'activisme en ligne n'exigent aucune compétence particulière des participants, l'activisme de données relève de personnes qui possèdent ou développent des compétences en science des données (Baack, 2015). Ce phénomène sociotechnique se caractérise par la manière dont les activistes interagissent par et avec les données pour soutenir une cause ou aborder un problème social (Milan et Van der Velden, 2016), dans une perspective politique (Milan et Gutiérrez, 2015). Mais c'est en mettant la main à la pâte que les activistes de données développent un sens de l'engagement dans un projet à la fois technologique et politique (Dunbar-Hester, 2012).

Milan et Van der Velden (2016) répertorient deux grandes logiques d'engagement avec des données. La première logique, dite « réactive », s'inscrit dans l'idée d'une résistance à la *datafication*. Les données y sont vues comme une technique de contrôle et de surveillance qu'il faut contrer en bloquant ou en contournant leurs mécanismes de génération et de collecte. La seconde logique, dite « proactive », et dans laquelle s'inscrit le collectif DFG, correspond à des initiatives de *data justice* (justice numérique) pour produire des données plus représentatives de réalités négligées ou biaisées dans les jeux de données existants. Les données y sont comprises comme une technique d'exposition des inégalités et des facettes de la réalité peu ou mal documentées.

Méthodologie

Notre démarche méthodologique a consisté en une ethnographie basée sur une observation non participante et sur des entretiens complétés par une analyse documentaire. L'observation des activités de la section montréalaise s'est déroulée de février 2019 à avril 2021 : nous avons assisté aux réunions mensuelles du collectif *in situ*, tout en suivant, entre les réunions, les activités de leurs bénévoles sur leurs plateformes collaboratives⁶. Dans la tradition de l'ethnographie dite « connective » (Hine, 2015), nous avons suivi les activités des bénévoles en ligne et hors ligne en empruntant les mêmes modes de communication. Ces observations non participantes (Soulé, 2007) nous ont permis de documenter le travail avec les données ainsi que l'organisation des tâches au sein du collectif, en plus de nous renseigner sur les participants à interroger.

De 2019 à 2022, nous avons mené douze entretiens semi-dirigés avec des bénévoles, dont sept appartiennent à la section montréalaise de DFG. Ces participants font partie du noyau de bénévoles ayant contribué directement au projet avec le Réseau. Les entretiens visaient à cerner la nature du travail sur les données au sein de DFG, les motivations des bénévoles et la signification de leur engagement au sein de l'organisme. Nous avons aussi interviewé quatre bénévoles d'autres sections ailleurs au pays (Calgary, Toronto, Waterloo), dont deux membres fondateurs de DFG. Les entretiens ont mis au jour des spécificités locales du travail au sein des différentes sections, notamment en lien avec le choix des projets des organismes communautaires.

En parallèle de l'observation et des entretiens, nous avons réalisé une analyse documentaire afin de mieux cerner la raison d'être et le fonctionnement de l'organisme DFG, d'une part, et le projet spécifique qui a mobilisé les bénévoles de la section montréalaise, d'autre part. Le corpus documentaire incluait les documents produits à l'intention des bénévoles et des organismes participants, une douzaine de cas provenant des différentes sections – auxquels se sont ajoutés les présentations produites par la section montréalaise de DFG à chaque réunion –, ainsi que les documents de travail liés au projet du Réseau, à l'exception de ceux qui contenaient des données sensibles. Pour comprendre DFG, nous avons aussi dû mener une collecte des traces en ligne liées à DFG Montréal et aux autres sections : recension

6. À compter d'avril 2020, pandémie oblige, les réunions mensuelles en présence ont migré en ligne ; nos observations ont naturellement emboîté le pas.

des comptes Twitter (maintenant X) et LinkedIn, et des contenus sur des plateformes collaboratives comme GitHub ou GitLab.

Résultats et discussion : les publics constitués autour de DFG

Une fois par mois, la section montréalaise se réunit dans un local de la maison Notman, un tiers-lieu hébergeant des *startups* du milieu technologique à Montréal. Les rencontres rassemblent de 20 à 40 personnes, dont un noyau dur d'une dizaine de membres, durant une heure et demie environ. Ces rencontres mensuelles s'accompagnent d'une activité en ligne en continu, soutenue par un ensemble d'outils collaboratifs, principalement MeetUp (pour convoquer les réunions), Slack (pour coordonner le collectif), Trello (pour organiser le travail sur les projets) et GitHub (pour entreposer des données). Le travail des bénévoles s'organise autour des données du projet en cours, dans ce cas, celles du Réseau.

La gestion du projet repose sur une structure de coordination du travail se déclinant en trois rôles : le responsable des données (*data lead*) planifie les tâches et constitue la personne de référence pour tout ce qui concerne les analyses de données ; l'ambassadeur des données (*data ambassador*) assure le lien avec l'organisme communautaire « client », ici le Réseau. C'est-à-dire que tous les échanges – de données, mais aussi les communications – entre DFG et le Réseau passent par cette personne. Finalement, le responsable de la communauté (*community lead*) assure la cohésion du collectif en intervenant auprès des bénévoles.

Le mandat que le Réseau a déposé auprès de la section montréalaise de DFG consistait d'abord à organiser et à standardiser l'ensemble des données amassées au sein de ses différentes maisons d'hébergement depuis plus de 10 ans, puis à analyser les ressources du Réseau comme les besoins en lits, la durée des séjours, etc. Le défi de ce mandat était de taille, puisque les données tenaient à la fois sur des documents papier et dans des fichiers numériques, principalement en format Excel. Non seulement les données sont consignées de manière variable d'une maison d'hébergement à l'autre, voire d'une personne à l'autre, mais la manière de les enregistrer et de les organiser a également changé dans le temps. Concrètement, ces ensembles de données sont les informations que les intervenantes recueillent lorsqu'elles accueillent les femmes dans les maisons d'hébergement : nom, nombre d'enfants, langue d'usage, existence ou non d'un rapport de police ou d'autres démarches juridiques, type de violence, pays ou région d'origine,

statut juridique, etc⁷. Or, les intervenantes ne sont pas toujours formées à de telles démarches, et, bien souvent, elles sont débordées par la demande et le manque de ressources. Les jeux de données du Réseau sont donc hétérogènes, incomplets, complexes à interpréter.

S'impliquer en tant que bénévole dans DFG suppose de travailler avec des ensembles de données plus ou moins volumineux, arrivant dans toutes sortes de formats et sur toutes sortes de supports, qui se présentent le plus souvent dans un état de désordre (« *messy* », selon le terme mentionné dans plusieurs entretiens). Il s'agit alors de nettoyer, de standardiser, d'analyser et d'interpréter ces données pour les faire « parler ». Le public des bénévoles de DFG agit directement sur ces données dans le sens où il les manipule pour les rendre signifiantes et utiles dans les contextes où elles seront mobilisées.

Une fois traitées par DFG, ces données permettront d'améliorer la gestion et la prévision des ressources au sein du Réseau, par exemple en fournissant des informations sur la clientèle la plus fréquente et ses besoins particuliers. Après des instances gouvernementales ensuite, elles serviront d'appui à des demandes précises sur les ressources manquantes dans le Réseau, notamment en montrant les disparités selon les régions du Québec où le Réseau offre ses services. À ce titre, le public « agissant » des bénévoles de DFG participe à la fois à mettre en forme et à rendre visible le problème de la violence faite aux femmes, d'une part, et la nature des ressources nécessaires au Réseau pour soutenir les victimes, d'autre part.

Replacer les données dans leur contexte

Si les publics se constituent par la reconnaissance d'un problème, le passage à l'action et la rectification de la situation, les femmes représentent un public dans la mesure où elles subissent la violence et, en mobilisant les ressources du Réseau, elles passent aussi à l'action. Les femmes représentent le public minorisé à partir duquel on génère les données en début de processus et qui se trouve *in fine* « touché » par les retombées du travail de DFG et du Réseau en ce qui a trait au financement et à l'amélioration de l'offre de services.

7. Les principaux champs du formulaire d'inscription aux services du Réseau ont été présentés par une bénévole dans une réunion de DFG, le 16 octobre 2019. L'ambassadeur de données avait alors souligné que la région d'origine et le statut juridique sont des éléments importants puisqu'ils influencent souvent la présence ou non, autour d'une femme, d'un filet social comme la famille ou des amitiés de longue date.

Lorsqu'elles ont recours au soutien du Réseau, ces femmes se trouvent dans une situation particulièrement critique et précaire. Des questions de sécurité et même de survie se posent, mais aussi de logistique, notamment au sujet du logement. D'une certaine manière, à ce moment précis, elles n'ont pas le privilège de se mobiliser collectivement parce qu'elles luttent chacune pour répondre à des besoins immédiats et urgents : retrouver un foyer sûr, recouvrer la santé, offrir stabilité et sécurité aux enfants, etc.

Dans un des entretiens, un bénévole qui s'intéresse à l'activisme de données depuis plusieurs années explique que les données avec lesquelles DFG œuvre ont toujours une dimension à la fois locale et temporelle. Pour lui, chaque projet met en données une facette et un moment de la vie. Dans le cas des femmes, cela se limite à leur passage dans le Réseau. Certaines bénéficient de soutien sans résider dans une maison d'hébergement, mais dès la cessation du service, la collecte de données cesse. Les données analysées par DFG montrent d'ailleurs que certaines femmes reviennent à plusieurs reprises, ce qui correspond aussi à la réalité documentée ailleurs et aux difficultés de sortir du cycle de la violence conjugale (LaPierre *et al.*, 2022). L'une des questions des bénévoles du chapitre de Montréal concernait d'ailleurs les facteurs pouvant influencer le retour des femmes. Or le travail d'analyse a pu montrer que la présence d'un réseau social soutenant autour de la personne joue un rôle clé. Les femmes n'ayant pas de famille ou de proches dans leur entourage immédiat vivent davantage de précarité, et le Réseau devient alors une ressource importante dans la réorganisation de leur vie hors du domicile toxique.

Ainsi, la mise en données de l'expérience des femmes – qui sont les premières concernées par le problème – reste située dans le temps, à un moment de leur trajectoire de vie. Les données figent cette expérience et, par la collecte manuelle faite au cas par cas dans chaque établissement du Réseau, un jeu de données substantiel se constitue.

L'intervention de DFG transforme ces données abondantes, mais non standardisées, en données probantes afin d'agir sur le problème de la violence. DFG organise les données et programme des analyses qui illustrent des fluctuations, des tendances, des récurrences. Cela donne une force épistémique aux données (Milette et Millerand, en évaluation), c'est-à-dire la capacité de les « faire parler » pour révéler avec davantage de force argumentative les répercussions du problème.

DFG au Canada : l'importance de l'ancrage local

L'ancrage local de l'organisme DFG est au cœur de sa mission et de son fonctionnement. Structuré en sections dans les principales villes canadiennes, le collectif répond d'abord et avant tout à des problèmes socioéconomiques portés par les organismes communautaires locaux auxquels ils rendent service. Bien que ces problèmes puissent se répéter d'une grande ville à l'autre, chaque section attire des publics situés, constitués autour d'enjeux ou de questions d'importance locale.

Ainsi, la section de Waterloo, en Ontario, entretient des liens privilégiés avec des partenaires du secteur technologique et avec les universités de sa région, comme l'indique une bénévole, directrice de projets : « Nous avons une communauté techno très forte aussi en données et qui est très engagée auprès des organismes à but non lucratif qui [eux aussi] souhaitent développer des partenariats avec la communauté techno. » Selon elle, ce sont les valeurs de Data for Good en général, mais aussi des préoccupations propres à la région de Waterloo, qui orientent le travail quotidien de la section.

La prégnance d'enjeux locaux se reflète également dans les profils et les motivations des bénévoles. Un membre de la section de Calgary, en Alberta, explique que les bénévoles souhaitent travailler en priorité sur les questions liées à la transition écologique et ses effets sur l'économie locale, dans un contexte où la région est marquée par une industrie pétrolière en crise :

Une chose que nous remarquons et c'est probablement unique à Calgary [c'est que] nous vivons une imposante transition énergétique, nous devons diversifier nos sources d'énergie, pétrole et gaz [...]. Donc beaucoup d'emplois dans le secteur des pétrolières ont été éliminés au cours des dernières années. Il en résulte que beaucoup de gens cherchent aussi à transitionner vers de nouvelles professions, de nouvelles carrières.

Le collectif attire donc désormais des spécialistes du domaine des sciences, des technologies, de l'ingénierie, de l'informatique et des mathématiques (STIM) qui souhaitent se former et développer une expertise en science des données sur des questions environnementales en particulier, dans l'optique d'une réorientation de carrière.

L'importance de l'ancrage local se fait également sentir dans la manière dont les projets arrivent à DFG. Généralement, les liens entre les organismes et la section locale de DFG prennent appui sur des réseaux de contacts et de connaissances. À la section montréalaise, la collaboration avec le Réseau a

pu se faire par l'entremise d'un lien d'amitié entre une des personnes responsables et l'un des bénévoles de DFG. À la section de Toronto, une bénévole précise que presque tous les projets sont le fruit de contacts interpersonnels, c'est-à-dire que les organismes partenaires sont le plus souvent proposés par des bénévoles sympathiques à leur cause et leur mission.

Situer les données

En sciences sociales, l'idée de « situer » un phénomène, une personne ou une pratique n'est pas anodine. Dans différentes littératures scientifiques, par exemple en approches féministes ou encore en sociologie, « situer quelque chose » signifie faire l'effort de le replacer dans son contexte original. Dans le cas d'une recherche sur des personnes, il s'agira, sur le plan macro, de relever les informations pertinentes afin de restituer le contexte social, politique et historique qui marque leur expérience. Cela permettra de prendre en considération, dans l'analyse et la discussion, des données les rapports d'oppression en jeu, le sexisme, le racisme ou d'autres éléments utiles pour construire un rapport nuancé et critique aux données. Sur le plan micro, l'effort pour situer les données d'enquête consistera à éviter d'essentialiser les résultats et de réifier les personnes concernées⁸, mais aussi à privilégier les nuances et l'hétérogénéité des expériences uniques par l'examen des saillances et des motifs récurrents (Maillé, 2017; Clair, 2016).

Pour la section montréalaise de DFG, la dimension située des données du Réseau s'est avérée essentielle à leur nettoyage et à leur organisation. En effet, les bénévoles de DFG ont dû comprendre les réalités de la collecte de ces données, sur le terrain, pour les organiser correctement. Un exemple marquant concerne le nombre de lits et le taux d'occupation. Dans les données initialement remises par le Réseau, le taux d'occupation des lits dépassait régulièrement 100 %, ce qui aurait pu s'interpréter comme une erreur ou une anomalie à rectifier. Or, ce taux indiquait plutôt que telle maison avait accueilli plus de femmes et d'enfants qu'elle ne possédait de lits. Situer

8. Il y a plusieurs manières de réifier ou d'essentialiser. La réification peut, par exemple, se faire en généralisant à outrance, ce qui gomme des éléments importants de l'expérience subjective et diversifiée, ou en traitant les personnes comme des objets dans la manière d'analyser les données – un phénomène d'ailleurs décrié par les critiques postcoloniales et les féministes intersectionnelles. Plaquer une lecture romantique ou misérabiliste sur les données est aussi une pratique qui peut réifier ou essentialiser les femmes, particulièrement les femmes précarisées.

la donnée montrait donc un besoin criant à mettre en avant, alors que des personnes avaient dû dormir dans les lits de fortune, et non une colonne de chiffres à rectifier.

La *datafication* évoque souvent un imaginaire des grands jeux de données et un savoir de haut vol, voire une science abstraite (Kitchin, 2014). Notre cas montre que malgré la présence d'un volume relativement important de données, la réalité mise en données ne prend son sens que dans la mesure où les données sont recontextualisées, autrement dit « situées ». Un tel travail de *re-situation* dans le milieu urbain, les enjeux d'immigration et de pauvreté, la question du soutien social des femmes et d'autres réalités de terrain permet ainsi d'éviter de réifier les personnes et les expériences mises en données (Maillé, 2017 ; D'Ignazio et Klein, 2020).

Ce rapport situé aux données s'insère dans la perspective de l'enquête du public compétent chez Dewey. Les allers-retours entre les bénévoles de DFG et le Réseau, par l'entremise de l'ambassadeur de données, servent notamment à fournir ce précieux contexte de « la situation » aux personnes qui ne sont pas sur le terrain. La contribution de DFG aux données vient ainsi soutenir un public déjà actif et mobilisé face au problème des violences envers les femmes, soit le personnel du Réseau, et participer à mieux reformuler le problème, à en saisir les répercussions fines dans les maisons d'hébergement et sur le plan de l'accompagnement offert ou à développer.

Repenser le « faire public »

Le cas de DFG invite à repenser la notion de public. Pour Dewey (2003), celui-ci se crée dans l'action et l'enquête : il s'agit de comprendre un problème et de s'outiller pour y répondre. Dans notre étude, au moment précis où leur expérience est mise en données, les femmes ont des priorités plus urgentes qui limitent leur capacité à s'organiser en public ou à joindre d'autres publics concernés par la lutte contre les violences. Le combat ou la militance pour éradiquer les violences faites aux femmes s'inscrira peut-être un jour dans leur trajectoire, mais leur statut distinct en tant que publics minorisés appelle à infléchir la définition de Dewey. Afin de reconnaître l'apport essentiel de ces publics, particulièrement dans le cas où des problèmes découlent d'enjeux systémiques comme le sexisme et les violences genrées, il est nécessaire de passer d'une conceptualisation du public où la compétence repose sur un apprentissage, dans le cadre plutôt strict de la raison et de la démarche scientifique d'enquête que propose Dewey, à une conceptualisation du public

qui inclut les contributions des personnes concernées, y compris dans les cas où la capacité de mobilisation et d'enquête est limitée, voire impossible pour ces personnes. Autrement dit, comment faire public *avec* les femmes directement concernées par la violence ?

Une piste consiste à prendre en considération les actions posées pour signaler le problème et reconnaître l'expertise expérientielle des personnes concernées directement par la situation. Dans une perspective pragmatiste inspirée de celle de Dewey, cela revient à élargir la définition du travail d'enquête pour reconnaître l'apport essentiel et même déclencheur des actions des femmes qui ont recours au service du Réseau. Ces femmes ne se trouvent pas dans une posture leur permettant de se saisir des données et de se constituer en public par et avec celles-ci. En revanche, directement concernées par le problème, elles jouent un rôle déterminant de deux manières. D'abord, elles rendent visible le problème, elles le subissent et y résistent. Leurs actions, leurs expériences et leurs récits sont essentiels pour formuler le problème. Deuxièmement, elles contribuent à la *datafication* du problème de la violence conjugale ; leurs expériences sont mises en données par l'entremise du Réseau et, potentiellement, par d'autres actions comme le recours aux tribunaux ou des démarches auprès de la police. Les autres publics peuvent ensuite se saisir de ces données, de ces récits et de ces expériences pour mener l'enquête, comprendre et agir sur des rouages du problème. Le Réseau s'en saisit, puis DFG. Chaque maillon apporte une contribution spécifique. Le Réseau offre un soutien direct de première ligne aux femmes, puis utilisera les rapports préparés par DFG pour faire pression afin d'obtenir du financement plus généreux.

Le cadrage théorique de Dewey ne rend pas pleinement justice aux apports des femmes qui sollicitent les services du Réseau. La compétence des femmes à établir ce qui fait problème et ce qui est digne d'intérêt pour appréhender ce problème se fonde plutôt sur l'expérience même de l'oppression et de la violence. Le cas empirique étudié invite ainsi à revoir les manières de concevoir le « faire public » afin d'intégrer les compétences expérientielles des personnes directement concernées par un problème, mais dont la capacité d'agir est limitée par leur situation de minorisation.

Le cas de DFG met en relief la dimension politique de la mise en données : l'invisible est difficilement considéré comme un problème ou, pour le dire autrement, ce qui n'est pas compté ne « compte pas » (D'Ignazio et Klein, 2020). L'expertise mise à disposition par DFG offre des occasions opportunes pour que les groupes historiquement minorisés « comptent » plus et mieux.

La représentation au moyen de données qui rendent visibles des groupes, des injustices et des problèmes constitue un puissant levier pour faire reconnaître les problèmes que subissent les personnes minorisées (D'Ignazio et Klein, 2020). Cela recoupe d'ailleurs l'idéal sous-jacent à la mission de DFG : « faire que les données comptent » (*make data count*) afin de contribuer à la justice sociale.

La conception de Dewey sur les publics et celle de Laplanche-Servigne et Sa Vilas Boas sur les publics minorisés se révèlent mutuellement fécondes pour penser la situation entourant les données traitées par DFG et les différents publics mobilisés. Discerner les publics minorisés, reconnaître leur agentivité et leur contribution à la définition même d'un problème représente une manière porteuse d'envisager les publics de manière critique et située. De plus, cet arrimage théorique enrichit la dimension politique du concept de public en prenant en compte les luttes de pouvoir à même l'expérience des publics, et particulièrement des publics minorisés. Sur le terrain, les analyses de DFG permettent au Réseau d'améliorer son efficacité et d'optimiser ses actions – ce qui était d'ailleurs une demande explicite de la part du réseau au moment de contacter les bénévoles de DFG. Sur le plan sociétal, ces données outillent les dirigeantes du Réseau dans leurs démarches auprès des élus afin d'obtenir davantage de fonds, mais aussi de mieux soutenir les actions communautaires et préventives. Ainsi, de la même manière que chaque public se constitue de personnes hétérogènes, les manières d'agir avec et par les données se distinguent pour chacun des publics concernés par le problème des violences faites aux femmes.

Références

- BAACK, Stefan, « Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism ». *Big Data & Society*, 2, n° 2, 2015.
- CLAIR, Isabelle, « Faire du terrain en féministe », *Actes de la recherche en sciences sociales*, vol. 213, n° 3, 2016, p. 66-83.
- DESCARRIES, Francine, « L'antiféminisme "ordinaire" », *Recherches féministes*, vol. 18, n° 2, 2005, p. 137-151.
- DEWEY, John, *Le public et ses problèmes. Œuvres philosophiques*, vol. 2, traduit par Joëlle Zask, Pau, Publ. de l'Université de Pau, Farrago L. Scheer, 2003.

- D'IGNAZIO, Catherine et KLEIN, Lauren F., *Data Feminism*, Cambridge, The MIT Press, 2020.
- DUNBAR-HESTER, Christina, « Soldering Toward Media Democracy: Technical Practice as Symbolic Value in Radio Activism », *Journal of Communication Inquiry*, vol. 36, n° 2, 2012, p. 149-169.
- ERMOSHINA, Ksenia, « Civic Hacking. Redefining Hackers and Civic Participation », *TECNOSCIENZA: Italian Journal of Science & Technology Studies*, vol. 9, n° 1, 2018, p. 79-101.
- FRASER, Nancy, « Recognition without Ethics? », *Theory, Culture & Society*, vol. 18, n° 2-3, 2001, p. 21-42.
- GUILLAUMIN, Colette, « The Masculine: Denotations/Connotations ». *Feminist Issues*, 5, n° 1, 1985, p. 65-73.
- HINE, Christine, *Ethnography for the Internet: Embedded, Embodied and Everyday*, Londres, Routledge, 2015.
- KATZ, Elihu, et DAYAN, Daniel, « On publics, non-publics, former publics, future publics, almost publics, and their students and genealogies », dans JACOBI, Daniel et LUCKERHOFF, Jason (dir.), *Looking for Non-Publics*, Québec, Presses de l'Université du Québec, 2012, p. VII-XV.
- KITCHIN, Rob, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*, Thousand Oaks, Sage Publications, 2014.
- LAPIERRE, Alicia, PEDNEAULT, Chloe I. et NUNES, Kevin L., « Factor Structure and Distinctiveness of the Violent Behavior Vignette Questionnaire », *Journal of Aggression, Maltreatment & Trauma*, 31, n° 7, 2022, p. 926-941.
- LAPLANCHE-SERVIGNE, Soline et SA VILAS BOAS, Marie-Hélène, « Introduction. Les catégorisations des publics minorisés en questions », *Participations*, vol. 25, n° 3, 2019, p. 5-31.
- LIVINGSTONE, Sonia, « Du rapport entre audiences et publics », *Réseaux*, vol. 126, n° 4, 2004, p. 17-55.
- MAILLÉ, Chantal, « Approche intersectionnelle, théorie postcoloniale et questions de différence dans les féminismes anglo-saxons et francophones », *Politique et sociétés*, vol. 36, Numéro spécial, 2017, p. 163-182.
- MARRES, Noortje, « Issues spark a public into being: A key but often forgotten point of the Lippmann-Dewey debate », dans LATOUR, Bruno et WEIBEL, Peter (dir.), *Making Things Public: Atmospheres of Democracy*, Cambridge, The MIT Press, 2005, p. 208-217.
- MILAN, Stefania, « Data activism as the new frontier of media activism », dans PICKARD, Victor et YANG, Guobin (dir.), *Media Activism in the Digital Age*, Londres, Routledge, 2017, p. 151-163.
- MILAN, Stefania et GUTIÉRREZ, Miren, « Citizens' media meets big data: The emergence of data activism », *Mediaciones*, vol. 11, n° 14, 2015, p. 120-133.
- MILAN, Stefania et VAN DER VELDEN, Lonneke, « The Alternative Epistemologies of Data Activism », *Digital Culture & Society*, vol. 2, n° 2, 2016, p. 57-74.
- MILLETTE, Mélanie et Florence MILLERAND, « Data Community Service as an Indirect Form of Activism: An Ethnography of the Collective 'Data for Good' », *The Information Society*, en évaluation.

SOULÉ, Bastien, « Observation participante ou participation observante? Usages et justifications de la notion de participation observante en sciences sociales », *Recherches qualitatives*, 27, n° 1, 2007, p. 127-40.

ZASK, Joëlle, « Le public chez Dewey : une union sociale plurielle », *Tracés, Revue de sciences humaines*, 15, 2008, p. 169-189.

CHAPITRE 8

L'inclusion des groupes minorisés dans les politiques publiques de données

Lucie Delias

Cette étude s'intéresse à un type de public particulier : celui constitué par les personnes appartenant aux groupes minorisés (Laplanche-Servigne et Sa Vilas Boas, 2019), c'est-à-dire les groupes historiquement discriminés sur les plans symbolique et matériel en raison de leur genre, de leur orientation sexuelle, de leur statut socioéconomique ou ethnoracial, ou encore de leur handicap¹. Qu'il s'agisse de la Ville de Montréal, avec son portail de données ouvertes, de la plateforme de santé durable Pulsar ou du collectif Data for Good, les initiatives mentionnées dans cet ouvrage ont pour point commun de produire, d'utiliser et de partager des données dans une visée non lucrative et pour le bien commun. Cette mobilisation des données a pour prétention de produire des retombées sociales, économiques et politiques positives pour l'ensemble de la population de la ville, de la région ou de la communauté concernée. Les projets ont donc pour vocation de s'adresser également aux groupes marginalisés.

La reconnaissance des conséquences spécifiques de l'utilisation des données sur les plus vulnérables, mais aussi de l'occasion favorable qu'elles peuvent représenter en termes de justice sociale, s'inscrit également dans un contexte où les organisations (publiques comme privées) investissent de plus en plus le terrain de l'encadrement éthique de leurs pratiques afin de

1. Au Canada, il existe des catégories spécifiques de populations dont la minorisation est liée au contexte sociohistorique du territoire, comme la langue parlée et le statut d'autochtone, ainsi que le montrent par exemple les catégories utilisées par Statistique Canada pour mesurer « le genre, la diversité et l'inclusion » dans ses enquêtes (Statistique Canada, 2023).

répondre aux critiques croissantes de la société civile qui font suite à la médiatisation de plusieurs scandales touchant à la surveillance et aux discriminations algorithmiques. On observe que cette « inflation éthique » donne surtout lieu à des déclarations d'intention qui s'accompagnent rarement d'actions concrètes et réellement transformatrices (Zacklad et Rouvroy, 2022).

Données publiques, discriminations et encadrement éthique

L'un des apports principaux des travaux se situant dans le champ des *Critical Data Studies* est de considérer les jeux de données comme des artefacts issus d'une construction sociale et donc nécessairement situés, à l'encontre des discours enthousiastes (portant surtout sur le *big data*) qui les présentent comme neutres, efficaces, et totalisants (Gitelman, 2013). Cette perspective permet notamment de penser les rapports de pouvoir historiquement asymétriques qui traversent les données et les infrastructures qui les encadrent au désavantage des groupes minorisés (D'Ignazio et Klein, 2020). Révélées par les journalistes et les chercheurs, les controverses autour de la reproduction et du renforcement des inégalités sociales par les données et leur exploitation automatisée ne concernent pas seulement les entreprises privées, mais également les organisations publiques. Les institutions qui utilisent des algorithmes afin de prendre des décisions qui s'appliquent aux administrés (comme l'attribution ou le retrait d'allocations sociales) sont accusées d'utiliser des jeux de données où les populations traditionnellement discriminées en raison de leur origine sociale sont surreprésentées ou sous-représentées, ce qui crée une « boucle de rétroaction négative » (O'Neil, 2016). Les données ouvertes, malgré les valeurs de démocratie, de transparence et de participation citoyenne qui ont porté leur conception, n'y échappent pas non plus. Les jeux de données publiés en licence ouverte qui se rapportent à certains groupes peuvent leur nuire : c'est le cas des données sur les populations autochtones qui, imprégnées d'un héritage colonial et raciste, ne décrivent les personnes qu'à partir de leurs « manques » (faible niveau d'éducation, prévalence de l'alcoolisme, etc.), reproduisant ainsi un « récit déficitaire » (*deficit narrative*) préjudiciable (Walter *et al.*, 2021). Au-delà des enjeux de représentation, les réutilisations des données publiques peuvent aller directement à l'encontre des intérêts des groupes minorisés, comme en témoigne l'exemple largement commenté de Bangalore, qui, en rendant public le cadastre des terres agricoles de la commune, a permis aux riches proprié-

taires terriens d'identifier et d'exproprier les petits paysans dalits (Gurstein, 2011), ce qui a révélé ainsi un *data divide* (fracture numérique) impensé, lié aux inégalités entre les acteurs ayant les moyens matériels et intellectuels d'exploiter les données et les autres.

La médiation croissante de ce genre d'affaire auprès de l'opinion publique a produit ces dernières années des injonctions de plus en plus pressantes pour les organisations qui travaillent avec les données à réguler leurs activités. Cette autorégulation passe en partie par une prolifération de chartes et de déclarations énonçant des principes et de bonnes pratiques à respecter. Jobin, Ienca et Vayena (AI, (2019) ont rassemblé un corpus de 84 chartes éthiques portant sur les données et l'intelligence artificielle, dont 64 font référence aux concepts de « *justice, fairness, and equity* » (justice, impartialité et équité) – démontrant que la question de la lutte contre les biais et les discriminations est désormais un thème majeur dans la régulation des conséquences de la *datafication*. Cependant, d'aucuns dénoncent ces prises de position comme opportunistes et relevant d'une « démarche marketing », puisque les organisations qui s'affichent comme « responsables » peuvent transformer leur engagement en « avantage compétitif » sur le marché (Hasselbalch et Tranberg, 2016). Un autre ensemble de critiques se concentre sur l'aspect « incantatoire » (Zacklad et Rouvroy, 2022) de ces documents. D'une part, en l'absence habituelle de vocation contraignante, ils peuvent être assimilés à de simples déclarations d'intention. D'autre part, certains de ces principes sont considérés par les chercheurs en *Critical Data Studies* comme insuffisants, voire contre-productifs. C'est le cas par exemple de la rhétorique de la « *bad apple* », qui fait reposer la responsabilité des discriminations technologiques sur des individus défaillants et isolés, invisibilisant ainsi les inégalités systémiques et l'ubiquité des rapports de pouvoir (Hoffmann, 2019). Plus généralement, ce qu'on entend par « éthique » dans le positionnement des organisations demeure souvent flou.

Définir l'« inclusion » : l'éthique des données observée par le prisme des politiques de la différence

Dans le domaine des politiques de lutte contre les inégalités liées à la diffusion et à l'utilisation des outils numériques, dont les données, on observe ces dernières années un glissement en faveur du terme d'« inclusion » pour décrire à la fois l'approche employée et les actions mises en place pour rapprocher les publics en situation de discrimination ou de non-usage de ces

outils. Comme l'« éthique », l'inclusion est rarement définie par les acteurs qui mobilisent ce terme, et la Ville de Montréal ne fait pas exception. Nous proposons d'analyser le positionnement des acteurs à partir des apports de la philosophie politique féministe, qui nous aideront notamment à clarifier les concepts d'inclusion et de participation des publics minorisés à la vie publique.

Selon Young (1990) et Fraser (2011), l'idéal démocratique traditionnel qui prévaut dans les sociétés occidentales se fonde sur un déni des différences sociales et culturelles entre les citoyens et les groupes de citoyens, au profit d'un universalisme qui ne permet pas la réalisation de la justice sociale et de l'égalité participation de tous à la vie publique. Ces autrices critiquent notamment la conception habermassienne de l'espace public, dont les conditions d'existence impliquent un consensus autour des sujets d'intérêts généraux à mettre en discussion et une égale maîtrise des outils de délibération par les participants. Dans ce que Fraser appelle les « sociétés stratifiées » qui sont les nôtres, c'est-à-dire dans lesquelles certains groupes sont systématiquement désavantagés par rapport à d'autres sur la base de leur position sociale et de leur identité², il apparaît utopique, voire contre-productif, de mettre en place des buts et d'exiger des conditions d'expression qui soient les mêmes pour tous. En ce qui concerne les discriminations liées aux données, on peut considérer par exemple que la production de données sur les personnes transgenres ou non binaires peut servir l'intérêt des chercheurs pour faire avancer les connaissances, mais aller à l'encontre de celui des individus concernés en leur donnant une forme de visibilité qui pourrait en faire les cibles de stigmatisation ou d'attaques transphobes. De plus, dans un contexte où, on l'a vu, les moyens techniques et l'expertise attendus pour s'appropriier les données ne sont accessibles qu'à un nombre réduit d'individus et d'organisations, il est peu raisonnable d'attendre de l'ensemble des citoyens qu'ils soient en mesure de saisir les occasions que peut fournir la maîtrise des données, notamment ouvertes. Sans prise en compte de ces déséquilibres, les politiques mises en place relèvent alors bien plus de l'intégration que de

2. Par exemple, Fraser montre que les définitions de ce qui relève de « l'intérêt commun » ont longtemps laissé de côté les problèmes liés à la sphère privée, ce qui bloquait les discussions autour de la domination des femmes (notamment l'exploitation de leur travail domestique et reproductif) avant que les mouvements féministes des années 1970 n'imposent l'idée que « l'intime est politique ». De la même façon, la raison délibérative habermassienne est définie à partir des normes discursives des hommes éduqués, ce qui disqualifie des modes d'expression utilisant par exemple le recours aux émotions, associés aux femmes et aux classes populaires.

l'inclusion, où l'on exige des individus une adaptation aux normes qui, loin d'être « universelles », sont en réalité celles des groupes dominants, « naturalisées » comme étant consensuelles (Pillant, 2019). À l'inverse, l'inclusion comme résultat d'une « politique des différences » consiste à reconnaître les besoins et les situations différenciées de chacun pour garantir sa participation, avec l'horizon théorique d'une société, débarrassée des inégalités sociales, où des solutions universelles pourraient effectivement être appliquées.

En complément de cette reconnaissance, il apparaît nécessaire de créer des conditions favorables pour que les citoyens puissent s'emparer des questions liées aux données, dans un contexte contemporain de mise en données de la vie sociale où les données sont à la fois un potentiel instrument d'oppression et un outil de lutte pour les groupes minorisés. Ces conditions exigent notamment que ces groupes aient accès aux informations les concernant, mais aient aussi la liberté et les moyens de se former, d'enquêter et d'exercer leur jugement afin d'émerger en publics actifs au sens proposé par Dewey (Zask, 2008 ; Millette, Millerand et Seraiocco, cet ouvrage).

Le cas de la Charte des données numériques de la Ville de Montréal

La Ville de Montréal lance son portail de données ouvertes en 2011 et développe, au fil du temps et de ses administrations municipales, une politique de plus en plus volontariste en matière d'ouverture de données et, plus globalement, de ville intelligente (Millerand et Meunier, cet ouvrage). Ce n'est cependant qu'en octobre 2020 que la Ville se dote d'une charte explicitement dédiée à l'encadrement éthique des données municipales (et plus largement de l'utilisation du numérique) par les services et organismes associés à la Ville. Cette Charte a été pensée en collaboration avec la Ville de Nantes³, en France, et dans le sillage de la publication de la déclaration de Montréal pour une IA responsable⁴. Elle est le résultat d'un processus entamé en mai 2019, que l'équipe responsable du portail veut collaboratif, puisque maints intervenants associés à la Ville sont amenés à participer à son élaboration, notamment différents services municipaux et des chercheurs⁵. Le résultat final est un texte composé de trois grands engagements

3. Nantes Métropole, *Charte métropolitaine de la donnée*, 2019.

4. Publiée en 2017 à la suite d'un processus de rédaction participatif par l'Université de Montréal.

5. Une version non définitive du texte de la Charte a été ouverte aux commentaires publics en ligne, mais le dispositif a suscité peu de participation. En outre, ce sont principa-

principaux déclinés en 13 principes couvrant de nombreuses thématiques associées à la gestion des données numériques, allant de la protection de la vie privée à la sobriété énergétique ; deux principes portent directement sur les questions d'inégalités et de discriminations liées aux données. La Charte n'est pas contraignante et les équipes concernées travaillent actuellement à l'opérationnalisation de ces principes.

Afin d'interroger les conceptions de l'inclusion des groupes minorisés déployées par les différents acteurs et la façon dont la Charte actualise ces valeurs d'égalité et d'inclusion, nous avons mené des entretiens semi-directifs et des analyses documentaires avec une approche qualitative. Nous nous sommes basée sur les données d'une enquête visant à décrire les publics « imaginés » du portail de données ouvertes (Millerand *et al.*, 2023) et consistant en l'analyse d'un corpus de documents officiels et techniques et de publications sur l'ouverture des données (N=81), mais aussi d'entretiens avec les principaux agents municipaux concernés (N=10). À cette enquête primaire s'est ajoutée une enquête secondaire sur la Charte, dans laquelle nous avons effectué des entretiens (N=5) avec sept professionnels impliqués dans la création et l'opérationnalisation de la Charte au sein de l'appareil municipal.

D'un point de vue méthodologique, la Charte apparaît comme un cas intéressant à étudier en sciences de l'information et de la communication et des *Science and Technology Studies*. Son texte, bien que déjà en circulation, est amené à évoluer au fil du temps en fonction des retours des « parties prenantes », tandis que les acteurs des services de la Ville sont actuellement en train de réfléchir aux moyens d'opérationnaliser les principes qui la sous-tendent. Ce texte est le produit de discussions entre des acteurs aux activités, aux intérêts et aux visions parfois antagonistes en ce qui concerne les données et l'encadrement de leurs usages. La Charte apparaît ainsi comme une sorte d'« objet-frontière » (Leigh Star, 2010) qui amène les professionnels à dialoguer autour de leurs pratiques et de leurs attentes, même si les termes de ce dialogue semblent contradictoires et les principes irréalisables – une critique d'ailleurs parfois adressée à la Charte en interne. Le processus d'écriture et d'opérationnalisation du texte a donc un double enjeu : en interne, mobiliser et faire coopérer les agents autour de la question des données ; à l'externe, afficher l'implication de la Ville dans les questions de

lement des collaborateurs professionnels du service responsable du portail qui ont rédigé le petit nombre de commentaires.

régulation éthique dans une perspective politique et communicationnelle. Cette étude, encore en cours, entend donc suivre la trajectoire de la Charte et son devenir.

Compter pour visibilité et caractériser les inégalités : représenter les publics minorisés dans les jeux de données

Le second principe énoncé dans la Charte des données numériques s'intitule « Inclusion ». Il y est expliqué qu'« afin d'éviter l'introduction de possibles biais et d'identifier les angles morts qui pourraient exclure des groupes de populations vulnérables et renforcer les inégalités existantes, la gestion de données doit permettre la mise en œuvre des principes de l'analyse différenciée selon les sexes dans une perspective intersectionnelle (ADS+) ». Ainsi, la Ville « s'engag[e] à intégrer l'application de l'ADS+ dans la collecte, le traitement, l'analyse et la diffusion des données ». L'analyse différenciée selon les sexes (ADS+) ne figurait pas dans les premières versions de travail du texte de la Charte; elle y a été ajoutée à la suite de la consultation du service responsable des questions d'inclusion sociale. Il s'agit à la fois d'« un processus, d'un outil, et d'une méthode » (Entretien 4, personne du service de l'inclusion sociale) de management de la diversité développé dans les années 1990⁶, qui vise à encourager les institutions publiques à sortir de la vision universaliste des citoyens comme un groupe homogène, composé d'individus moyens qui auraient les mêmes ressources et les mêmes expériences, pour adapter les politiques publiques en fonction des particularités et des besoins des différentes catégories de personnes qui composent la société. Si l'ADS+ comporte des outils classiques comme les formations à destination des collaborateurs, elle repose principalement sur les données – décrites comme le « premier réflexe en ADS+ » (Ville de Montréal, 2020). Il s'agit ainsi « d'avoir recours à des données quantitatives et qualitatives incluant des données ventilées selon le sexe et à une désagrégation plus fine selon l'âge, le revenu, l'appartenance ethnique, la région ou toute autre variable significative pour l'analyse » (Massé, Laberge et Massé, 2002). Par exemple, le principe de l'ADS+ appliqué aux données municipales révélerait

6. Alors que l'ADS commence à être appliquée au Québec dès 1997, le processus est initié à Montréal en 2008 lorsque la Ville se dote d'une politique d'égalité femmes-hommes. C'est seulement depuis 2018 que le Service de la diversité et de l'inclusion sociale (SDIS) a explicitement pour but de diffuser l'ADS+ de manière transversale dans l'administration. (Entretien 4, personne du SDIS)

si les entreprises qui composent la liste des fournisseurs de la Ville de Montréal (liste actuellement publiée sur le portail de données ouvertes) appartiennent ou sont gérées par des personnes issues de groupes minorisés, afin que les contrats puissent être attribués de manière plus équitable.

Cette vision se rapproche des conceptions de l'inclusion issues de la philosophie politique féministe ; la meilleure représentation chiffrée des particularités, et donc de la complexité des populations, est par ailleurs une des voies recommandées par les chercheuses en *Feminist Data Studies*, dont les avantages sont résumés par la formule « Ce qui est comptabilisé compte⁷ » (D'Ignazio et Klein, 2020). Une des personnes du service de l'inclusion sociale résume ainsi cette position :

Quand on quantifie pas les choses, on sait pas. Par exemple, aux RH, c'est pas clair, c'est quoi les populations ? Qui est là ? À quel niveau ? En fait, c'est des choses qui sont sues, mais qui ne sont pas publiques. Mais c'est la même chose [...], quand on donne des subventions, qui n'y a pas accès ? C'est vraiment difficile, voire impossible, de voir comment on s'améliore, comment est-ce qu'on lutte contre les discriminations si on a aucun chiffre. (Entretien 5, personne du service de l'inclusion sociale)

Les données « riches » apparaissent comme une base à la fois pour la connaissance et pour l'action. Elles permettent de décrire une réalité de manière systématique et, plus encore, dans un contexte où la valeur sociale et rhétorique du chiffre est particulièrement forte, de conférer aux phénomènes décrits une visibilité et une légitimité importantes auprès des pouvoirs et de l'opinion publique. Articulées à cette dimension symbolique, elles servent aussi de levier de légitimation pour justifier la mise en place de programmes. Une personne parmi les agents du SDIS souligne l'importance de disposer de « données probantes » : « Ça veut dire qu'on est capable de prendre des décisions sur ces données-là. C'est de la donnée de qualité [...], mais c'est les mêmes définitions de la donnée probante en médecine, c'est la donnée qui nous permet de faire des tests statistiques, de montrer une certaine représentativité, de prendre des décisions là-dessus. » (Entretien 6, personne du service de l'inclusion sociale)

La Charte apparaît également comme un instrument utile au service de l'inclusion sociale pour mener à bien une de ses missions principales :

diffuser la culture de l'ADS+ à l'ensemble des services de la Ville, en espérant que les données produites à l'avenir par les services respecteront ces principes,

7. « *What gets counted counts* » (traduction libre).

puisque ses agents ont une mission transversale qui doit toucher le plus de professionnels possible et engager, pour reprendre les mots d'une personne rencontrée, un processus de transformation organisationnelle. (Entretien 4, personne du service de l'inclusion sociale)

Cependant, ce programme rencontre pour l'instant deux limites. D'abord, dans son application concrète, le principe de l'ADS+ entre en contradiction avec d'autres principes de la Charte, notamment celui de la « sobriété numérique », qui stipule que « conformément au cadre légal, toutes collectes d'informations personnelles sont limitées à celles strictement nécessaires à la prestation d'un service ». Cette contradiction pose des défis d'opérationnalisation, comme le souligne une personne du service des technologies et de l'information (TI) :

Nous, aux TI, [...] on va chercher une information que si elle est nécessaire. C'est un peu un paradoxe. Quand tu arrives avec l'approche ADS+ par exemple sur des logiciels de bibliothèque, c'est un peu difficile de demander l'orientation sexuelle de quelqu'un pour pouvoir l'autoriser à accéder en ligne [...] déjà, on avait énormément de données sur le citoyen et là on touche au cœur de leur intimité, c'est encore un niveau au-dessus, donc ça va nous demander certainement un niveau au-dessus de protection, de technique. (Entretien 2, personne du service des TI)

C'est tout l'enjeu politique de la définition de ce qu'est une donnée essentielle qui émerge ici : la lutte contre les discriminations est-elle un enjeu prioritaire qui justifie les efforts demandés aux autres services pour produire et protéger ces données ? On voit ici que le programme et les conceptions de ce qu'est une donnée « éthique » divergent au sein de la municipalité.

La seconde limite réside dans le fait que très peu de jeux de données répondant aux principes de l'ADS+ existent, qu'ils soient accessibles en licence ouverte ou destinés à un usage interne. De manière significative, ces données ont été collectées ou rendues publiques grâce à la mobilisation de groupes de la société civile. C'est le cas, par exemple, des chiffres sur la diversité des profils des cadres intermédiaires et supérieurs employés par la Ville, rendus publics après la mobilisation du Conseil des Montréalaises qui a demandé à les connaître (Entretien 5, personne du service de l'inclusion sociale)⁸. Cette situation fait écho au déséquilibre dans les sujets représentés

8. C'est également le cas du seul jeu de données disponible sur le portail de données ouvertes sous l'étiquette « ADS » : il s'agit des résultats d'un sondage commandé à un prestataire et mené auprès des habitants de la Ville sur l'inclusion des personnes immigrantes, qui trouve sa place dans le plan d'action *Solidarité, équité et inclusion* mis en place après une

dans les jeux de données publiés sur le portail : on y observe une forte invisibilisation des sujets liés aux thématiques sociales comme l'itinérance ou encore la garde d'enfants, ce qui entraîne la non-utilisation du portail par certains groupes dont les intérêts sont absents. La Charte, cependant, fait le choix de mettre l'accent sur une autre cause de non-utilisation des données : le manque de compétences numériques des citoyens.

***Former les publics à la compréhension et à l'utilisation des données :
une approche universelle***

Le dixième principe de la Charte des données numériques vise à s'attaquer à la « fracture numérique » et s'intitule « Universalité d'accès ». Il y est stipulé ceci :

La Ville souhaite s'assurer que la population ait les moyens et soit outillée pour accéder aux données et les comprendre. [...] Ainsi nous nous engageons à déployer l'ensemble des leviers déterminants pour garantir la démocratisation et l'équité numérique, entre autres en développant la littératie numérique autour des données pour faciliter leurs utilisations dans la vie citoyenne et l'émergence de solutions innovantes.

Ce principe met clairement l'accent sur la littératie de données (Frank *et al.*, 2016) en lien avec l'utilisation citoyenne des données ouvertes. Sa rédaction est principalement le fruit de l'équipe responsable du portail, dont une personne qui en est membre précise :

On le voit au fil du temps [...], c'est bien les données ouvertes, mais on réalise que des données brutes, ce n'est pas digeste pour tous, [ce qui] fait qu'on a un gros travail à faire pour les rendre plus accessibles. Les rendre plus accessibles, c'est outiller les gens à mieux comprendre qu'est-ce qu'une donnée, ça va jusqu'à mieux mâcher la donnée, la valoriser pour que les gens s'approprient le contexte de la donnée. (Entretien 1, personne de l'équipe responsable du portail de données ouvertes)

La nécessité de diffuser une « culture de la donnée » (Casemajor, cet ouvrage) à l'externe (et non plus en interne comme c'est le cas avec l'ADS+) comporte un volet connexe qui préoccupe les porteurs des données ouvertes publiques depuis de nombreuses années : susciter des usages pour les données mises à disposition de la population sans nécessairement qu'ait été pensé en amont leur intérêt potentiel pour les citoyens (Dymytriva et Paquiénéguy,

grande consultation sur le racisme et les discriminations à Montréal déclenchée par la mobilisation de groupes communautaires en 2018.

2017). Il s'agit d'abord de rendre accessibles à tous sans distinction les ressources produites par la Ville dans le cadre de son mandat de service public – et *a fortiori* les données ouvertes que l'on peut considérer comme un bien commun. Cependant, l'horizon d'une population devenue plus compétente en matière de données peut justifier la pertinence de l'activité d'ouverture des données et laisser espérer l'actualisation des idéaux de l'*open data* dans l'aspect démocratique de la participation politique («la vie citoyenne») et économique («l'émergence de solutions innovantes»).

Ici encore, l'opérationnalisation de ce principe éthique reste incertaine au moment de l'enquête. Cependant, les leviers évoqués par les professionnels rencontrés sont très semblables à ceux utilisés dans d'autres contextes et relevant de stratégie de médiation, d'une part, et de stratégie d'éducation, d'autre part (Lehmans, 2018). Après une période où les efforts de médiation étaient surtout dirigés vers les citoyens «experts» sous la forme, par exemple, de hackatons, ce sont les visualisations de données qui sont présentées comme l'outil de vulgarisation par excellence (ou par défaut) des enjeux liés aux données. Ainsi, l'équipe responsable du portail travaille à l'intégration de visualisations de certains jeux de données jugés d'intérêt pour les citoyens :

En ce moment, on se doute que les gens qui utilisent le portail des données ouvertes, ça ne sera pas ma mère ni mon père qui vont aller chercher un fichier pour l'explorer. Mais ça serait intéressant que justement si ma mère souhaite savoir c'est quoi la qualité de l'air à Montréal, elle puisse avoir un outil pour la visualiser. (Entretien 3, personne de l'équipe responsable du portail)

C'est également le cas du service de l'inclusion sociale, dont les membres souhaitent médiatiser les données sur les inégalités produites en son sein grâce à des «tableaux de bord», afin de créer «un intermédiaire entre l'intelligence [produite] et le *raw data* [donnée brute]» (Entretien 6, personne du service de l'inclusion sociale). Dans de nombreux domaines (éducatif, politique, scientifique), le recours à la visualisation apparaît comme un outil idéal de vulgarisation, sans que l'on remette toujours remis en question le fait que la lecture de ce genre de graphique demande déjà un niveau initial de littératie (Kennedy et Engebretsen, 2020). Le projet énoncé dans la Charte laisse également la porte ouverte à des stratégies d'éducation qui, au-delà de la médiation, relève plus directement du développement d'une littératie de données. Cependant, au-delà de l'évocation de la mise en place de «programmes en bibliothèque» destinés à proposer des ateliers de quartier autour de ces questions (Entretien 1, personne responsable de l'ouverture des données), la forme que ces stratégies pourraient prendre demeure floue.

Ici, le type de public visé et la conception de l'inclusion qui se déploie se rattachent plutôt à la logique de l'intégration. On convoque la figure du citoyen « ordinaire », envisagé à partir de son universalité – il pourrait être n'importe quelle personne non experte – et non de caractéristiques identitaires vectrices de discriminations. De manière intéressante, la personne professionnelle de l'équipe du portail citée plus haut incarne ce citoyen dans la figure de parents âgés (« ma mère » ou « mon père ») qui représentent un groupe de population (les aînés) dont la minorisation est spécifique, puisque les difficultés liées au vieillissement concernent potentiellement l'ensemble des citoyens à un moment donné de leur existence, quel que soit leur profil par ailleurs. La seule caractéristique commune qui réunit le public ciblé par ces stratégies est donc le manque de compétences supposées en matière de compréhension et de manipulation des données. Dans cette version de l'inclusion, il s'agit donc moins de reconnaître une injustice que de donner l'occasion à un ensemble homogène de citoyens de s'intégrer aux normes et de profiter des circonstances d'un environnement caractérisé par l'ubiquité des données.

La Charte des données numériques constitue un effort de la Ville de Montréal pour amorcer une réflexion sur un encadrement éthique commun des pratiques municipales en la matière. Elle apparaît comme un objet à la frontière de plusieurs projets et visions. À l'externe, elle traduit une question politique en matière d'image, qui vise à montrer que l'institution agit sur ces questions dans un contexte de forte demande sociale de régulation. En interne, elle apparaît comme une base de discussion pour mobiliser les professionnels de son administration pour qu'ils mettent en place des pratiques plus vertueuses, ce qui explique le caractère hétéroclite et la prétention à l'exhaustivité des principes proposés.

Ainsi, deux conceptions de l'« inclusion » et des publics qu'elle vise cohabitent dans le texte. En interne, il s'agit de développer une culture de la donnée « désagrégée et intersectionnelle » dans une logique de reconnaissance des différences, où les publics minorisés sont envisagés à partir de leur variété. Les données servent alors à un projet de visibilisation (elles viennent décrire la réalité des inégalités sociales) et d'action (elles sont une base pour justifier la mise en place de politiques publiques de luttes contre les discriminations). À l'externe, le but est de développer une « culture de la donnée » auprès des

citoyens, par des stratégies de médiation et d'éducation où les publics ne sont pas imaginés à partir des besoins spécifiques liés à leur position sociale, mais plutôt selon la nécessité, pensée comme universelle, de s'adapter à un monde où les données sont omniprésentes. Les publics minorisés se trouvent alors dissous dans le groupe large des « non-experts ».

Il sera intéressant de suivre la façon dont les principes éthiques se traduiront en actions concrètes – on a vu les frictions que les premières tentatives ont générées – et d'élargir le périmètre du terrain à d'autres institutions publiques, afin d'observer, à des fins de comparaison, comment la question de l'« inclusion » s'y définit.

Références

- D'IGNAZIO, Catherine, et Lauren F. Klein, *Data Feminism*, Cambridge, The MIT Press, 2020.
- DYMYTROVA, Valentyna et PAQUIENSÉGUY, Françoise, « La réutilisation et les réutilisateurs des données ouvertes en France : une approche centrée sur les usagers », *Revue internationale des gouvernements oerts*, vol. 5, 2017, p. 117-132.
- FRANK, Mark, WALKER, Johanna, ATTARD, Judie et TYGEL, Alan, « Data Literacy – What is it and how can we make it happen? », *The Journal of Community Informatics* 12, n° 3, 19 septembre 2016, p. 4-8.
- FRASER, Nancy, *Qu'est-ce que la justice sociale?*, Paris, La Découverte, 2011.
- GITELMAN, Lisa, « *Raw Data* » *Is an Oxymoron*, Cambridge, The MIT Press, 2013.
- GURSTEIN, Michael B., « Open Data : Empowering the Empowered or Effective Data Use for Everyone? », *First Monday*, vol. 2, n° 16, 2011.
- HASSELBALCH, Guy et TRANBERG, Pernille, *Data Ethics. The New Competitive Advantage*, Copenhagen, Publishare, 2016.
- HOFFMANN, Anna Lauren, « Where fairness fails : Data, algorithms, and the limits of antidiscrimination discourse », *Information, Communication & Society*, vol. 22, n° 3, 2019, p. 900-915.
- JOBIN, Anna, IENCA, Marcello et VAYENA, Effy, « The Global Landscape of AI Ethics Guidelines », *Nature Machine Intelligence*, vol. 1, n° 9, 2019, p. 389-399.
- KENNEDY, Helen et ENGBRETSSEN, Martin, « Introduction : The relationships between graphs, charts, maps and meanings, feelings, engagements », dans ENGBRETSSEN, Martin et KENNEDY, Helen (dir.), *Data Visualization in Society*, Amsterdam, Amsterdam University Press, 2020, p. 19-32.
- LAPLANCHE-SERVIGNE, Soline et SA VILAS BOAS, Marie-Hélène, « Les catégorisations des publics minorisés en questions », *Participations*, vol. 25, n° 3, 2019, p. 5-31.
- LEHMANS, Anne, « Les réinventions de la démocratie à l'aune de l'ouverture des données : du discours de la participation aux contraintes de la gouvernance », *Les Enjeux de l'information et de la communication*, vol. 2, n° 19/2, 2018, p. 135-146.

- LEIGH STAR, Susan, « Ceci n'est pas un objet-frontière ! Réflexions sur l'origine d'un concept », *Revue d'anthropologie des connaissances*, vol. 4, n° 1, 2010, p. 18-35.
- MASSÉ, Hélène, LABERGE, Michèle et MASSÉ, Ginette, « L'analyse différenciée selon les sexes au gouvernement du Québec : vers une mobilisation interne et des alliances stratégiques pour l'égalité », *Lien social et politiques*, n° 47, 2002, p. 43-54.
- MILLERAND, Florence *et al.*, « À la recherche du citoyen "ordinaire", les publics imaginés de l'ouverture des données publiques au niveau municipal », *Les Enjeux de l'information et de la communication*, vol. 23, n° 4, 2023, p. 109-129.
- O'NEIL, Catherine, *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, New York, Penguin, 2016.
- PILLANT, Yves, « Le paradigme de l'inclusion ou la société des singularités pour horizon », *Pensée plurielle*, vol. 49, n° 1, 2019, p. 11-24.
- STATISTIQUE CANADA, *Normes du genre, la diversité et l'inclusion*, <https://www.statcan.gc.ca>.
- VILLE DE MONTRÉAL, *ADS+ 101*, <https://ocpm.qc.ca>.
- WALTER, Maggie *et al.*, « Indigenous Data Sovereignty in the era of Big Data and Open Data », *Australian Journal of Social Issues*, vol. 56, n° 2, 2021, p. 143-156.
- YOUNG, Iris Marion, *Justice and the Politics of Difference*, Princeton, Princeton University Press, 1990.
- ZACKLAD, Manuel et ROUVROY, Antoinette, « L'éthique située de l'IA et ses controverses », *Revue française des sciences de l'information et de la communication*, n° 25, 2022.
- ZASK, Joëlle, « Le public chez Dewey : une union sociale plurielle », *Tracés. Revue de sciences humaines*, n° 15, 2008, p. 169-189.

CHAPITRE 9

Les publics de données quotidiens

*Jean Burgess, Kath Albury, Anthony McCosker
et Rowan Wilken*

Dans une ère où les médias sociaux ont ouvert la voie à une convergence sans précédent de nos univers sociaux (Burgess, 2017) et à un « effondrement des contextes » (Marwick et Boyd, 2011), des pratiques et cultures de données de la vie quotidienne prennent forme non seulement dans la sphère individuelle privée, mais aussi dans les collectivités et la sphère publique¹. Tout comme auparavant, les publics sont multiples ; ils sont constitués tout autant par les médias et les données que par les pratiques de gouvernance formelles. Là réside la différence fondamentale entre l'idée contemporaine d'un public regroupé autour d'un enjeu ou d'un intérêt commun et la notion relativement statique de « sphère publique », laquelle n'est pas nécessairement inclusive, ou de l'idée de « public » pour désigner l'ensemble des membres d'une organisation à l'échelle d'un État – notion qu'affectionnent les diffuseurs nationaux et les administrations. À partir du moment où nous nous définissons et agissons collectivement sur la base de données, ou en rapport avec elles, on peut dire qu'on appartient à des *publics de données*.

Avant d'accorder à un agrégat le statut de « public », cependant, il faut examiner l'éventail des termes qui peuvent décrire les groupes d'individus

1. Texte traduit par Margot Lacroix avec la collaboration de Guillaume Latzko-Toth et de Florence Millerand, d'abord publié sous le titre « Everyday data public », dans *Everyday Data Cultures*, Polity Press, 2022. Ce chapitre est reproduit avec la permission de Polity Press. Les auteurs y ont proposé et développé la notion de « cultures de données de (ou dans) la vie quotidienne » (*everyday data cultures*). Selon les occurrences, le mot « *everyday* » (qui peut être un nom ou un adjectif) a été traduit par « quotidien », « vie quotidienne » et « vie de tous les jours ».

qui produisent des données ou s'adonnent au quotidien à des pratiques autour des données, depuis le comptage des pas jusqu'aux tentatives de manipulation d'un algorithme de TikTok. Pour les désigner, les chercheurs qui étudient l'Internet, dont les auteurs de ces lignes, emploient souvent le terme « usager » – soit d'une infrastructure comme le Wi-Fi, soit d'une application, d'une plateforme ou d'un dispositif particulier. N'étant associé à aucune idée de valeur ou de motivation, ce vocable a démontré son utilité dans les débats sur certaines pratiques numériques controversées comme l'usage d'applications de rencontres et d'aventures d'un soir. Par ailleurs, il ne connote pas de relation commerciale, car il n'évoque pas la dépendance implicite envers les marchés associée au terme « consommateur ». Mais tout en ayant son utilité, le terme « usager » ne rend pas très bien compte de la dimension relationnelle des cultures de données.

Pour leur part, les plateformes numériques utilisent fréquemment le mot « communauté » pour désigner des groupes d'utilisateurs. Dans leurs énoncés de politique, toutefois, ces groupes sont rarement décrits en fonction de leur agentivité ou de leur capacité d'action collective. Comme le fait observer Tarleton Gillespie, certaines plateformes emploient dans leurs directives (règles de participation) un vocabulaire libertaire, invoquant des principes comme la liberté de parole et d'expression, tout en prenant soin de préciser que la « communauté » ainsi formée n'a d'existence qu'au sein du contenant mis à disposition par la plateforme. Dans un tel contexte, les politiques communautaires précitées laissent entendre que les restrictions en matière de contenu ne seront appliquées qu'avec parcimonie et peu souvent (Gillespie, 2018, p. 48). Ces textes, relève le chercheur, présentent le plus souvent la plateforme comme un espace qui permet à « une communauté diversifiée, mais fragile » de se créer (p. 49). Lorsqu'une communauté n'a pas d'existence reconnue à l'extérieur de la plateforme qui lui a pourtant donné naissance, et qu'il faut la « protéger pour qu'elle survive » (p. 49), il y a de fortes chances pour que les « normes communautaires » qui lui seront imposées soient celles de la plateforme elle-même.

Cette conception de la communauté diffère quelque peu des définitions juridiques classiques des normes sociétales inscrites dans la législation et les politiques – notamment dans les lois sur l'obscénité, en vertu desquelles ce sont les « normes sociales » qui dictent s'il est permis ou non de faire circuler ou de vendre publiquement des textes particuliers. Dans ce genre de cas, tant les textes législatifs que la jurisprudence de *common law* renvoient habituellement à la figure abstraite de la « personne raisonnable » (sa race,

son sexe et sa classe étant néanmoins implicites), dont les valeurs et les points de vue sont considérés comme représentatifs de la population générale à l'extérieur de la salle d'audience (Moran, 2003).

Les champs d'études sur les médias et la culture s'intéressent depuis longtemps au processus par lequel des groupes en viennent à se considérer comme des acteurs ou des agents sur la base de leur expérience commune des technologies médias – en tant que public, audience ou autre chose. Comme l'explique Sonia Livingstone, en langage courant, un public se constitue autour « d'une interprétation commune du monde, d'une identité collective, d'une prétention à l'inclusivité, d'un consensus sur l'intérêt commun », auxquels s'ajoute « un forum collectif d'un genre ou d'un autre, ouvert, auquel la population participe » (2005, p. 9). La chercheuse rappelle que les recherches qualitatives menées par le Birmingham Centre for Contemporary Cultural Studies dans les années 1970 ont montré que les façons dont les audiences s'intéressent aux contenus populaires en matière de nouvelles et de divertissement correspondent de près aux définitions courantes et théoriques de *public* et de *caractère public* (Livingstone, 2019).

Par ailleurs, bien que les mots comme « public » et « citoyen » soient traditionnellement associés à l'État-nation, ils acquièrent un sens nouveau dans le contexte des plateformes mondialisées que sont Apple, Google, Microsoft, Amazon, Tencent et Facebook, lesquelles semblent être partout et nulle part à la fois. Ces termes recourent d'autres catégories, comme « témoin » ou « foule », porteuses de leur propre bagage en lien avec les hiérarchies historiques qui privilégient certaines formes d'action collective comme le dialogue civique, tout en rejetant d'autres formes qu'elles assimilent à du militantisme mou (*slacktivism*) ou à des campagnes de dénigrement. Déjà, dans les années 1920 et 1930, John Dewey cherchait une solution pragmatique à des questionnements du même genre sur l'association humaine et l'espace pluraliste que représente « le public ». « L'âge de la machine, écrit-il, a si considérablement déployé, multiplié, intensifié et compliqué la portée des conséquences indirectes [de l'association humaine] que le public qui en résulte ne parvient pas à s'identifier et à se discerner lui-même² » (2016, p. 157). En d'autres termes, les publics ne sont pas pré-existants ou stables, mais dynamiques et émergents – et l'on peut donc en dire tout autant des publics de données (McCosker et Graham, 2018).

2. Traduction tirée de Dewey (2010, p. 217).

Dans leur cartographie des travaux de recherche s'intéressant au rapport entre données et publics, Jannie Møller Hartley et ses collègues dégagent quatre axes principaux : premièrement, l'incidence des données et de la datafication sur la sphère publique (en tant que site où se façonnent les notions collectives de citoyenneté et d'agentivité civique) ; deuxièmement, un regard porté sur les problèmes (« *issues* »), qui s'intéresse à la formation de publics de données dans la vie quotidienne en réaction à des préoccupations ou à des événements précis ; troisièmement, l'analyse des publics en tant que sites d'interaction réseautés et façonnés par les potentialités offertes par les plateformes ; et quatrièmement, l'étude des « publics algorithmiques » engendrés par les systèmes de tri et de recommandation des plateformes comme la section « Vous connaissez peut-être » de Facebook (2021, p. 4-8).

On peut donc affirmer que l'Internet a provoqué un réexamen du concept de public et de la notion de caractère public. Soulignons entre autres le modèle conceptuel de « publics en réseau » proposé par Danah Boyd (2010), destiné à analyser le rôle des premiers sites de réseautage social dans la formation de collectifs ; soulignons aussi les descriptions de la formation et de la participation de publics autour de mots-clics, comme dans le cas des « publics affectifs » de Zizi Papacharissi (2015) ou des publics *ad hoc* de Bruns et Burgess (2015), que l'avènement de la recherche algorithmique et de la curation des fils d'information a transformés, du moins en partie, en « publics calculés ». Parmi les nombreux exemples de ce type de public, en lien avec la politique des cultures de données de la vie quotidienne, figure #datamustfall, lequel aura marqué, soulignent Moyo et Munoriyarwa (2021), « le début d'une action citoyenne concertée en Afrique du Sud pour protester contre une tarification des données comparée à de l'extorsion » (p. 375). Enfin, sur la question des dynamiques intensément réflexives et performatives des publics en ligne dans un contexte caractérisé par la culture algorithmique et la microcélébrité, mentionnons la notion de « publics réfractés » proposée par Crystal Abidin (2021) pour décrire la dispersion de l'activité sociale entre différentes plateformes.

On trouvera certainement des éléments de tous ces modèles conceptuels dans la notion de « publics de données ». Toutefois, il faudrait d'abord s'interroger sur la façon dont les données représentent les individus et les amènent à se considérer comme appartenant à différentes sortes de publics (ou exclus de ceux-ci). Ces représentations des gens en tant que publics constitués *par* les données sont reflétées également dans leurs rapports avec les institutions et les organisations qui fournissent et présentent ces données.

Durant les premiers mois de la pandémie de COVID-19, le rôle accordé aux données épidémiologiques, au fur et à mesure de leur publication, a fait l'objet d'un vaste débat et d'un fort courant de contestation à la fois dans les organes de presse traditionnels et dans les médias sociaux. Les mises à jour quotidiennes sur les nouveaux cas diagnostiqués, les hospitalisations et les décès étaient reproduites et diffusées dans les forums publics et privés tandis que les autorités de santé publique s'efforçaient d'expliquer leurs directives (du lavage des mains au port du masque en passant par le confinement), visualisations de données et terminologie statistique à l'appui. Sur des plateformes de blogage comme *Medium*, des fils Twitter et des messages ont commencé à apparaître qui mettaient en doute les prédictions officielles tout en proposant des modélisations épidémiologiques « alternatives » (Muir, 2020). Certains de ces modèles avaient été réalisés par des spécialistes en science des données, mais sans expérience en épidémiologie, alors que d'autres avaient été « bricolés » par des individus ne possédant que des notions de base en statistique (Muir, 2020). Les épidémiologistes et les mathématiciens professionnels n'ont pas manqué de condamner ces incursions assimilables à de « l'épidémiologie de salon », les qualifiant, au mieux, de malavisées, et au pire, d'opportunistes (Obungu, 2020). Même si le traitement amateur des données est devenu moins fréquent à mesure qu'évoluait la pandémie, les débats incessants sur la réponse optimale à privilégier face au virus ont montré le rôle central des données, et de leur remise en question, dans la définition générale de ce que constituent l'intérêt public et le « bien commun ».

Ces débats illustrent bien la tension entre ce que Kennedy et Moss (2015) appellent « les publics connus et les publics connaissants ». Il s'agit ici de distinguer entre le recours aux données pour définir, mesurer et décrire des populations et des publics, et la connaissance que peuvent acquérir des individus, au moyen des données, sur leur situation collective. Les publics de données quotidiens forment des cercles de pratique plus larges, loin des relations individuelles et intimes avec les données que nous avons abordées dans d'autres chapitres de l'ouvrage dont est tiré celui-ci (Burgess *et al.*, 2022) – même s'il existe d'importants recoupements. Épidémiologistes, politiciens et fonctionnaires s'appuient sur des données pour formuler des recommandations fondées sur leur connaissance de la démographie, de la géographie, des contacts entre les cas positifs connus et des données biomédicales sur la transmission virale. Autrement dit, ils tentent de définir des « publics de la COVID-19 ». De la même façon, les praticiens amateurs (*do-it-yourself*) de

la science des données et autres commentateurs publics tentent d'exercer leur agativité ou de mettre en valeur des formes d'expertise non élitistes en proposant leur propre interprétation des données – aussi partielle ou incomplète soit-elle. Cet exemple met bien en lumière le tiraillement entre une connaissance experte des données et une volonté de réfuter celle-ci sur la base de connaissances populaires fondées sur l'expérience quotidienne; il montre aussi à quel point ce genre de conflit peut s'amplifier lorsque des pratiques individuelles autour des données s'articulent à des publics.

Les rituels numériques

Même les pratiques les plus banales en lien avec les données peuvent nous relier à la vie publique, puisqu'elles relèvent de ce qu'on pourrait appréhender comme des rituels de communication ou de médiatisation. Dans les années 1980, James Carey rappelait à ses lecteurs que cette conception de la communication en tant que rituel était inscrite depuis le XIX^e siècle dans la culture (américaine), aux côtés de l'idée dominante de transmission de l'information, essentielle au paradigme des communications de masse (2009, p. 12). De la même façon, on peut concevoir les données non seulement comme de l'information à transmettre, à interpréter et à utiliser dans la prise de décision, mais aussi comme un matériau destiné aux rituels communicationnels et sociaux – ces pratiques habituelles et répétées par lesquelles s'enracinent, se maintiennent et sont contestées les structures et les significations sociales et culturelles.

Les rituels quotidiens concernant les données relèvent ainsi de l'ordinaire: des pratiques «routinisées» et répétées, qui structurent nos vies et maintiennent nos relations. Au début des années 2010, le fait de pouvoir «signaler sa présence» quelque part au moyen des services mobiles de géolocalisation³ a suscité un énorme engouement; une grande part de l'attrait résidait dans le fait que cette nouvelle fonctionnalité permettait d'exprimer des «attachements collectifs» et d'«attester de notre lien étroit avec les lieux auxquels nous sommes particulièrement attachés» (Schwartz, 2015, p. 96). Dans certains cas, des interactions quotidiennes hautement ritualisées, mais systématisées, renforçaient ces lieux-liens: par exemple, suivre un trajet précis à pied pour signaler qu'on se trouve devant un immeuble public favori ou programmer son application avant de quitter la maison afin d'être prêt

3. À l'instar de la fonction *check-in* de Foursquare (N.D.T.).

au moment où l'on traversera le pont Golden Gate en voiture (Wilken, 2019, p. 141).

Prenant appui sur les travaux de l'anthropologue Catherine Bell, Lee Humphreys présente une série d'exemples allant du blogue au journal intime pour montrer que les pratiques quotidiennes de documentation médiatique – qu'elle désigne par le terme de *media accounts* [comptes rendus médias] – ont le pouvoir de convertir des pratiques ordinaires en rituels médiatiques (Humphreys, 2018, p. 35). La chercheuse s'inspire également des recherches de Karin Becker (1995) sur les rituels médiatiques quotidiens pour montrer que la médiatisation peut directement transformer ces rituels à caractère privé en rituels à caractère public. Ce faisant, ces rituels établiraient un lien entre des expériences individuelles ou privées en apparence, comme celles qui ont lieu au sein d'une maisonnée, et la vie publique. Les dernières années nous ont fourni quelques exemples de rituels médiatiques convergents qui associent consommation quotidienne des médias, spectacle public et public engendré par mot-clic : ils vont du concours Eurovision de la chanson (Highfield, 2017), durant lequel des groupes de téléspectateurs en salle retransmettent des égoportraits via Instagram et Twitter à l'aide du mot-clic #eurovision, à l'effusion de chagrin exprimé dans les récits de vie entourant la mort de célébrités adulées comme David Bowie (Burgess, Mitchell et Münch, 2019).

L'événement annuel Spotify Wrapped traduit en « publics calculés » des rituels quotidiens courants. Ici, les usagers de la plateforme sont non seulement invités, à l'aide de boutons générant des messages automatiques formatés, à afficher leurs propres statistiques annuelles sous forme de tableau de bord, mais aussi à publier leur palmarès sur des plateformes populaires. L'événement et les visualisations de données qui l'accompagnent produisent bien davantage qu'un portrait annuel, interactif et attrayant des habitudes d'écoute de l'auditeur : la dynamique connective des mots-clics intégrés (#spotifywrapped et ses déclinaisons, comme #2020wrapped, utilisé en 2020) convertit ce profil personnel en rituel médiatique fabriqué à partir des données qu'il renferme.

À l'instar de la célèbre description donnée par Dayan et Katz (1994) des grands rassemblements publics que sont les inaugurations et les mariages royaux, qualifiés par ceux-ci de « grande fête des médias radiotélévisés », l'initiative Spotify Wrapped tente de recréer un événement de grande envergure dans l'environnement médiatique des plateformes. Ici, les consommateurs de musique sont explicitement invités – et vivement encouragés – à

s'associer à des publics de médias sociaux sur la base de l'expérience partagée, mais diversifiée, de leur engagement avec ces « égoportraits en données » (*data selfies*) que constituent en quelque sorte les tableaux de bord et les indicateurs, ce qui engendre ainsi au profit de la plateforme un second rayonnement médiatique auprès du grand public (Galant, 2020 ; Weiss, 2018). On retrouve là un parfait exemple du passage d'un public *ad hoc* – organique en apparence, émergent – généré par mot-clic à un public *calculé* (Bruns et Burgess, 2015), la plateforme incitant les usagers à profiter d'une potentialité temporaire permettant d'afficher publiquement leurs résultats individuels dans un format générique partagé par des millions d'autres personnes. L'usage collectif d'un mot-clic nous fait ainsi sauter de la sphère privée, dont relèvent nos données hautement personnalisées, à l'appartenance à un public de données, aussi fugace soit-elle.

Le caractère spectaculaire et synchrone de l'événement médiatique Spotify Wrapped favorise en outre un apprentissage social, lequel s'opère au moment où journalistes et utilisateurs s'efforcent de comprendre la recette algorithmique de production de tableaux personnalisés visualisables, et s'interrogent sur l'usage que pourrait faire la plateforme de toutes ces analyses de données – un apprentissage qu'encouragent d'ailleurs Spotify Spotify et ses experts et auquel ils participent activement. Ainsi, la revue *Springboard*, spécialisée en science des données à des fins de marketing, a souligné le lancement de Spotify Wrapped en publiant un article ayant pour but de « décortiquer son fonctionnement, les raisons de son efficacité et la façon dont Spotify mène le bal *en tissant au moyen des données un lien émotionnel avec ses consommateurs* » (Galant, 2020, c'est nous qui soulignons). L'article passe d'abord en revue la « solide » conception de la campagne sur le plan affectif, l'interactivité dynamique des tableaux et la simplicité des potentialités de partage automatisé sur les réseaux sociaux ; il aborde ensuite l'exploitation des données qui sous-tend non seulement l'événement en soi, mais aussi plus largement les systèmes de recommandation de Spotify, fournissant chemin faisant une explication des techniques de base en science des données, soit le filtrage collaboratif, le traitement automatisé du langage naturel et l'analyse des signaux audio au moyen des techniques d'IA.

Comme le montrent ces quelques exemples, les rituels médiatiques numériques constituent des moyens routiniers et théâtralisés d'expérimenter de nouvelles configurations sociales associées à la transformation numérique, de les mettre à l'épreuve et de s'en distancier. En ce sens, en prenant part à ces rituels numériques et en devenant membres de publics de données,

les gens participent du même coup à un effort collectif motivé par une volonté d'apprendre à cohabiter avec les données et les machines.

L'activisme de données

Les plateformes traitent les contenus déposés par leurs utilisateurs comme des données destinées à être quantifiées et évaluées par leurs systèmes automatisés. Ce faisant, ces contenus sont soupesés à la lumière des préoccupations qui façonnent les « normes » (tant juridiques que prescriptives) régissant ce qui est présenté sur une plateforme et à quel public. En dépit de la dispense accordée aux plateformes au motif qu'elles ne sont que des « hébergeurs » de contenu, les règles de communauté mises en place concernent des questions sociales qui touchent au traitement des races et des genres, des opinions politiques, de la mésinformation, de la désinformation et de la malinformation, ainsi que du bien-être personnel. Les médias sociaux sont en quelque sorte devenus des paratonnerres au cœur des débats publics portant sur des fractures sociales et des problèmes sociaux aux racines profondes. Pour parvenir à cerner et à dénoncer les répercussions sociales du traitement algorithmique des contenus dans une plateforme, il faut souvent un effort concerté de la part des utilisateurs.

En avril 2019, la plateforme Instagram a confirmé qu'elle allait déployer une politique ayant pour but de « rétrograder » les messages jugés comme « suggestifs », mais qui ne violaient pas explicitement ses lignes directrices (Constine, 2019). L'annonce a été faite à l'occasion d'un événement organisé par Facebook, avec cette déclaration de Will Ruben, membre du personnel d'Instagram : « Nous avons commencé à nous servir de l'apprentissage machine pour déterminer si un message particulier pouvait être recommandé à notre communauté » (Constine, 2019). Comme le rapporte Josh Constine du site *TechCrunch*, les créateurs abonnés à Instagram ont exprimé leur insatisfaction à l'égard du libellé vague de sa politique, qu'ils ont assimilé à une politique de censure et de suppression. Ces instagrameurs formaient un groupe diversifié. Certains étaient des travailleurs du sexe qui se servaient de leur égoportrait pour diriger de nouveaux publics vers des plateformes au contenu plus explicite comme OnlyFans. D'autres étaient des instructeurs de mise en forme et des artistes de scène qui publiaient des photos d'eux-mêmes habillés de vêtements de danse et d'entraînement. Certains artistes visuels et photographes dont les œuvres montraient des nus non explicites (souvent dans un but avoué de positivité corporelle) ont eux aussi connu des

baisses de fréquentation, tout comme d'ailleurs un éventail de comptes axés sur la naissance et le parentage, dont ceux de doulas et de conseillères en allaitement maternel.

Même si ces instagrameurs s'adressaient, dans certains cas, à des publics complètement différents, leur contenu a été soumis aux mêmes règles de modération, qui semblaient traiter toutes les images de nudité ou de semi-nudité comme des contenus à caractère « suggestif ». Ils se sont plaints d'être victimes de « bannissement furtif » (*shadow banning*) – leur contenu n'apparaissait plus dans les résultats de recherche et les suggestions (y compris la page « Explorer » de la plateforme) – et ce, malgré le fait qu'ils avaient pris soin de lire et de respecter les règles. Tout au long de l'année 2019, des membres de ce public émergent et leurs abonnés discutent des pratiques d'Instagram dans des blogues et des fils de commentaires. Après avoir constaté des variations du taux d'engagement à l'égard de quelques-unes de leurs publications, certains ont dit soupçonner qu'on avait dissimulé leur contenu à leur insu.

Les activistes qui se sont adressés directement à Instagram dans le but de savoir quel genre de pose ou de mot-clic pourrait être en jeu ont d'abord été informés que leur contenu n'avait pas, en fait, été invisibilisé (Are, 2019). Durant la même période (au milieu de 2019), des instagrameurs issus des communautés noires, queers et diversifiées ont exprimé une inquiétude grandissante à l'égard d'une politique apparente de modération automatique ayant pour effet de signaler de façon erronée que leur contenu était « explicite ». Par exemple, une application de publicité payante destinée à promouvoir la revue queer et féministe *Salty*, montrant une image d'un modèle noir trans vêtu de la tête aux pieds, avait été refusée par Instagram dans un message citant une politique concernant les « services d'escorte⁴ ». Durant la même période, des entreprises LGBTQ+ se sont vu refuser l'affichage de publicités au motif qu'il s'agissait de « contenus destinés aux adultes » (O'Hara, 2019). Aux États-Unis, des activistes du secteur du cybersexe regroupés sous la bannière Hacking//Hustling ont établi un lien entre l'augmentation des cas de bannissement furtif de leur compte et l'adoption des lois FOSTA-SESTA, une modification du *Communications Decency Act* qui a rendu les plateformes numériques juridiquement responsables de tout contenu pouvant s'assimiler à du « racolage » (Blunt et Wolf, 2020 ; Blunt *et al.*, 2020).

4. On peut consulter le rapport de Salty, *An investigation into algorithmic bias in content policing on Instagram*, à l'adresse : <https://saltyworld.net>.

Tout au long de 2019, des producteurs de contenus se sont concertés de plus en plus souvent pour faire circuler de l'information et « tester » les cas suspectés de bannissement furtif en lançant des recherches sur leurs comptes respectifs et en colligeant leurs résultats. Parallèlement, ceux-ci ont invité des abonnés à les assister en leur demandant de rechercher des publications et des mots-clics, et de vérifier si les contenus « problématiques » étaient visibles dans leurs fils. Une équipe d'instagrameurs et d'activistes, dont la sociologue et danseuse de *pole dance* Carolina Are (2020, 2021) a lancé la campagne *EveryBODYVisible*, conviant les fournisseurs de contenus et leurs alliés à publier des contenus protestataires, accompagnés de la déclaration :

Notre mouvement comprend des artistes, des photographes, des athlètes, des danseurs, des yogis, des comédiens, des effeuilleuses, des travailleurs et travailleuses du sexe, la communauté LGBTQIA, la communauté BBW, des féministes, des militants de la cause des personnes handicapées, des éducateurs spécialisés sur les questions de la positivité corporelle, du sexe et de l'accouchement, des entraîneurs de mise en forme et un grand nombre d'autres personnes⁵.

Le groupe militant a produit une série d'images de semi-nudité ainsi que des versions multiples du logo *EveryBODYVisible* que les alliés de la campagne pouvaient superposer à leurs publications Instagram. On les a en outre invités à publier des contenus à l'occasion de la Journée internationale de l'Internet, le 29 octobre 2019 (en « taguant » le personnel de Facebook et d'Instagram). Des images de la campagne ont été largement diffusées par des instagrameurs très en vue, dont la danseuse burlesque Dita Von Teese, certains ayant toutefois vu leurs propres publications soumises à la modération et retirées de la plateforme. Documentant la campagne sur son blogue, Carolina Are décrit les diverses stratégies déployées par *EveryBODYVisible* :

Depuis les manifestations solidaires de femmes ayant modifié de « femme » à « homme » le sexe indiqué dans leur profil, à la pratique de « taguer » des cadres haut placés chez Facebook et Instagram durant la campagne, de façon à ce que la section « identiquetée » de leur profil finisse par afficher une variété de postérieurs féminins, nous avons fait tant de bruit que le PDG d'Instagram, Adam Mosseri, a dû admettre dans un message que nos revendications étaient raisonnables. (Are, 2020, p. 742)

Des questions similaires concernant le dispositif de modération basé sur l'IA d'Instagram ont fait surface en lien avec le traitement de la taille

5. Pour connaître le travail du groupe militant à l'origine d'une série d'images et de versions du logo *EveryBODYVisible* : <https://everybodyvisible.com>.

corporelle, donnant lieu également à des réactions collectives (Richman, 2019). Au cours de l'année 2019, la fondatrice du groupe *The Visible Collective*, Jessica Richman, a répondu à quantité de messages dénonçant des cas de discrimination non intentionnelle envers des personnes de forte taille sur Instagram. L'hypothèse a circulé que les systèmes automatisés de la plateforme calculaient le pourcentage de chair exposée et identifiaient des parties du corps « d'apparence féminine » comme les seins ou les mamelons. Bien que Facebook, société mère d'Instagram, ait nié que ses systèmes effectuassent ce genre de calcul – et désavantageât de façon disproportionnée les femmes de grande taille –, tout indique que la politique de modération des « contenus pour adultes » contrecarrait la promotion de la positivité corporelle sur la plateforme. La mannequin de taille plus et artiste de scène Carina Shero (@unskinnyhero sur Instagram en ce moment) a rapporté avoir vu plusieurs de ses comptes repérés et bannis par Instagram et s'être ainsi retrouvée à devoir faire preuve d'ingéniosité pour contourner les systèmes de modération de la plateforme (Richman, 2019).

On peut voir ces réponses collectives comme des exemples de publics de données quotidiens en action. Le groupe *Adult Performers Actors Guild* (APAG) a notamment créé une page dans le but d'aider les personnes qui avaient l'impression « d'avoir été injustement bannies [...] à riposter à des tactiques inéquitables et discriminatoires⁶ ». En recueillant leurs expériences et les contenus rejetés, l'APAG entend établir un jeu de données dans le but de dénoncer et de corriger les pratiques de modération qui ciblent injustement ses membres. En 2020, Instagram a répondu aux demandes de révision de sa politique déposées par la mannequin de taille plus Nyome Nicholas-Williams, le photographe Alex Cameron et la militante Gina Martin ; Instagram a accepté de laisser les femmes présenter leur corps nu en serrant leurs seins contre leur poitrine ou en les tenant dans le creux de la main – une pose auparavant signalée comme pornographique.

On peut considérer les actions collectives autour des données, souvent observées dans les plateformes, comme des stratégies de contournement des méthodes de modération de différents types de « contenus limites ». Les plateformes les perçoivent comme des cas problématiques, mais qu'il ne leur appartient pas de bannir purement et simplement. Elles peuvent à tout moment, et sans consulter le public, restreindre la visibilité de mots-clics ou

6. Les politiques de modération des « contenus pour adultes » de Facebook sont accessibles à : <https://transparency.fb.com>.

de mots-clés signalant certains types de contenus jugés indésirables, en masquant ou en étouffant leur présence dans les moteurs de recherche et les fils. La suppression de contenus, avance Tarleton Gillespie, fait partie des tâches fondamentales et en quelque sorte courantes des plateformes et de leurs algorithmes de recommandation et de chronologie pour ordonner et organiser les contenus⁷. Ces pratiques sont à distinguer du bannissement furtif, car elles se concentrent sur les types de contenus partagés, plutôt que sur des comptes ou des profils désignés comme ayant transgressé des normes ou des règles. Il est difficile de prévoir ou de prédire les contenus qui seront minimisés ou promus par les modèles de tri – puisqu'ils évoluent au gré de la politique d'une plateforme, des considérations politiques et des événements.

À la question des contenus suggestifs s'ajoute celle des thèmes relatifs à la maladie mentale, qu'Instagram a tendance à traiter par défaut comme des contenus problématiques ou « limites » (McCosker et Gerrard, 2021). Le mot-clic *#depressed* (« déprimé ») et d'autres du même genre ont périodiquement été bannis ou supprimés des recherches. C'est aussi le cas des mots relatifs à la scarification et à d'autres formes d'automutilation, celui des contenus évoquant directement les troubles de l'alimentation et celui du mot-clic *#thinspo* (employé pour désigner des contenus « motivants » en matière d'amaigrissement, qui encouragent en soi une alimentation déséquilibrée). Dans ce genre de cas, les stratégies de modération de contenu constituent de louables tentatives de promotion du bien-être. Des travaux empiriques sur le sujet nous apprennent toutefois que si le mot-clic *#depressed* sert de point de contact pour un nombre considérable d'utilisateurs, très vraisemblablement de jeunes instagrameurs, on se rend aussi compte que ces derniers savent se montrer très créatifs dans la composition de messages pouvant à la fois signaler une dépression – quoiqu'anonymement – et jouer sur le sentiment de tristesse ou le subvertir à l'aide de mêmes tristes et « grinçants » (des blagues d'initiés populaires, pastichant des mêmes surutilisés et des tendances dépassées), qui visent à provoquer des « rires auxquels on peut s'identifier » (McCosker et Gerrard, 2021). En cherchant collectivement des moyens de contourner les marqueurs conduisant au retrait ou à la dissimulation de contenus sur la santé mentale, les cultures adaptatives caractéristiques de la fréquentation des plateformes inventent sans cesse de nouvelles façons de communiquer sur le sujet.

7. Le billet de Tarleton Gillespie sur la suppression de contenus figure à : <https://threadreaderapp.com/thread/1387170486988247042.html>.

Des pratiques collectives autour des données et des pratiques algorithmiques réactives peuvent également prendre forme par la logique mémétique et par les genres communicationnels. Les mêmes deviennent ce mécanisme dont on se sert pour contourner les déclencheurs de modération sur les plateformes, tout en cultivant des publics connaissants grâce à des modes de fabrication du sens multiples et culturellement riches. Les mêmes *#depressed* sont courants sur Instagram. Certains chercheurs considèrent ces mêmes et « l'humour noir » des contenus comme du « bruit » généré par l'usage des plateformes (Ging et Garvey, 2018). D'autres ont démontré que les mêmes permettent de tisser des liens autour des contenus en santé mentale, grâce à la manipulation collective des signaux qui déclenchent, ou contournent, la modération algorithmique (McCosker et Gerrard, 2021). Prenant appui sur la notion de « politique de données contentieuses » analysée par Beraldo et Milan (2019, p. 3), on peut affirmer que ces méthodes illustrent bien « les pratiques quotidiennes de résistance, de subversion et d'appropriation inventive mises en œuvre par des individus » face aux restrictions imposées par les plateformes.

Les dispositifs de mesure, d'automatisation et de curation des données de Twitter ont fait l'objet d'une surveillance similaire de la part du public. Après s'être rendu compte que le système visionique servant au recadrage des visages dans les gazouillis incorporait des biais racistes et sexistes manifestes, des usagers de la plateforme ont décidé de le prouver. Un grand nombre d'entre eux se sont livrés à des expériences publiques. Ils ont constaté, par exemple, que leur propre visage avait été recadré de manière à mettre en valeur les hommes blancs attrayants qui figuraient sur la même photo. C'est le doctorant Colin Malard qui a découvert le biais en question après s'être adressé à Twitter pour se plaindre que la fonction d'affichage d'un « arrière-plan virtuel » dans Zoom n'avait pas reconnu l'un de ses collègues, noir de peau, et qu'elle avait effacé son image dans une visioconférence (Hearn, 2020). Au moment d'afficher la capture d'écran sur Twitter, Malard a constaté que son collègue avait disparu une seconde fois. Sa découverte a suscité des échanges soutenus sur un sujet devenu tendanciel et a conduit à des expériences publiques subséquentes sur le système de recadrage automatique. Tony Arcieri, un usager de Twitter, a par exemple publié une image de Barack Obama et de Mitch McConnell avec la mention : « Je vais tenter une expérience affreuse... Qui des deux l'algorithme de Twitter choisira-t-il : Mitch McConnell ou Barack Obama ? » (Hearn, 2020).

Comme l'explique Sarah Williams dans *Data Action* (2020), le fait de participer à des actions portant sur les données, comme celles que nous venons d'aborder, permet aussi de former des publics ou de renforcer les communautés ou les groupes intéressés. En se servant des données pour défendre des intérêts communs, des publics de données émergent et évoluent rapidement ; leur mobilisation force parfois des changements dans les dispositifs techniques qui façonnent leur vie de tous les jours. Réagissant à la forte publicité négative engendrée par le biais raciste de son algorithme de recadrage, Twitter s'est excusé assez rapidement, admettant qu'il lui restait du travail à faire pour améliorer ses systèmes automatisés. L'entreprise a même proposé des mesures pour inciter le public à contribuer à cette tâche. Malgré tous les moyens dont Twitter disposait pour faire des tests et des analyses, ce fut la mise au jour par le public des inégalités découlant des données qui a pesé dans la balance.

L'intensité des réactions au triage des données visuelles par Twitter ainsi qu'au traitement des usagers sur Instagram démontre à quel point les inégalités relevées sont perçues comme des problèmes par les communautés touchées. Elles témoignent aussi des capacités affectives dont disposent les contre-publics pour riposter. Toutefois, vu l'évolution rapide des technologies et des outils voraces en données, il n'est pas toujours facile de contrer efficacement les politiques dommageables des plateformes. Pendant que nous apprenons comme société à vivre avec les données, les risques de voir émerger des applications malveillantes et perturbatrices ne cessent d'augmenter.

L'apprentissage social

Les plateformes de médias sociaux ont créé les conditions requises pour circonscrire de nouveaux modes d'apprentissage social et les appliquer à la création de publics et de contre-publics. Les données et leur manipulation en constituent le fondement et sont souvent à l'origine d'un combat collectif, comme les exemples précédents le démontrent. Ces réactions à la manière dont les plateformes façonnent l'action sociale et les expériences nous enseignent que l'apprentissage social et l'apprentissage computationnel sont en interaction constante. Pour Fourcade et Johns, cette relation dépasse la simple acquisition de compétences et de connaissances. L'apprentissage se prolonge « par le biais de l'imitation, des emprunts stylistiques, de la moquerie, de la fabrication de mèmes, de l'échantillonnage, de l'acculturation, de l'identification, de la modélisation, de la priorisation, de l'appréciation et de la

propagation et la mise en application de toutes sortes de pédagogies informelles» (Fourcade et Johns, 2020, p. 806). L'apprentissage social élargit la formation de publics et en engendre de nouveaux, tout en créant des conditions propices à de possibles manipulations – à la fois productives et malveillantes.

Ces manières de connaître et d'agir sont imbriquées dans les systèmes techniques favorisant l'échange ainsi que dans les métriques qui déterminent si une interaction prend un sens nouveau en tant que donnée. Pensons à la vitesse à laquelle des tendances émergent et disparaissent dans les médias sociaux. Les plateformes comme TikTok se fondent sur la proposition d'outils conçus pour propager des actions et des réactions sous forme mémétique en tant que tendances qui acquièrent une existence propre. Dans la même veine, des pédagogies informelles ont proliféré dans YouTube sous forme de guides pratiques et de tutoriels. En outre, on trouve sur Reddit et ailleurs des fils consacrés à l'échange d'information sur les façons de « brouiller une localisation ». Elles concernent habituellement l'usage de VPN, d'applications de brouillage des localisations et d'autres méthodes permettant de masquer le lieu où l'on se trouve au moment de se servir de son téléphone pour accéder à certaines applications. Ce savoir informel peut servir de différentes façons, certaines bienveillantes et espiègles (utiliser Pokémon GO pour « les attraper tous », par exemple) et d'autres, plus préoccupantes (comme dans l'usage d'applications de rencontre telles que Tinder). Ailleurs⁸, nous avons examiné ces pédagogies par les pairs sous l'angle des littératies en matière de données et d'algorithmes. Les compétences et les habiletés s'acquièrent par la connaissance des données, de la manière dont sont organisés contenus et interactions sous forme de métriques, ainsi que des opérations algorithmiques effectuées par les plateformes. Ce sont là des compétences et des connaissances personnelles et individuelles, mais qui peuvent s'avérer riches par rapport au contexte social et à la création collective de sens. De plus, en pratiquant les rituels de l'interaction numérique et de l'usage des données, nous apprenons collectivement à cohabiter avec les données et les machines.

Le concept d'apprentissage social est mobilisé depuis plusieurs décennies pour faire contrepoids aux théories cognitives axées sur la connaissance et l'individu; il met en lumière l'importance que revêtent l'observation, l'imitation et la modélisation (Bandura, 1969; Vygotsky, 1980). Associé aux développements en apprentissage machine et en conception algorithmique,

8. Voir Burgess *et al.* (2022).

le flux de données découlant de l'usage quotidien des technologies numériques signifie que les interactions entre apprentissage social et computationnel sont d'ores et déjà considérables et complexes. S'appuyant sur une tradition d'un demi-siècle en théorie de l'apprentissage, dont les travaux de Vygotsky, Foucade et Johns font observer que « les machines sont mal outillées pour composer avec le fait que tout apprentissage humain est culturel, c'est-à-dire qu'il n'est pas ancré dans les psychés individuelles, mais plutôt dans des systèmes de signification collectifs et des histoires relationnelles sédimentées » (2020, p. 807). De nos jours, l'apprentissage social prend plusieurs formes et se déroule sur un vaste éventail de sites et de plateformes (dont GitHub, YouTube, TikTok et Reddit), formels et informels, par le truchement de la recherche universitaire, mais aussi de « l'éducation populaire » (Pangrazio et Sefton-Green, 2020).

Pensons par exemple aux hypertrucages, ces documents vidéo et audio synthétiques, produits à partir de vastes jeux de données et de systèmes d'apprentissage profond, capables de synthétiser ou de créer des extraits entièrement nouveaux. Le terme *deepfake* [ou « hypertrucage » en français], s'est immiscé dans le discours populaire en 2017 dans la foulée des activités d'un utilisateur de Reddit appelé u/deepfakes, créateur également du sous-reddit r/deepfakes, banni par la suite. Grâce aux codes sources accessibles dans certains entrepôts de données – principalement dans la plateforme de partage de codes GitHub –, la création d'hypertrucages est devenue chose possible pour les personnes dotées de compétences modestes en codage et en gestion des données (Winter et Salter, 2020). Des heures et des heures de vidéos numériques en ligne offrent des données de base prêtes à être décomposées en dizaines de milliers d'images fixes, à être nettoyées et manipulées, pour ensuite servir à synthétiser un visage sur d'autres corps ou à modifier les expressions d'une personne, ainsi que son sexe, son âge ou tout autre marqueur de son identité personnelle. La circulation, dans Reddit et d'autres forums, de codes, de données, d'instructions et de conseils, et leurs premières utilisations en pornographie synthétique sur des plateformes comme Pornhub démontre que les hypertrucages étaient ancrés, dès le départ, dans le « macrocontexte de l'inégalité des sexes » (Öhman, 2020, p. 133 ; Winter et Salter, 2020).

Les hypertrucages alimentent des usages et des risques qui posent problème, mais aussi un assortiment de techniques qui leur font contrepoids. Ils constituent un outil idéal pour renforcer la littératie numérique (Vee, 2017), socialiser les systèmes d'IA et se former à la manipulabilité des données

visuelles. Apprendre à créer des hypertrucages, à extraire des données visuelles, à les manipuler et à les retravailler est devenu une façon d'apprendre collectivement *au moyen* des modèles d'apprentissage machine. En tant que produits des pédagogies informelles et des dépôts de codes et de données visuelles, les hypertrucages sont des artefacts culturels. Ils jouent très pesamment sur les références culturelles – comme les mimiques distinctives de l'acteur Tom Cruise, parodiées à répétition sur TikTok dans les créations du spécialiste belge des effets visuels Chris Ume, alias @deptomcruise (Vincent, 2021). Réalisés en collaboration avec Miles Fisher, personnificateur de Cruise, ces hypertrucages sont le fruit d'entreprises chronophages qui exigent un degré d'expertise élevé. Malgré tout, la quantité d'interfaces faciles à utiliser et de tutoriels YouTube sur le sujet a donné lieu à une activité substantielle dans les médias sociaux. Le genre a évolué dans de multiples directions, qui vont du divertissement à une volonté de nuire et de causer du tort notamment par la manipulation et le partage non consensuel d'images, en passant par des tentatives informatiques visant à détecter et à perturber les vidéos modifiées artificiellement. En d'autres termes, nous avons affaire à des pratiques liées aux données fondées sur des codes sociaux, eux-mêmes issus des cultures des plateformes, qui transforment en domaines sociaux et collaboratifs ces littératies de la vie de tous les jours que les individus élaborent sur l'IA et les données (McCosker et Wilken, 2020).

De la même façon, il est possible de voir, dans le travail collectif des acteurs malveillants et des réseaux de bots sur des plateformes comme Twitter et Reddit, une forme d'apprentissage collectif, s'opérant dans la manipulation des systèmes, des signaux et des données, destinée à déclencher des métriques et des tris algorithmiques et à y réagir. Les attaques coordonnées de bots servent à contrecarrer les mesures de modération ainsi qu'à imiter et à subvertir les normes organiques de distribution de l'information. Par leur activité coordonnée et perturbatrice, les réseaux de bots produisent des contenus et distribuent ou amplifient des informations controversées dans le but de déstabiliser des systèmes politiques, ou encore, de discréditer les messages de la santé publique (Broniatowski *et al.*, 2018; Graham *et al.*, 2020). On ne connaît pas toujours les motivations qui sous-tendent ces réseaux; leurs activités impliquent habituellement un amalgame d'acteurs malveillants – des personnes réelles – et de comptes automatiques créés pour ressembler à ceux de vrais utilisateurs.

Les réseaux coordonnés de ce genre ont le pouvoir de pervertir les formations publiques qui émergent autour de questions importantes et sen-

sibles. Leur capacité d'exploiter et de dénaturer les flux d'information excède celle des individus, de telle sorte que les métriques en viennent à les comptabiliser comme des sujets tendance. Leurs efforts coordonnés engendrent également un certain degré de visibilité pouvant attirer encore davantage d'attention et s'amplifier s'ils sont entretenus suffisamment longtemps. Enfin, ils suscitent en réponse à leurs provocations des réactions favorables ou défavorables, et fortement émotionnelles (McCosker et Graham, 2018 ; McCosker, 2014). On pourrait les considérer en soi comme des publics de données surhumains.

Certains comptes (bots) automatisés sur des plateformes comme Twitter sont conçus pour agir comme s'ils étaient des êtres humains : ils publient des contenus souvent construits artificiellement, publient des liens, aiment des messages et les font circuler, émettent des commentaires. En ce sens, ils activent et manipulent artificiellement tous les signaux utilisés par une plateforme pour mesurer le niveau d'activité et promouvoir des tendances et des informations à leur sujet. Exploitée à l'encontre du récit dominant, la conversion en métriques de l'activité d'une plateforme (accomplie par enchâssement, puis par mise en données de potentialités inventées à l'origine par ses utilisateurs – comme les mots-clics et les réponses @) appelle ces activités et les récompense (Burgess et Baym, 2020). L'examen des publics de données, par conséquent, doit tenir compte de l'influence exercée par ces réseaux d'acteurs coordonnés.

Avant même que survienne la pandémie de COVID-19, des réseaux coordonnés de bots et d'individus visaient déjà délibérément à perturber le débat public sur les bienfaits de la vaccination (Broniatowski *et al.*, 2018 ; Yuan, Schuchard et Crooks, 2019). Regroupés autour de croyances sur ses dangers ou autour du droit de refuser d'obéir aux consignes de santé publique, les antivaccinalistes recourent à des publics de données coordonnés pour s'opposer aux messages sur la vaccination, en utilisant les mêmes outils de piratage des tendances et métriques sur la visibilité, à la façon d'activistes cherchant à redresser les biais racistes ou le ciblage de groupes marginalisés. Dans un examen de 25 millions de gazouillis publiés sur une période de dix jours durant la pandémie, Graham et ses collègues ont dénombré 5 752 comptes dédiés à la propagation de mésinformation et de désinformation « dans des buts commerciaux ou politiques » (Graham *et al.*, 2020). Ils attribuent cette activité principalement à une volonté de promouvoir des partis politiques ou des gouvernements de droite. En créant des contre-publics nuisibles susceptibles de perturber les efforts déployés par la santé

publique pour protéger les populations contre ce coronavirus mortel, les réseaux de bots ont contribué dans le même temps à la naissance de groupes de scientifiques, de limiers amateurs et de pirates qui ont mis au point des dispositifs de détection afin de dénoncer les méfaits des bots et les contrecarrer.

Durant la pandémie, certaines campagnes antivaccinales se sont empêtrées dans une vague d'opposition aux technologies 5G (Bruns, Harrington et Hurcombe, 2020). Ce mouvement a pris plusieurs formes. On a vu apparaître des efforts coordonnés de mésinformation et de désinformation, notamment la diffusion d'une théorie du complot selon laquelle le rayonnement émis par le réseau 5G causerait directement la COVID-19 (Meese, Frith et Wilken, 2020). Des manifestations populaires se sont tenues pour empêcher ou ralentir la construction d'infrastructures 5G – au Royaume-Uni, elles ont pris la forme d'actions directes : en avril et mai 2020, pas moins de 77 stations cellulaires ont été incendiées (Asher Hamilton, 2020). Les opposants ont aussi dirigé des efforts coordonnés contre les décisionnaires et les gouvernements pour tenter de ralentir ou de stopper le déploiement de la 5G en invoquant principalement des motifs sanitaires (Meese, Frith et Wilken, 2020). Toutes ces initiatives ont eu des répercussions concrètes. En Australie, les autorités de santé publique, à l'échelon fédéral et à celui des États, ont été forcées de publier des déclarations pour lutter contre la mésinformation sur les causes de la COVID-19 (Meese, Frith et Wilken, 2020). En Europe, certains pays ont suspendu le déploiement du réseau 5G en réaction aux craintes sanitaires de la population ; la Suisse a même annoncé un « moratoire indéfini » sur l'usage de cette technologie (Jones, 2020).

Poussant plus avant l'idée d'une « politique de données contentieuse », Beraldo et Milan (2019) attirent notre attention sur des initiatives transformatrices venant de la base et capables d'entraver ou de détourner des processus dominants de mise en données. S'alignant sur la notion de publics émergents défendue par Dewey plutôt que sur celle d'un « public » préexistant, leurs travaux nous rappellent que la mise en œuvre de la politique des données peut être le fait d'individus ou de collectifs, mais que l'action collective n'est jamais acquise : il faut constamment y travailler. De la même manière, l'activisme façonne les données tout en étant lui-même façonné par les données et les plateformes. En d'autres termes, les données peuvent être l'objet d'une intervention et susciter un besoin d'action collective, ou elles peuvent fournir un moyen déterminant ou stratégique permettant de mener une action collective ou de défendre des droits. Dans le premier cas,

ce sont les données mêmes qui sont en jeu ; dans le second cas, elles constituent le médium et le répertoire des actions collectives (Beraldo et Milan, 2019, p. 6). Comme le soutiennent Flyverbom et Murray, d'une part, « les moyens largement invisibles et en apparence techniques de traitement des données ont une incidence sur la façon dont les individus, les organisations et les sociétés accèdent aux mondes et aux réalités, organisent leur présence et effectuent des interventions » (2018, p. 2) ; d'autre part, on observe un regain et un élargissement de l'intérêt porté à ces questions dès lors que les données sont le médium par lequel la vie quotidienne se déploie.

L'art public et la visualisation des données

Les données sont au cœur de certaines pratiques d'art public ayant pour but de sensibiliser la population au phénomène de la mise en données. Parallèlement, de nombreuses activités, comme le journalisme « numérique », recourent à la visualisation de données créative ou interactive non seulement pour produire des connaissances sur des publics, mais aussi pour cultiver des « publics *connaissants* » (Kennedy et Moss, 2015, c'est nous qui soulignons). Cependant, comme le démontrent les projets abordés plus loin, on constate au sein des cultures de données une tension persistante entre l'expertise formelle du spécialiste en sciences des données ou de l'artiste et l'expérience quotidienne du sujet des données.

Dans son ouvrage *Living in Data* (2021), Jer Thorp, artiste spécialisé en visualisation des données, explique l'évolution de sa démarche en près de vingt ans de travail auprès des collectivités. Tout comme d'autres intervenants du milieu communautaire, il soutient que l'analyse des données sociales devrait « viser moins à chercher des réponses qu'à trouver de l'agentivité ». À ses yeux, les données sont toujours intriquées avec le quotidien, influant sur la manière dont les gens vivent et travaillent. Thorp expose sa démarche de travail *avec* les communautés, plutôt que de se limiter à les représenter en recourant à l'analyse et à la visualisation des données. Il y voit l'occasion de « susciter un questionnement », en ce sens que sa méthode ne lui sert pas à schématiser le monde, mais bien à en révéler les complexités et à mettre au jour des phénomènes sociaux jusque-là invisibles et insoupçonnés (Thorp, 2021). Néanmoins, la visualisation des données s'accompagne toujours d'une prétention à l'objectivité et à une puissance explicative que les conventions en matière de communication des statistiques amènent l'observateur à tenir pour acquise (Kennedy *et al.*, 2016), même lorsqu'il s'agit

de politiques de données contentieuses et de modes d'analyse complexes ou opaques.

Les diagrammes, les graphiques, les cartes et les tableaux associés à la visualisation ont traditionnellement un rôle social, celui de rendre les données accessibles au « public ». Toutefois, les vastes jeux de données, l'éventail d'outils d'analyse et de visualisation existants et l'apparition de nouvelles professions comme le journalisme de données font en sorte que la production « de formats et d'usages inédits [a] un effet déstabilisant sur notre compréhension de leur fonctionnement comme phénomènes sémiotiques et esthétiques, et de leur incidence, positive ou négative, sur l'agentivité personnelle et sociale » (Kennedy et Engebretsen, 2020, p. 20). La visualisation des données prend des formes multiples, mais grâce aux outils computationnels et aux images vectorielles, elle est devenue précise, voire esthétique (Halpern, 2015). Des sites Web comme *Flowing Data*⁹ ou le sous-reddit *r/DataIsBeautiful* illustrent magnifiquement l'esthétisme des techniques contemporaines de visualisation de données, susceptibles de capter l'intérêt du public.

Dans les faits, la représentation visuelle des données publiques fait peut-être de plus en plus partie de l'expérience médiatique quotidienne de la plupart des gens – comme l'ont montré notamment les diagrammes et les tableaux représentant l'évolution des cas de COVID-19 et des taux de vaccination –, mais cela ne réduit pas pour autant la complexité des données massives, pas plus que cela ne règle le problème de leur mésutilisation ou de leur mésinterprétation par le public. Il existe encore une tension entre la lisibilité des visualisations de données et leur rôle de substituts des opérations mathématiques et statistiques, ou de calculs informatiques, qui dépassent l'interprétation « ordinaire ». Sous forme visuelle, les *big data* opèrent entre le sublime et le schématique – entre leur réduction au statut de symbole et leur matérialisation en quelque chose pouvant servir à formuler des questions ou à résoudre des problèmes (McCosker et Wilken, 2014). En effet, la « sensation des chiffres » est un aspect essentiel de la façon dont les publics interagissent avec les données (Kennedy et Hill, 2017) et de leur pouvoir d'action.

En inventant des méthodes similaires en collaboration avec des organisations de la société civile, des journalistes et des citoyens, les centres de recherche et les équipes de projet peuvent contribuer à combler le fossé entre l'opacité des données quotidiennes et les sujets que ces données concernent.

9. Pour en savoir davantage sur *Flowing Data*, consulter : <https://flowingdata.com>

Instigateurs du projet Art/Data/Health, Folotopoulou et ses collègues ont cherché à « bâtir un pont entre données, créativité et récits expérientiels, en rapprochant les communautés, les artistes et les organisations de la société civile dans le but d'améliorer le pouvoir d'action collective et l'appropriation des données sur la santé et le bien-être ». À l'Université de Sheffield, dans le cadre du projet *Living With Data*, Kennedy et ses collègues se fondent sur les « cheminements de données » élaborés par Jo Bates pour rendre visibles des pratiques particulières à différents moments dans le temps, repérer les données quotidiennes et étudier ce que les gens savent et pensent à leur sujet¹⁰.

Les rapports de pouvoir inhérents à la tension entre public « connu » et public « connaissant » ne sont pas toutefois une question simple. Par exemple, Dorothy Kidd (2019) s'est intéressée à l'activisme de données et à la souveraineté des données dans les pratiques de contre-cartographie des peuples autochtones au Canada. Rappelant que ce genre de militantisme s'inscrit toujours dans un écosystème de relations économiques, sociales et culturelles, et que la cartographie critique comme forme de justice des données doit être « fusionnée avec de plus vastes projets axés sur la justice redistributive, transformatrice et restauratrice », Kidd documente la façon dont les Inuits, les Dénés, les Wet'suwet'en et d'autres Premières Nations au Canada recourent « à des pratiques de cartographie conjuguées à des récits, à des échanges et à des prières pour réinscrire leurs mémoires historiques et revivre leurs relations réciproques avec leur territoire et leurs modes de vie » (2019, p. 966).

Il importe d'aborder les processus évolutifs par lesquels on se représente la connaissance de soi *en tant que* publics et on *agit en tant que* publics par l'entremise des données. Les pratiques liées aux données ne sont pas seulement personnelles et individuelles, mais aussi, fondamentalement, collectives et sociales. De plus en plus souvent, les données sont utilisées et présentées sous forme visuelle dans le but de décrire des groupes, des populations, des utilisateurs et des collectivités. Toutefois, elles permettent aussi de constituer des « publics connaissant » (Kennedy et Moss, 2015). Les exemples que nous avons présentés font état de réponses collectives aux données qui entourent les gens ainsi que de pratiques personnelles en matière de données qui relient ces mêmes gens à la vie publique et aux publics en ligne, pratiques qui jouent le rôle de rituels de communication médiatique importants, comme l'illustre le cas de Spotify Wrapped.

10. Le site Art/Data/Health loge à l'adresse : <https://www.artdatahealth.org>.

C'est dans l'effort collectif déployé par les usagers des plateformes qui s'unissent pour cerner et critiquer les répercussions sociales du traitement algorithmique des contenus dits « controversés » qu'il faut chercher les publics de données quotidiens. Les plateformes numériques créent des conditions d'observation et de développement de nouvelles formes d'apprentissage social, qui engendrent elles-mêmes par la suite de nouveaux publics et contre-publics. Malgré l'enthousiasme des débuts sur le rôle des publics en réseau en tant que moteurs de changement social, ces pratiques ont aussi propulsé des groupes aux intentions néfastes. Le regroupement des données d'activité par les plateformes et leur interprétation sous forme de tendances facilite l'action collective, qu'elle soit à bon ou à mauvais escient. La visualisation des données et l'art numérique public, en ce qu'ils participent à la constitution et au maintien de publics connus et connaissants, ont un rôle social. Il faut continuer à observer les cultures de données qui se créent sur les plateformes et se dispersent, et à mettre en évidence le pouvoir – positif et négatif, générateur et destructeur – que recèlent les publics de données quotidiens.

Références

- ABIDIN, Crystal, « From “networked publics” to “refracted publics”: A companion framework for researching “below the radar” studies », *Social Media + Society*, vol. 7, n° 1, 2021, p. 1-13.
- ARE, Carolina, « Instagram denies censorship of pole dancers and sex workers ». *Blogger on Pole*, <https://bloggeronpole.com/2019/07/instagram-apologises-to-pole-dancers-about-the-shadowban/>, 2019.
- , « How Instagram's algorithm is censoring women and vulnerable users but helping online abusers », *Feminist Media Studies*, vol. 20, n° 5, 2020, p. 741-744.
- , « The Shadowban Cycle: An autoethnography of pole dancing, nudity and censorship on Instagram », *Feminist Media Studies*, vol. 22, n° 8, 2021, p. 2002-2019.
- ASHER HAMILTON, Isobel, « 77 cell towers have been set on fire so far due to a weird coronavirus 5G conspiracy theory », *Business Insider Australia*, <https://www.businessinsider.com/77-phone-masts-fire-coronavirus-5g-conspiracy-theory-2020-5>, 2020.
- BANDURA, Albert, « Social-learning theory of identificatory processes » dans GOSLIN, D.A. (dir.), *Handbook of Socialization Theory and Research*, Chicago, Rand McNally & Company, 1969, p. 213-262.
- BECKER, Karin, « Media and the ritual process », *Media, Culture & Society*, vol. 17, n° 4, 1995, p. 629-646.
- BERALDO, Davide et MILAN, Stefania, « From data politics to the contentious politics of data », *Big Data & Society*, vol. 6, n° 2, 2019.

- BLUNT, Danielle et WOLF, Ariel, «Erased: The Impact of FOSTA-SESTA and the Removal of Backpage», *Anti-Trafficking Review*, n° 14, 2020, p. 117-121.
- BLUNT, Danielle *et al.*, «Posting into the Void», *Hacking//Hustling*, hackinghustling.org, 2020.
- BOYD, Danah, «Social network sites as networked publics: Affordances, dynamics, and implications», dans PAPACHARISSI, Zizi, (dir.), *A Networked Self: Identity, Community, and Culture on Social Network Sites*, New York, Routledge, 2010, p. 39-58.
- BRONIATOWSKI, David A. *et al.*, «Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate», *American Journal of Public Health*, vol. 108, n° 10, 2018, p. 1378-1384.
- BRUNS, Axel et BURGESS, Jean, «Twitter hashtags from ad hoc to calculated publics», dans N. Rambukkana (dir.), *Hashtag Publics: The Power and Politics of Discursive Networks*, New York, Peter Lang, 2015, p. 13-27.
- BRUNS, Axel, HARRINGTON, Stephen et HURCOMBE, Edward, «“Corona? 5G? Or both?”: the dynamics of COVID-19/5G conspiracy theories on Facebook», *Media International Australia*, vol. 177, n° 1, 2020, p. 12-29.
- BURGESS, Jean, «Convergence», dans OUELETTE, Laurie et GRAY, Jonathan (dir.), *Keywords in Media Studies*, New York, New York University Press, 2017, p. 47.
- BURGESS, Jean et BAYM, Nancy, *Twitter: A Biography*, New York, New York University Press, 2020.
- BURGESS, Jean, MITCHELL, Peta et MÜNCH, Felix, «Social media rituals: The uses of celebrity death in digital culture», dans Papacharissi, Zizi (dir.), *A Networked Self and Birth, Life, Death*, New York, Routledge, 2019, p. 224-239.
- CAREY, James W., *Communication as Culture: Essays on Media and Society* (éd. révisée), New York, Routledge, 2009.
- CONSTINE, Josh, «Instagram now demotes vaguely “inappropriate” content», *Tech Crunch*, <https://techcrunch.com/2019/04/10/instagram-borderline>, 2019.
- DAYAN, Daniel et ELIHU, Katz, *Media Events: The Live Broadcasting of History*, Cambridge, Harvard University Press, 1994.
- DEWEY, John, *Le Public et ses problèmes*, traduit par Joëlle Zask, Paris, Gallimard, 2010.
- DEWEY, John, *The Public and its Problems: An Essay in Political Inquiry*, Athens, Swallow Press, 2016.
- FLYVERBOM, Mikkel et MURRAY, John, «Datastructuring – Organizing and curating digital traces into action», *Big Data & Society*, vol. 5, n° 2, 2018.
- FOURCADE, Marion et JOHNS, Fleur, «Loops, ladders and links: The recursivity of social and machine learning», *Theory and Society*, vol. 49, n° 5, 2020, p. 803-832.
- GALANT, Simona, «Data spotlight: How Spotify Wrapped makes music data feel personable», *Springboard Blog*, 2020.
- GILLESPIE, Tarleton, *Custodians of the Internet*, New Haven, Yale University Press, 2018.

- GING, Debbie et GARVEY, Sarah, « “Written in these scars are the stories I can’t explain” : A content analysis of pro-ana and thinspiration image sharing on Instagram », *New Media & Society*, vol. 20, n° 3, 2018, p. 1181-1200.
- GRAHAM, Timothy *et al.*, *Like a virus: The coordinated spread of coronavirus disinformation*, The Australia Institute, Canberra, A.C.T., 2020.
- HALPERN, Orit, *Beautiful Data: A History of Vision and Reason since 1945*, Durham, Duke University Press, 2015.
- HEARN, Alex, « Twitter apologises for “racist” image-cropping algorithm », *The Guardian*, <https://www.theguardian.com/technology/2020/sep/21/twitter-apologises-for-racist-image-cropping-algorithm>, 2020.
- HIGHFIELD, Tim, *Social Media and Everyday Politics*, Cambridge, Polity Press, 2017.
- HUMPHREYS, Lee, *The Qualified Self: Social Media and the Accounting of Everyday Life*, Cambridge, The MIT Press, 2018.
- JONES, Sam, « Switzerland halts rollout of 5G over health concerns », *Financial Times*, <https://www.ft.com/content/848c5b44-4d7a-11ea-95a0-43d18ec715f5>, 2020.
- KENNEDY, Helen et ENGBRETSSEN, Martin, « Introduction: The relationships between graphs, charts, maps and meanings, feelings, engagements », dans ENGBRETSSEN, Martin et KENNEDY, Hélène (dir.), *Data Visualisation in Society*, Amsterdam, Amsterdam University Press, 2020, p. 19-32.
- KENNEDY, Helen et MOSS, Giles, « Known or knowing publics? Social media data mining and the question of public agency », *Big Data & Society*, vol. 2, n° 2, 2015.
- KENNEDY, Helen *et al.*, « The work that visualisation conventions do », *Information, Communication & Society*, vol. 19, n° 6, 2016, p. 715-735.
- KENNEDY, Helen et HILL, Rosemary L., « The pleasure and pain of visualizing data in times of data power », *Television & New Media*, vol. 18, n° 8, 2017, p. 769-782.
- KIDD, Dorothy, « Extra-activism: Counter-mapping and data justice », *Information, Communication & Society*, vol. 22, n° 7, 2019, p. 954-970.
- LIVINGSTONE, Sonia, « On the relation between audiences and publics », dans LIVINGSTONE, Sonia (dir.), *Audiences and Publics: When Cultural Engagement Matters for the Public Sphere*, Bristol, Intellect Books, 2005, p. 17-41.
- , « Audiences in an age of datafication: Critical questions for media research », *Television & New Media*, vol. 20, n° 2, 2019, p. 170-183.
- MARWICK, Alice et BOYD, Danah, « I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience », *New Media & Society*, vol. 13, n° 1, 2011, p. 114-133.
- MCCOSKER, Anthony, « Trolling as provocation: YouTube’s agonistic publics », *Convergence*, vol. 20, n° 2, 2014, p. 201-217.
- MCCOSKER, Anthony et GRAHAM, Timothy, « Data publics: Urban protest, analytics and the courts », *M/C Journal*, vol. 21, n° 3, 2018.
- MCCOSKER, Anthony et GERRARD, Ysabel, « Hashtagging depression on Instagram: Towards a more inclusive mental health research methodology », *New Media & Society*, vol. 23, n° 7, 2021, p. 1899-1919.

- MCCOSKER, Anthony et WILKEN, Rowan, « Rethinking 'big data' as visual knowledge: The sublime and the diagrammatic in data visualization », *Visual Studies*, vol. 29, n° 2, 2014, p. 155-164.
- , *Automating Vision: The Social Impact of the New Camera Consciousness*, New York, Routledge, 2020.
- MEESE, James, FRITH, Jordan et WILKEN, Rowan, « COVID-19, 5G conspiracies and infrastructural futures », *Media International Australia*, vol. 177, n° 1, 2020, p. 30-46.
- MØLLER HARTLEY, Jannie *et al.*, « Researching publics in datafied societies: Insights from four approaches to the concept of "publics" and a (hybrid) research agenda », *New Media & Society*, vol. 25, n° 7, 2021, p. 1668-1686.
- MORAN, Mayo, *Rethinking the Reasonable Person: An Egalitarian Reconstruction of an Objective Standard*, Oxford, Oxford University Press, 2003.
- MOYO, Dumisani et MUNORIYARWA, Allen, « Data must fall: Mobile data pricing, regulatory paralysis and citizen action in South Africa », *Information, Communication & Society*, vol. 24, n° 3, 2021, p. 365-380.
- MUIR, Becca, « "Armchair epidemiologists" and "data bros": Inside the DIY world of Covid-19 research », *Prospect*, <https://www.prospectmagazine.co.uk/ideas/technology/40329/armchair-epidemiologists-and-data-bros-inside-the-diy-world-of-covid-19-research>, 2020.
- OBUNGU, C. Brandon, « Don't be fooled by Covid-19 carpetbaggers », *Wired*, <https://www.wired.com/story/opinion-dont-be-fooled-by-covid-19-carpetbaggers/>, 2020.
- O'HARA, Mary Emily, « Queer and feminist brands say they are being blocked from running ads on Instagram and Facebook », MTV News, www.mtvnews.com, 2019.
- ÖHMAN, Carl, « Introducing the pervert's dilemma: A contribution to the critique of Deepfake Pornography », *Ethics and Information Technology*, vol. 22, n° 2, 2020, p. 133-140.
- PANGRAZIO, Luci, et SEFTON-GREEN, Julian. « The social utility of "data literacy" ». *Learning, Media and Technology* 45, n° 2 (2 avril 2020) : 208-220.
- PAPACHARISSI, Zizi, *Affective Publics: Sentiment, Technology, and Politics*, Oxford, Oxford University Press, 2015.
- RICHMAN, Jessica, « This is the impact of Instagram's accidental fat-phobic algorithm », *Fast Company*, <https://www.fastcompany.com/90415917/this-is-the-impact-of-instagram-accidental-fat-phobic-algorithm>, 2019.
- SCHWARTZ, Raz, « Online place attachment: Exploring technological ties to physical places », dans DE SOUZA E SILVA, Adriana, et SELLER, Mimi, (dir.), *Mobility and Locative Media: Mobile Communication in Hybrid Spaces*, New York, Routledge, 2015, p. 85-100.
- THORP, Jer, *Living in Data: A Citizen's Guide to a Better Information Future*, New York, MCD/Picador, 2021.
- VEE, Annette, *Coding Literacy: How Computer Programming is Changing Writing*, Cambridge, The MIT Press, 2017.
- VINCENT, James, « Tom Cruise deepfake creator says public shouldn't be worried about "one-click fakes" », The Verge, <https://www.theverge.com/2021/3/5/22314980/tom-cruise-deepfake-tiktok-videos-ai-impersonator-chris-ume-miles-fisher>, 2021.

- VYGOTSKY, Lev S., *Mind in Society: The Development of Higher Psychological Processes*, Cambridge, Harvard University Press, 1980.
- WEISS, Haley, «Why people love Spotify's annual wrap-ups», *The Atlantic*, <https://www.theatlantic.com/technology/archive/2018/12/spotify-wrapped-and-data-collection/577930/>, 2018.
- WILKEN, Rowan, *Cultural Economies of Locative Media*, Oxford, Oxford University Press, 2019.
- WILLIAMS, Sarah, *Data Action: Using Data for Public Good*, Cambridge, The MIT Press, 2020.
- WINTER, Rachel et SALTER, Anastasia, «DeepFakes: Uncovering hardcore open source on GitHub», *Porn Studies*, vol. 7, n° 4, 2020, p. 382-397.
- YUAN, Xiaoyi, SCHUCHARD, Ross J. et CROOKS, Andrew T., «Examining emergent communities and social bots within the polarized online vaccination debate in Twitter», *Social Media + Society*, vol. 5, n° 3, 2019.

POSTFACE

Les données et leurs mondes

Jérôme Denis

« *Nothing comes without its world* », a écrit Donna Haraway en 1997, condensant par une de ces formules simples et vertigineuses dont elle a le secret l'un des fondements théoriques et politiques des études des sciences et des techniques : l'existence d'une entité, quelle qu'elle soit, tient toujours à l'existence de nombreuses autres. Aucune chose ne peut « arriver » sans un monde qui en assure les conditions de félicité. La tournure de la phrase, toutefois, invite à aller un peu plus loin et assumer le caractère transitif de la relation. Reconnaître aussi que chaque chose fait advenir le monde qui, dans le même mouvement, la fait tenir. On retrouve dans ce geste l'héritage de la philosophie de Gilles Deleuze, dont Karen Barad (2003) a rappelé les implications « relationnalistes ». Il est aussi très proche de la sémiotique matérielle que Madeleine Akrich (1987) a cristallisé dans la notion de « script ». La *de-scriptio* des objets techniques est en effet une opération de cartographie du monde que ces objets contiennent : des usagers, leurs compétences, leurs partenaires, leur environnement (Akrich, 1992). Un monde avec et par lequel ils existent, mais qu'ils participent aussi, avec un succès plus ou moins grand, à actualiser. C'est l'apport précieux des enquêtes d'Akrich, condensé dans l'expression d'Haraway : il n'y a rien de mécanique dans cette symétrie, bien au contraire. L'aventure de l'existence des mondes et des choses peut échouer, ou mal tourner. Akrich et Haraway rappellent, chacune à sa manière, qu'en cherchant à décrire des ontologies relationnelles, elles en soulignent les dimensions politiques. Elles invitent à mettre en lumière les interdépendances vitales aussi bien que les formes d'exclusion qui s'instaurent dans la relation entre les choses et leurs mondes, entre les mondes et leurs choses.

C'est à une exploration semblable que se sont attelés les autrices et les auteurs assemblés dans le présent livre, en se penchant sur ces objets particuliers que sont les données, et en interrogeant une partie du monde qui va avec elles : leurs publics. Des infrastructures matérielles qui en assurent le stockage, le calcul et la circulation (Blanchette, 2011) aux entités qu'elles font exister en les nommant et en les comptant (Bowker et Star, 1999), en passant par les travailleuses et les travailleurs qui les produisent, les corrigent, les formatent ou les entretiennent (Denis, 2018), le monde qui « va avec » les données est densément peuplé. Le public y tient toutefois une place à part. Ou plutôt des places à part. Imaginé, mesuré, recomposé, anticipé, impensé... il est à la fois omniprésent et difficile à appréhender. C'est précisément cette ubiquité et l'absence de contours nets, que permet d'assumer le terme même de « public », qui placent l'analyse au milieu du gué, *in media res*, sans rien avoir à présupposer à l'avance de la forme et des positions que prend, selon les configurations, le collectif assemblé par, avec ou pour les données. On a pu ainsi croiser au fil des chapitres qui composent ce volume des usagers, des auditoires, des citoyens, des patients et même des absents qui, tous, mettent en lumière à leur façon l'épaisseur sociopolitique de données dont on pourrait croire, au premier abord, qu'elles se réduisent à des artefacts informationnels « techniques », qu'elles circulent dans les entreprises, les administrations ou les laboratoires scientifiques. Aucune donnée n'existe sans public, peut-on ainsi affirmer en refermant ce livre, dans les traces d'Haraway et d'Akrich, et l'enquête sur ces publics qui existent et font exister les données est une formidable porte d'entrée pour analyser ces dernières comme des *choses politiques*.

Pour suivre la piste de ces publics relationnels et multiformes, la plupart des chapitres de ce livre s'inscrivent dans la lignée des travaux de John Dewey (1938, 1939), dont la philosophie pragmatiste a profondément marqué la sociologie, les sciences politiques et les sciences de l'information et de la communication. Ils en retiennent deux principaux enseignements, qui nourrissent leurs propres questionnements. Le premier repose sur l'idée de « *concernement* ». Celle-ci est essentielle pour penser les données au-delà de leur capacité à instituer des réalités objectives et à établir des « *matters of facts* ». L'entrée par les publics permet en effet de comprendre comment, et à quelles conditions, les données peuvent faire émerger des « *matters of concern* » (Latour, 2004), c'est-à-dire des problèmes (Marres, 2005), ou plutôt des problématisations, et autour d'elles, avec elles, des publics qui s'agrègent en enquêtant. Comme l'explique Evelyn Ruppert à propos des politiques de

transparence gouvernementale, et reprenant les analyses de Steven Shapin (1985), ce qu'elle nomme les « data publics » sont « incités à faire leurs propres analyses par des moyens matériels, sociaux et littéraires (essentiellement visuels) » (Ruppert, 2013, p. 53). Se joue ici une question politique essentielle qui nourrit l'enquête sur les données et les processus de *datafication* en l'invitant à rester attentive, pour chaque situation étudiée, à la diversité des formes de problématisation que les données permettent de cultiver, et à celles qu'elles entravent.

Ce livre nous rappelle aussi, avec Dewey, que les publics, fussent-ils concernés, ne sont jamais des entités stabilisées dont la présence serait préalable à la situation d'enquête ou, en l'occurrence, à la récolte ou la diffusion de données. C'est en tant qu'ils sont toujours en émergence qu'il est important de les étudier. Non pas au titre de structures qui permettraient d'en décrire des caractéristiques arrêtées, mais à l'aune des processus qui nourrissent leur existence. L'attention à cette dimension émergente des publics a une portée politique qui est peut-être un peu moins évidente que celle consacrée aux dynamiques de problématisation. Elle ouvre pourtant à des questionnements dont nombre des chapitres de ce livre illustrent l'importance.

On peut en retenir deux. Le premier est un classique de l'analyse critique des données dès lors qu'elle se préoccupe de leurs publics au sens large et renvoie au vocabulaire diplomatique des « parties prenantes » : qui, exactement, est représenté par tel ou tel jeu de données ? Sous quelle forme ? Comment les données peuvent-elles participer à rendre visibles certains collectifs, à les faire entrer à la « table des négociations » (Stengers, 2018) ? Le second porte sur les formes d'action que les données peuvent accompagner. Il n'est pas tant question de visibilité ici que d'équipement et d'encapacitation. Qu'est-ce que les données aident à faire ? Quelles prises sur le réel offrent-elles dans la formation des publics ? Comment permettent-elles de mobiliser les publics en question, ou les instances qu'elles interpellent, parfois malgré des qualités techniques et scientifiques insuffisantes (Gabrys, Pritchard et Barratt, 2016) ? Ce sont plus généralement les possibilités qu'offrent les données en matière d'enquête qui sont au centre de ces analyses, voire leur capacité à outiller la réflexivité des publics « récurifs » qu'évoque Chris Kelty (2008), ces publics qui se constituent dans un souci constant d'assurer les conditions techniques et infrastructurelles de leur propre existence.

Si le caractère émergent des publics et plus encore celui des deux questionnements qu'il nourrit sont politiques, c'est aussi au regard de leur versant

néгатif. En miroir de chaque interrogation se dessine en effet un risque. Celui de l'invisibilité, d'abord, et des publics réduits au silence par des jeux de données dont la structuration en catégories ou les modalités de calcul qu'ils intègrent produisent des absences. C'est une des leçons essentielles des études des infrastructures informationnelles (Bowker et Star, 1999 ; Star, 1991) et de la sociologie de la quantification (Espeland, 1993 ; Porter, 1996) : dans le même mouvement qu'elles rendent visibles des aspects de la réalité qu'elles font importer, les données, qu'elles soient scientifiques ou administratives, produisent des invisibilités d'autant plus difficiles à appréhender qu'elles les « naturalisent ». Quant au risque en miroir des possibilités d'encapacitation, il est celui de voir des publics « empêchés » par des jeux de données qui ne donnent aucune prise à certaines parties de la population, par leur technicité, leur opacité ou encore l'absence de pertinence des problèmes dont ils sont porteurs.

Ce sont à ces points critiques que les différents chapitres de ce livre contribuent à nous rendre sensibles. Et la plus grande qualité de l'ouvrage est sans aucun doute sa capacité à embrasser les ambivalences de la *datafication* et l'ontologie relationnelle des données et de leurs publics. Sans jamais tomber ni dans l'enthousiasme béat ni dans la dénonciation systématique, les autrices et les auteurs y cultivent une forme d'espoir : celui de voir dans les données un instrument de multiplication et de diversification de publics en devenir. De même qu'en science, les données ne sont pas condamnées à rigidifier et à simplifier les cadres d'appréhension de la réalité, par exemple de la biodiversité (Bowker, 2000), ce livre montre dans la variété même des situations qu'il explore que les données peuvent se défaire des pièges de l'universalisme hégémonique (Star, 1991 ; Blaser et de la Cadena, 2017) et devenir un bon terreau pour la multiplicité. C'est à l'aune de cet espoir qu'elles méritent toute l'attention des disciplines représentées ici. La critique politique des publics qui « vont avec » les données, et plus généralement des mondes qui les soutiennent et qu'elles font advenir, a tout à gagner à se nourrir d'analyses précises et empiriquement riches qui reconnaissent, et accompagnent, la part générative de la *datafication*. Car aucune enquête, aucun concept n'est jamais innocent (Haraway, 1988). Et s'il faut dénoncer les coups de force, les effets d'exclusion et les appauvrissements du réel, il est aussi vital de distinguer les voies qui dégagent l'horizon, celles par qui les données et leurs publics parviennent à faire œuvre d'irréduction (Latour, 1984).

Références

- AKRICH, Madeleine, « Comment décrire les objets techniques ? », *Techniques et culture*, n° 9, 1987, 49-64.
- , « The De-description of Technical Objects », dans *Shaping Technology/Building Society. Studies in Sociotechnical Change*, Cambridge, The MIT Press, 1992, p. 205-224.
- BENKLER, Yochai, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, New Haven, London, Yale University Press, 2006.
- BLASER, Mario, et DE LA CADENA, Marisol, « The Uncommons: An Introduction ». *Anthropologica*, vol. 59, n° 2, 2017, p. 185-193.
- BOWKER, Geoffrey C. et LEIGH STAR, Susan, *Sorting Things Out: Classification and Its Consequences*, Cambridge, The MIT Press, 1999.
- DEWEY, John, *Logic: The Theory of Inquiry*, New York, Henry Holt and Company, 1938.
- , *Theory of Valuation*, Chicago, University of Chicago Press, 1939.
- GABRYS, Jennifer, PRITCHARD, Helen et BARRATT, Benjamin, « Just Good Enough Data: Figuring Data Citizenships through Air Pollution Sensing and Data Stories », *Big Data & Society*, vol. 3, n° 2, 2016.
- HARAWAY, Donna, « Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective », *Feminist Studies*, vol. 14, n° 3, 1988, p. 575-599.
- , *Modest_Witness@Second_Millennium. FemaleMan_Meets_OncoMouse: Feminism and Technoscience*, 2^e éd., New York, Routledge, 2018.
- KELTY, Christopher M., *Two Bits: The Cultural Significance of Free Software*, New York, Duke University Press, 2008.
- LATOUR, Bruno, *Les microbes: guerre et paix. Suivi de Irréductions*, Paris, A.M. Métailié, 1984.
- , « Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern », *Critical Inquiry*, vol. 30, n° 2, 2004, p. 225-248.
- MARRES, Noortje, « Issues spark a public into being: A key but often forgotten point of the Lippmann-Dewey debate », dans LATOUR, Bruno, et WEIBEL, Peter (dir.), *Making Things Public: Atmospheres of Democracy*, Cambridge, The MIT Press, 2005, p. 208-217.
- PORTER, Theodore M., *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*, Princeton, Princeton University Press, 1996.
- RUPPERT, Evelyn, « Doing the transparent state: Open government data as performance indicators », dans ROTTENBURG, Richard, MERRY, Sally E., PARK, Sung-Joon, et MUGLER, Johanna (dir.), *The World of Indicators: The Making of Governmental Knowledge through Quantification*, Cambridge, Cambridge University Press, coll. « Cambridge Studies in Law and Society », 2015, p. 127150.
- SHAPIN, Steven, « Une pompe de circonstance. La technologie littéraire de Boyle ». *Culture technique*, n° 14, 1985, p. 70-87.
- STENGERS, Isabelle, « The Challenge of Ontological Politics », dans *A World of Many Worlds*, New York, Duke University Press, 2018, p. 83-111.
- , *Cosmopolitiques*, Paris, La Découverte, coll. « Les empêcheurs de penser en rond », 2022.

Notices biographiques

Kath Albury est professeure au Department of Media and Communication à la Swinburne University of Technology, Australie. Elle est membre du Centre of Excellence for Automated Decision-Making and Society (ADM+S) du Conseil australien de la recherche.

Dominique Boullier est professeur des universités émérite à Sciences Po, Paris, et chercheur au Centre d'études européennes et de politique comparée.

Jean Burgess est professeure à la School of Communication et chercheuse au Digital Media Research Centre (DMRC) à la Queensland University of Technology, Australie. Elle est directrice adjointe du Centre of Excellence for Automated Decision-Making and Society (ADM+S) du Conseil australien de la recherche.

Nathalie Casemajor est professeure agrégée au Centre Urbanisation Culture Société de l'Institut national de la recherche scientifique, Montréal. Elle est membre du Laboratoire sur la communication et le numérique (LabCMO) et de l'Observatoire des médiations culturelles (OMEC).

Alexandre Coutant est professeur agrégé au Département de communication sociale et publique à l'Université du Québec à Montréal. Il est membre du Laboratoire sur la communication et le numérique (LabCMO).

Lucie Delias est maîtresse de conférences en sciences de l'information et de la communication à l'Université Paul Valéry Montpellier 3. Elle est chercheuse au Laboratoire d'études et de recherches appliquées en sciences sociales (LERASS).

Jérôme Denis est professeur et directeur du Centre de sociologie de l'innovation à l'École des Mines de Paris de l'Université Paris Sciences et Lettres.

Maxime Harvey est doctorant en communication et chargé de cours en cinéma à l'Université du Québec à Montréal. Il est membre étudiant du Laboratoire sur la communication et le numérique (LabCMO).

Sarah Labelle est professeure des universités en sciences de l'information et de la communication à l'Université Paul Valéry Montpellier 3. Elle est chercheuse au sein du Laboratoire d'études et de recherches appliquées en sciences sociales (LERASS).

Guillaume Latzko-Toth est professeur titulaire au Département d'information et de communication de l'Université Laval. Il codirige le Laboratoire sur la communication et le numérique (LabCMO) et est membre du Centre interuniversitaire de recherche sur la science et la technologie (CIRST).

Mary Elizabeth Luka est professeure adjointe au Department of Arts, Culture and Media à l'Université de Toronto, Scarborough. Elle est membre cofondatrice du Critical Digital Methods Institute.

Anthony McCosker est professeur au Department of Media and Communication à la Swinburne University of Technology, Australie. Il dirige le Centre of Excellence for Automated Decision-Making and Society (ADM+S) du Conseil australien de la recherche.

Sarah Meunier est docteure en sociologie de l'Université de Toulouse et en science, technologie et société (STS) de l'Université du Québec à Montréal.

Florence Millerand est professeure titulaire au Département de communication sociale et publique à l'Université du Québec à Montréal. Elle dirige le Centre interuniversitaire de recherche sur la science et la technologie (CIRST) et codirige le Laboratoire sur la communication et le numérique (LabCMO).

Mélanie Millette est professeure titulaire au Département de communication sociale et publique de l'Université du Québec à Montréal. Elle est membre du Laboratoire sur la communication et le numérique (LabCMO) et du Réseau québécois en études féministes (ReQEF).

Nadia Seraio est docteure en communication de l'Université du Québec à Montréal. Elle est coordonnatrice au Réseau Hexagram et membre associée de l'Institut de recherches et d'études féministes (IREF).

Stéphane Vial est professeur agrégé à l'École de design de l'Université du Québec à Montréal. Il est titulaire de la Chaire Diamant en design pour la cybersanté mentale et membre du Centre de recherche de l'Institut universitaire en santé mentale de Montréal.

Rowan Wilken est professeur agrégé à la School of Media and Communication au Royal Melbourne Institute of Technology, Australie. Il est membre du Centre of Excellence for Automated Decision-Making and Society (ADM+S) du Conseil australien de la recherche.

Remerciements

Le thème de cet ouvrage a pris forme dans un projet de recherche intitulé *La « mise en données » de la société : enjeux sociopolitiques de la production et de l'usage de jeux de données publiques et privées*, subventionné par le Conseil de recherches en sciences humaines du Canada (CRSH, 2018-2022). La majorité des textes sont issus de travaux produits dans le cadre de ce projet. Ces travaux se sont enrichis des apports de collègues qui travaillent sur des thématiques similaires ; plusieurs de ces collègues ont d'ailleurs participé à la journée d'étude *Interroger la mise en données de la société à travers les « publics de données »* qui s'est tenue le 29 avril 2023 à l'Université du Québec à Montréal. Nous tenons à exprimer notre gratitude à Sarah Meunier ainsi qu'à Camille Nicol qui ont été nos assistantes éditoriales pour ce projet de livre ; la qualité de l'ouvrage doit beaucoup à leur travail rigoureux. Nous remercions également Jean Burgess, Kath Albury, Anthony McCosker, Rowan Wilken et les éditions Polity Press d'avoir permis de présenter une traduction française inédite de leur chapitre. Enfin, nous remercions l'équipe des PUM, et particulièrement Nicolas Lévesque, pour son accueil chaleureux et son accompagnement dans ce projet éditorial.

Table des matières

Préface	7
<i>Dominique Boullier</i>	
Introduction	11
<i>Florence Millerand, Alexandre Coutant, Guillaume Latzko-Toth et Mélanie Millette</i>	

PREMIÈRE PARTIE

L'ouverture

CHAPITRE 1	
Les publics « impensés » des données ouvertes	25
Quand les fournisseurs désignés sont aussi les usagers <i>Florence Millerand et Sarah Meunier</i>	
CHAPITRE 2	
Les figures du public dans les politiques d'ouverture des données	41
<i>Sarah Labelle</i>	

DEUXIÈME PARTIE

La santé

CHAPITRE 3	
La représentation des citoyens dans une plateforme de données en santé	59
Défis et ambiguïtés <i>Maxime Harvey et Guillaume Latzko-Toth</i>	
CHAPITRE 4	
Le codesign d'un service numérique de santé mentale	75
Concilier les imaginaires et les attentes autour des données <i>Alexandre Coutant et Stéphane Vial</i>	

TROISIÈME PARTIE

La culture

CHAPITRE 5

Les cultures de données 93

Genèse d'une notion, enjeux épistémologiques et implications pratiques

Nathalie Casemajor

CHAPITRE 6

La mobilisation de publics de niche 111

Création d'une base de données dans le secteur culturel

Mary Elizabeth Luka

QUATRIÈME PARTIE

L'inclusion

CHAPITRE 7

Data for Good 125

La science des données au service des publics minorisés et des milieux communautaires

Mélanie Millette, Florence Millerand et Nadia Seraiocco

CHAPITRE 8

L'inclusion des groupes minorisés dans les politiques publiques de données 141

Lucie Delias

CHAPITRE 9

Les publics de données quotidiens 155

Jean Burgess, Kath Albury, Anthony McCosker et Rowan Wilken

POSTFACE

Les données et leurs mondes 183

Jérôme Denis

Notices biographiques 189

Remerciements 193