Les partenariats de données numériques

LES PARTENARIATS DE DONNÉES NUMÉRIQUES:
METTRE LES BASES D'UNE GOUVERNANCE DE DONNÉES COLLABORATIVE DANS L'INTÉRÊT DU PUBLIC

Sarah Gagnon-Turcotte
Miranda Sculthorp
Steven Coutts

Février 2021

Grâce au soutien de :

SYNAPSE C

Laboratoire d'innovation urbaine de Montréal

NordOuvert

REMERCIEMENTS

Le présent rapport a vu le jour grâce au soutien de deux organismes engagés en faveur des initiatives de partenariats des données au Québec : Synapse C et le Laboratoire d'innovation urbaine (LIUM) de la Ville de Montréal. Nous souhaitons également remercier toutes les personnes qui ont accepté d'être interviewées dans le cadre de ce projet pour leur temps et leur générosité. Finalement, tous nos remerciements à nos collègues chez Nord Ouvert, notamment Megan Wylie et Lauriane Gorce pour leur soutien dans le cadre de ce projet.

À PROPOS DE NOS PARTENAIRES

Synapse C

Synapse C a pour vocation de développer une culture de la donnée au sein du milieu culturel. Pour ce faire, l'organisme à non but lucratif développe et met en commun l'expertise en valorisation des données pour les arts et la culture au Québec et au Canada, tout en travaillant en collaboration avec les secteurs culturel, académique, et entrepreneurial pour devenir une référence internationale dans l'exploitation de ces données au profit de l'écosystème culturel.

Au cours des dernières années, Synapse C a travaillé à définir les meilleures pratiques en collecte et en mutualisation des données. Dans le cadre de groupes de données mutualisées, Synapse C a ainsi contribué à mettre en commun plusieurs jeux de données et procédé à leur analyse, tout en assurant leur hébergement sécuritaire. Plus de 60 organismes du secteur culturel ont pu bénéficier jusqu'à maintenant de ces actions, principalement au Québec. En respect de ses valeurs de transparence et de collaboration, Synapse C s'assure de transmettre les apprentissages au plus grand nombre possible d'organismes.

Fort de son expérience dans la conduite et l'accompagnement de partenariats de données dans le secteur des arts et de la culture, Synapse C est désireux d'explorer et de contribuer à des modèles collaboratifs de gouvernance des données qui servent les intérêts publics. C'est pourquoi l'organisme s'est associé à Nord Ouvert pour explorer plus en profondeur la gouvernance des partenariats de données.

Le Laboratoire d'innovation urbaine de la Ville de Montréal

Le Laboratoire d'innovation urbaine de la Ville de Montréal (LIUM) favorise et accompagne l'émergence de solutions innovantes provenant de tous horizons. À la Ville, le LIUM est un espace voué à l'innovation, une zone franche où l'on explore, expérimente et imagine un futur adapté aux enjeux d'aujourd'hui. C'est un espace où la population montréalaise, les entreprises, le personnel municipal et les partenaires sont invités à co-créer une ville plus humaine, plus créative, plus ouverte et plus efficiente.

Le LIUM est l'unité responsable de mener à bien le vaste programme *Montréal en commun*, le volet montréalais du Défi des villes intelligentes du Canada, un concours du Gouvernement du Canada. Réunissant plus de 22 partenaires animés par le désir de repenser la ville, 13 projets lauréats serviront à propulser le développement et l'expérimentation de solutions innovantes dans 3 grands domaines : 1. *Alimentation*, 2. *Mobilité* et 3. *Données et expérimentation réglementaire*. Pour le troisième volet, la volonté est d'utiliser le partage des données et la gouvernance collaborative comme leviers pour mettre en œuvre l'ensemble des projets, dégager une compréhension plus fine des besoins, mesurer l'impact et prendre des décisions plus éclairées. Le rôle de Nord Ouvert s'inscrit dans l'appui et l'encadrement de ce processus.

Les actions et les moyens déployés d'ici 2024 avec *Montréal en commun* utiliseront l'innovation et les nouvelles technologies pour améliorer la qualité de vie urbaine dans toutes ses dimensions : efficacité des services, relations humaines riches, environnement sain et stimulant, milieu de vie où chaque personne se sent bien et incluse.

À PROPOS DE NOUS

Nord Ouvert

Fondé en 2011, Nord Ouvert est un organisme à but non lucratif montréalais qui a fait ses débuts dans le domaine des données ouvertes et de la technologie civique. Aujourd'hui, son équipe interdisciplinaire travaille avec une grande diversité d'administrations publiques et d'intervenants communautaires innovateurs dans les domaines clés de la gestion et de la gouvernance des données et des technologies.

Nos activités de recherche appliquée, de renforcement des capacités et nos services-conseils sont propulsés par nos valeurs de transparence, d'autonomie et de responsabilité. Notre mission est de donner aux communautés les moyens de réinventer la manière dont elles utilisent et gèrent les données et les technologies.

Nord Ouvert a développé une expertise relative à l'opérationnalisation de la gouvernance des données dans les villes intelligentes ainsi que sur le partage et la mutualisation des données dans divers secteurs (culture, santé, mobilité). Nord Ouvert joue actuellement un rôle clé dans la mise en place d'une gouvernance collaborative des données à Montréal dans le cadre de son programme *Montréal en commun*.

Les projets poursuivis dans le cadre de ce programme aborderont de multiples enjeux liés à la gouvernance des données : la collecte, l'accès et la mutualisation des données; la protection des renseignements personnels; le consentement; la propriété, le contrôle et la sécurité des données; l'ouverture des données; la conformité aux lois en vigueur, etc. L'objectif est de définir des cadres de gouvernance clairs dans lesquels la vision des données est celle d'un Commun.

La présente recherche sur les partenariats de données vise à établir des bases solides pour l'action de Nord Ouvert dans le cadre de son mandat au sein de *Montréal en commun*, en plus de consolider des apprentissages utiles qui pourront soutenir les travaux d'organismes comme Synapse C qui cherchent à accélérer la mise en place de projets de valorisation des données porteurs au Québec.

CRÉDITS

Recherche et rédaction

Sarah Gagnon-Turcotte, Directrice, Laboratoire de recherche appliquée de Nord Ouvert

Miranda Sculthorp, Analyste de recherche principale, Laboratoire de recherche appliquée de Nord Ouvert

Steven Coutts, Analyste de recherche, Laboratoire de recherche appliquée de Nord Ouvert

Mise en page

Tatev Yesayan, Consultante en design et communications

Notice bibliographique recommandée

Sarah Gagnon-Turcotte, Miranda Sculthorp et Steven Coutts (2021). *Les partenariats de données numériques: mettre les bases d'une gouvernance de données collaborative dans l'intérêt du public*. Nord Ouvert.

TABLE DES MATIÈRES

LISTE DES FIGURES, TABLEAUX ET ENCADRÉS

Figures

Tableaux

Encadrés


BUREAUX
A LOUER
(514)
282-1155
CANPRO
MASSAGE
DUNNS
AVIS
PARC
Sainte-Catherine
CLUB

PRÉFACE

Mus par l'envie d'instituer de bonnes pratiques de gestion et de gouvernance de la donnée, le Laboratoire d'innovation urbaine de Montréal (LIUM) et Synapse C ont mandaté Nord Ouvert d'effectuer une recherche sur les modalités entourant son partage.

Au cours de la dernière décennie, les nombreuses applications de la donnée ont pu démontrer l'apport utilitaire et décisionnel que celle-ci peut avoir dans une organisation municipale, et plus largement dans les transformations de la société. Ces développements récents nécessitent d'interroger les concepts de partage, de gouvernance et de gestion de la donnée, mais également d'établir des notions communes afin de faciliter la collaboration, renforcer la résilience et l'agilité, tout en soutenant une éthique responsable. Alors que des oligopoles de données se développent dans certains domaines d'activité, le LIUM, Synapse C et Nord Ouvert, conjointement avec leurs partenaires de données, veulent positionner les données comme un bien partagé et en faire un moteur de notre développement collectif. Faire de certaines données un bien partagé nécessite de développer de nouveaux cadres d'usage respectant les droits de la personne autant que la propriété intellectuelle de certains contributeurs.

Partant de ces constats, cette étude a été motivée par plusieurs intentions communes :

- Renforcer les capacités des acteurs de l'écosystème de donnée à suivre les mutations des usages et des nouvelles réglementations ;
- Améliorer la connaissance sur ces enjeux, explorer et faire un état des lieux des modèles existants dans l'objectif de définir les mécanismes propres à la réalité montréalaise ;
- Ouvrir le dialogue au plus grand nombre afin de développer un cadre de collaboration fluide avec des processus décisionnels bien définis.

Particulièrement complexes, ces intentions ont été traduites avec brio par l'équipe de Nord Ouvert dans ce document. Les risques et les défis associés à la gouvernance des données méritent en effet une grande attention. Le haut potentiel de répercussion à l'échelle individuelle et collective, tout comme dans les sphères professionnelles et organisationnelles se doivent d'être adressées afin de créer un véritable espace pour l'innovation civique, la participation citoyenne, la prise de décision et la collaboration multipartite.

Au cours des deux dernières années, Synapse C a travaillé à définir les meilleures pratiques en valorisation collective de données tout en les transmettant par la suite au plus grand nombre possible d'organismes. De son côté, la Ville de Montréal permet depuis 2011 l'ouverture de jeux de données pour lesquelles l'équipe du LIUM travaille depuis 2015 à faire émerger un maximum de valeur sociale et économique, tout en respectant la charge humaine et citoyenne que celles-ci supportent. Nos deux équipes sont particulièrement stimulées et enthousiasmées à l'idée que cette recherche devienne un document clé à ajouter à la trousse à outils de toute entité qui souhaite valoriser de la donnée dans une perspective individuelle et collective.

Diane De Courcy
Directrice générale, Synapse C

Stéphane Guidoin
Directeur, LIUM

SOMMAIRE EXÉCUTIF

À l'ère de la transformation numérique et de l'intelligence artificielle, les données et les enjeux qu'elles soulèvent sont sous les feux de la rampe, qu'il s'agisse de données ouvertes ou de données massives (*big data*). Motivé par le potentiel de leur mise en commun, un nombre croissant d'acteurs publics, privés et de la société civile s'intéresse au partage de données numériques entre tiers pour atteindre des objectifs d'intérêt public ou résoudre des problèmes sociaux complexes. De nouvelles formes de collaboration interorganisationnelle émergent chaque jour en vue de partager, combiner, croiser et valoriser des jeux de données.

Ces partenariats de données numériques demandent toutefois temps, efforts, ressources et une collaboration soutenue. Leur succès exige également la mise en place d'une solide gouvernance des données apte à protéger le public et maintenir sa confiance.

Les organisations désireuses de se lancer dans de telles initiatives trouveront dans le présent rapport une discussion des différents facteurs de succès et d'activation des partenariats de données numériques ainsi que des informations pratiques pour les guider dans la construction d'une gouvernance des données partagées qui soit collaborative, responsable, efficace et imputable.

Les partenariats de données numériques : définitions et concepts

Au cours de nos recherches, il est rapidement apparu qu'il n'existait pas de modèles clé en main en matière de partenariats de données et de leur gouvernance, qui pourraient être facilement dupliqués et mis à l'échelle. Malgré l'intérêt suscité par certains concepts comme les fiducies de données ou les communs numériques, les expériences pratiques, matures et documentées, demeurent encore trop rares.

Nous avons toutefois discerné que certains facteurs organisationnels ainsi que le contexte social et politique dans lequel se situent les partenariats de données numériques influencent grandement le cadre de gouvernance des données que devraient privilégier les organisations qui partagent et échangent leurs données. En effet, malgré la diversité des configurations des organisations privées, publiques et de la société civile qui les composent, du type de données qu'ils valorisent et des objectifs qu'ils poursuivent, les partenariats de données réussis sont généralement caractérisés par le rôle décisif de la collaboration dans leur succès. Par ailleurs, en reconnaissance des préoccupations des citoyens quant aux risques associés à de nouvelles avancées technologiques et en l'absence d'un cadre réglementaire éprouvé pour protéger leurs droits, il semble que les partenariats de données qui inscrivent leurs visées dans la recherche du bien commun ont une plus grande légitimité et capacité d'action.

Ces constats sont documentés dès le premier chapitre de notre rapport, puisqu'ils ont orienté l'ensemble de notre travail d'analyse et d'exploration des meilleures pratiques en matière de gouvernance des données.

Les principales composantes de la gouvernance des données

Alors que les partenariats de données numériques se multiplient et que les preuves de leur potentiel s'accumulent, un nombre croissant d'acteurs œuvrant dans des secteurs autres que celui des technologies de l'information s'y intéressent. Le second chapitre offre donc une base de connaissances utile et un vocabulaire commun pour définir ce en quoi consiste la gouvernance des données.

Définie simplement, la gouvernance des données détermine qui prend les décisions, comment elles sont prises et comment les décideurs sont tenus responsables en ce qui a trait à la collecte, l'utilisation, le partage ou le contrôle des données d'une organisation ou d'un groupe.

Pour faciliter la compréhension de cette définition, nous avons eu recours à un cadre conceptuel développé par Abraham, Schneider et vom Brocke (2019) pour circonscrire les éléments constitutifs de la gouvernance des données. Ce cadre est d'abord descriptif et non normatif. Il souligne l'influence des conditions préexistantes sur la gouvernance (le cadre législatif ou la culture d'entreprise par exemple), puis identifie trois éléments clés permettant de délimiter le périmètre de la gouvernance des données, soit le niveau de gouvernance organisationnel, les caractéristiques des données (partagées) et le champ d'application de la gouvernance, lesquels en retour influencent les mécanismes concrets à travers lesquels s'opérationnalise la gouvernance au quotidien.

Trois principes phares pour une gouvernance des données dans l'intérêt du public

Le troisième chapitre est consacré à une discussion approfondie de plusieurs catégories de ces mécanismes de gouvernance. Nous abordons notamment les notions de consentement éclairé, d'anonymisation, d'évaluation des risques, de qualité des données, de standardisation et d'interopérabilité, de gestion des accès, de contrôle de la conformité et d'auditabilité des décisions. Les mécanismes pouvant être déployés sont aussi nombreux que les enjeux qu'ils cherchent à encadrer. Leur choix doit prendre en considération non seulement les conditions préexistantes, mais aussi le contexte de chaque partenariat et le périmètre de la gouvernance établi.

Afin d'aider les organisations à orienter les choix qu'elles font en matière de gouvernance vers des finalités moralement et socialement désirables, nous avons structuré ce chapitre autour de trois grands principes phares. Ainsi, nous proposons que le cadre de gouvernance mis en place par les partenariats de données numériques soit guidé par les principes suivants :

- **Responsabilité : Valoriser les données de manière responsable et éthique**
- **Efficacité : Gérer les données de manière efficace et cohérente**
- **Imputabilité : Évaluer en continu la conformité et l'impact**

Perspectives montréalaises

Finalement, dans le dernier chapitre du rapport, nous présentons le résultat d'entretiens menés avec des représentants d'organisations montréalaises engagées dans diverses initiatives de partage de données. Les expériences et les perspectives concrètes de ces acteurs issus du milieu des arts et de la culture ou participant au programme de ville intelligente *Montréal en commun*, ont joué un rôle incontestable dans l'évolution de ce rapport. Ils ont notamment permis de confirmer l'influence des facteurs organisationnels comme condition d'activation et facteur de succès des partenariats de données.

En effet, les entretiens ont mis en évidence le rôle de la culture de la donnée au sein de l'organisation dans le degré d'adhésion à une initiative de partage de données, le manque de capacité organisationnelle et les coûts de la production des données comme freins à la participation et l'importance du soutien de tiers (experts légaux, initiatives gouvernementales, fonds publics, etc.) en réponse à la complexité des enjeux soulevés ou des exigences techniques, pour ne citer que ceux-là.

Malgré ces importants obstacles, nous avons néanmoins constaté l'intérêt réel des participants à explorer et développer des modèles alternatifs de gouvernance des données, qui sont adaptés à leurs besoins et leurs ambitions.

Conclusion

Ce rapport a été rédigé avec l'intention de faire œuvre utile en contribuant au corpus théorique et pratique existant sur la gouvernance des données. Nous espérons qu'il soutiendra également et de manière concrète, le mouvement en faveur du partage et de la mutualisation des données au Québec, où depuis quelques années déjà des acteurs clés, comme Synapse C et le Laboratoire d'innovation urbaine de la Ville de Montréal, se sont lancés dans l'exploration et l'expérimentation de nouvelles approches en matière de gouvernance des données.

Pour ceux qui seront tentés de se lancer dans l'aventure, nous concluons ce rapport en résumant quelques-uns de nos apprentissages, espérant qu'ils sauront contribuer à leur succès futur.

- **Reconnaître que l'intérêt public est défini et négocié par les citoyens**
- **Investir du temps dans votre processus collaboratif et d'expérimentation**
- **Créer une gouvernance des données adaptée à vos besoins**
- **Documenter votre impact et partager vos succès**

INTRODUCTION

Depuis l'avènement de l'ordinateur, la production de données numériques croît à une vitesse exponentielle. Source d'information, d'innovation, d'avantage concurrentiel, les données s'avèrent être un levier de plus en plus important aux yeux des organisations pour accroître leur impact économique et social. Dans tous les secteurs, les organisations, publiques comme privées, explorent de nouvelles manières d'utiliser leurs données afin d'en tirer le plein potentiel. Les possibilités de combinaison et de partage des données semblent presque infinies.

Il n'est donc pas surprenant de constater un engouement pour les **partenariats de données numériques**. Ces initiatives, réunissant plusieurs organisations, notamment publiques, qui joignent leurs forces pour collecter, échanger, combiner et mutualiser leurs données, se multiplient à travers le monde. Toutefois, loin de n'avoir que des bénéfices pour la société, l'utilisation et le partage des données soulèvent également des enjeux et des risques importants auxquels ces nouveaux types de partenariats se retrouvent rapidement confrontés.

Protection de la vie privée, consentement éclairé, utilisation responsable et éthique, privatisation et accès aux données, prise de décisions algorithmiques biaisées ou discriminatoires, participation citoyenne à la prise de décision : les défis sont nombreux et complexes. Tandis que la modernisation du cadre légal requis pour y répondre, tant au Canada comme à l'international, tarde à répondre aux préoccupations grandissantes du public.

Ce vide juridique a peu à peu favorisé l'émergence du concept de **gouvernance des données** numériques comme moyen pour mieux encadrer l'utilisation et le partage des données. Ce concept, issu de la traditionnelle gestion des entreprises, est désormais utilisé par une multitude de praticiens et de chercheurs désireux de définir un cadre et des principes capables de rebâtir la confiance du public. Ces réflexions ont mené à l'élaboration de divers modèles prometteurs de gouvernance des données : fiducie, coopérative, communs numériques, etc.

Néanmoins, les exemples où ces modèles de gouvernance ont été mis en pratique demeurent rares. Les facteurs de succès des partenariats de données numériques demeurent relativement peu documentés, tout comme les éléments fondamentaux d'une gouvernance des données axée sur la responsabilité, l'efficacité et l'imputabilité. C'est pourquoi les travaux visant à discerner quelles pratiques et techniques novatrices permettent d'utiliser, de combiner et de partager les données de manière responsable et éthique dans le cadre de nouveaux modèles de collaboration demeurent indispensables.

Le présent rapport entend contribuer à cet effort pour soutenir la création et le succès de partenariats de données numériques dont les visées sont dans l'intérêt du public et à proposer des recommandations concrètes quant à la mise en place des mécanismes de gouvernance des données nécessaires pour ce faire. Il s'inscrit dans un mouvement plus large en faveur de la mutualisation des données à Montréal, où depuis quelques années des acteurs clés se sont engagés à explorer et expérimenter de nouvelles approches en matière de gouvernance des données.

Afin de formuler nos recommandations, nous avons d'abord recensé les différents modèles de partenariats des données discutés dans la littérature et cherché à trouver les mécanismes de gouvernance qu'ils adoptent ainsi que les facteurs de réussite, les risques courus et les obstacles surmontés dans le cadre de ces initiatives. Nous avons analysé plus de 100 articles universitaires et rapports tirés de la littérature grise concernant la gouvernance, le partage et la mutualisation des données ainsi que

Le présent rapport entend contribuer à cet effort pour soutenir la création et le succès de partenariats de données numériques dont les visées sont dans l'intérêt du public et à proposer des recommandations concrètes quant à la mise en place des mécanismes de gouvernance des données nécessaires pour ce faire. Il s'inscrit dans un mouvement plus large en faveur de la mutualisation des données à Montréal, où depuis quelques années des acteurs clés se sont engagés à explorer et expérimenter de nouvelles approches en matière de gouvernance des données.

la gouvernance interorganisationnelle. Nous avons ensuite complété nos résultats en menant une série d'entretien auprès d'experts du milieu des arts et de la culture montréalais ainsi que des représentants d'organismes qui sont impliqués dans le programme de la ville intelligente *Montréal en commun*. Nos résultats feront ensuite l'objet d'expérimentations qui seront documentées et, nous l'espérons, nourriront un jour une nouvelle recherche permettant de valider nos résultats.

Le rapport est constitué de trois grands chapitres. Le premier, plus théorique, présente les définitions et les concepts qui sont au cœur de ces nouveaux types de partenariats. Sa lecture familiarise le lecteur néophyte avec un vocabulaire parfois technique afin qu'il gagne une vision globale des composantes de la gouvernance des données. Le second chapitre offre des exemples concrets, illustrant comment les mécanismes de gouvernance des données incarnent les principes clés de responsabilité, efficacité et imputabilité au sein des partenariats de données numériques. Finalement, le dernier chapitre synthétise les apprentissages que nous avons tirés de nos entretiens avec des acteurs engagés dans divers types de partage de données, offrant une perspective concrète des défis de telles initiatives. En conclusion, nous offrons quelques recommandations destinées aux organisations qui seraient tentées de se lancer dans l'aventure des partenariats de données numériques.

CHAPITRE 1

LES PARTENARIATS DE DONNÉES NUMÉRIQUES : DÉFINITIONS ET CONCEPTS

LES NOUVEAUX TYPES DE PARTENARIATS DE DONNÉES NUMÉRIQUES

À sa plus simple expression, le partage de données numériques est un simple échange de données entre entités dans un but particulier (Thuermer et coll., 2019). Depuis des décennies, de tels échanges ont lieu sous diverses formes dans les secteurs public, privé et universitaire.

Pour de nombreuses organisations, le partage des données constitue une fonction opérationnelle essentielle. Les ministères et les organismes publics par exemple échangent des données pour faciliter leur planification interne et améliorer leur prestation de services. Dans le secteur privé, les modèles d'affaires fondés sur l'agrégation, le partage, le courtage et l'analyse de données se sont multipliés ces dernières années (D'Addario et coll., 2020). Dans le secteur universitaire, les centres de recherche s'unissent depuis longtemps pour obtenir de nouveaux résultats grâce à l'agrégation et l'analyse de vastes ensembles de données anonymes, telles que les données génomiques (Byrd et coll., 2020). Aujourd'hui, toutefois, le partage de données numériques entre différentes organisations se fait désormais à une échelle encore jamais atteinte, et de plus en plus fréquemment entre partenaires publics et privés (Verhulst et coll., 2019).

Ces **partenariats de données numériques** ont en effet connu une croissance substantielle ces dernières années. Cette tendance n'est pas surprenante. L'économie mondiale contemporaine dépend de plus en plus des flux de données entre les individus et les organisations. Les données sont utilisées dans un éventail croissant d'activités sociales et économiques et de très nombreuses entités contribuent désormais à la production et au partage des données. De nouvelles formes de collaborations entre organisations à but non lucratif, entreprises privées, centres de recherches universitaires et administrations publiques voient désormais le jour afin de valoriser ces nouveaux flux.

Nous entendons par partenariat de données numériques toute initiative où au moins deux organisations s'unissent autour d'un objectif commun, lequel requiert le partage et la valorisation de données.

ENCADRÉ 1: QUELQUES EXEMPLES DE PARTENARIAT DE DONNÉES

PULSAR

Mis en œuvre par l'Université Laval et Alliance santé Québec, PULSAR est un espace collaboratif de recherche et d'innovation en santé durable, première initiative du genre au Québec. À la fois virtuel et réel, l'espace PULSAR rassemble des acteurs de tout horizon qui voient la recherche en santé autrement dans le but d'améliorer de façon significative et durable la santé et le bien-être de la population. Le projet implique directement le citoyen dans la démarche, en invitant les membres du public à s'inscrire à la plateforme et participer aux études en santé durable menées par les partenaires du réseau. Ultimement, les données générées par les recherches entreprises nourriront une Banque de données numériques en santé durable, une ressource informationnelle précieuse pour l'étude de la santé dans toutes ses dimensions.

IDAHO HEALTH DATA EXCHANGE

L'Idaho Health Data Exchange (IHDE), une société à but non lucratif, est l'organisme d'échange d'informations sur la santé de l'État de l'Idaho. Pour atteindre ces objectifs, l'IHDE travaille avec un large éventail de parties prenantes et construit activement une infrastructure technologique de pointe pour donner accès à des données et des informations fiables, en combinant les données traditionnelles sur les soins de santé avec d'autres sources de données. L'échange d'informations de santé permet aux médecins, aux infirmières, aux laboratoires et aux autres prestataires de soins médicaux d'accéder en toute sécurité et rapidement aux informations de santé électroniques de leurs patients, afin d'améliorer la rapidité, la qualité, la sécurité et le coût des soins aux patients.

APIDAE TOURISME

Apidae Tourisme est un réseau d'acteurs du secteur du tourisme, né en 2004, dans la région Rhône-Alpes. La plateforme est utilisée pour gérer de façon collaborative les informations touristiques de l'ensemble des territoires couverts par le projet. Les membres du réseau produisent et partagent leurs données à la plateforme, et ces mêmes producteurs sont les premiers utilisateurs des données. Cette plateforme permet de saisir, stocker et exploiter les informations touristiques pour renseigner les clients sur l'offre des destinations des membres du réseau. Le réseau compte aujourd'hui 23 départements français, 1 collectivité d'outre-mer et plus de 23 800 utilisateurs de la plateforme.

YORKINFO PARTNERSHIP

Le YorkInfo Partnership se décrit comme un « marché gouvernemental » pour les professionnels des données et de l'analyse. Ce partenariat coordonne le partage des données de système d'information géographique (SIG), y compris la photographie aérienne, les infrastructures d'eau et de déchets et les réseaux routiers — entre neuf municipalités locales et une municipalité régionale. Ces données sont accessibles à tous les partenaires via un portail en ligne et soutiennent la planification et le développement, les services d'urgence, les services sociaux et le développement économique dans toute la région.

TUI'KN PARTNERSHIP *STRENGTH IN NUMBERS PROJECT*

Dans le cadre du partenariat Tui'kn, les Premières Nations de la Nouvelle-Écosse travaillent avec des partenaires provinciaux et fédéraux pour améliorer leur accès à des informations fiables sur la santé dans le cadre du projet Strength in numbers. Cette initiative a mené à la création du Nova Scotia First Nations Client Linkage Registry, un registre de la population des Premières Nations en Nouvelle-Écosse directement relié aux données provinciales sur la santé, offrant ainsi aux Premières Nations une meilleure capacité de suivi d'un ensemble d'indicateurs de santé de leur population. L'une des pierres angulaires de ce projet est un accord de partage de données entre les Premières Nations et le gouvernement de la Nouvelle-Écosse.

GLOBAL FISHING WATCH

Global Fishing Watch est une collaboration entre SkyTruth, Oceana et Google pour cartographier et mesurer l'activité de pêche dans le monde en utilisant les données du système d'identification automatique, un système de suivi des navires utilisé par les grands navires de pêche. Une carte de ces données est disponible pour toute personne disposant d'une connexion Internet. Elle permet aux utilisateurs de suivre quand et où la pêche commerciale est pratiquée dans le monde. En accédant à cette carte, les gouvernements peuvent par exemple identifier et prendre des mesures contre les bateaux qui ne sont pas autorisés à pêcher dans leurs eaux, ou qui pêchent illégalement dans des zones protégées.

Nos recherches à ce jour nous ont permis d'identifier un grand nombre d'initiatives où des données sont partagées, liées ou mutualisées, dans le cadre d'un partenariat multipartite. Le site Data Collaboratives Explorer maintenu par The GovLab, par exemple, recense plus de 500 initiatives de partage de données, dont plusieurs dizaines d'initiatives multipartites. Cela illustre à quel point les partenariats de données numériques peuvent prendre diverses formes, se retrouvent dans différents secteurs et utilisent divers types de données. Les objectifs qui les guident sont également diversifiés.

ENCADRÉ 2 : RAISONS DE RECOURIR AU PARTAGE DES DONNÉES

Le *Data Sharing Toolkit* indique cinq raisons principales pour lesquelles des organisations sont intéressées à partager leurs données dans une dynamique de partenariat (Smart, Dubai et Nesta, 2020, p.17) [traduction libre] :

- Découvrir de nouvelles perspectives et identifier les problématiques clés
- Déceler de nouvelles sources de création de valeur à partir des données et l'innovation de tiers
- Fournir une compréhension plus complète et précise d'enjeux complexes pour une prise de décision plus rapide
- Améliorer l'exactitude des prévisions
- Renforcer la coordination et l'efficacité des processus opérationnels

Cette nouvelle tendance est si importante que plusieurs organisations, dont le Open Data Institute (Royaume-Uni), The GovLab (États-Unis), Nesta (Royaume-Uni) et plus récemment le Data Futures Lab de Mozilla, ont engagé d'importantes ressources pour documenter les modèles de partenariats de données numériques, et faire progresser leur adoption.

Parmi les modèles les plus couramment identifiés, on retrouve les collectifs de données (*data collaboratives*), les coopératives de données (*data cooperatives*), les communs numériques (*data commons*), les fiducies de données (*data trusts*) et le concept de souveraineté numérique sur les données personnelles (*personal data sovereignty*).

Comme les partenariats de données numériques sont un phénomène émergent, plusieurs de ces termes ne font toutefois toujours pas l'objet de définition établie. Certains termes sont parfois utilisés de manière interchangeable dans la littérature, ce qui les rend difficiles à distinguer. Par exemple, comme Bass et Old (2020) le font remarquer, « les communs numériques peuvent comprendre une structure de type coopératif ou fiduciaire, et on utilise parfois le terme "fiducie de données" pour désigner quelque chose qui ressemble à un modèle de communs numériques [traduction libre] » (p. 11).

ENCADRÉ 3 : LES MODÈLES ALTERNATIFS DE GOUVERNANCE DE DONNÉES

COLLECTIF DE DONNÉES

The GovLab définit les collectifs de données comme des formes de partenariats qui regroupent des entreprises privées, des institutions de recherche et des agences gouvernementales, et qui visent à combiner des données et à générer de la valeur publique (Verhulst et Sangokoya, 2015).

COOPÉRATIVES DE DONNÉES

Les coopératives de données s'apparentent aux coopératives traditionnelles : il s'agit aussi d'un groupe de personnes se rassemblant pour atteindre des objectifs communs dans une organisation jointe. Les coopératives de données peuvent être définies comme des mutuelles « détenues et contrôlées démocratiquement par des membres, qui délèguent le contrôle des données les concernant [traduction libre] » (Hardinges et coll., 2019, p. 9).

COMMUNS NUMÉRIQUES

Les communs numériques quant à eux sont des initiatives dans le cadre desquelles les données sont partagées comme des ressources communes entre individus ou organisations qui établissent collectivement les règles qui en régissent l'accès (Bass et Old, 2020).

FIDUCIE DE DONNÉES

Les fiducies de données sont définies comme des structures juridiques dont le mandat est d'assurer une intendance indépendante des données, au bénéfice de ses fiduciaires (Hardinges et coll., 2019). Ce modèle a fait l'objet de beaucoup d'attention ces dernières années. Par exemple, l'Open Data Institute a commencé à explorer le concept avec le gouvernement britannique en 2018. Elles se distinguent ainsi des autres modèles en ce sens où la fiducie est un intermédiaire distinct des membres de l'initiative de partage de données.

SOUVERAINETÉ NUMÉRIQUE SUR LES DONNÉES PERSONNELLES

La principale caractéristique d'une approche fondée sur la souveraineté numérique est que les personnes concernées ont un contrôle direct sur leurs informations personnelles. De nouvelles plateformes et initiatives numériques permettent désormais aux individus de « stocker leurs données personnelles, de collecter les données diffusées sur différentes plateformes, et de contrôler leur partage avec des tiers [traduction libre] » (Micheli et coll., 2020, p. 9). Le mouvement en faveur de la souveraineté numérique a notamment été renforcé par le droit à la portabilité des données prévu par l'article 2 de la directive 20 du Règlement général sur la protection des données (RGPD) adopté par l'Union européenne.

Comme nous le verrons dans ce chapitre, cette confusion entre les différents modèles de gouvernance s'explique en partie du fait que chaque initiative de partenariat de données est unique et qu'elle inscrit dans une dynamique et un contexte particulier, engendrée par les interactions entre les divers acteurs impliqués, leurs attentes et leurs niveaux d'expertise, leurs motivations et les relations qui leur sont propres ainsi que les lois et règles qui les régissent.

Nous avons pu constater des différences considérables entre les définitions conceptuelles et « idéalisées » utilisées pour décrire ces modèles de gouvernance, et leur mise en œuvre dans le monde réel.

De plus, malgré les travaux réalisés à ce jour pour distinguer plus précisément les différents modèles de gouvernance de données, nos recherches ont également démontré des lacunes importantes au niveau de la documentation d'exemples probants, matures et réussis (Coutts et Gagnon-Turcotte, 2020). Ces nouvelles approches demeurent nouvelles et émergentes et nécessitent donc une analyse prudente.

S'appuyant sur ces considérations, le présent rapport ne cherche pas à promouvoir ou définir un modèle de partenariat de données numériques en particulier. Nous nous sommes plutôt intéressés aux éléments fondamentaux qui caractérisent la gouvernance des données lorsqu'elle se déploie dans le cadre de tels partenariats en général. Nous avons souhaité, par cette approche, offrir une perspective plus opérationnelle, centrée sur la gouvernance des données et pouvant être adaptée et mise à l'échelle peu importe la forme et la structure privilégiées.

Avant d'aborder directement la gouvernance des données, il nous semble néanmoins essentiel d'aborder plus en détail deux facteurs de succès qui, malgré la diversité des partenariats de données numériques, nous sont rapidement apparus comme incontournables à leur réussite : la collaboration et la recherche de l'intérêt public.

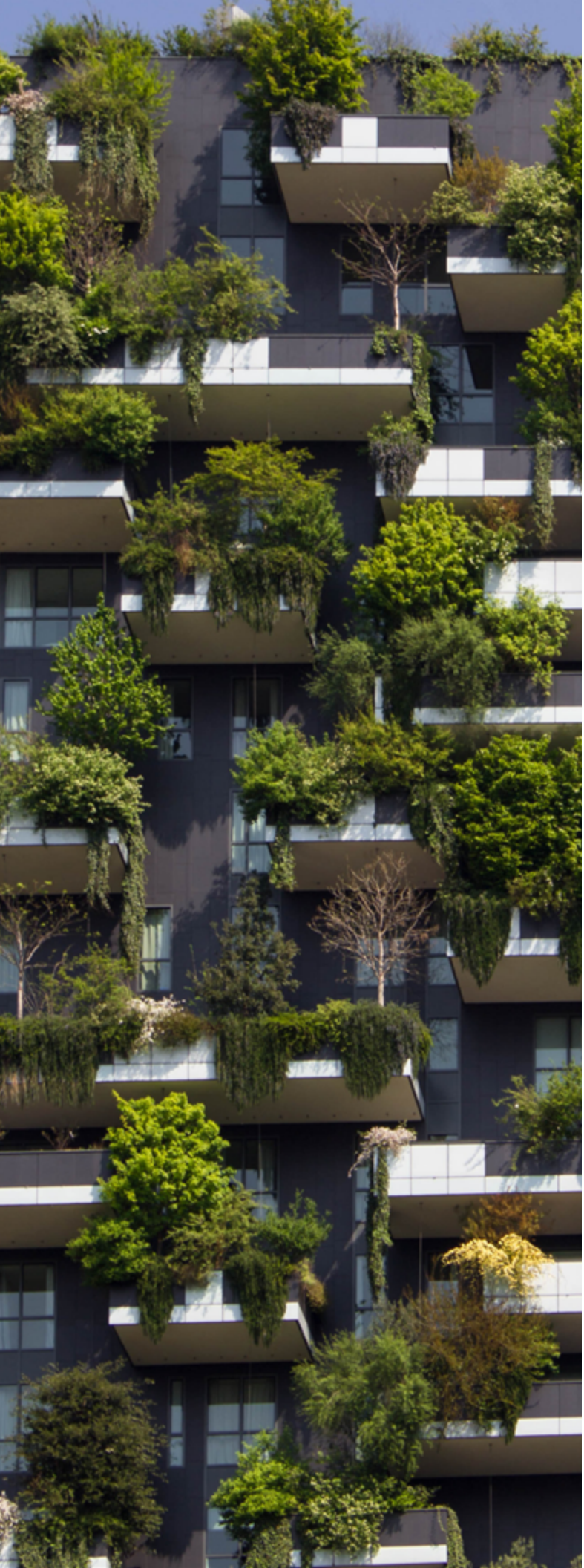
FONDEMENTS D'UN PARTENARIAT DE DONNÉES NUMÉRIQUES RÉUSSI

Au cours de notre recherche, en plus de tenter de mieux cerner les caractéristiques propres de la gouvernance des partenariats des données, nous avons également cherché à identifier quels facteurs influencent leur succès. Quoiqu'ils soient nombreux (voir notamment le chapitre 4 qui recense ceux identifiés lors des entrevues), nous avons identifié deux facteurs qui nous sont apparus comme étant des fondements incontournables des partenariats de données numériques dans le contexte actuel : **la création d'un climat de confiance propice à la collaboration et la recherche d'un impact social positif dans l'intérêt du public.**

Le processus collaboratif

Le rôle de la collaboration dans la réussite d'un partenariat peut sembler de prime abord une évidence. Toutefois, les difficultés rencontrées par les participants à des partenariats de données numériques sont telles que seule une réponse collaborative est à même de les surmonter. Parmi celles-ci, on retrouve un cadre juridique flou et en évolution rapide, la difficulté d'anticiper les impacts négatifs des nouvelles manières d'utiliser les données, le poids des compétences techniques (en particulier pour les partenariats issus de secteurs ou d'industries non technologiques) et le regard du public. Dans un contexte à si haut risque, **la confiance et la collaboration** sont parmi les principales caractéristiques des interactions dans le cadre des initiatives réussies de partenariats de données numériques.

Tel que le démontrent Ansell et Gash (2007) dans leur *processus collaboratif*, la confiance et la collaboration forment un cycle continu et vertueux. La confiance favorise l'engagement dans le processus, ce qui contribue à développer une vision commune des enjeux et des objectifs, ce qui favorise ensuite l'atteinte de résultats et encourage le dialogue, lequel démontre la valeur du processus collaboratif, ce qui renforce à son tour la confiance.



Les obstacles à la participation

Avant de s'engager dans un processus collaboratif, les relations entre acteurs peuvent être caractérisées par des dynamiques de concurrence, dont certaines de tensions sont décrites dans la littérature : concurrence, perte de contrôle et objectifs divergents.

Concurrence

Les organisations peuvent être réticentes à partager des données qui, selon elles, offrent un avantage concurrentiel à leurs partenaires (Klievink et coll., 2018). Par exemple, deux entreprises peuvent avoir investi beaucoup de temps et de ressources dans la création de bases de données contenant des informations sur leurs clients respectifs, telles que des données démographiques, des historiques d'achat et des préférences. Dans une perspective de maintien du statu quo, ces données seraient considérées comme un avantage concurrentiel unique. Toutefois, un partenariat de données numériques a peu de chances de voir le jour si les deux entreprises se considèrent comme engagées dans un jeu à somme nulle. Il a plus de chances de réussir si les deux entreprises ont conclu le partenariat en pensant qu'une plus grande valeur peut être obtenue en mettant en commun leurs données, par exemple en « élargissant leur bassin de clients potentiels ou en développant des produits et services complémentaires [traduction libre] » (D'Addario et coll., 2020, p.14).

Perte de contrôle

De manière similaire à l'idée de concurrence, selon la théorie des ressources, pour créer un avantage concurrentiel, les entreprises doivent disposer d'une ressource précieuse et rare, et éviter l'utilisation de cette ressource par d'autres, par le biais de l'imitation, du transfert ou de la substitution (Wade et Hulland, 2004). Partager des données dans le cadre

d'un partenariat donc signifie renoncer à un certain contrôle sur une ressource organisationnelle fondamentale (Klievink et coll., 2018), ce qui peut laisser aux acteurs un sentiment de vulnérabilité. Dans un environnement collaboratif, les membres d'un partenariat doivent alors accepter que leurs actifs en données soient utilisés dans les processus qu'ils ne contrôlent pas entièrement, mais en cédant une certaine autonomie aux collaborateurs, ils ouvrent la voie à la création de nouvelles informations, services ou produits (Klievink et coll., 2018).

Objectifs divergents

Même si les organisations se dotent de différents objectifs (Vangen et Huxham, 2012), ils peuvent créer des synergies en combinant leurs ressources et leurs capacités en matière de données. Il est possible que différentes organisations s'accordent sur une vision commune, mais diffèrent sur la manière exacte d'y parvenir. Par exemple, la vision de la création de systèmes alimentaires locaux équitables et durables peut être servie par une organisation qui concentre ses efforts sur l'offre de services de banque alimentaire, tandis qu'une autre se concentre sur la réduction du gaspillage alimentaire dans les chaînes d'approvisionnement (Bolychevsky et coll., 2019). Une divergence quant au moyen à privilégier peut entraver la création d'un partenariat de données numériques.

Ces dynamiques peuvent néanmoins être surmontées lorsque les parties sont fermement engagées dans un processus collaboratif et grâce à la définition et à l'adhésion à un ensemble d'objectifs communs.

L'adhésion à une vision commune

En effet, Ansell et Gash (2007) identifient le développement d'une vision commune parmi les parties prenantes comme une des étapes critiques dans le processus collaboratif. Cela implique que les parties s'accordent sur la définition du problème à résoudre, s'alignent autour d'un ou des objectifs communs et s'entendent sur les connaissances et les moyens nécessaires à la résolution du problème. Ces bases doivent être établies avant même d'aborder la question des capacités ou des ressources requises au sein d'une organisation pour pouvoir pleinement s'engager dans un partenariat de données numériques.

Les partenaires doivent se faire une idée précise du problème qu'ils cherchent à résoudre, discerner exactement la valeur et l'utilité des données pour y répondre et déterminer l'avantage spécifique qu'un partenariat de données numériques peut offrir. À ce stade, les parties doivent chercher à répondre à des questions telles que : *Pourquoi ce projet ? Quels sont nos objectifs ? Pourquoi ces données ? À quoi serviront-elles ?* De cet exercice doit émerger une vision commune, à l'aune de laquelle les partenaires pourront ultimement mesurer leur progrès. Sans quoi, le partenariat est à risque d'une dérive coûteuse des objectifs ou encore de voir vaciller l'adhésion des partenaires au fil du temps.

Appliquant les concepts tirés du processus collaboratif d'Ansell et Gash (2007) à un exemple concret de partenariat de données aux Pays-Bas, Klievink et coll. (2018) ajoute que la rétroaction positive entre la confiance et la collaboration favorise l'institutionnalisation de cette dernière, ce qui permet aux partenariats de traverser plus facilement des épisodes difficiles ou de surmonter les tensions entre les parties lorsqu'elles émergent, dont les dynamiques concurrentielles. Il constate également que la présence d'un historique de collaboration entre les parties peut contribuer à soutenir l'émergence et le succès des partenariats de données numériques.

Les cas d'usage comme outil d'expérimentation

En présence de nombreux acteurs aux intérêts divers et compte tenu de la diversité des données pouvant être partagées, l'adhésion à une vision commune peut toutefois être un défi. Dans le contexte d'un partenariat de données, en particulier ceux axés sur l'innovation, où les parties sont intéressées à découvrir de nouvelles façons de valoriser des données n'ayant encore jamais fait l'objet de partage, cela peut être encore plus difficile.

Dans ces circonstances, la désignation des différents **cas d'usage** justifiant le partage des données facilite grandement la création de cette vision commune. Les cas d'usage viennent préciser pour les parties les objectifs qu'ils souhaitent atteindre et ciblent les données à partager pour les atteindre, puis spécifient la manière dont elles seront utilisées. Cette approche facilite l'identification des obstacles et des risques au partage ainsi que la compréhension partagée des enjeux.

Considérant les importantes ressources, notamment techniques, qui peuvent être engagées avant d'atteindre les résultats d'un partenariat de données numériques, il est utile de favoriser les approches itératives centrées sur l'expérimentation. Ce type d'approches est largement utilisé dans le secteur des technologies de l'information, car il favorise les cycles de développement courts et le prototypage, de manière à découvrir si une idée ou une utilisation de données est viable et désirable. L'utilisation de cas d'usage précis comme base au processus collaboratif se prête particulièrement bien à ces méthodes.

Cette philosophie est prometteuse dans le cadre des partenariats de données numériques, parce qu'elle encourage les parties à s'adapter et innover face aux nouvelles données et aux environnements numériques, lesquels sont en perpétuel changement. De plus, le prototypage est un bon moyen d'instaurer la confiance autour de petits succès qui peuvent ensuite être transposés à plus grande échelle. En ce sens, l'expérimentation est un élément essentiel du *processus collaboratif* comme il est décrit par Ansell et Gash (2007).

ENCADRÉ 4 : VALORISATION ET CAS D'USAGE

Un cas d'usage (ou cas d'utilisation, *use case* en anglais) fait référence à une manière d'utiliser un logiciel ou, en l'occurrence, des données, en vue d'atteindre une utilité précise pour les acteurs impliqués. L'identification de cas d'usage précis est essentielle à la valorisation des données dans le cadre d'un partenariat de données numériques.

Un même jeu de données peut donner lieu à plusieurs cas d'usage. Chaque cas d'usage définit précisément le périmètre d'utilisation des données en vue de réaliser sa finalité. Il permet ainsi de décrire les exigences fonctionnelles et d'identifier les mécanismes de gouvernance requis par le système. En général, les cas d'usage sont construits en adoptant le point de vue de l'utilisateur final (définition inspirée de Jacobson et coll., 2011).

Pour concevoir un cas d'usage, il est utile de commencer en repérant et en documentant les données disponibles et utiles, par exemple au moyen d'un audit ou d'une cartographie des données.

La recherche de l'intérêt public

Notre présentation des partenariats de données numériques a jusqu'à présent mis l'accent sur plusieurs dimensions clés : le partage et la valorisation des données impliquant différents acteurs, ainsi que la mise en place d'une dynamique de collaboration à partir de l'adhésion aux objectifs communs et l'élaboration des cas d'usage. Une question se pose alors, quels types d'objectifs peuvent être poursuivis par le biais de ces partenariats ? Bien que de nombreux types d'objectifs différents puissent être poursuivis, nous pensons que le potentiel d'impact social positif est le plus important lorsque les partenariats de données numériques poursuivent des objectifs dans l'intérêt du public.

Bien plus qu'une simple bonne pratique, la recherche de l'intérêt public est nécessaire pour le succès des partenariats de données numériques. En effet, l'appropriation de la valeur des données publiques et celles des individus à des fins commerciales et lucratives est de plus en plus mal perçue par le public et peut mener à des enjeux de légitimité sérieux (Artyushina, 2020). La crédibilité du projet et les relations avec les partenaires et les bénéficiaires locaux peuvent également être mises à mal lorsque les parties prenantes ont l'impression d'être exploitées pour leurs données ou encore lorsque les projets ont des conséquences négatives et imprévues. Les conséquences pour les organisations impliquées dans un projet qui ne reçoit pas l'aval du public peuvent être sérieuses et affecter jusqu'à son financement ou sa responsabilité juridique. Ce faisant, un tel ancrage est de plus en plus largement adopté.

En effet, un examen rapide du discours et des projets récents en matière de gouvernance des données révèle différentes façons dont l'intérêt public s'incarne actuellement dans les initiatives de partenariats de données numériques. Premièrement, une majorité d'acteurs reconnaissent qu'il existe des raisons éthiques de protéger les données et de veiller à ce qu'elles ne soient pas utilisées à mauvais escient, des raisons qui vont au-delà du simple respect des lois et réglementations existantes. Par exemple, le *Data Management Body of Knowledge* (DAMA-DMBOK) (Earley et coll., 2017, p.49), un des guides les plus réputés dans le domaine, souligne que la gestion des données doit être guidée par les principes éthiques de respect des personnes, de bienfaisance et de justice.

Deuxièmement, de plus en plus d'initiatives sont désormais centrées sur le concept des « données pour le bien public », à savoir que « les données sont créées par la société » et doivent donc d'abord être mises à profit dans l'intérêt de la collectivité (Commission européenne, 2020, p.8). Cette approche reconnaît l'apport des institutions publiques et des gouvernements dans la production de données (sur la démographie, les infrastructures, la mobilité, la santé, etc.). En d'autres termes, les données ont une valeur publique ou comme l'indique la nouvelle Charte des données numériques de la Ville de Montréal :

« Les données recueillies par les organisations ayant une portée publique ou servant l'intérêt public, le sont au nom du citoyen. Elles représentent un actif partagé et donc un bien commun. Selon cette logique, nous avons le devoir de permettre à chaque personne de bénéficier de la valeur de ces données en les rendant disponibles. »

Finalement, d'autres discours vont plus loin et considèrent les données comme un bien public en elles-mêmes ou une ressource commune qui doit non seulement servir le public, mais être gérée collectivement. À l'origine de cette idée, on retrouve notamment les théories proposées par Elinor Ostrom, lauréate du prix Nobel d'économie, dans le cadre de son travail sur la gestion des communs (Ostrom, 1990).

> **Les données recueillies par les organisations ayant une portée publique ou servant l'intérêt public, le sont au nom du citoyen. Elles représentent un actif partagé et donc un bien commun. Selon cette logique, nous avons le devoir de permettre à chaque personne de bénéficier de la valeur de ces données en les rendant disponibles.**
>
> **- Charte des données numériques de la Ville de Montréal (2020)**

Concrètement, comment un partenariat de données numériques peut-il s'assurer que ses objectifs sont bel et bien ancrés dans la poursuite du bien commun ? Nous avons identifié trois piliers complémentaires à cet effet : la création de bénéfices tangibles pour le public, la participation citoyenne et l'adhésion à des principes de bonne gouvernance des données.

Générer des bénéfices tangibles pour le public

Un partenariat de données numériques dans l'intérêt du public doit tout d'abord s'assurer que l'utilisation qu'il fait des données publiques et personnelles génère des bénéfices réels pour la collectivité. Il peut s'agir d'avantages directs pour les individus ou encore d'impacts sociaux positifs plus larges. Involve, Understanding Patient Data and the Carnegie UK Trust (2018), lors une série d'ateliers organisés dans des collectivités locales en Angleterre, a examiné la définition et l'évaluation des bénéfices qui peuvent être générés à partir du partage des renseignements personnels parmi des prestataires de services publics. Les résultats de leur étude permettent d'identifier un certain nombre de critères pour évaluer les retombées d'une initiative de partage de données : le nombre de personnes qui en bénéficieront, la gravité du besoin (par exemple, améliorer le sort d'un groupe vulnérable tel que les sans-abri) ou traiter des problèmes sociaux clés (tel que l'isolement social) et la durée des retombées à long terme sur les individus et les services offerts. Ce ne sont là que quelques indicateurs applicables au secteur public.

En plus d'évaluer les avantages tangibles de leurs initiatives, les membres d'un partenariat de données numériques pourront réfléchir à la manière dont ces avantages peuvent être distribués de façon équitable, par exemple, en application du principe de justice identifié dans le DAMA-DMBOK (Earley et coll., 2017, p.52).

Faire participer le public de manière significative

Les citoyens ont des attentes, des aspirations, voire des exigences quant à la manière dont leurs données (personnelles) et les données du domaine public, devraient être collectées, utilisées et gérées. La contribution du public à ces questions est donc essentielle à la légitimité des partenariats de données numériques et donc à leur succès.

D'une part, les gouvernements (à toutes les échelles) jouent un rôle clé dans ce domaine d'intervention, car ils ont la responsabilité d'amorcer et de mener des discussions et des débats sociétaux sur les données et leur usage responsable, par le biais de consultations publiques ou d'autres forums consultatifs. Ils jouent également un rôle critique dans le développement et le renforcement des compétences et des connaissances dont les citoyens et les parties prenantes ont besoin pour participer à ces discussions et à la prise de décision dans le cadre des partenariats des données (voir encadré 5).

Mais de manière générale, tout partenariat de données numériques, même entre organisations privées[1], a intérêt à se doter de solides processus consultatifs permettant à ses parties prenantes, mais également aux citoyens, de se prononcer sur l'initiative et la façon dont les données y sont utilisées. Ces démarches, pour être légitimes, doivent inclure une grande variété de voix et de perspectives, ce qui implique la mise en place de stratégies de mobilisation adaptées aux différentes populations, en particulier celles qui sont typiquement sous-représentées dans les processus consultatifs. En outre, les citoyens et les parties prenantes devraient être impliqués et engagés dès le départ, avant même que la collecte de données n'ait lieu.

Un des moyens pour favoriser la participation est de créer des canaux de communication explicites ou des instances permanentes, tels que des comités consultatifs citoyens, où le public est encouragé à participer à la gouvernance de l'initiative elle-même. Ces différents modes de participation citoyenne contribuent à maintenir la confiance du public sur une base continue.

Adhérer à des principes forts pour orienter la gouvernance des données

Finalement, afin d'atteindre des finalités socialement responsables, les partenariats de données numériques doivent s'assurer que les données qu'ils utilisent sont gérées de manière responsable et efficace et que des mécanismes de reddition de compte sont mis en place. C'est là que la gouvernance des données entre en jeu.

La gouvernance des données est un cadre qui permet de réfléchir aux pratiques encadrant la gestion des données et aux processus décisionnels qui les sous-tendent. La gouvernance des données existe pour garantir la manière dont les données sont collectées, traitées, consultées, utilisées, stockées, partagées, etc., par les parties servent ultimement à atteindre les objectifs établis en commun dans le cadre du partenariat. Elle existe également pour garantir que le traitement des données par les parties est conforme à la loi et répondre à des principes plus larges en gouvernance des données.

Le reste de ce rapport sera entièrement consacré à explorer la gouvernance des données et son rôle dans le cadre de partenariats de données numériques. Dans le chapitre 2, nous décortiquons les éléments fondamentaux de la gouvernance des données afin d'offrir une vue d'ensemble sur le sujet et de développer un vocabulaire commun. Puis dans le chapitre 3, nous examinons plus en détail quels mécanismes de gouvernances concrets nous permettent d'atteindre les principes de responsabilité, efficacité et imputabilité qui sont au cœur d'un projet dans l'intérêt du public.

1 Une organisation privée peut être une organisation à but non lucratif, une association, une coopérative, etc.

ENCADRÉ 5 : LA LITTÉRATIE DES DONNÉES

Pour donner aux citoyens les moyens de participer pleinement dans notre société qui accorde de plus en plus un rôle central aux données dans la prise de décision, de nombreux auteurs (Ridsdale et coll., 2015; Bhargava et D'Ignazio, 2015; Calzada Prado et Marzal, 2013; Wolff et coll., 2016) soulignent l'importance de bâtir les compétences en *littératie des données* du grand public. Wolff et coll. (2016) définissent ce concept comme suit :

« La littératie des données est la capacité de poser des questions concrètes et d'y répondre à partir de vastes et de petits ensembles de données au moyen d'un processus de requête, en tenant compte de l'utilisation éthique des données. Elle repose sur des compétences de base pratiques et créatives, avec une capacité d'étendre les connaissances des compétences de traitement des données spécialisées selon les objectifs. Cela comprend les capacités de sélectionner, de trier, d'analyser, de visualiser, de critiquer et d'interpréter les données, de même que de communiquer les histoires à partir des données et d'utiliser les données dans le cadre d'un processus de conception [traduction libre] » (p.23).

ENCADRÉ 6 : ÉTAPES D'UN PROCESSUS DE PARTAGE DE DONNÉES

En misant sur la confiance et la collaboration et en se dotant d'objectifs clairs fondés sur la recherche de l'intérêt public, un partenariat de données numériques a plus de chance de réussir. Nous sommes toutefois conscients que tout projet d'envergure nécessite de nombreuses autres démarches de la part des partenaires impliqués, telles que l'identification d'un modèle d'affaires qui assurera la viabilité financière du partenariat ou le choix des technologies à déployer.

Voici un exemple des principaux points de décision identifiés par The Royal Academy of Engineering (2019) lors de la constitution d'une initiative de partage de données [traduction libre] :

1. Définir la possibilité
2. Déterminer le périmètre des données à partager et leur mode d'utilisation
3. Développer le modèle d'affaires qui permet de générer et de partager de la valeur
4. Développer le modèle de partage des données et le partenariat
5. Assurer la participation de personnes aux compétences et à l'expertise appropriées
6. Déterminer les contraintes, dont les exigences juridiques et réglementaires, qui s'appliquent au partage, au traitement et à l'utilisation des données
7. Déterminer les architectures et les technologies requises pour permettre le partage des données
8. Développer les mécanismes appropriés de gouvernance et de supervision qui assureront un partage sécuritaire des données

La suite de ce rapport se consacre principalement à l'étape 8 : l'établissement d'une gouvernance des données solide et capable d'atteindre les objectifs du partenariat.

CHAPITRE 2

LES PRINCIPALES COMPOSANTES DE LA GOUVERNANCE DES DONNÉES

Plus une organisation recueille de données, plus il lui est nécessaire de régir continuellement et systématiquement leur utilisation. Ce besoin est encore plus crucial lorsque les données sont détenues ou utilisées par plusieurs organisations, puis partagées et agrégées pour en tirer de la valeur. Par conséquent, la **gouvernance des données** doit être au cœur de toute initiative de partenariat de données numériques.

QU'EST-CE QUE LA GOUVERNANCE DES DONNÉES? UN CADRE CONCEPTUEL

La gouvernance des données et de ses champs d'application est l'objet d'interprétations disciplinaires diverses de sorte qu'il existe de multiples conceptions de cette notion. Pour bien comprendre ce que signifie la gouvernance des données, il est utile de se référer aux définitions issues du domaine des technologies de l'information (TI). Selon Weill, la *gestion* des TI concerne les décisions prises, tandis que la *gouvernance* des TI réfère aux personnes qui prennent ces décisions et comment elles en sont tenues responsables (Weill, 2004). Ce point de vue sépare deux niveaux d'abstraction : le « quoi » (la gestion) et le « comment » (la gouvernance).

En effet, la gouvernance des données ne doit pas être confondue avec la *gestion* des données. Par exemple, le DAMA-DMBOK (Earley et coll., 2017) décrit la gouvernance des données comme «l'exercice de l'autorité et du contrôle (planification, suivi et application) sur la gestion des actifs de données [traduction libre] » (p.67).

En somme, la gouvernance des données détermine qui prend les décisions, comment elles sont prises et comment les décideurs sont tenus responsables en ce qui a trait à la collecte, l'utilisation, le partage ou le contrôle des données d'une organisation ou d'un groupe.

En somme, la gouvernance des données détermine qui prend les décisions, comment elles sont prises et comment les décideurs sont tenus responsables en ce qui a trait à la collecte, l'utilisation, le partage ou le contrôle des données d'une organisation ou d'un groupe.

Cette clarté en matière de responsabilités et de processus décisionnels associés à un ensemble particulier de données est essentielle pour assurer le succès d'un partenariat de données numériques. Le cadre de gouvernance permet également d'enchâsser, puis d'évaluer le respect des principes de responsabilité, d'efficacité et d'imputabilité qui assure la défense du bien commun au sein des projets.

Pour faciliter la compréhension de la gouvernance des données et de son opérationnalisation, nous avons utilisé dans le cadre de notre analyse un cadre conceptuel développé par Abraham, Schneider et vom Brocke (2019). L'utilisation de ce cadre conceptuel, décrit dans l'encadré 7, facilite la distinction des principaux éléments constitutifs de la gouvernance des données ainsi que leur interrelation. Il permet ainsi de mieux comprendre le rôle et les impacts de la gouvernance des données dans le cadre d'un partenariat de données numériques.

Il est important de noter que ce cadre décrit uniquement ces éléments tels qu'ils apparaissent dans la littérature ; il ne s'agit pas d'un cadre normatif qui indique comment la gouvernance des données devrait fonctionner ou quelles valeurs devraient la guider. Le cadre se veut d'abord descriptif, plutôt que normatif, c'est pourquoi le présent rapport discute également de l'importance d'une vision, de principes et d'objectifs communs aux partenariats de données numériques.

ENCADRÉ 7 : COMPRENDRE LA GOUVERNANCE DES DONNÉES AU MOYEN D'UN CADRE CONCEPTUEL

Le cadre conceptuel suivant, développé par Abraham, Schneider et vom Brocke synthétise les principaux éléments constitutifs de la gouvernance des données sur la base d'une analyse documentaire recensant les ouvrages et articles publiés au cours des deux dernières décennies sur la gouvernance des données :

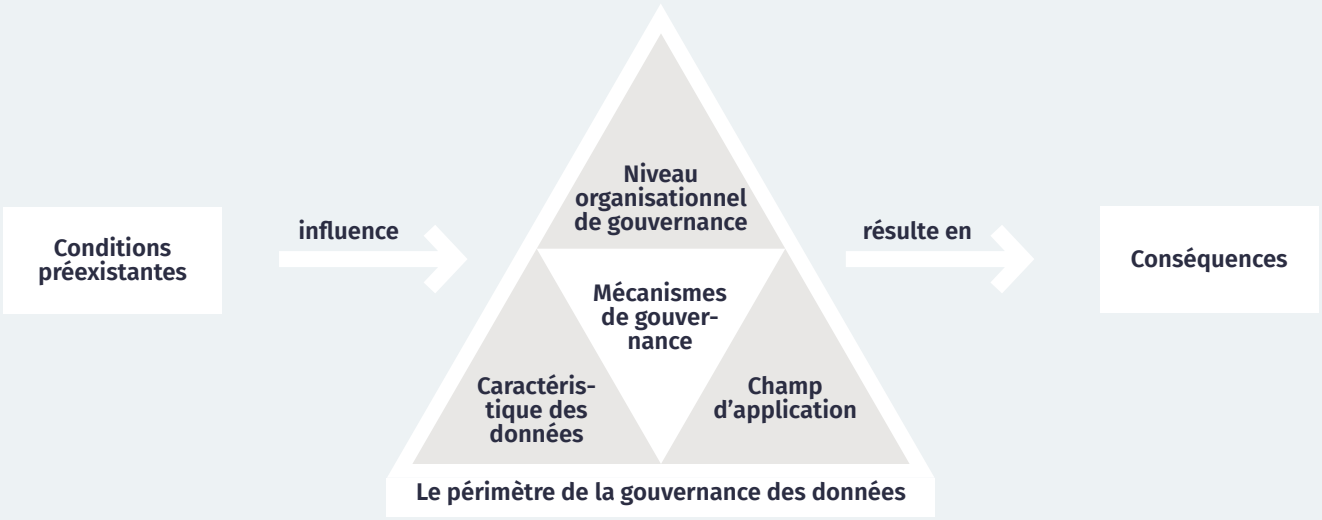


FIGURE 1 : SCHÉMA DES ÉLÉMENTS FONDAMENTAUX DE LA GOUVERNANCE DES DONNÉES

Reproduction tirée de l'*International Journal of Information Management*, vol. 49, Rene Abraham, Johannes Schneider, et Jan vom Brocke, « Data governance: A conceptual framework, structured review, and research agenda » p. 428, Copyright (2019), avec l'autorisation d'Elsevier.

Ils concluent que cette gouvernance s'exprime au moyen d'un ensemble complexe de **mécanismes** explicites ou implicites, qui peuvent prendre de nombreuses formes : politiques, procédures, pratiques, etc., encadrant la collecte, l'utilisation, le partage et le contrôle des données.

Ces mécanismes assurent l'opérationnalisation de la gouvernance des données et sont façonnés par diverses **conditions préexistantes** au niveau politique, juridique, réglementaire, organisationnel ou même culturel.

Ces mécanismes dépendent également du **périmètre** de la gouvernance des données, laquelle comporte trois dimensions c'est-à-dire de son **champ d'application**, le **niveau de gouvernance** auquel elle est appliquée et les **caractéristiques des données** qui font l'objet de la gouvernance.

Finalement, les choix en matière de gouvernance des données ont des **conséquences** mesurables. Il peut s'agir d'améliorer l'efficacité opérationnelle à court terme d'une organisation ou d'un groupe, d'atténuer certains risques (tel que l'atteinte à la vie privée) ou, à plus long terme, d'accroître la confiance du public à l'égard de la gouvernance des données.

LES CONDITIONS PRÉEXISTANTES

Le premier élément de la gouvernance des données consiste en une variété de conditions préexistantes qui viennent déterminer et influencer la gouvernance des données. Selon Abraham, Schneider et vom Brocke (2019), il existe des conditions préexistantes internes (organisationnelles) et d'autres externes (principalement réglementaires).

Les **conditions internes**, lesquelles concernent directement le mode de fonctionnement et les priorités des organismes qui mettent en œuvre la gouvernance des données, peuvent être d'ordre stratégique, organisationnel ou culturel. Par exemple, les objectifs de profitabilité d'une organisation, son niveau de centralisation, son type de leadership peuvent tous influencer le cadre de gouvernance privilégié. La présence de silos, la culture d'entreprise (axée sur l'innovation ou non, par exemple) ou encore le niveau d'adhésion de la direction au projet entrent aussi en ligne de compte. Ces aspects organisationnels, quoiqu'importants, ne seront pas abordés en détail dans le présent rapport, mais doivent tout de même être gardés à l'esprit par ceux qui s'apprêtent à lancer un partenariat de données numériques.

Les **conditions externes** quant à elles font d'abord référence aux lois et règlements en vigueur, ainsi qu'aux normes et standards existants. Tant le cadre juridique que les normes peuvent varier d'une région et d'une industrie à l'autre. En effet, le type de données, la juridiction applicable ou encore le champ d'application concerné par la gouvernance peuvent déterminer l'applicabilité d'une loi plutôt qu'une autre, tandis que le régime juridique en vigueur peut désigner certain type de mécanismes requis en pour assurer la conformité. Il est donc requis de porter une attention particulière à ces conditions dès les premiers développements d'un partenariat de données numériques.

Le contexte juridique

Au Canada, il existe de nombreuses lois qui servent de conditions préexistantes à un partenariat de données numériques. Nous soulignons ici brièvement celles qui sont les plus pertinentes pour ces partenariats, notamment celles qui régissent les données personnelles et la propriété intellectuelle.

La protection de la vie privée

Au niveau fédéral, la *Loi sur la protection des renseignements personnels et les documents électroniques* (LPRPDE) est la principale loi qui régit les données du domaine personnel. Elle s'applique à toute organisation qui recueille, utilise et diffuse des renseignements personnels dans le cadre d'activités commerciales (veuillez voir tableau 1). Elle stipule les règles (ainsi que leurs exceptions), que les organisations doivent suivre pour utiliser des renseignements personnels des consommateurs.

Plusieurs provinces, comme l'Alberta, la Colombie-Britannique et le Québec, ont quant à elles leurs propres lois provinciales sur la protection de la vie privée, lesquelles sont semblables à la loi fédérale correspondante (Commissariat à la protection de la vie privée du Canada, 2017).

Au Québec, s'ajoutent deux grandes lois décrivant comment les données personnelles doivent être traitées. Les organismes du secteur public sont soumis à la *Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels* tandis que les organisations du secteur privé sont soumises à la *Loi sur la protection des renseignements personnels dans le secteur privé*.

En juin 2020, le gouvernement québécois a déposé un nouveau projet de loi visant à renforcer le cadre juridique en vigueur, le projet de loi n° 64, intitulé *Loi modernisant des dispositions législatives en matière de protection des renseignements personnels*. Ce projet de loi vise à moderniser le cadre législatif

TABLEAU 1 : OBLIGATIONS LÉGALES FONDÉES SUR LA *LOI SUR LA PROTECTION DES RENSEIGNEMENTS PERSONNELS ET LES DOCUMENTS ÉLECTRONIQUES*

1. Responsabilité	Une organisation est responsable des renseignements personnels dont elle a la gestion. Elle doit nommer une personne qui devra s'assurer de sa conformité à ces principes relatifs à l'équité.
2. Détermination des fins de la collecte des renseignements	Les fins auxquelles des renseignements personnels sont recueillis doivent être déterminées par l'organisation avant la collecte ou au moment de celle-ci.
3. Consentement	Toute personne doit être informée de toute collecte, utilisation ou communication de renseignements personnels qui la concernent et y consentir, à moins qu'il ne soit pas approprié de le faire.
4. Limitation de la collecte	L'organisation ne peut recueillir que les renseignements personnels nécessaires aux fins déterminées et doit procéder de façon honnête et licite.
5. Limitation de l'utilisation, de la communication et de la conservation	À moins que la personne concernée n'y consente ou que la loi ne l'exige, les renseignements personnels ne doivent être utilisés ou communiqués qu'aux fins auxquelles ils ont été recueillis. On ne doit conserver les renseignements personnels qu'aussi longtemps que nécessaire pour répondre à ces fins.
6. Exactitude	Les renseignements personnels doivent être aussi exacts, complets et à jour que possible afin de satisfaire aux fins auxquelles ils sont destinés.
7. Mesures de sécurité	Les renseignements personnels doivent être protégés au moyen de mesures de sécurité correspondant à leur degré de sensibilité.
8. Transparence	Une organisation doit faire en sorte que des renseignements précis sur ses politiques et ses pratiques concernant la gestion des renseignements personnels soient facilement accessibles au public.
9. Accès aux renseignements personnels	Une organisation doit informer toute personne qui en fait la demande de l'existence de renseignements personnels qui la concernent, de l'usage qui en est fait et du fait qu'ils ont été communiqués à des tiers, et lui permettre de les consulter. Il sera aussi possible de contester l'exactitude et l'intégralité des renseignements et d'y faire apporter les corrections appropriées.
10. Possibilité de porter plainte à l'égard du non-respect des principes	Toute personne doit être en mesure de se plaindre du non-respect par une organisation des principes énoncés ci-dessus. La plainte doit être adressée au responsable de la conformité à la LPRPDE au sein de l'organisation concernée, en l'occurrence, le chef de la protection des renseignements personnels.

Reproduction tirée de « Principes relatifs à l'équité dans le traitement de l'information de la LPRPDE » par le Commissariat à la protection de la vie privée du Canada (2019).

relatif à la protection des renseignements personnels à la fois pour le secteur privé et pour le secteur public, de façon à ce qu'il soit plus adapté aux réalités technologiques d'aujourd'hui et mieux aligné avec le cadre législatif international. La réforme législative proposée repose sur deux grands principes : redonner aux citoyens le plein contrôle de leurs renseignements personnels et responsabiliser les organisations qui collectent de tels renseignements (Du Perron, 2020a).

Le projet de loi n° 64, s'il entre en vigueur, introduira un changement important dans la manière dont les données personnelles sont définies. En effet, ce nouveau projet de loi introduit une distinction novatrice entre la « dépersonnalisation » et « l'anonymisation ». Le projet de loi n° 64 considère qu'un renseignement personnel est dépersonnalisé au moment où il ne permet plus d'identifier *directement* une personne (Du Perron, 2020b). Toutefois, les données dépersonnalisées comportent toujours le risque d'identifier un individu *indirectement* grâce aux techniques avancées d'analyse de données qui sont largement disponibles aujourd'hui. Ces renseignements demeurent donc assujettis à la législation (Du Perron, 2020b; Rocher, Hendrickx et de Montjoye, 2019). Un renseignement personnel est considéré comme *anonymisé* seulement lorsqu'il ne permet plus d'identifier, directement ou indirectement, une personne, et ce, de façon irréversible (Projet de loi n° 64, article 111). Ainsi, le projet de loi n° 64 prévoit que seul un renseignement anonymisé peut se libérer du droit à la protection des renseignements personnels, une situation analogue à ce que prévoit le Règlement (UE) 2016/679 (le Règlement général sur la protection des données [RGPD]) (Du Perron, 2020b).

Le paysage législatif en matière de gouvernance des données pourrait continuer d'évoluer considérablement dans les années à venir et pas seulement au Québec. En réponse à la législation plus stricte en matière de protection de la vie privée introduite dans d'autres juridictions, notamment en Europe, le gouvernement fédéral canadien revoit actuellement ses propres lois afin de déterminer comment elles peuvent mieux répondre aux opportunités et aux défis d'une société numérique.

De plus, certaines lois internationales peuvent également avoir un impact sur des partenariats de données numériques locaux. Prenons par exemple le RGPD adopté par l'Union européenne en 2016, qui a une portée extraterritoriale. Il s'applique aux organisations non communautaires qui offrent des biens ou des services ou collectent des données sur les résidents de l'Union européenne (et non seulement ses citoyens). Par exemple, une université canadienne qui recrute des étudiants étrangers de l'UE peut être soumise au RGPD dans la mesure où elle traite les informations personnelles des étudiants (Commissaire à l'information et à la protection de la vie privée de l'Ontario, 2018). Des changements juridiques tant au Canada qu'à l'international pourraient ainsi avoir des impacts importants sur les organisations désireuses de valoriser leurs données.

La propriété intellectuelle

Les données sont devenues une source de valeur incontournable dans la nouvelle économie numérique. Les partenariats en matière de données se heurteront presque certainement à la question de savoir qui possède les données et ce que cette propriété implique (The British Academy et The Royal Society, 2018). Les droits de propriété sur les données, comme le note Teresa Scassa (2018a), « fournissent un outil de contrôle puissant [traduction libre] » (p.2). Tout comme une organisation peut souhaiter posséder ses données afin de les commercialiser, un gouvernement peut faire valoir son droit de propriété sur ses données pour en tirer des revenus ou, à l'inverse, les rendre disponibles en tant que données ouvertes. Tandis que du côté du public, un nombre grandissant de voix s'élèvent en faveur de nouveaux

ENCADRÉ 8 : PROPRIÉTAIRE VS DÉTENTEUR DES DONNÉES

Comme l'Office québécois de la langue française (OQLF) l'explique, le terme « propriétaire » dans l'expression « propriétaire d'une donnée ou d'un fichier » fait généralement référence à une « personne nommément désignée, responsable de la gestion et de la protection d'un ou de plusieurs fichiers informatiques et habilitée à prendre toute décision concernant ce ou ces fichiers, en vue d'assurer leur intégrité et leur confidentialité ». Donc, le terme propriétaire ne fait donc pas nécessairement référence à la personne qui est titulaire du droit de propriété sur la donnée ou le fichier selon le cadre juridique en vigueur.

L'OQLF ajoute d'ailleurs que « [c]'est pour cette raison que son emploi est parfois contesté et que l'on parle également, dans le milieu, de *détenteur de fichier*. Malgré ces réserves, *propriétaire* fait l'unanimité et est solidement implanté. Il s'agit d'un usage spécifique au milieu de la sécurité informatique ».

Afin de bien distinguer les deux concepts, nous utiliserons le terme « détenteur » en référence aux individus qui ont la responsabilité de gérer les données et « propriétaire » pour ceux qui détiennent des droits de propriété sur les données.

droits numériques qui reconnaissent les droits de propriété des citoyens sur leurs renseignements personnels (Bass et coll., 2018; Mozilla Insights, 2020).

Bien qu'il existe plusieurs sources de droits de propriété en vertu du droit canadien, dont la *Loi sur le droit d'auteur*, la *Loi sur les brevets* et la *Loi sur les marques de commerce*, il n'est pas clair si les données en général sont soumises à des droits de propriété. Les données sont différentes des autres types de biens à plusieurs égards qui affectent leur capacité à être possédées. Premièrement, elles sont non rivales, ce qui signifie que la distribution et le partage de données n'en diminuent pas la quantité. Ainsi, le créateur original d'une donnée peut en donner une copie exacte à une autre partie sans perdre aucune partie de l'original. Deuxièmement, certaines données peuvent concerner plusieurs personnes à la fois. Considérons, par exemple, les informations génétiques. Celles-ci concernent non seulement un individu, mais également toute sa famille — ce qui rend difficile pour un individu d'en revendiquer la propriété exclusive. Au Canada, les tribunaux ont jugé que les individus peuvent avoir le droit d'accéder à leurs informations personnelles et de les corriger, mais ces droits ne vont pas jusqu'à la reconnaissance d'un droit de propriété.

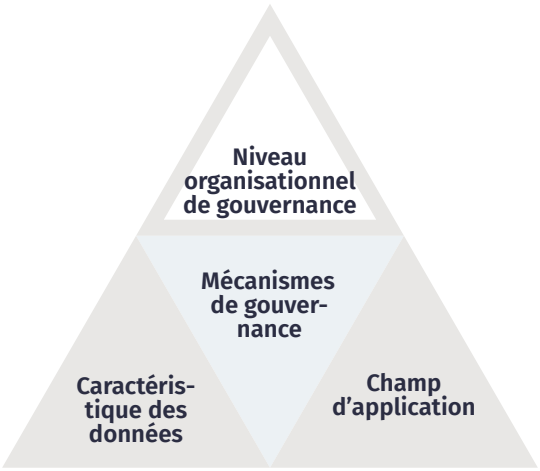
La législation actuelle sur les droits d'auteur établit une distinction entre les « faits » ou « idées » dans l'abstrait (qui ne sont pas protégés) et l'expression originale des idées (qui sont protégées). Ainsi, une compilation de faits, y compris un ensemble de données auquel de nouvelles informations sont continuellement ajoutées, peut ne pas être couverte par la protection sur le droit d'auteur, malgré l'acte de curation qu'elle requiert. Lorsqu'il s'agit de données déduites ou dérivées par le biais de l'analyse et du traitement de vastes quantités de données, cela dépend si les données sont considérées comme des faits (en quel cas elles sont non protégées) ou des expressions originales d'idées (elles sont alors protégées) (Scassa, 2018a). Les données sont donc intimement liées aux systèmes qui les ont produites, ce qui rend difficile toute généralisation sur la question de la propriété des données (Scassa, 2018a).

En conclusion, au vu de la complexité du paysage juridique actuel et du fait que les règles et leurs interprétations sont encore en transformation dans le secteur du numérique, il est préférable que les partenaires d'un partenariat de données aient recours à des experts pour vérifier qu'ils respectent les lois en vigueur ou pour rédiger des accords clairs où la propriété et le contrôle des données sont clairement spécifié.

LE PÉRIMÈTRE DE LA GOUVERNANCE DES DONNÉES

La gouvernance des données est définie par un périmètre qui consiste en trois éléments selon Abraham, Schneider et vom Brocke 2019 (figure 1) : 1) le niveau de gouvernance, 2) les caractéristiques des données, et 3) le champ d'application de la gouvernance.

Le niveau de gouvernance : Qui participe aux partenariats de données numériques ?



EXTRAIT FIGURE 1

Le **niveau (organisationnel) de gouvernance** fait référence à l'unité de référence à laquelle s'applique la gouvernance des données. L'unité de référence peut consister en une seule organisation, plusieurs organisations, tel que dans les partenariats de données numériques, voire même se situer à l'échelle d'un écosystème. C'est pourquoi Abraham, Schneider et vom Brocke distinguent entre les niveaux intraorganisationnel (une seule entité) et interorganisationnel (partenariat et écosystème). Le niveau de la gouvernance est donc déterminé par les acteurs engagés dans la gouvernance et influence en retour l'organisation de la gouvernance, les interactions qu'elle génère et les mécanismes de gouvernance des données privilégiés.

ENCADRÉ 9 : PRINCIPAUX ACTEURS ET LEURS RÔLES

Les partenariats des données sont organisés autour d'une variété de relations et de flux de données numériques entre différents acteurs. Par conséquent, la définition du rôle des acteurs qui participent à ces initiatives est un complément utile au cadre de gouvernance des données développé par Abraham, Schneider et vom Brocke. Un aperçu des principaux rôles pouvant être présents dans un partenariat est présenté ci-dessous (OCDE, 2019 ; IMDA et PDPC, 2019) [traduction libre].

- **Détenteurs des données.** Ils ont la compétence requise pour décider comment les données sont utilisées et partagées. Ils sont parfois appelés « propriétaires des données » même s'ils n'ont aucun droit de propriété juridique concernant les données qu'ils détiennent. Ces détenteurs de données choisissent de partager leurs données ou de les enrichir avec d'autres acteurs.
- **Utilisateurs des données.** Ces acteurs sont les destinataires des données qui font l'objet du partenariat. Ils s'efforcent généralement de produire de la valeur en analysant et traitant les données partagées de manière à les transformer en information utile.
- **Agents de soutien.** Ces entités offrent les moyens et l'assistance nécessaires pour partager les données. Leurs responsabilités peuvent notamment comprendre l'apport d'expertise en matière de préparation et d'analyse des données, d'assistance juridique, de gestion des risques, de financement de l'initiative, et même la conduite d'activités de renforcement des capacités.
- **Organes directeurs.** Un ou plusieurs organes peuvent être habilités à exercer une fonction de direction, de coordination, de supervision et de contrôle de la conformité de l'initiative ou du cadre de gouvernance des données mis en place.
- **Autorités et organismes de réglementation.** Ces acteurs influencent les partenariats de données en créant des lois et des normes en matière d'utilisation et de protection des données, et en publiant des directives ou des codes de pratique. Ces autorités peuvent comprendre des gouvernements, des associations et des institutions.
- **Bénéficiaires.** Il s'agit des entités qui tirent des avantages ou des bénéfices générés par le partenariat de données. Par exemple, les bénéficiaires peuvent être des citoyens, des communautés ou des groupes, entre autres.

Il convient de remarquer que ces rôles ne s'excluent pas mutuellement et que les activités qui leur sont associées peuvent se chevaucher.

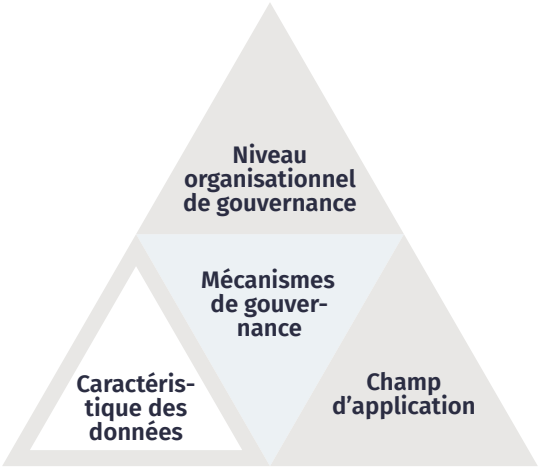
Il est utile de s'intéresser aux rôles joués par les différentes parties prenantes d'un partenariat de données numériques. Selon les objectifs poursuivis, les besoins identifiés et les capacités des partenaires, ceux-ci peuvent être appelés à jouer des rôles différents (voir encadré 9). Par exemple, une organisation pourra décider de se joindre à un partenariat afin de faire bénéficier ses partenaires de ses importantes bases de données. À l'inverse, une autre organisation pourrait n'avoir aucune donnée en sa possession, mais pouvoir offrir des capacités techniques d'analyse et de traitement.

Traditionnellement, le niveau organisationnel auquel s'exerce la gouvernance des données est ***intra*organisationnel**, c'est-à-dire restreint à une organisation unique. Un objectif courant de la gouvernance des données au sein d'une entreprise, par exemple, consiste à accroître l'efficacité opérationnelle en définissant et en mettant en œuvre des principes, des politiques et des pratiques (tous des mécanismes de gouvernance) concernant, par exemple, le stockage et la qualité des données. Cela peut aider l'organisation à améliorer la cohérence entre ses unités, à réduire les redondances, à améliorer les recherches dans les données et l'accès aux résultats, et donc de fonctionner plus efficacement. Une même organisation peut regrouper tous les types d'acteurs sous un même toit : détenteurs de données, utilisateurs, bénéficiaires, agents de soutien et organes directeurs.

Dans le cadre des partenariats de données numériques, le niveau de la gouvernance des données est alors ***inter*organisationnel**. Ce type de partenariats exige des mécanismes structurels qui alignent les parties intéressées sur des objectifs communs et structurent leur participation, mais également des mécanismes procéduraux visant à standardiser et à coordonner les pratiques de gestion des données à l'échelle du groupe, à améliorer l'accès aux données ou encore à réduire les risques. Les objectifs visés dans le cadre des partenariats de données numériques consistent par exemple à développer des usages novateurs de l'information ou à résoudre des problèmes au moyen d'efforts interdisciplinaires ou intersectoriels de collaboration et de réflexion.

Le niveau de gouvernance des données peut même s'étendre à **l'échelle d'un écosystème**. Les municipalités, en particulier, engagées dans divers projets de villes intelligentes, se retrouvent notamment au cœur de ce type de gouvernance, cherchant à établir des cadres de gouvernance globaux où s'imbriquent des processus décisionnels pouvant influencer la collecte, l'utilisation et le partage des données de leur administration, de leurs citoyens, de leurs partenaires et d'acteurs actifs à l'échelle de tout leur territoire. La mise en place de coalitions (Cities Coalition for Digital Rights), l'adoption de déclarations (Déclaration de Montréal IA responsable) ou de chartes énonçant des principes fondateurs (Charte de la donnée métropolitaine par la Ville de Nantes, France) ou encore l'établissement de liens transnationaux avec des mouvements actifs dans d'autres juridictions sont tous de bons exemples de mécanismes de gouvernance de portée écosystémique.

Les caractéristiques des données : Quels types de données sont partagées ?



EXTRAIT FIGURE 1

La gouvernance des données doit être adaptée aux **caractéristiques des données** auxquelles elle s'applique. En effet, selon Abraham, Schneider et vom Brocke, le type de données doit être déterminé avec précision puisqu'il conditionne en grande partie les mécanismes de la gouvernance des données, voire même le cadre réglementaire applicable. Les spécificités des données, telles que leur origine, par exemple, peuvent soulever des enjeux complètement distincts.

ENCADRÉ 10 : QU'EST-CE QU'UNE DONNÉE NUMÉRIQUE ?

Les données sont la représentation des faits sous forme de texte, de chiffres, d'images, de son ou de vidéo (Earley et coll., 2017, p.19). Aujourd'hui, lorsque nous parlons de données numériques, nous faisons souvent référence à celles « codée[s] dans un format permettant son traitement par ordinateur » (Office québécois de la langue française, 2004). Il est important de noter que les données ne sont pas simplement des faits sur le monde, mais des interprétations de faits qui nécessitent un contexte pour être significatives (Earley et coll., 2017, p.19). Les mêmes informations sous-jacentes peuvent être représentées de différentes manières selon l'objectif pour lequel elles doivent être utilisées.

Il n'existe toutefois aucune taxonomie dominante établissant des types précis de données ou encore pour en classifier les caractéristiques. Ainsi, Abraham, Schneider et vom Brocke font uniquement la distinction entre *données traditionnelles* (telles que les données transactionnelles et les données de référence de l'entreprise) et *données massives* (telles que les données du Web et des médias sociaux). Alors que la gouvernance des données traditionnelles vise principalement à en garantir l'utilisation cohérente à travers l'organisation, celle des données massives vise plutôt à soutenir l'innovation tout en réduisant les risques (Abraham, Schneider et vom Brocke, 2019, p.431).

Les données peuvent être catégorisées de plusieurs autres manières : leur domaine (personnel, privé, public); leur degré d'ouverture (fermé, partagé ou ouvert); leur origine (données fournies librement, données observées); leur fonction (maître, référence, métadonnée, transactionnelle); leur degré de sensibilité (dépersonnalisé, anonymisé); leur format (textuel, numérique, multimédia); ou même leur sujet (données de mobilité, données sociales). Dans cette section, nous explorons de plus près quelques-unes de ces catégories afin d'aider les organisations qui se lancent dans un partenariat de données numériques à mieux cerner les caractéristiques des données qu'ils utilisent et comprendre les enjeux qu'ils soulèvent.

Le domaine des données : personnel, privé ou public

Les données peuvent être considérées comme appartenant à l'un des trois domaines suivants, qui déterminent les règles (c'est-à-dire les lois et règlements) concernant leur utilisation (OCDE, 2019) :

- Le **domaine personnel**, qui englobe toutes les données personnelles « relatives à une personne identifiée ou identifiable [traduction libre] » à l'égard de laquelle les sujets des données ont des intérêts en matière de protection. Ce domaine est généralement régi par des cadres de réglementation de la protection de la confidentialité des données.
- Le **domaine privé**, qui englobe toutes les données exclusives généralement protégées par des droits de propriété intellectuelle (dont les droits d'auteur et les secrets commerciaux) ou par des privilèges d'accès et de contrôle (prévus, par exemple, par le droit des contrats et le droit pénal appliqué à la cybercriminalité), et dont l'intérêt économique interdit généralement le partage.
- Le **domaine public**, qui englobe toutes les données non protégées par des droits de propriété intellectuelle ou tous autres droits ayant des effets similaires. Ainsi, le terme « domaine public » doit être interprété de façon plus large que l'ensemble des données non protégées par le droit d'auteur, et qui sont donc librement accessibles et réutilisables (en général, on considère que les données du domaine public doivent être accessibles comme des données ouvertes).

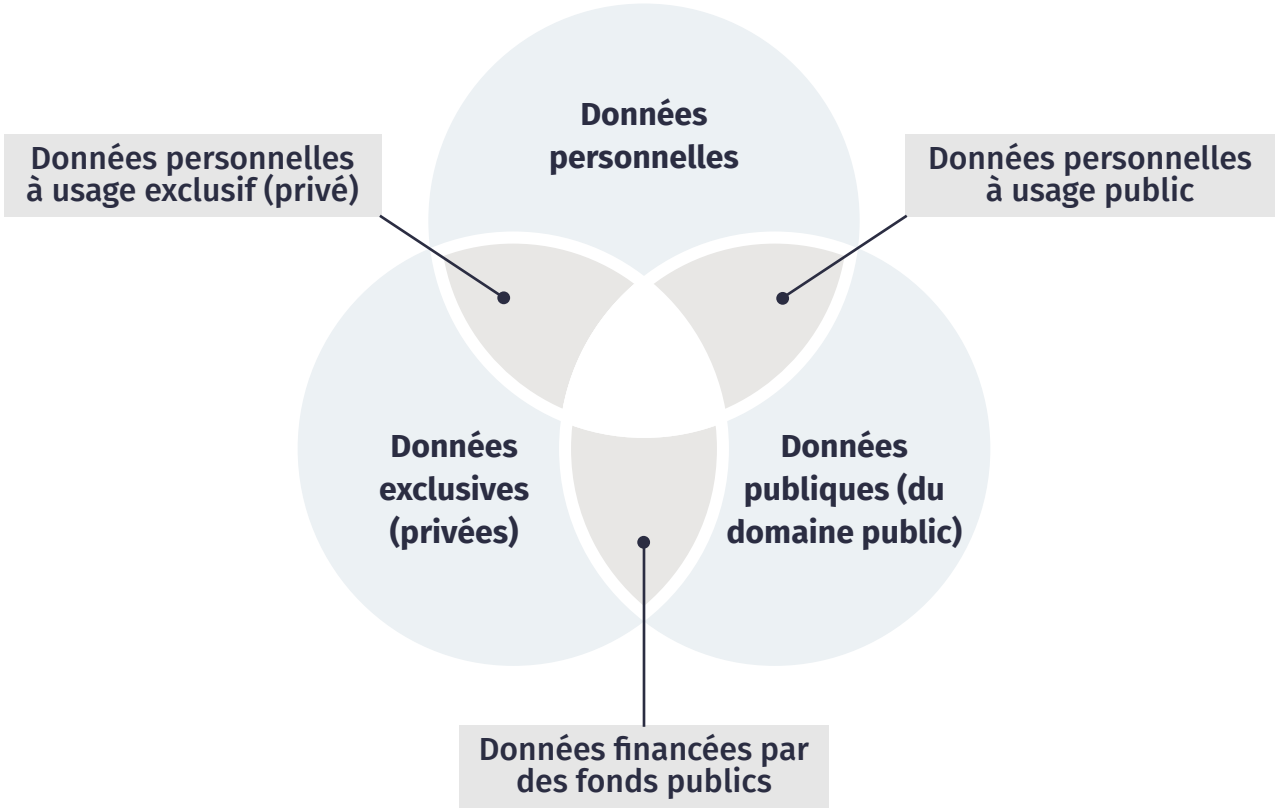


FIGURE 2 : DOMAINES DES DONNÉES

Reproduction tirée de « Enhancing Access to and Sharing of Data » par l'OCDE (2019). La reproduction de cette figure et sa traduction a été autorisée. La traduction n'a pas été créée par l'OCDE et ne constitue pas une traduction officielle de l'OCDE. L'OCDE ne peut être tenue responsable en cas d'erreur ou d'exactitude dans cette traduction.

Il est important de réaliser que les qualificatifs « public » et « privé » correspondent aux données; ils ne correspondent pas au type de l'organisation qui produit ou utilise les données correspondantes. Par exemple, de nombreuses organisations du secteur privé sont de plus en plus intéressées par la publication de leurs données comme des « données ouvertes », c'est-à-dire des données du domaine public. De même, un organisme du secteur public peut produire des données qu'il réserve à son usage, c'est-à-dire des données du domaine privé (OCDE, 2019).

Ces domaines peuvent se chevaucher, comme le montre la figure 2. Les *données financées par le secteur public* y constituent l'intersection des domaines public et privé. Ces données pourraient être créées, par exemple, dans le cadre de partenariats public-privé. Ces chevauchements pourraient être causés par des points de vue et des intérêts potentiellement contradictoires de certaines parties qui cherchent à valoriser leurs données en les partageant dans le cadre d'un partenariat (OCDE, 2019).

En général, il faudra toujours examiner attentivement chaque jeu de données dans son contexte pour déterminer les cadres législatif et réglementaire qui peuvent lui être appliqués. En effet, les cadres juridiques qui régissent ces domaines varient selon la juridiction applicable au partenariat, les caractéristiques des données concernées et la finalité de leur utilisation. Cette réglementation est généralement fondée sur les lois relatives à la protection de la vie privée et à la propriété intellectuelle discutées dans la section sur les conditions existantes ci-dessus.

ENCADRÉ 11 : LES DONNÉES PERSONNELLES

Les données personnelles sont inextricablement liées à la notion de vie privée. La vie privée est considérée comme un pilier de notre liberté et est un droit protégé par la constitution en vertu de la Charte canadienne des droits et libertés. Le droit individuel à la vie privée a été interprété dans la jurisprudence canadienne comme incluant des protections pour « la vie privée personnelle, la vie privée territoriale et la vie privée informationnelle » (*R. v. Jarvis*, 2019 CSC 10, [2019] 1 R.C.S. 488). La vie privée informationnelle est la plus pertinente pour les données personnelles, car elle fait référence au droit individuel de contrôler avec qui, combien et dans quel but les informations personnelles sont divulguées.

En vertu de ce cadre réglementaire, les renseignements personnels sont ceux qui portent sur une personne physique et permettent de l'identifier (La Commission d'accès à l'information du Québec, s.d.).

Selon la LPRPDE (LC 2000, c 5, s 2 [1]), les renseignements personnels comprennent « tout renseignement factuel ou subjectif, consigné ou non, concernant une personne identifiable » :

- l'âge, le nom, un numéro d'identification, le revenu, l'origine ethnique ou le groupe sanguin ;
- une opinion, une évaluation, un commentaire, le statut social ou une mesure disciplinaire ;
- le dossier d'un employé, un dossier de crédit ou de prêt, un dossier médical, l'existence d'un différend entre un consommateur et un commerçant ou le projet d'une personne (par exemple, l'intention d'acquérir des biens ou des services ou de changer d'emploi).

Certaines données, telles que les données nominatives, permettent facilement l'identification directe des personnes, parfois à partir d'une seule donnée. Même des données apparemment non personnelles, peuvent pourtant permettre l'identification indirecte, des individus, lors par exemple d'une mise en corrélation d'un ensemble de données et ainsi avoir un impact sur la vie privée.

Les données à caractère personnel bénéficient d'une plus grande protection que les données représentant des sujets environnementaux ou d'autres sujets non humains, car elles sont uniquement sujettes à un certain nombre de violations de la vie privée et de préjudices qui peuvent potentiellement résulter de diverses pratiques inappropriées (voir le tableau 2 à la page 57).

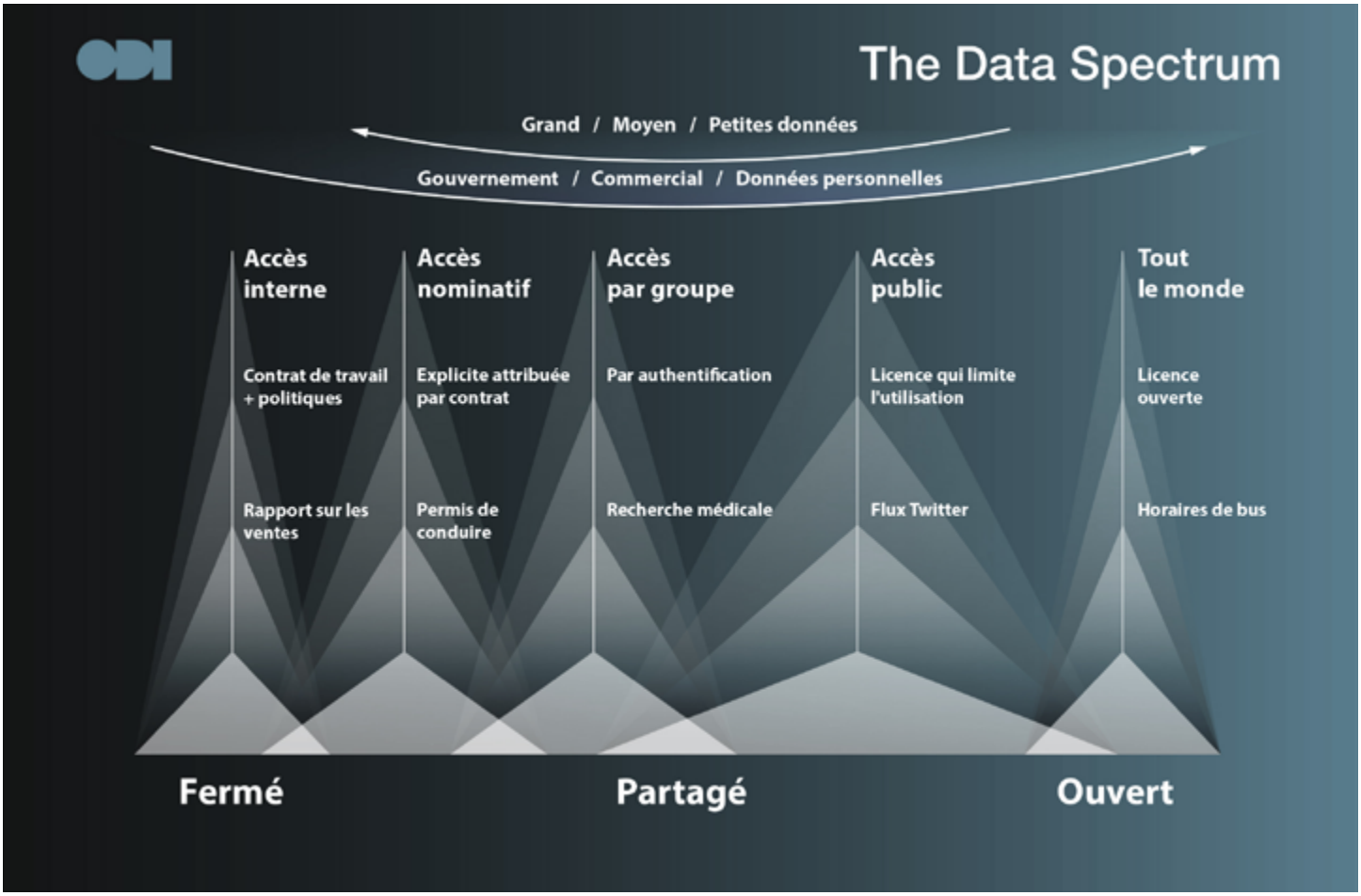


FIGURE 3 : LE SPECTRE DES DONNÉES

Reproduction tirée de « The Data Spectrum » par l'Open Data Institute (s.d.). Cette figure est soumise à une licence CC BY 4.0. Des modifications lui ont été apportées.

Le degré d'ouverture des données : fermé, partagé ou ouvert

Une autre caractéristique des données est leur degré d'ouverture. L'Open Data Institute (ODI) place les différents degrés d'ouverture d'un jeu de données sur une échelle allant de *fermé* à *ouvert* en passant par *partagé* (figure 3). Ce degré d'ouverture correspond au niveau d'accès aux données privilégié par une organisation. En effet, les organisations ont la possibilité d'octroyer certains droits d'utilisation, tout en se réservant d'autres droits (par le biais d'un contrat ou d'une licence Creative Commons, par exemple). Le cadre législatif qui s'applique à un domaine de données particulier peut également déterminer si un ensemble de données donné peut ou non être accessible et utilisable.

Du côté **Fermé** du spectre se trouvent les données que les organisations collectent et utilisent à l'interne, sans que des parties externes puissent y accéder. Les informations sensibles concernant les employés, les finances, les opérations ou les secrets commerciaux sont des exemples de données fermées. Du côté **Ouvert** du spectre se trouvent les données librement accessibles à tous, dont les données dans le domaine public ou les données sous licence ouverte. Entre les extrémités **Fermé** et **Ouvert** du spectre se trouve la zone **Partagé**, où se situent les données qui peuvent être partagées. Par opposition aux données ouvertes, ces données sont accessibles de façon *contrôlée* ou *restreinte*.

L'origine des données

Lorsqu'un partenariat de données numériques envisage de partager des données *personnelles*, il est essentiel d'examiner **l'origine** des données. Comme Abrams (2014) le fait remarquer, dans notre environnement numérique actuel « de plus en plus souvent, les données ne sont pas directement recueillies auprès de la personne, mais plutôt à distance, sans que celle-ci ait conscience de l'origine et des utilisations ultérieures de ces données [traduction libre] » (p.1). La façon d'aborder les problèmes de protection de la confidentialité des données pourrait donc dépendre du degré de sensibilisation de la personne qui recueille les données.

Une taxonomie des données est utile pour établir un lien entre l'origine des données et le degré de sensibilisation de la personne en matière de collecte et de traitement des données. Selon Abrams (2014, p.5), les quatre principales catégories taxonomiques sont les suivantes :

- Les **données fournies librement** sont des données qu'un individu partage activement et délibérément (par exemple, en créant un profil sur un réseau social ou fournissant des renseignements figurant sur sa carte de crédit pour faire des achats en ligne).
- Les **données observées** sont des données saisies et enregistrées concernant des activités. Les données de géolocalisation des téléphones mobiles et les données qui décrivent les comportements des utilisateurs du Web sont des exemples de données observées.
- Les **données dérivées** sont des données générées à partir d'autres données. Elles deviennent de nouveaux éléments de données concernant une personne particulière. La cote de crédit calculée au moyen de l'historique financier d'une personne constitue un exemple de donnée dérivée.
- Les **données déduites** sont générées par analyse ou mises en relation avec des données concernant une personne. La cote de crédit d'une personne, qui est obtenue à partir de son historique de paiement observé, est un exemple de donnée déduite.

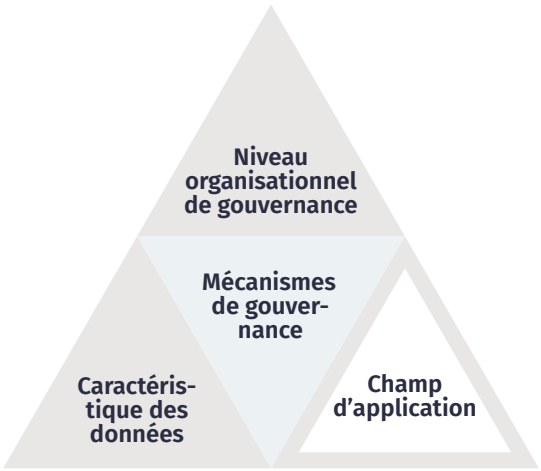
Dans le contexte des partenariats de données numériques, l'ajout d'une catégorie, comme le propose l'OCDE (2019), peut se révéler pertinent :

- Les **données acquises** (également appelées données achetées ou données sous licence) sont obtenues auprès de tiers (par exemple, des courtiers de données) au moyen de contrats de licence commerciale ou par des moyens non commerciaux (par exemple, dans le cadre d'initiatives gouvernementales ouvertes). En conséquence, les obligations contractuelles et les obligations légales peuvent influencer la réutilisation et le partage de ces données.

Des données sont généralement fournies librement lorsqu'une personne adhère à un service. Lorsqu'une personne accepte les modalités de service, on admet qu'elle autorise librement la collecte et le traitement de ses données personnelles. Dans les cas des quatre autres catégories, les personnes n'ont que peu ou pas de possibilités d'accorder un consentement valable, car elles pourraient ne pas savoir que des données sont recueillies sur elles.

Dans des circonstances précises où le consentement est impossible à obtenir, mais où il existe un fort intérêt du public pour la collecte et l'utilisation des données, des alternatives au consentement ou encore des mesures supplémentaires peuvent être nécessaires (Jones et coll., 2017a). En général, cette taxonomie des types de données nous offre un point de départ utile pour déterminer non seulement les données qui peuvent être partagées, mais aussi les mécanismes de gouvernance des données à appliquer pour faciliter la circulation des données entre parties en s'assurant que les personnes sont protégées et informées (Abrams, 2014).

Le champ d'application de la gouvernance



EXTRAIT FIGURE 1

Finalement, le dernier élément qui contribue à définir le périmètre de la gouvernance des données est son **champ d'application** (Abraham, Schneider et vom Brocke, 2019, p.431-32). Au quotidien, la gestion des données relève de différents champs. Certains visent à assurer la qualité des données, d'autres à mettre en place les infrastructures numériques requises à leur stockage, partage, etc. La gouvernance des données doit s'intéresser à l'ensemble des champs requis à l'atteinte de ses objectifs.

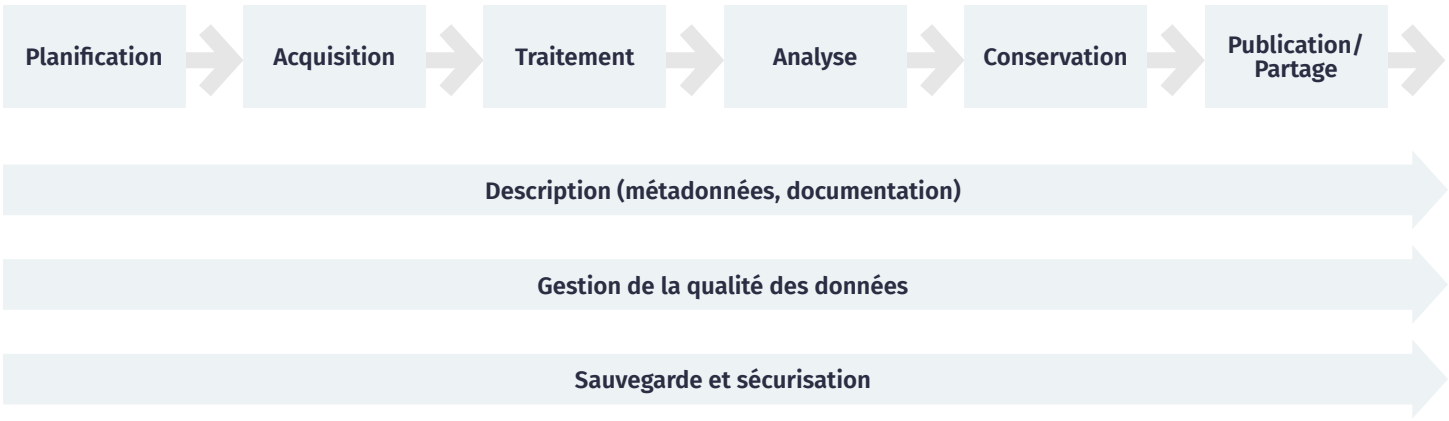
Abraham, Schneider et vom Brocke (2019) identifient six grands champs d'application :

- **La qualité des données :** mettre en place de techniques de gestion de la qualité pour mesurer, évaluer et améliorer la qualité des données afin qu'elles puissent être utilisées comme prévu dans un contexte donné ;
- **La sécurité des données :** contrôler les accès internes et externes, protéger la vie privée et assurer l'authenticité, la disponibilité, la confidentialité, l'intégrité et la fiabilité des données ;
- **L'architecture des données :** développer et maintenir le modèle et le plan de gestion des données d'entreprise ainsi que les politiques, directives et standards à suivre ;
- **Les métadonnées :** documenter et classifier les données, les flux et les modèles et toutes autres informations essentielles à la compréhension des données et des systèmes à travers lesquels elles sont créées, maintenues et accédées.
- **L'infrastructure de stockage des données :** offrir les capacités matérielle et logicielle pour répondre aux besoins en matière de fonctionnalités, fiabilité, capacité, etc.
- **Le cycle de vie des données :** instaurer des processus et des procédures qui déterminent ce qu'il advient des données de leur collecte à leur suppression.

Le cycle de vie des données est un cas intéressant, car il peut également être utilisé comme un cadre de gestion. Habituellement, le cycle de vie est présenté sous forme de schéma pour aider à visualiser les différentes phases traversées par les données et les transformations qu'elles subissent tout au long de leur vie utile. Dans ce schéma, il est possible d'intégrer l'ensemble des autres champs d'application de manière à visualiser et comprendre les différents aspects que la gouvernance des données peut être appelée à régir.

La figure ci-dessous présente le modèle de cycle de vie des données scientifiques utilisé par l'agence *United States Geological Survey*, développé par Faundeen et coll. (2014, p. 2). Ce schéma représente un bon exemple d'un cycle de vie intégrant les principaux champs d'application de la gouvernance des données.

FIGURE 4 : MODÈLE DE CYCLE DE VIE DES DONNÉES SCIENTIFIQUES



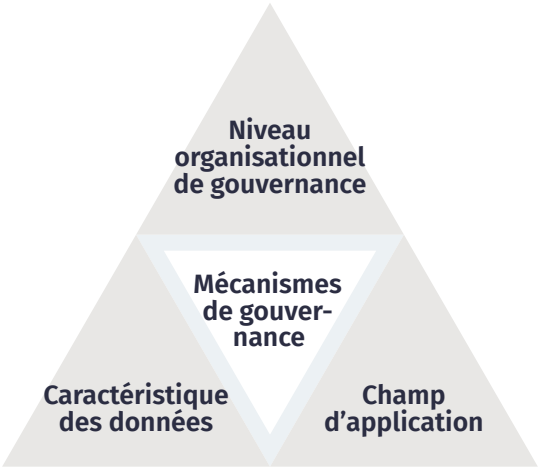
Reproduction tirée de « The United States Geological Survey Science Data Lifecycle Model » par Faundeen et coll. (2014, p. 2).

LES MÉCANISMES DE GOUVERNANCE DES DONNÉES

Ensemble, le niveau organisationnel, le champ d'application ainsi que les caractéristiques des données forment le périmètre de la gouvernance des données. Ce périmètre circonscrit et détermine en retour les mécanismes de gouvernance des données qui doivent être déployés.

Selon Abraham, Schneider et vom Brocke, les mécanismes de gouvernance consistent en la *dimension fondamentale* de la gouvernance des données. Les auteurs les regroupent en trois catégories (2019, pp.428-430) :

- Les **mécanismes structurels** déterminent les organes décisionnels, les structures hiérarchiques et les responsabilités en place. Ils définissent les rôles et les responsabilités et attribuent le pouvoir décisionnel.
- Les **mécanismes procéduraux** régissent les politiques, les normes, les processus, les procédures, les contrats, les mesures de performance, le contrôle de la conformité, les stratégies en matière de données et la gestion des problèmes. Cela garantit que les données sont enregistrées avec précision, conservées en toute sécurité, utilisées efficacement et partagées de façon appropriée.
- Les **mécanismes relationnels** sont les différentes pratiques qui facilitent la collaboration entre les parties intéressées. Ils comprennent la communication, la formation, la coordination et la prise de décisions.



EXTRAIT FIGURE 1

Ces mécanismes sont la manière dont la gouvernance s'opérationnalise et est mise en pratique dans le quotidien des organisations engagées dans les partenariats de données numériques. C'est pourquoi l'ensemble du chapitre suivant sera consacré à l'exploration d'exemples concrets de mécanismes de gouvernance des données.

Dans le chapitre suivant, nous examinerons les principaux principes et mécanismes que les parties chercheront à atteindre et à mettre en œuvre dans un partenariat de données numériques. Ensuite, nous présentons les résultats des entretiens, qui permettent de mieux comprendre comment les individus de l'écosystème de Montréal sont actuellement confrontés avec certains risques, tensions et enjeux en gouvernance de données dans leur propre travail et dans leurs organisations.


HealthMap | Novel Coronavirus (COVID-19)
1,436,794
North Atlantic Ocean
South Atlantic Ocean
Indian Ocean
South Pacific Ocean

CHAPITRE 3

TROIS PRINCIPES PHARES POUR UNE GOUVERNANCE DES DONNÉES DANS L'INTÉRÊT DU PUBLIC ET LEUR MISE EN OEUVRE PRATIQUE

Comme nous l'avons vu précédemment, la gouvernance des données comporte différentes facettes qui s'entre-déterminent. Le cadre juridique en vigueur, les organisations impliquées, leurs objectifs et leur culture d'entreprise et le type de données qu'elles partagent, etc. influenceront tous le cadre de la gouvernance dont elles devront collectivement se doter pour mener un partenariat de données numériques fructueux. Mais concrètement en quoi consiste la gouvernance des données ?

Le présent chapitre sera consacré à l'exploration de plusieurs grandes catégories de mécanismes de gouvernance à travers lesquels s'incarne la gouvernance des données au quotidien.

Avant d'explorer plus en détail les différentes manières d'opérationnaliser la gouvernance, il nous faut d'abord mettre en évidence trois grands principes sous lesquels nous avons décidé de regrouper les mécanismes de gouvernance discutés. Grâce à ces principes, les organisations sont en mesure de mieux orienter les choix qu'elles font en matière de gouvernance vers des finalités moralement et socialement désirables. Ce sont la responsabilité, l'efficacité et l'imputabilité.

Ces **principes** se traduisent en objectifs dont l'impact influencera l'ensemble de la gouvernance des données.

- **Responsabilité : Valoriser les données de manière responsable et éthique**
- **Efficacité : Gérer les données de manière efficace et cohérente**
- **Imputabilité : Évaluer en continu la conformité et l'impact**

Ces principes se traduisent en objectifs dont l'impact influencera l'ensemble de la gouvernance des données.

Responsabilité : **Valoriser les données de manière responsable et éthique**

Efficacité : **Gérer les données de manière efficace et cohérente**

Imputabilité : **Évaluer en continu la conformité et l'impact**

RESPONSABILITÉ

D'abord, pour préserver la confiance du public et la légitimité de leurs initiatives, les partenariats de données numériques doivent impérativement mettre en place toutes les mesures nécessaires pour **valoriser les données de manière responsable et éthique**. Cela commence par la nécessité de respecter les lois et de protéger les droits des citoyens. Ensuite, ce principe reconnaît également que dans notre environnement numérique actuel, où les lois ne suivent pas toujours le rythme de la collecte de données et l'émergence de nouvelles technologies, de sorte que les organisations ont un devoir de vigilance supplémentaire, où elles doivent évaluer les risques possibles et minimiser les impacts préjudiciables.

EFFICACITÉ

Ensuite, dans un esprit d'inclusion et de renforcement de l'autonomie du public, les partenariats de données numériques doivent mettre en place des mécanismes de gouvernance qui favorisent une **gestion efficace et cohérente des données**. La recherche de la qualité, de l'interopérabilité et de l'accessibilité des données permet en effet de partager, relier et combiner les données avec plus de facilité, ce qui en retour stimule l'innovation et génère de nouvelles opportunités pouvant bénéficier au public.

IMPUTABILITÉ

Finalement, l'imputabilité est une caractéristique incontournable d'un partenariat de données numériques dans l'intérêt public. Pour ce faire, l'initiative doit mettre en place des mécanismes capables d'**évaluer la conformité et l'impact** de ses décisions tout au long du cycle de vie des données.

Nous croyons que ces trois grands principes permettent d'approcher la gouvernance des données de manière globale en reconnaissant qu'elle s'inscrit dans le temps et dans la pratique en plus d'incarner une certaine vision éthique.

Ce chapitre est donc divisé en trois parties, chacune traitant un de ces principes. Chaque section commence par une mise en contexte où nous expliquons l'importance du principe et de l'objectif recherché et décrivons les principaux risques, enjeux et défis qui y sont associés. Ensuite, nous illustrons la manière dont ces objectifs peuvent être atteints par le biais de divers mécanismes de gouvernance.

Nous donnerons des exemples tirés des trois catégories de mécanismes : structurel, procédural et relationnel. Comme nous le verrons, ces catégories ne sont pas exclusives. Plusieurs mécanismes peuvent se chevaucher ou se compléter afin d'atteindre un même objectif. De plus, notre sélection n'a pas la prétention d'être exhaustive. Des mécanismes différents ou encore d'autres natures pourraient être tout aussi utiles ou efficaces selon les circonstances du partenariat de données numériques. Nos choix aspirent néanmoins à répondre aux principales préoccupations soulevées par les citoyens et les organisations ainsi que dans la littérature.

Finalement, il est important de noter que nous ne discuterons pas de la dimension juridique des différents mécanismes proposés ici, puisque, comme nous l'avons indiqué dans le chapitre précédent, les règles et lois applicables aux partenariats de données numériques varient en fonction de leur industrie, des caractéristiques des données et de leurs champs d'application. Ainsi, certaines catégories de mécanismes décrites ci-dessous sont parfois exigées par la loi, d'autres sont simplement volontaires. Par exemple, le consentement éclairé peut être une exigence stricte et formelle lors de la collecte de données personnelles par une entreprise privée. Néanmoins, même en l'absence d'une telle exigence, la collecte de données non personnelles peut bénéficier de l'obtention d'un consentement éclairé, puisque même les données qui ne représentent pas les individus peuvent être utilisées pour prendre des décisions qui affectent la vie des personnes (Earley et coll., 2017). Il va de soi que la conformité aux lois est une condition *sine qua non* de tout partenariat de données responsable et éthique.

RESPONSABILITÉ : VALORISER LES DONNÉES DE MANIÈRE RESPONSABLE ET ÉTHIQUE

Les administrations publiques, les organisations privées et les citoyens produisent et utilisent collectivement une quantité grandissante de données numériques dans l'intérêt de déceler de nouvelles sources de création de valeur et de générer de nouvelles connaissances susceptibles d'informer la prise de décision. Les données personnelles sont commodités courantes, les données publiques ouvertes se multiplient et les données massives (*big data*) sont exploitées dans tous les secteurs de l'économie. De plus en plus fréquemment, ces données font l'objet d'analyse et de croisements automatiques et algorithmiques avancés, dont les résultats sont désormais intégrés dans les processus décisionnels publics. Données et algorithmes offrent de vastes et intéressantes possibilités de générer des informations susceptibles d'améliorer la vie des citoyens ; toutefois, ces possibilités s'accompagnent également de défis éthiques importants.

Les préjudices qui peuvent découler d'une utilisation négligente des données sont désormais bien documentés : atteinte à la vie privée, perte d'autonomie et d'agentivité individuelles, présence de biais interprétatif et de discrimination (Shamsi et Khojaye, 2018 ; Järvinen et coll., 2014 ; Peña Gangadharan et Niklas, 2019).

Les conséquences de ces préjudices sur les individus sont réelles, ils peuvent en effet être lésés physiquement, émotionnellement ou financièrement. Lorsque des données identifiables sont divulguées dans des contextes délicats, par exemple, cela peut déclencher des violences, de la discrimination ou des politiques d'exclusion. Des groupes entiers peuvent également être lésés, et ce, sans même que les individus soient identifiés, par le biais de la mise en place de politiques discriminatoires sur la base de données de mauvaise qualité ou de relations faussement perçues par exemple (The Engine Room, 2016).

Daniel Solove (2006) a créé une liste des préjudices potentiels qui peuvent affecter les individus, groupes ou communautés en cas d'atteinte à leur vie privée (veuillez voir tableau 2). Elle démontre l'ampleur des risques encourus et la nécessité de mettre en œuvre des mesures pour traiter et manipuler les données de manière sécuritaire, sur la base des principes éthiques solides en plus des limites établies par les cadres législatifs.

L'adoption du principe de responsabilité pour orienter la gouvernance du partenariat de données numériques vise donc à s'assurer que toute valorisation des données est effectuée de manière à prévenir et empêcher de tels préjudices tout en anticipant et minimisant les risques que des impacts néfastes non prévus au départ surgissent en cours de route. Bien sûr, une utilisation responsable et éthique des données commence avant tout par le respect des lois et la protection des droits des citoyens.

Nous proposons ici d'explorer cinq types de mécanismes de gouvernance visant à enchâsser la responsabilité dans la gouvernance des données. Premièrement, l'adoption d'une **déclaration de principes** pour définir la vision éthique du projet, ensuite le **consentement**, pour reconnaître le droit de l'individu à décider s'il partage ou non ses propres données, puis les **options de recours** et l'**anonymisation** afin de protéger les droits et la vie privée des individus, la confidentialité programmée, et finalement l'**évaluation des risques** pour pallier au flou juridique.

TABLEAU 2 : UNE TAXONOMIE DES ATTEINTES À LA VIE PRIVÉE CRÉÉE PAR DANIEL SOLOVE (2006)

Domaine	Atteinte à la vie privée	Description
Collecte d'information	Surveillance	Observer une personne, l'écouter ou enregistrer ses activités
	Interrogation	Diverses formes d'interrogation ou de recherche d'information
Traitement de l'information	Agrégation	Combinaison de diverses données concernant une personne
	Identification	Associer de l'information à des personnes particulières
	Insécurité	Négligence en matière de protection des données stockées contre les fuites et les accès abusifs.
	Utilisation secondaire	Utilisation des données dans un but différent du but prévu sans le consentement du sujet des données.
	Exclusion	Ne pas permettre au sujet des données d'accéder à l'information le concernant que d'autres personnes détiennent, de participer au traitement et à l'utilisation de cette information et de corriger les erreurs qui y figurent.
Diffusion de l'information	Violation de la confidentialité	Violation d'une promesse de protection de renseignements personnels
	Divulgation	Révélation de renseignements au sujet d'une personne qui ont des effets sur la façon dont le caractère de cette personne est jugé
	Exposition	Révélation de la nudité, du chagrin ou des fonctions corporelles d'une autre personne
	Accessibilité accrue	Augmentation de l'accessibilité à l'information
	Chantage	Menace de divulgation de renseignements personnels
	Appropriation	Utilisation de l'identité du sujet des données pour servir les objectifs et les intérêts d'une autre
	Distorsion	Diffusion de renseignements faux ou trompeurs au sujet de personnes
Ingérence	Intrusion	Actes invasifs qui perturbent la tranquillité ou la solitude d'une personne
	Interférence décisionnelle	Intervention dans la prise de décisions du sujet des données à propos de ses intérêts privés

Mécanismes de gouvernance

La déclaration de principes

L'adoption d'un ensemble de principes par les différentes parties prenantes impliquées dans le partenariat de données numériques, que ce soit sous la forme d'une déclaration, d'un manifeste, d'une charte de projet ou d'une entente de partage de données constitue un moyen efficace pour articuler et démontrer l'adhésion à un ensemble de valeurs ou de positions sur la gouvernance des données (Coutts et Gagnon-Turcotte, 2020).

Par exemple, bien qu'elles n'aient aucun statut juridique ni aucune force, les déclarations de principe telles que la Charte numérique du Canada ou la toute récente Charte des données numériques de la ville de Montréal démontrent l'engagement public à défendre les droits numériques existants des citoyens et à introduire de nouveaux droits. Dans le domaine des données, il existe plusieurs déclarations importantes qui permettent de guider l'action des acteurs du numérique, dont notamment en données ouvertes (la Charte internationale des données ouvertes) et en intelligence artificielle (la Déclaration de Montréal pour un développement responsable de l'intelligence artificielle). L'adhésion à de tels textes ou l'adoption d'une charte interne ou à tout le moins de principes et de valeurs peuvent renforcer la légitimité du projet et la confiance entre les partenaires et favoriser l'adéquation de leurs intérêts.

En outre, la mise en œuvre d'exercices d'engagement des parties prenantes, voire de participation publique, lors de l'élaboration de telles déclarations de principes favorise à la prise en considération des besoins et des préoccupations d'un public élargi, ce qui peut mener à une plus grande acceptabilité sociale du projet.

Toutefois, il est nécessaire de souligner que la conformité volontaire à des valeurs et des normes communes en matière de gouvernance des données peut s'avérer insuffisante si elle n'est pas intégrée dans un cadre réglementaire et de gouvernance plus large (Bennett et Raab, 2018).

Le consentement éclairé

Le consentement éclairé est un mécanisme majeur pour la protection de la vie privée (Groupe des politiques et de la recherche du Commissariat à la protection de la vie privée du Canada, 2016). En effet, « le consentement représente un moyen pour les individus de protéger leur vie privée en exerçant un contrôle sur leurs renseignements personnels—il s'agit de déterminer quels renseignements personnels les organisations peuvent recueillir, comment elles peuvent les utiliser et à qui elles peuvent les communiquer » (Groupe des politiques et de la recherche du Commissariat à la protection de la vie privée du Canada, 2016). La recherche du consentement éclairé reconnaît également le droit des individus à décider si et quelles données ils partagent.

ENCADRÉ 12 : LA PROTECTION DE LA VIE PRIVÉE

L'Office québécois de la langue française (1999) définit la protection de la vie privée comme, « la mise en vigueur d'un ensemble de mesures administratives, techniques et physiques visant à prévenir les intrusions dans la vie privée des personnes ou dans les affaires privées des personnes et des organisations, lesquelles intrusions découlent spécifiquement de la collecte, du traitement, de la dissémination et de la divulgation d'informations ayant trait à ces personnes ou à ces organisations ».

Pour accorder un consentement éclairé, les personnes concernées doivent comprendre la nature, le but et les conséquences du partage de leurs renseignements personnels. Selon les *Lignes directrices pour l'obtention d'un consentement valable* publié par le Commissariat à la protection de la vie privée du Canada, les organisations qui souhaitent obtenir un consentement éclairé doivent s'assurer que les individus comprennent les éléments clés qui pourraient avoir une incidence sur leur décision. Ainsi, lorsque les individus prennent des décisions en matière de consentement, les organisations elles mettront l'accent sur fourniront les informations éléments suivantes :

- Renseignements personnels qui seront recueillis;
- Tiers auxquels les renseignements personnels seront communiqués;
- Fins auxquelles les renseignements personnels seront recueillis, utilisés ou communiqués;
- Risque de préjudice et autres conséquences.

La présentation de ces éléments constitue une exigence informationnelle fondamentale que les organisations doivent respecter pour obtenir un consentement éclairé au moment de la collecte de données.

Toutefois, le consentement éclairé pose plusieurs défis. D'une part, il peut être difficile à mettre en œuvre dans la pratique, car de nombreuses personnes donnent leur accord aux politiques de protection de la vie privée sans les lire (Scassa, 2018b). Ensuite, même si une personne accorde son consentement pour une utilisation particulière de ses données, il est presque impossible de remonter jusqu'à elle après coup pour obtenir son consentement pour leur « réutilisation » à des fins différentes des objectifs de la collecte initiale, une situation probable dans le cadre d'un partenariat de données numériques (Curty et Qin, 2014). Finalement, la limitation de l'utilisation des données à une finalité déterminée est remise en cause par les applications d'apprentissage automatique, qui fonctionnent en analysant un maximum de données pour des finalités qui évoluent au fur et à mesure de leur traitement (Gellert, 2016).

ENCADRÉ 13 : LA RÉUTILISATION DES DONNÉES

Curty et Qin, (2014), définissent la réutilisation des données comme « la ré-analyse d'un ensemble de données ou une combinaison de différents ensembles de données dans le but de répondre aux questions de recherche originales avec une nouvelle méthode d'analyse, ou de répondre à de nouvelles questions basées sur d'anciennes données qui n'étaient pas nécessairement l'objet de la collecte de données originale [traduction libre] » (p.1).

Les mécanismes d'obtention de consentement pour des fins de réutilisation des données ont récemment suscité beaucoup d'attention des auteurs, en particulier ceux du domaine de la recherche en santé. Ces derniers s'intéressent à de nouveaux modèles de consentement tels que le consentement dynamique et le métaconsentement.

Le **consentement dynamique** permet d'obtenir un consentement au cours de multiples phases de la collecte et du traitement des données (Budin-Ljøsne et coll. 2017 ; Kaye et coll., 2015). Le processus de consentement dynamique offre généralement des options granulaires à différents « points de contact ». Ces points de contact autorisent l'utilisation (ou la réutilisation) d'un même ensemble de renseignements personnels avec le consentement éclairé des individus chaque fois que les raisons de la collecte, de l'utilisation ou de la divulgation de ces données changent (Budin-Ljøsne et coll., 2017). Dans leur rapport intitulé *Trusted Data Sharing Framework*, Infocomm Media Development Authority of Singapore (IMDA) et Personal Data Protection Commission (PDPC) (2019) suggèrent que lorsqu'une organisation a l'intention de partager les données pour une finalité différente de celle pour laquelle le consentement avait été obtenu, elle devrait informer les personnes concernées, en soulignant tout nouveau risque résultant du partage secondaire. Les personnes doivent avoir la possibilité de retirer leur consentement si elles ne sont pas à l'aise avec le partage secondaire de leurs données.

Le **métaconsentement**, d'autre part, est une approche dans laquelle les participants à l'étude sont invités à faire connaître leurs préférences « concernant le(s) type(s) et la fréquence des décisions de consentement — ce qui leur donne un contrôle présumé sur la manière précise dont le consentement continuera à leur être demandé sur une base individuelle [traduction libre] » (Sheehan et coll., 2019, p.227). Ils pourraient décider, par exemple, qu'ils préfèrent

que leurs données ne soient utilisées qu'à des fins non commerciales ou même qu'ils préfèrent ne pas être contactés du tout. Toutefois, il reste à voir si ces nouvelles approches de la gestion du consentement offrent des avantages tangibles par rapport aux pratiques générales de consentement que les chercheurs connaissent bien (Sheehan et coll., 2019).

Les options de recours

Bien que le consentement éclairé soit une condition de base pour collecter et utiliser les données personnelles en respectant le droit à la vie privée de l'individu, il est toujours possible que la personne concernée souhaite à un moment donné restreindre l'accès à ses données, notamment s'il estime que ses données ont été utilisées à mauvais escient. Offrir aux individus des possibilités de recours est donc un élément important d'une utilisation responsable et éthique des données. En fournissant des canaux par lesquels les individus peuvent soumettre et résoudre des plaintes ou des problèmes concernant leurs données, un partenariat de données numériques démontre un respect pour les données de l'individu et son droit de déterminer comment elles doivent être utilisées.

Les mécanismes de recours peuvent prendre différentes formes allant du retrait du consentement individuel, jusqu'à l'obtention d'une compensation pour une utilisation abusive des données en passant par la modification d'une donnée personnelle.

À la suite de l'adoption du Règlement général sur la protection des données (RGPD) par l'Union européenne, la reconnaissance des droits numériques individuels gagne du terrain et la mise en place d'options de recours dans le domaine des renseignements personnels est de plus en plus reconnue comme une pratique responsable en matière de gouvernance des données. L'article 21 du RGPD a accordé aux individus le droit de s'opposer à tout moment au traitement de leurs données personnelles. Cela leur permet effectivement d'imposer à l'organisation l'arrêt ou l'interdiction du traitement de leurs renseignements personnels. De plus, en vertu du RGPD, les particuliers peuvent déposer une plainte auprès de leur commission nationale de protection de la vie privée s'ils pensent que leurs droits en matière de données ont été violés. Ils peuvent également obtenir une indemnisation si une entreprise ou une organisation n'a pas respecté la loi sur la protection des données (Règlement [UE] 2016/679 du parlement européen et du conseil, 2016).

L'anonymisation

En plus d'obtenir le consentement éclairé et d'offrir des options de recours, les partenariats de données numériques devraient appliquer un ensemble de mécanismes procéduraux visant à protéger la confidentialité des données. Une des techniques les plus couramment décrites dans la littérature sur la gouvernance des données est l'anonymisation. L'anonymisation est un processus visant à rendre négligeable le risque qu'une personne soit identifiée au moyen des données (Elliot et coll., 2020, p.10). Effectivement, les techniques d'anonymisation peuvent viser différents degrés d'identifiabilité des données. La norme ISO/IEC 19441 (développée pour assurer l'interopérabilité et la portabilité des données dans le cadre de services d'infonuagique) distingue cinq catégories à cet effet :

- **Données identifiables :** Données qui peuvent être associées sans ambiguïté à une personne particulière parce que des renseignements permettant d'identifier cette personne y sont observables.
- **Données pseudonymisées :** Données dans lesquelles tous les identificateurs sont remplacés par des pseudonymes, la fonction d'attribution étant telle que les remplacements ne peuvent pas être inversés au moyen d'efforts raisonnables par une personne différente de celle qui les fait.

- **Données pseudonymisées non liées :** Données dans lesquelles tous les identificateurs sont effacés ou remplacés par des pseudonymes, la fonction d'attribution étant effacée ou irréversible, de sorte que les liens ne puissent pas être rétablis par des efforts raisonnables, y compris de la part de l'entité qui a effectué l'opération.
- **Données anonymisées :** Données non liées dont les attributs sont modifiés (par exemple, par randomisation ou généralisation de leurs valeurs) de façon à ce que ces données seules ou combinées à d'autres données ne permettent pas d'identifier directement ou indirectement une personne avec un niveau de confiance raisonnable.
- **Données agrégées :** Données statistiques qui ne contiennent pas d'entrées de niveau individuel et qui sont classées au moyen de renseignements sur tellement de personnes différentes que les attributs de niveau individuel ne sont pas identifiables.

Comme le soulignent Elliot et coll. (2020), l'objectif de l'anonymisation est de rendre la réidentification plus difficile. Il ne s'agit pas d'une solution infaillible. Sur ce sujet, les auteurs marquent l'importance d'examiner attentivement l'environnement dans lequel vous partagez ou diffusez des données (Elliot et coll., 2020). Cette prudence est très importante dans notre environnement numérique actuel, puisqu'on témoigne actuellement de développements en matière d'analyse des données (et d'intelligence artificielle) permettent d'établir des liens entre des renseignements apparemment non personnels et une personne identifiée ou identifiable de plus en plus facilement (OCDE, 2019).

Par exemple, une étude de 2019 a montré que des données anonymisées peuvent être réidentifiées et associées à une personne identifiable avec succès à un taux de 99,98 % en utilisant quinze facteurs démographiques (Rocher, Hendrickx et de Montjoye,

2019). Une étude précédente a montré que, dans un ensemble de données de 1,5 million de personnes collectées sur six mois en utilisant des points de localisation triangulés à partir de tours de téléphonie mobile, 95 % des individus pouvaient être identifiés de manière unique sur la base de seulement quatre points horodatés et géolocalisés (de Montjoye et coll., 2013).

En fin de compte, si l'anonymisation peut être une technique qui améliore la protection de la vie privée dans la mesure où elle supprime les éléments identifiables d'un ensemble de données, des risques importants pour la vie privée subsistent même après l'anonymisation des données. Par conséquent, l'application de techniques d'anonymisation n'est qu'une étape du processus de protection des renseignements personnels — d'autres mécanismes de protection des données doivent être intégrés pendant tout leur cycle de vie (voir la section Efficacité : Gérer les données de manière efficace et cohérente du chapitre 3 pour revisiter la notion du cycle de vie).

La confidentialité programmée

Une des approches visant à protéger la vie privée qui s'intègre à l'ensemble du cycle de vie est la confidentialité programmée (*privacy by design*), qui se traduit par l'idée que « l'assurance de la protection de la vie privée doit idéalement devenir le mode de fonctionnement par défaut d'une organisation [traduction libre] » (Cavoukian, 2009, p.1). L'application des principes de la confidentialité programmée (voir encadré 14) vise à « adapter toute la structure d'une entreprise ou d'une organisation, incluant ses technologies de l'information, ses pratiques et ses processus opérationnels, sa conception physique et son infrastructure réseau [traduction libre] » autour des exigences requises pour la protection de la vie privée (Cavoukian et Dixon, 2013, p.6).

En général, l'obtention d'un consentement éclairé et l'application de techniques d'anonymisation peuvent être considérées comme des mécanismes de gouvernance procédurale des données qui peuvent servir à protéger la confidentialité des renseignements personnels. Cependant, ces procédures ne sont pas suffisantes par elles-mêmes. Lorsqu'on comprend que l'anonymisation n'est pas une solution infaillible, le besoin de mécanismes supplémentaires se fait sentir pour gérer les risques et disposer de voies de recours en cas d'atteinte à la vie privée.

ENCADRÉ 14 : LES PRINCIPES DE LA CONFIDENTIALITÉ PROGRAMMÉE (*PRIVACY BY DESIGN*)

Selon Cavoukian et Dixon (2013) les 7 principes de la confidentialité programmée sont les suivants [traduction libre] :

1. Prendre des mesures proactives et non réactives, préventives et non correctives
2. Assurer la protection implicite des renseignements personnels
3. Intégrer la protection des renseignements personnels dans la conception des systèmes et des pratiques
4. Assurer une fonctionnalité complète selon un paradigme à somme positive et non à somme nulle
5. Assurer la sécurité de bout en bout, pendant toute la période de conservation des renseignements
6. Assurer la visibilité et la transparence — Faciliter les vérifications
7. Assurer la protection des renseignements personnels des utilisateurs — Privilégier l'utilisateur

L'évaluation des risques

L'évaluation des risques en matière de protection des données utilise des outils et des méthodologies d'analyse des risques calculés et contextuels pour évaluer et gérer les risques associés aux activités de traitement des données prévues par le partenariat de données numériques. Ces outils ont pour objectif de calibrer et d'opérationnaliser les obligations légales des organisations, par exemple en matière de protection de la vie privée, contenus dans les lois et règlements, en fonction des risques et avantages réels posés par l'utilisation des données proposées (Centre for Information Policy Leadership, 2014).

L'évaluation des risques cherche à identifier en amont les menaces pesant sur les données personnelles ou les préjudices pouvant découler du traitement des données et de retracer leurs causes. Selon le Centre for Information Policy Leadership (2014), « la question devrait être de savoir s'il existe une probabilité significative qu'une menace identifiée puisse conduire à un dommage reconnu avec un degré de gravité important [traduction libre] » (p.4).

À l'instar de la confidentialité programmée, l'approche d'évaluation des risques vise à intégrer la protection des données dans le *mode opérationnel* d'une organisation. C'est-à-dire que l'organisation adopte des pratiques dans le cadre desquelles la gestion des risques est considérée comme partie intégrante du processus décisionnel, et non comme une contrainte technique ou juridique distincte (OCDE, 2019).

Ainsi, avant de procéder au traitement de données collectées dans le cadre d'un partenariat de données numériques, la conduite d'une analyse des risques peut faciliter la détermination de potentiels préjudices qui pourraient découler de la divulgation de renseignements personnels. Il suffit de répondre à des questions comme (The Engine Room, 2016, p. 92) :

- Des personnes ou groupes peuvent-ils avoir intérêt à découvrir ou dévoiler l'identité des sujets des données ?
- Le croisement avec d'autres jeux de données disponibles pourrait-il permettre de découvrir l'identité des sujets des données qui sont diffusés ?
- Quelles pourraient être les conséquences de la dé-anonymisation des données pour les personnes ou les groupes concernés ?

En plus d'anticiper les risques liés au traitement des données, « la catégorisation des données selon leur type, leur valeur, leur sensibilité et leur niveau critique [traduction libre] » pour l'organisation est essentielle pour gérer les risques (IMDA et PDPC, 2019, p.45). Une telle catégorisation permet de définir différents niveaux de risque et de planifier des mesures de sécurité appropriées pour chacun. Comme nous l'avons souligné dans la section sur les caractéristiques des données, différents types de données conduisent à différentes exigences juridiques, réglementaires, juridictionnelles et même contractuelles, lesquelles doivent être examinées, gérées et vérifiées et prises en considération lors de l'évaluation des risques (IMDA et PDPC, 2019, p.45).

En somme, l'évaluation et la gestion des risques requièrent la mise en place de structures et de processus précis. Un des cadres d'évaluation et de gestion des risques les plus couramment mentionnés dans la littérature sur la gouvernance des données est le cadre des *Cinq éléments de la sécurité* (*Fives Safes Framework*), développé à l'origine au sein de l'organisme *Office for National Statistics* du Royaume-Uni en 2003 (voir encadré 15).

ENCADRÉ 15 : LE CADRE DES CINQ ÉLÉMENTS DE LA SÉCURITÉ

Le cadre des cinq éléments de la sécurité (*Five Safes Framework*) est fondé sur une approche multidimensionnelle de gestion du risque de divulgation. Chaque « élément de la sécurité » correspond à un aspect indépendant, mais lié au risque de divulgation. Le cadre pose des questions particulières pour faciliter l'évaluation et la description qualitative de chaque aspect du risque (ou de la sécurité). Cela permet aux dépositaires de données d'instaurer des mesures de contrôle appropriées, non seulement des données, mais aussi des modes d'accès à ces données. Ce cadre est conçu pour faciliter la diffusion sécuritaire des données et pour éviter une réglementation excessive (Ritchie, 2017). Les cinq éléments du cadre sont les suivants (Ritchie, 2017) [traduction libre] :

- Personnes fiables : la personne qui conduit des recherches est-elle dûment autorisée à accéder aux données et à les utiliser ?
- Projets fiables : les données seront-elles utilisées dans un but convenable ?
- Environnement fiable : l'environnement d'accès empêche-t-il l'utilisation non autorisée ?
- Données fiables : les données sont-elles protégées de façon appropriée et suffisante ?
- Résultats fiables : la confidentialité des données est-elle respectée dans les résultats statistiques ?

EFFICACITÉ : GÉRER LES DONNÉES DE MANIÈRE EFFICACE ET COHÉRENTE

Malgré les opportunités à leur portée, plusieurs organisations demeurent réticentes à s'engager dans des partenariats de données numériques d'envergure. Dans le chapitre 1, nous avons vu que des dynamiques concurrentielles, la peur d'une perte de contrôle ou encore des objectifs divergents peuvent créer des obstacles à de telles collaborations. Toutefois, une des entraves les plus fréquemment mises de l'avant par les organisations, notamment celles que nous avons interviewées dans le cadre de ce rapport, est tout simplement les coûts engendrés par la participation à de telles initiatives.

En effet, la collecte, le traitement et l'analyse d'importantes quantités de données peuvent rapidement engendrer des coûts élevés, en particulier pour les petites organisations ne possédant que de faibles capacités techniques. **L'adoption d'un principe d'efficacité vise à sensibiliser les acteurs au bénéfice d'une gestion efficace et cohérente des données.** Comme nous le verrons dans ce chapitre, des mécanismes clairs permettent de structurer, partager, lier, combiner et analyser les données avec plus de facilité. En limitant les efforts dédiés à la préparation et au traitement des données et en maximisant les analyses et les informations exploitables qui peuvent en découler, l'efficacité accroît l'utilité et réduit les barrières à la participation à des initiatives collectives. Par ailleurs, soulignons que l'efficacité est également importante du point de vue du citoyen. Après tout, les citoyens sont aussi des utilisateurs des données et l'accessibilité à des données de bonne qualité peut encourager l'innovation en plus de renforcer la transparence et l'imputabilité d'un projet.

Concrètement, le principe d'efficacité vise à encourager la collecte et la production de données de qualité, interopérables et facilement accessibles aux acteurs légitimes, et ainsi d'en simplifier le traitement tout en réduisant les risques de mener à des interprétations fautives ou discriminatoires.

Concrètement, le principe d'efficacité vise à encourager la collecte et la production de données de qualité, interopérables et facilement accessibles aux acteurs légitimes, et ainsi d'en simplifier le traitement tout en réduisant les risques de mener à des interprétations fautives ou discriminatoires.

Un manque de rigueur dans le traitement des données peut donner lieu à des enjeux de taille pour un partenariat de données numériques :

- **Données de mauvaise qualité :** Les données inexactes, incomplètes ou obsolètes présentent des risques, car elles peuvent être mal comprises et utilisées à mauvais escient pouvant conduire à des interprétations erronées ou l'émergence de biais et de préjudices sérieux pour les individus (Earley et coll., 2017).
- **Jeux de données incompatibles :** Les organisations ont toutes leurs propres manières d'utiliser et de structurer les données en plus d'utiliser parfois des logiciels différents. Cela signifie qu'il peut être difficile de lier les systèmes informatiques entre eux pour partager des données (Information Commissioner's Office, 2019). Cette qualité des systèmes est connue sous le nom d'interopérabilité, telle que définie par l'IEEE (1990) comme « la capacité de deux ou plusieurs systèmes ou des composants pour échanger des informations et utiliser les informations qui ont été échangées [traduction libre] ». Au niveau des données, l'interopérabilité est la capacité de deux ou plusieurs ensembles de données à les relier, les combiner et les traiter (Janssen et coll., 2014). La faible interopérabilité des données signifie qu'elles ne peuvent pas être combinées efficacement pour générer des analyses utiles.
- **Accès non autorisé aux données :** Le partage des données dans le cadre d'un partenariat exige de définir clairement les privilèges d'accès aux données qui sont accordés aux différentes parties. En l'absence de limites claires et de mesures de protection des accès à l'information, des entités non autorisées pourraient accéder aux données et les utiliser de façon abusive.

Afin de favoriser le partage des données au sein d'un écosystème, que ce soit à l'échelle d'un territoire, d'un secteur ou d'un groupe de partenaires, la qualité, l'interopérabilité et l'accessibilité des données sont donc des considérations importantes. Ces éléments sont de véritables leviers capables de débloquer le potentiel des données détenues ou produites collectivement. Pour atténuer ces risques et minimiser des problèmes potentiels, les partenariats de données numériques doivent donc établir des mécanismes clairs et cohérents et s'assurer que chaque partie comprend bien son rôle et ses responsabilités quant à leur mise en œuvre.

Mécanismes de gouvernance

Le cycle de vie des données

Rappelons d'abord, comme nous l'avons vu au chapitre 2, que le cycle de vie des données est un cadre de gestion utile qui permet de visualiser facilement les différentes étapes que traversent les données dans le cadre d'un projet, de leur collecte à leur destruction. Le recours au cycle de vie permet ainsi de planifier en amont, puis de structurer en continu la gouvernance des données de manière à mettre en place les mécanismes requis à chacune des étapes. Il permet également de réfléchir à l'efficacité de ceux-ci selon une approche systémique, qui tient compte de leur interaction entre eux. C'est pourquoi sans être un mécanisme en lui-même, il est important que la qualité, l'interopérabilité et la gestion des accès soient évaluées et réévaluées à chacune des étapes du cycle de vie.

La qualité des données

Dans la littérature sur la gouvernance des données, les décisions qui visent à mettre en place des techniques pour mesurer, évaluer et améliorer la qualité des données reçoivent une attention considérable et sont généralement perçus comme critiques (Earley et coll., 2017 ; Khatri et Brown, 2010).

Cette préoccupation pour la qualité est encore plus présente dans la littérature s'intéressant plus spécifiquement à la réutilisation et au partage des données (Yoon, 2017 ; Peer, Green, et Stephenson, 2014 ; Sposito, 2017). Dans le cadre de partenariats, avant d'utiliser les données d'autres partenaires, les utilisateurs doivent avoir l'assurance que les données sont de haute qualité. Ils doivent évaluer leur pertinence, leur fiabilité et leur validité (Faniel et Jacobsen, 2010). Autrement dit, la qualité des données est un facteur important en raison de son lien avec la confiance.

La qualité des données se mesure au moyen de divers

ENCADRÉ 16 : LES DIMENSIONS DE LA QUALITÉ DES DONNÉES

Ces définitions sont des traductions libres des définitions proposées par le DAMA-DMBOK (Earley et coll., 2017, pp. 458-459).

La précision fait référence à la justesse ou l'exactitude avec laquelle les données représentent le réel.

L'exhaustivité fait référence à la présence ou non de toutes les données requises.

La cohérence fait référence au fait que les valeurs des données sont représentées de manière uniforme dans et entre les ensembles de données, ainsi qu'à la taille et à la composition des ensembles de données entre les systèmes ou dans le temps.

L'intégrité fait généralement référence soit à l'intégrité référentielle (c'est-à-dire la cohérence entre bases de données via une clé de référence), soit à la cohérence interne au sein d'un ensemble de données, de sorte qu'il n'y ait pas de données manquantes.

La plausibilité (*reasonability*) fait référence à la question de savoir si un modèle de données répond aux attentes basées, par exemple, sur des données de référence.

L'actualité (*timeliness*) fait référence à savoir si les données sont à jour (la rapidité de leur publication), et à leur latence, c'est-à-dire le temps écoulé entre le moment où les données sont créées et celui où elles sont mises à disposition pour utilisation.

Le caractère unique fait référence au fait qu'aucune entité n'existe plus d'une fois dans un ensemble de données.

La validité fait référence au fait que les valeurs des données sont cohérentes avec un domaine défini de valeurs tel qu'un ensemble de référence ou une gamme de valeurs par exemple.

critères, lesquels peuvent varier selon les contextes. Le DAMA-DMBOK (Earley et coll., 2017, pp. 458-459) recense à lui seul des dizaines de dimensions de la qualité des données ainsi que plusieurs manières de les regrouper et d'y réfléchir. Selon Earley et coll. (2017), les plus importantes dimensions de la qualité des données, sur lesquelles un consensus est établi sont les suivantes : la précision, l'exhaustivité, la cohérence, l'intégrité, la plausibilité, l'actualité, le caractère unique et la validité (pp. 458-459) (voir encadré 16).

En pratique, les organisations qui partagent des données doivent mettre en place des normes claires et adaptées à leur contexte et à leur cas d'usage pour assurer la qualité de leurs données. Il est préférable que le choix et la mise en œuvre de ces mécanismes soient faits de manière collective afin d'assurer l'adhésion de tous les membres du partenariat aux règles en place et afin de faciliter l'imputabilité des acteurs impliqués en cas de conséquences néfastes.

Il existe une diversité de mécanismes pouvant être mis en œuvre pour atteindre les objectifs recherchés et les trois types de mécanismes (structuraux, procéduraux et relationnels) peuvent être utiles. Voici quelques exemples utiles (Information Commissioner's Office, 2019, p.28) :

- afin de s'assurer que les données partagées sont exactes, on peut procéder à un exercice d'échantillonnage périodique de la base de données communes ;
- pour garantir que les données sont enregistrées de manière homogène et cohérente, le partenariat peut se doter de gabarits détaillant avec précision comment certaines données doivent être enregistrées, par exemple les dates de naissance ;
- afin de ne pas réutiliser des données qui ne sont plus d'actualité, il peut être nécessaire de convenir de périodes de conservation et de modalités de suppression communes ;
- une offre de formation peut être un mécanisme puissant pour assurer la réduction des erreurs dans le traitement des données.

La gestion des métadonnées

Les métadonnées sont des données qui fournissent de l'information sur d'autres données (Riley, 2017). Cela peut sembler simpliste, mais les métadonnées sont d'une importance capitale, car au sein d'une même organisation ou entre plusieurs organisations, aucun individu ne peut avoir une connaissance complète de toutes les données produites ou utilisées. Comme l'explique clairement Sebastian-Coleman (2018), « sans métadonnées fiables, une organisation ne sait pas quelles données elle possède, ce que les données représentent, d'où elles proviennent, comment elles se déplacent dans les systèmes, qui y a accès, ou ce que cela signifie pour les données d'être de haute qualité. Sans métadonnées, une organisation ne peut pas gérer ses données comme un actif [traduction libre] » (pp.142-3).

Concrètement, les métadonnées peuvent consister en des définitions cohérentes des différents éléments d'un jeu de données ou encore de l'information sur l'origine des données et leur lignée. On peut penser aux règles d'intégration et de traitement qui leur ont été appliquées par exemple. Idéalement, les métadonnées doivent offrir aux utilisateurs des réponses aux questions suivantes (Opendatasoft, 2020, p.5) :

- **Quoi :** De quoi parle le jeu de données ?
- **Qui :** Qui en est à l'origine ?
- **Pourquoi :** Pourquoi le jeu de données existe-t-il ?
- **Comment :** Comment utiliser le jeu de données ? Quelles règles s'appliquent ?
- **Quand :** Dans quelle temporalité le jeu de données s'inscrit-il ?
- **Où :** Dans quel territoire le jeu de données se situe-t-il ?

Plus le volume de données qu'un partenariat utilise et partage augmente, plus il est essentiel de mettre en place des stratégies de gestion des métadonnées. La plupart des systèmes qui traitent les données génèrent automatiquement des métadonnées, toutefois celles-ci peuvent manquer de cohérence ou de précision. Une stratégie de gestion délibérée permet en revanche aux partenaires de s'entendre au préalable sur les informations qu'ils doivent connaître sur les données qu'ils utilisent et ainsi de mettre en place les pratiques requises pour s'assurer que ces informations sont colligées et maintenues adéquatement.

Par exemple, lorsque des données personnelles qui se rapportent au genre, à l'ethnicité ou à l'orientation sexuelle sont utilisées, il est important de s'assurer de mettre en place des définitions et des règles de catégorisation inclusive, acceptées par tous les partenaires (à titre d'exemple de processus de création de métadonnées inclusives, veuillez voir l'encadré 17). Ou encore des inspections et des audits de qualité périodiques peuvent être déployés pour vérifier que les métadonnées sont à jour.

En plus d'améliorer la compréhension des données au sein des organisations, des métadonnées de bonne qualité et bien gérées permettent de maximiser leur découvrabilité et leur interopérabilité, en permettant aux utilisateurs de données de facilement repérer et récupérer des jeux de données (Opendatasoft, 2020). La disponibilité de métadonnées s'avère également un élément important de la traçabilité des données, un enjeu que nous examinons plus en détail dans le chapitre 3 Imputabilité : Évaluer en continu la conformité et les impacts.

ENCADRÉ 17 : LA CRÉATION DE MÉTADONNÉES INCLUSIVES

CultureBrew.Art (CBA) est une plateforme numérique qui fait la promotion et favorise l'interculturalisme intersectionnel dans tout le secteur des arts du spectacle et des médias au Canada. Pour y parvenir, son outil central est une base de données réunissant des informations sur des artistes indigènes et racisés auxquelles les entreprises et les agences peuvent accéder en tant qu'abonnés.

La banque de données CBA peut être consultée en utilisant les catégories suivantes : sexe, héritage racial/ethnique, langue, disciplines artistiques et d'autres domaines déterminés par les données de recherche recueillies lors de consultations communautaires (Visceral Visions, 2020). Les définitions pour chacune de ces catégories ont été déterminées de manière collective pour assurer leur exactitude.

L'usage de normes et l'interopérabilité

Afin de combiner, comparer ou relier des jeux de données provenant de sources variées, un partenariat de données numériques a besoin que ces jeux soient portables[2], standardisés et interopérables.

Ces différents piliers du partage de données sont bien mis en évidence par Michal Gal et Daniel Rubinfeld (2019), selon qui « la **portabilité** des données (la capacité [d'extraire ou] transférer des données sans en affecter le contenu) et **l'interopérabilité** (la capacité d'intégrer deux ou plusieurs ensembles de données) affectent considérablement l'utilisation efficace des données et, par conséquent, le bien-être public et privé [traduction libre] » (p. 739).

2 Le terme portabilité des données fait référence à la capacité de déplacer, copier ou transférer facilement des données d'un environnement informatique à un autre de manière sûre et sécurisée, sans affecter son utilité (Information Commissioner's Office, s.d.). Le terme a été popularisé par le Règlement général sur la protection des données (RGPD) qui stipule dans son article 20 le droit à la portabilité des données.

Ils ajoutent que c'est par le biais de la **standardisation** (ou la normalisation, les deux termes peuvent être utilisés) que la portabilité et l'interopérabilité peuvent être atteintes. Selon eux, « la normalisation est une condition préalable au fonctionnement des industries dans lesquelles les échanges de données entre entreprises et entre secteurs sont critiques [traduction libre] » (Gal et Rubinfeld, 2019, p.740). Autrement dit, la portabilité et l'interopérabilité sont obtenues par l'adhésion à des normes en matière de traitement des données qui permettent de les transférer facilement d'un système à l'autre et de les intégrer ensemble.

Il existe une grande variété de normes — plus d'un million de normes nationales et 330 000 normes internationales selon Michel Girard (2018) — et la majorité des partenariats de données numériques en adopteront quelques-unes. En matière de gestion numérique, ces normes peuvent affecter à peu près tous les champs d'application des données. Elles permettent par exemple de « clarifier les définitions, l'architecture des systèmes, la propriété des données, leur classement, leur regroupement, leur stockage et leur élimination. Ces normes visent aussi à établir des références en matière de confidentialité et d'agrégation [traduction libre] » (Girard, 2018, p.1). Elles peuvent encadrer la qualité des données, établir des formats informatiques précis ou délimiter les informations contenues dans les métadonnées.

Ces normes peuvent parfois être internes, c'est-à-dire propres au partenariat, mais en général ce sont des standards techniques externes établis par des organes de normalisation indépendants. Nombre d'entre elles ont été élaborées par des organismes de normalisation tels que l'Organisation internationale de normalisation (ISO) (par exemple la norme internationale ISO/IEC 27001 sur la gestion de la sécurité de l'information), des consortiums industriels (par exemple, le code HTML) ou des entreprises individuelles (par exemple le format informatique standardisé *General Transit Feed Specification* développé initialement par Google). Certaines normes exigent que les utilisateurs achètent une licence, tandis que d'autres sont des normes ouvertes disponibles gratuitement. De plus, les normes de données varient grandement d'une industrie à l'autre. Ils existent par exemple des normes bien établies en matière de publication de données ouvertes (Guidoin et coll., 2016).

L'adoption et l'application de normes numériques visent à améliorer la capacité des parties à relier et à combiner l'information, ce qui permet de créer des jeux de données de plus en plus complexes et donc de découvrir de nouvelles perspectives au moyen des données liées, mutualisées, partagées, etc. (Girard, 2018). Toutefois, dans certains domaines émergents, il n'existe encore aucune norme. Si un partenariat de données numériques estime qu'il n'existe pas de normes appropriées qui répondent à ses besoins, il peut alors développer ses propres pratiques de normalisation. Au fil du temps, le partenariat peut examiner s'il existe un besoin plus large de normalisation pour son secteur ou son industrie (par exemple, au fur et à mesure que des organisations rejoignent leur initiative) et s'il faut investir du temps et des ressources dans la création de nouvelles normes.

En somme, l'identification des normes à respecter, puis la mise en place de procédures et de directives pour en assurer l'application seront donc des éléments importants de la gouvernance des données des partenariats de données numériques. Cela pourrait impliquer la création d'outils pour soutenir l'adoption et l'utilisation des normes, de processus pour évaluer, tester et améliorer les normes sélectionnées, de produits certifiés qui répondent déjà aux normes de sécurité et d'interopérabilité telles que des interfaces de programmations (API) ou encore des politiques et structures de gouvernance pour assurer la conformité.

ENCADRÉ 18 : DES NORMES… POUR TOUS ?

Bien qu'il y ait des avantages avérés à l'élaboration de normes pour faciliter l'échange de données entre deux ou plusieurs systèmes, il est important de reconnaître que les normes de données sont parfois critiquées parce qu'elles donnent la priorité à l'interopérabilité technique sur la compréhension humaine des systèmes (Brandusescu, Canares et Fumega, 2020). Par exemple, Montenegro (2019), réfléchissant à l'adaptabilité des normes de métadonnées pour promouvoir la souveraineté des données autochtones, conteste « l'hypothèse la plus fondamentale concernant tout processus de normalisation [à savoir] qu'une norme qui réussit pour un groupe de personnes réussit pour tous, et que la norme adoptée fonctionne mieux que toute autre méthode de documentation et de gestion de l'information [traduction libre] » (Montenegro, 2019, p.736).

Cette vision fait fi selon lui des nombreuses tensions en jeu dans le processus de normalisation, entre le désir d'organiser et de diffuser les connaissances de manière systémique, le penchant idéologique en faveur de l'efficacité, l'existence de manières alternatives de connaître et de produire des connaissances et la nécessité pour les communautés locales de préserver une certaine souplesse dans la manière dont elles documentent leurs propres connaissances en fonction de leurs croyances. Ces tensions se manifestent aussi à travers différentes dynamiques du pouvoir, puisque « différents groupes de la société utilisent la connaissance et le contrôle de la connaissance et de sa signification afin d'exercer un pouvoir sur d'autres groupes [traduction libre] » (Battiste, 2008, p.5).

Brandusescu, Canares et Fumega (2020) mettent de l'avant trois recommandations visant à améliorer la conception des normes pour répondre à ces enjeux, soit : l'inclusion de multiples perspectives dans le processus de conception des normes, la prise en compte des contextes et des besoins de multiples utilisateurs dans la définition des normes, et le besoin d'être explicite sur les rôles et les relations pendant leur mise en œuvre.

La gestion des accès et des privilèges associés

La gestion des accès dans le cadre d'un partenariat de données numériques vise d'abord à protéger les données contre la divulgation ou la modification non autorisée ou abusive, tout en assurant leur disponibilité pour les utilisateurs légitimes. Pour garantir cette protection, « il faut donc que tout accès à un système et à ses ressources soit contrôlé et que tous et seulement les accès autorisés puissent avoir lieu. Ce processus est appelé [gestion ou] contrôle des accès [traduction libre] » (Samarati et de Vimercati, 2001, p. 137).

La gestion des accès aux données réfère ainsi au système d'autorisations requises pour accéder aux données afin de mener toutes activités liées au stockage, à la récupération ou au traitement des données hébergées dans une base de données (Earley et coll., 2017, p.197). À cet égard, l'octroi de droits d'accès minimaux est un principe de sécurité reconnu. Il signifie qu'« un utilisateur, un processus ou un programme ne devrait être autorisé à accéder qu'aux informations autorisées par son objectif légitime [traduction libre] » (Earley et coll., 2017, p. 232).

En plus de contrôler les accès, le système d'autorisations peut contenir diverses règles qui déterminent et limitent la manière dont les données peuvent être manipulées. Le système permet ainsi de moduler les privilèges associés aux autorisations d'accès.

Cela peut s'effectuer en fonction de divers critères selon le contexte et les besoins du partenariat de données. D'abord, à presque tout coup, les accès varient selon la **fonction de l'utilisateur**. Par exemple, en appliquant la technique du *contrôle d'accès basée sur les rôles*, certains utilisateurs pourront seulement consulter la base de données, alors que d'autres pourront en extraire des données ou la modifier (Conrad et coll., 2016, pp. 321-2). Il est fréquent que seuls les administrateurs système puissent effacer des données, par exemple. Ensuite, on retrouve souvent des catégories d'accès par **type d'utilisateur** (Samarati et de Vimercati, 2001, pp. 139-40). Dans le secteur public notamment, les entreprises privées, les chercheurs académiques et les représentants gouvernementaux n'ont généralement pas accès aux mêmes données. Finalement, un autre critère à prendre en compte est **le degré de sensibilité** des données elles-mêmes (Kum et Ahalt, 2013).

Par exemple, au Royaume-Uni, l'organisme Consumer Research Data Centre (CRDC) utilise des métadonnées pour classer les données selon trois niveaux de sensibilité : ouvertes, protégées ou contrôlées (CRDC, 2020). Selon la classification, les données ouvertes sont librement accessibles à tous. Les données protégées sont des renseignements non nominatifs accessibles dans les limites imposées par une licence et la loi. Finalement, les données sensibles sont des données contrôlées et à ce titre doivent être conservées dans les conditions les plus sécuritaires possibles, avec des privilèges d'accès extrêmement restreints.

Pour finir, le contrôle des accès aux données doit être assuré par l'application de diverses solutions techniques et de sécurité. De nombreux options et protocoles peuvent être envisagés, allant de processus relativement simple, tel qu'un formulaire d'inscription exigeant le nom de l'utilisateur, à des systèmes d'accès complexes, tel qu'un processus d'authentification de l'identité de l'utilisateur accompagné d'un contrôle de ses accès basé sur son rôle (pour un aperçu des techniques de gestion des identités et des accès, voir Conrad et coll., 2016).

Dans l'ensemble, toutefois, même lorsque les parties s'entendent et établissent un cadre de gestion des accès et des mécanismes précis pour garantir la sécurité des données, des risques subsistent que les données soient utilisées de manière non anticipée et préjudiciable (que ce soit de manière intentionnelle ou non) (OCDE, 2019). C'est pourquoi nous avons besoin de mécanismes d'imputabilité, qui feront l'objet de la prochaine section.

ENCADRÉ 19 : LES ENTENTES DE PARTAGE DE DONNÉES

Au sein d'un partenariat de données numériques, où plusieurs employés de différentes organisations ont accès à une base de données centralisée, il est important de mettre en place des règles d'accès précises, mais aussi de réfléchir plus largement aux conditions d'utilisation des données. Par exemple, qu'arrive-t-il si un employé contrevient au cadre de gouvernance en place ? Ou encore si les données extraites en fonction d'une autorisation d'accès valable sont toutefois utilisées à mauvais escient ? Ces conditions d'utilisation des données sont généralement articulées dans des ententes de partage de données (*data-sharing agreement*) entre les différents partenaires de l'initiative.

Bien qu'aucun format ne soit prescrit pour les ententes de partage de données, voici les modalités les plus importantes selon IMDA et PDPC (2019, p.37) [traduction libre] :

1. L'octroi d'une licence/autorisation d'utilisation des données dans le but prévu
2. L'imposition de restrictions à l'utilisation autorisée des données (le cas échéant), comme des limitations territoriales ou temporelles, des droits d'exclusivité ou des droits de commercialisation.
3. L'exigence de garanties ou d'autres assurances concernant les droits du fournisseur de données sur les données.
4. Un mode d'attribution de la responsabilité en cas de violation de l'entente, la liste des responsabilités entre les parties, les droits d'indemnisation et autres recours en cas de violation.
5. Un énoncé de confidentialité, la durée de l'entente, les lois applicables et le mode de règlement des litiges.

IMPUTABILITÉ : ÉVALUER EN CONTINU LA CONFORMITÉ ET LES IMPACTS

Le dernier principe clé que devraient chercher à mettre en œuvre les partenariats de données numériques préoccupés par la protection de l'intérêt public est l'imputabilité. En termes simples, cela signifie que les parties acceptent la responsabilité de leurs actions, qu'elles sont transparentes quant à la manière dont elles traitent, analysent et utilisent les données, et qu'elles s'engagent à évaluer et à répondre des impacts tant positifs que négatifs qui en découlent. À l'instar des deux principes précédents, ce n'est qu'à travers la mise en place de mécanismes concrets qu'il est possible d'incarner ce principe.

La littérature sur la gouvernance des données souligne à répétition l'importance de la répartition des responsabilités sur les actifs de données dans l'organisation (Abraham, Schneider et vom Brocke, 2019 ; Khatri et Brown, 2010 ; Otto, 2011). Dans les partenariats de données numériques, où les données franchissent les frontières organisationnelles, il devient encore plus important de définir clairement les rôles et les responsabilités de chacun, sachant que les risques d'utilisation ou d'exposition inappropriées (délibérées ou accidentelles) augmentent considérablement lorsque de multiples utilisateurs peuvent accéder aux données ou encore que des processus automatisés les traitent (OCDE, 2019).

L'imputabilité comporte d'abord une dimension interne qui vise à s'assurer que la gouvernance des données atteint ses objectifs. Elle s'appuie sur une structure décisionnelle claire et des mécanismes assurant la conformité et ouvrant la voie à l'auditabilité des systèmes. Toutefois, l'imputabilité comporte également une dimension externe, où le partenariat est appelé à faire une reddition de compte vis-à-vis du public.

Malgré son importance, cette seconde dimension ne sera pas abordée plus en détail dans cette section. Encore peu d'auteurs qui s'intéressent à la gouvernance des données intègrent les notions de participation citoyenne et d'évaluation d'impact dans leurs recherches, des éléments néanmoins incontournables, à notre avis, des projets de partage de données qui visent à servir l'intérêt public et que nous avons décrit plus en détail dans le chapitre 1 à la section Fondements d'un partenariat de données numériques réussi.

Ces deux dimensions de l'imputabilité (interne et externe) se rejoignent et se renforcent mutuellement : les efforts de transparence et d'engagement du public seront sans contredit soutenus par des structures de responsabilités claires en matière de gouvernance des données et par la mise en place de processus internes visant à assurer la conformité et documenter les décisions prises par rapport aux données tout au long de leur cycle de vie.

Mécanismes de gouvernance

Une répartition claire des responsabilités

Inévitablement, l'introduction d'un plus grand nombre d'acteurs ayant des intérêts variés crée le besoin de mécanismes de gouvernance plus complexes (Abraham, Schneider et vom Brocke, 2019). Comme l'expliquent The British Academy et The Royal Society (2017) « un des principaux défis de la gouvernance des données consiste à établir des mécanismes de répartition des responsabilités au sein d'une structure complexe, pour que toute action frauduleuse, non éthique, abusive ou discriminatoire puisse être détectée, corrigée et sanctionnée de manière appropriée [traduction libre] » (p.45). L'existence d'une autorité décisionnelle clairement définie est ainsi souvent citée comme un facteur déterminant du succès de la gouvernance des données et il en va de même pour les partenariats de données numériques (Earley et coll., 2017 ; Khatri et Brown, 2010 ; Otto, 2011).

Il n'existe toutefois pas de modèle ou de structure infaillible pour faciliter la prise de décision dans les partenariats de données numériques. Cela dépend avant toute chose soit du choix de structurer le partenariat autour d'ententes de partages de données dans lesquelles sont précisées à qui incombe les responsabilités de la gouvernance des données, soit de créer une entité externe, un intermédiaire, à laquelle est déléguée l'ensemble du pouvoir décisionnel et de la responsabilité de gouverner les données. Dans les deux cas, il est néanmoins essentiel que les responsabilités soient clairement assignées.

Dans les partenariats de données numériques **formés autour d'ententes**, il peut ainsi être quand même nécessaire de se doter d'une instance décisionnelle aux règles de représentation claires (c'est-à-dire qui y siègent) pour déployer et superviser la mise en œuvre

Il n'existe toutefois pas de modèle ou de structure infaillible pour faciliter la prise de décision dans les partenariats de données numériques. Cela dépend avant toute chose du choix soit de structurer le partenariat autour d'ententes de partages de données dans lesquelles sont précisées à qui incombe les responsabilités de la gouvernance des données, soit de créer une entité externe, un intermédiaire, à laquelle est déléguée l'ensemble du pouvoir décisionnel et de la responsabilité de gouverner les données. Dans les deux cas, il est néanmoins essentiel que les responsabilités soient clairement assignées.

d'un cadre commun de gouvernance des données. Les partenaires peuvent, par exemple, former un comité mixte responsable de concevoir et sélectionner conjointement les protocoles et les procédures définissant la manière dont ils partagent les données entre eux. Ou encore, ils peuvent se doter d'un comité éthique ou d'audit responsables de régler les disputes ou de tester la sécurité des systèmes des membres du partenariat.

Lorsque les partenaires utilisent des données publiques et spécialement des renseignements personnels toutefois, un arrangement basé sur un ensemble d'ententes de partages de données peut être insuffisant pour soulager les craintes du public quant à l'utilisation éthique et responsable des données et sur l'imputabilité des acteurs en cas d'utilisation inappropriée. Dans ce cas, certains partenariats de données numériques préfèreront confier l'autorité et la responsabilité de la gouvernance des données à **un intermédiaire de confiance**, externe à l'ensemble des parties.

Un tel intermédiaire de confiance peut prendre diverses formes. Le plus connu est sans aucun doute le « data trust », un concept développé en Angleterre dans le cadre de la common law (Open Data Institute, 2019). Dans le contexte juridique du Québec, cela pourrait être par exemple une organisation à but non lucratif, une coopérative ou encore une fiducie de protection de données, un nouveau véhicule juridique de droit civil à l'étude en ce moment (Marchand, 2019).

Malheureusement, il n'existe encore que peu d'études de cas réussis, matures et bien documentés en matière de gouvernance des données, de sorte que les avantages de recourir à un intermédiaire de confiance ainsi que les impacts que pourrait avoir une forme juridique plutôt qu'une autre sur la gouvernance des données et l'atteinte de résultats sont malheureusement à ce jour peu connus (Coutts et Gagnon-Turcotte, 2019).

Au niveau des avantages, on peut imaginer que le recours à un intermédiaire de confiance puisse aider à dépolitiser et à rationaliser les négociations entre partenaires, notamment en ce qui a trait à la propriété des données. Dans une situation où une asymétrie existe entre les partenaires au niveau des compétences techniques ou des ressources internes disponibles, l'introduction d'un intermédiaire de confiance peut aider à niveler tout déséquilibre dans leur pouvoir d'influence respectif. En revanche, le recours à un intermédiaire de confiance ne réduit pas complètement le risque qu'il puisse détourner l'intérêt de ses membres ou encore exercer un mauvais jugement au nom de ses fiduciaires (Porcaro, 2020).

Ainsi, un avantage recherché, soit celui de bâtir la confiance du public, peut être affecté par le modèle d'affaires choisi pour assurer la pérennité de l'intermédiaire. D'autres modalités peuvent également influencer la perception du public, voire même la gouvernance des données elle-même, telles que les conditions associées au partenariat ou bien le degré de contrôle individuel conservé par le détenteur des données ou cédé à l'intermédiaire. L'ensemble de ces considérations devront être évaluées avec mesures par les membres du partenariat de données numériques.

Peu importe la structure de gouvernance privilégiée par le partenariat, il n'en demeure pas moins que l'attribution du pouvoir décisionnel en matière de gouvernance des données doit être claire et transparente afin d'établir la confiance des partenaires comme du public. « Si un organe décisionnel est perçu comme n'ayant aucun pouvoir réel, cela pourrait affecter la légitimité perçue du mode de gouvernance [traduction libre] » (Coutts et Gagnon-Turcotte, 2019, p.48).

Le contrôle de la conformité

Un partenariat de données numériques doit s'assurer qu'il utilise les données conformément aux lois et règlements en vigueur (tels que ceux mentionnés au chapitre 2), mais aussi aux procédures et normes établies par le partenariat lui-même. Pour ce faire, les instances décisionnelles du partenariat doivent donc exercer une fonction de contrôle de la conformité.

La conformité peut être comprise au sens large, c'est-à-dire qu'elle peut se référer aux objectifs des parties, à leurs principes éthiques, à leurs obligations légales et contractuelles ainsi qu'à leur devoir envers le public. Le contrôle de la conformité peut être accompli par un individu ou encore par une instance précise (comité d'éthique ou d'audit).

Le **rôle de responsable de la protection des données** est un exemple de plus en plus courant où un individu exerce cette fonction. Ce rôle a d'abord été formalisé dans le cadre du Règlement général sur la protection des données (RGPD), lequel exige la désignation d'un délégué à la protection des données (DPD) pour les organes et les institutions de l'Union européenne. Ce délégué est une autorité indépendante et experte en matière de protection des données et occupe un rôle central dans l'assurance de la conformité aux lois relatives à la protection des renseignements personnels (Règlement [UE] 2016/679 du parlement européen et du conseil, 2016).

Le DPD a plusieurs responsabilités relatives à la protection de la vie privée au sein de l'organisation, le Contrôleur européen de la protection des données (s.d.) recense notamment :

- Faire en sorte que les responsables du traitement et les personnes concernées soient informés de leurs droits, obligations et responsabilités en matière de protection des données et qu'ils soient sensibilisés à ceux-ci;
- Fournir conseils et recommandations à l'institution quant à l'interprétation ou l'application des règles relatives à la protection des données;
- Créer un registre des traitements des données effectués au sein de l'institution;
- Veiller au respect de la protection des données au sein de son institution et aider celle-ci à rendre des comptes en la matière;
- Attirer l'attention de l'institution sur tout défaut de conformité avec les règles applicables en matière de protection des données.

Il est intéressant de noter qu'au Québec, les organisations du secteur privé pourraient bientôt devoir nommer des responsables de la protection des données. Bien qu'il soit encore dans sa phase initiale d'adoption, le projet de loi no 64 (2020) propose de « créer la fonction de responsable de la protection des renseignements personnels au sein des entreprises et d'exiger de ces dernières que les paramètres des produits ou services technologiques qu'elles utilisent pour recueillir des renseignements personnels assurent, par défaut, le plus haut niveau de confidentialité sans aucune intervention de la personne concernée » (p.3).

Au-delà de l'assurance du respect des lois, le contrôle de la conformité peut comporter d'autres aspects. Premièrement, la définition de la conformité pour l'ensemble des partenaires passe par l'adoption de codes de pratique, de codes d'éthique, de politiques d'utilisation, etc. puis à leur communication au sein du partenariat. Cet élément nécessite souvent l'exercice d'une vigie sur les développements juridiques et les normes en vigueur pour que les pratiques du partenariat soient constamment à jour. Ensuite, des mesures de contrôle doivent être mises en place à des points précis du cycle de vie des données pour évaluer les pratiques au quotidien ou les activités proposées pour voir si elles s'alignent sur les

politiques et les règles existantes. Cet élément peut également comporter la mise en place d'audits ou de tests de sécurité périodiques pour vérifier que les autorisations d'accès sont respectées, par exemple. Finalement, le renforcement des capacités par l'offre de formations et la mise en place des communautés de pratiques permettent de hausser le niveau de compétence technique et de littératie numérique des partenaires et ainsi réduire les risques de bris de conformité.

Une bonne pratique en matière de contrôle de la conformité est d'établir une séparation entre cette fonction et les instances décisionnelles stratégiques en particulier dans les partenariats public-privé. Par exemple, en ce qui concerne les données « à haute valeur » (telles que les données sur le comportement humain), des entreprises membres peuvent être fortement incitées à utiliser ces données sans tenir suffisamment compte des implications éthiques ou juridiques. Une telle séparation permet alors de garantir qu'il existe un lieu où ces dernières questions peuvent être débattues (Coutts et Gagnon-Turcotte, 2020).

En somme, les mécanismes structurels de gouvernance des données, tels que les organismes de surveillance de la conformité et les DPD, peuvent jouer un rôle essentiel dans un partenariat de données numériques pour rendre visible le principe d'imputabilité. Toutefois, pour remplir réellement leurs fonctions, ces organismes et ces personnes doivent être en mesure de remonter à la source des décisions et retracer le chemin parcouru par les données.

ENCADRÉ 20 : BANQUE DE DONNÉES SAIL

La banque de données SAIL est un dépôt de données anonymisées sur la santé des individus et de la population du Pays de Galles au Royaume-Uni couvrant une période pouvant atteindre deux décennies et auquel les chercheurs peuvent accéder. Sous réserve de garanties et d'approbations, les données peuvent être reliées entre elles pour répondre à des questions de recherche.

Le comité indépendant *Information Governance Review Panel (IGRP)* supervise les accès à la banque de données SAIL. Ce comité est composé de représentants de divers organismes et secteurs gouvernementaux ainsi que du public. En plus de fournir des orientations et des conseils indépendants sur les politiques, les procédures et les processus, l'IGRP examine toutes les propositions d'utilisation de la banque de données SAIL pour s'assurer qu'elles sont appropriées et d'intérêt public (Jones et coll., 2017b).

L'auditabilité des décisions

L'activité consistant à documenter les décisions en matière de collecte, traitement et analyse des données est souvent appelée « auditabilité » (*auditability*) (Zook et coll., 2017). Selon Zook et coll., (2017) « l'objectif de l'auditabilité est de documenter clairement le moment où des décisions sont prises afin de pouvoir, si nécessaire, revenir à un ensemble de données précédent et attaquer un problème à sa source (si les stratégies d'anonymisation des données ont été compromises, par exemple) [traduction libre] » (p.7). Retracer la prise de décisions dans les systèmes de données n'est pas une tâche facile, mais c'est néanmoins un domaine d'activité important pour un partenariat de données numériques qui vise à respecter le principe d'imputabilité.

L'auditabilité des décisions implique de documenter non seulement les décisions prises par les individus, mais aussi celles prises par les systèmes automatisés. Alors que de plus en plus d'organisations privées et publiques ont recours aux algorithmes dans le cadre de prise de décisions automatisées, l'auditabilité est une dimension clé d'une littérature émergente s'intéressant à la responsabilité en matière d'intelligence artificielle et de transparence algorithmique (Bertino et coll., 2019 ; Abiteboul et coll., 2016 ; Gasser et Almeida, 2017).

L'auditabilité est étroitement liée à la notion de traçabilité des données. Cette dernière désigne la documentation du chemin emprunté par les données à travers différents systèmes et logiciels ou à travers plusieurs manipulations. Autrement dit, elle permet de « comprendre comment cette donnée est devenue ce qu'elle est [traduction libre] » (Groth et coll., 2008, p. 250). L'identification de la provenance des données contribue à fournir une piste d'audit des données, à déterminer l'attribution et la propriété des données et à améliorer la qualité des données (de Lusignan et coll. 2011). Des métadonnées de bonne qualité sont essentielles à garantir la traçabilité des données et l'auditabilité générale du système, notamment lorsqu'elles permettent de suivre l'accès aux données et les modifications. Intégrer à des registres internes, ces informations aident à garantir que les données suivent leur cycle de vie comme prévu (Allen et Cervo, 2015).

Une autre approche décrite dans la littérature en gouvernance des données pour accroître le contrôle et l'auditabilité des données tout au long de leur cycle de vie est la création de « politiques collantes » (*sticky policies*) (Pearson et Casassa-Mont,

ENCADRÉ 21 : LES REGISTRES ALGORITHMIQUES

Les algorithmes jouent un rôle de plus en plus important dans la prestation des services publics dans les villes, de sorte qu'il est de plus en plus important pour les citoyens d'accéder aux informations sur les algorithmes utilisés et sur la manière dont ils sont utilisés. Sur ce front, Helsinki et Amsterdam sont actuellement les deux premières villes au monde à avoir des registres publics, dans lesquels les rouages internes de leurs algorithmes sont soigneusement expliqués. Les deux registres donnent un aperçu de chaque système, ainsi que des détails supplémentaires sur les données qu'ils utilisent, leur logique de fonctionnement et la gouvernance des applications. Les registres comportent actuellement une poignée d'algorithmes, mais elles représentent néanmoins un pas considérable vers une meilleure transparence et une plus grande responsabilité face au recours à la prise de décision automatisée (Johnson, 2020).

2011). Les politiques collantes sont des conditions et des contraintes associées (« collées ») aux données, lisibles par les systèmes et les logiciels et qui limitent la façon dont ces données peuvent être traitées. Les politiques collantes définissent les utilisations autorisées et les obligations liées aux données lorsque celles-ci sont transmises d'une partie à l'autre. Cela permet aux utilisateurs de mieux contrôler les données. Par exemple, un dossier médical transmis d'un hôpital à un institut de recherche, puis à une équipe de recherche, peut prendre une forme où certains de ses attributs (par exemple, les résultats médicaux et les renseignements personnels comme le nom, l'adresse, etc.) sont chiffrés, et où une politique collante explique comment certaines parties de ce dossier peuvent être utilisées (Pearson et Casassa-Mont, 2011).

Dans l'ensemble, nous constatons qu'il existe de nombreuses manières de garantir l'auditabilité de la gestion des données. Peu importe la méthode choisie, dans l'ensemble la mise en place de mesures visant à documenter la façon dont les partenaires choisissent de collecter, produire, traiter ou accéder les données permettent de renforcer l'imputabilité.

En fin de compte, la conformité et l'auditabilité ne sont que deux dimensions de l'imputabilité en matière de gouvernance des données. Alors que se multiplient les initiatives tournées vers le bien commun, nous pouvons espérer que d'autres dimensions de l'imputabilité telle que l'évaluation des impacts ainsi que la participation citoyenne prendront une place de plus en plus importante dans la littérature sur la gouvernance des données.

Comme nous l'avons vu dans ce chapitre, un partenariat de données numériques dans l'intérêt public repose sur trois principes, dont la responsabilité, l'efficacité et l'imputabilité. Une grande variété de mécanismes peuvent être utilisés pour traduire ces principes en pratique, qu'elles soient de nature structurelle, procédurale ou relationnelle. Dans le chapitre suivant, nous passons à des expériences concrètes d'organisations qui s'intéressent au partage de données ou qui l'ont déjà expérimenté. Ainsi, nous observons des liens entre les thèmes abordés dans ce chapitre et les expériences réelles en gouvernance des données d'acteurs montréalais.

CHAPITRE 4

PERSPECTIVES MONTRÉALAISES

Comme un de nos objectifs de recherche était la détermination des principaux facteurs de succès, des obstacles et des risques associés aux partenariats de données numériques, nous voulions entendre des personnes qui ont une expérience et un intérêt réel dans ce domaine.

Par conséquent, pendant l'été de 2020, nous avons organisé des entretiens avec 8 représentants et experts d'organisations montréalaises qui sont impliquées ou ont un intérêt dans les partenariats de données numériques. Les participants ont un large éventail d'expériences. Certains proviennent du secteur des arts et de la culture, tandis que d'autres sont directement impliqués dans *Montréal en commun* (l'annexe A présente une liste complète des participants).

L'objectif de cette recherche ne consistait pas à obtenir un échantillon représentatif de parties intéressées. Nous nous sommes plutôt lancés dans ces entretiens avec un esprit exploratoire, en utilisant une approche semi-structurée. Cette approche nous a permis de découvrir avec les participants divers points de vue et perspectives de la gouvernance des données dans l'écosystème de Montréal. Bien que les participants proviennent de domaines et de secteurs différents, nous avons constaté qu'il en émerge un certain nombre de thèmes communs. Ces thèmes et leur analyse constituent l'essentiel de ce chapitre.

Les lecteurs remarqueront que bon nombre des sujets abordés ici ont été évoqués dans les chapitres précédents de ce rapport, ce qui nous a permis de confirmer que les orientations de la recherche documentaire reflétaient les véritables enjeux et approches que les organisations de l'écosystème de Montréal traitent actuellement.

Différentes conceptions de la gouvernance des données

Pour commencer, la plupart de nos entretiens ont porté sur la gouvernance des données : la signification de la gouvernance des données dans le cadre du travail des personnes interrogées, des projets de leur organisation et de leur secteur.

Les discussions avec les participants ont révélé qu'**il n'y a pas de définition commune ou unique de la gouvernance des données**, et que les interprétations dépendent du secteur, de l'organisation et du rôle de l'individu dans l'organisation. Lorsqu'on a demandé aux participants « que signifie pour vous la gouvernance des données ? », un certain nombre de perspectives ont émergé. Les participants ont donné les définitions suivantes de la gouvernance des données :

- Cadres et processus internes de gestion des données qui permettent d'assurer la qualité des données ou de contrôler l'accès à l'information.
- Cadres et processus communs permettant de mutualiser les données ou de les partager entre plusieurs partenaires.
- Politiques, procédures et mesures de sécurité visant à assurer la conformité aux lois et protéger la confidentialité des données.
- Aptitude d'un ensemble d'acteurs à adhérer à un ensemble de règles d'utilisation des données et à prendre des décisions communes.

Nous avons constaté que certains participants (par exemple, ceux qui avaient des rôles de gestion des données dans leur organisation) ont souligné les éléments plus traditionnels ou corporatifs de la

gouvernance des données, c'est-à-dire la gestion des données et le respect de la conformité aux lois et aux politiques internes. D'autres, qui avaient de l'expérience en matière de partenariats de données numériques, présentaient une vision beaucoup plus large de la gouvernance des données, la décrivant en termes de cadres et de processus qui permettent à de multiples organisations de partager des données et de prendre des décisions collectives en la matière.

Culture de la donnée et capacité organisationnelle

Tous les participants ont souligné que la culture de l'organisation en matière de données est un facteur important lorsqu'on participe à des initiatives de partage des données impliquant plusieurs acteurs. Cette culture organisationnelle de l'utilisation des données est influencée par de nombreux facteurs, comme le secteur d'activité de l'organisation, sa taille, ses choix technologiques et l'attitude de ses employés envers les données.

En particulier, certains participants conçoivent la culture en matière de données comme un reflet de la culture organisationnelle. Ce fut le cas lors de notre conversation avec Jean-Sébastien Bélanger, chef du Service aux membres et à la clientèle, Musée des beaux-arts de Montréal, qui a souligné l'importance de réduire les silos de données au sein de l'organisation afin d'assurer un accès adéquat à l'information aux différents services et de créer plus de valeur.

> *« Ici, au musée, on a décloisonné l'accès des données [...] C'est sûr que s'il y a des silos très, très hermétiques, je vais prendre des exemples, des musées ont un logiciel de gestion de la fondation, ils ont un logiciel de gestion de la billetterie, avec des données à chaque fois qui sont générées de manière différente [...] Bref là tu te ramasses avec toute sorte de données, déjà là c'est un enjeu, nous on l'a pas, parce qu'on a pris la décision de ne pas faire ça [...] À cause des données justement, on a décidé d'en prendre juste un (logiciel) parce qu'on voulait avoir des données vraiment utilisables par tous, puis que tout le monde comprenne la structure des données, puis sache comment les interroger, les utiliser. »*

En outre, nos discussions ont permis de conclure que la culture organisationnelle est un facteur important qui crée des conditions favorables aux partenariats de données numériques. Audray Fontaine, coordonnatrice en transfert des savoirs, Centre de recherches interdisciplinaires en études montréalaises (CRIEM), a expliqué que le projet du Pôle de données sociales dans le cadre de *Montréal en commun* vise à partager et à superposer des données provenant de multiples sources officielles et non officielles afin de « mieux informer les prises de décisions de la Ville et de leurs partenaires sur différents enjeux sociaux ». Il sera difficile d'y parvenir, car la culture en matière de données varie d'une organisation à l'autre, ce qui signifie que des parties impliquées dans le projet peuvent être moins disposées ou moins ouvertes au partage des données en raison de la façon dont l'organisation les apprécie et les perçoit.

De plus, certains participants ont établi des liens entre la culture en matière de données et la compréhension que l'organisation a de ses propres données. Par exemple, des entretiens avec des acteurs du secteur des arts et de la culture ont permis de conclure qu'avant de participer à un projet de mutualisation de données numériques, les organisations doivent bien comprendre leur actif en matière de données (les caractéristiques des données à leur disposition, les mesures de protection à prendre, etc.). À l'occasion, des désaccords internes peuvent surgir quant à la définition même de ce qui est une donnée pour l'organisation. Cette compréhension commune dans l'organisation a été considérée comme une première étape importante pour découvrir la valeur des données dans le cadre d'un partenariat.

De notre côté, les questions sont nombreuses ; par exemple, si nous partageons nos données, même pour la cause la plus noble, n'y a-t-il pas toujours un risque de défaillance, de fuite de données ? Quels sont les garanties ou recours ? Et à partir du moment où l'on se fait voler des données, d'usage de surcroît, qu'advient-il ? Ce sont des enjeux très importants qui sont soulevés. Surtout si l'on considère que les données d'usage ne nous appartiennent pas à la base. Techniquement, une donnée de contact appartient à la personne qui a fourni cette information. À partir de là, pouvons-nous réellement partager ces données ?

- Anastasia Vaillancourt, directrice du développement, Culture pour tous

Des sujets complexes aux contours flous

Comme les renseignements personnels, la vie privée et la propriété des données font l'objet de débats importants dans la littérature sur la gouvernance des données, il n'est pas surprenant de constater que ces sujets ont fait l'objet de discussions clés au cours des entretiens. En raison de leur complexité, ces sujets ont été considérés comme des obstacles ou des freins à la participation aux partenariats de données numériques.

À ce sujet, les participants ont souligné *la nature dynamique des données personnelles*. Par exemple, Sophie Tremblay, avocate et cheffe de l'exploitation, Novalex, qui a fourni des conseils juridiques dans le cadre de projets de mutualisation de données dans le secteur des arts et de la culture, a souligné que même si certaines données ne permettent pas d'identifier facilement une personne, leur contexte et la façon dont elles sont liées à d'autres données peuvent créer des conditions qui permettent d'identifier des personnes.

Cette observation a de vastes implications pour les partenariats de données numériques. Par exemple, une étape cruciale pour les organisations qui participent à projet de mutualisation de données consiste à déterminer les cadres législatifs qui s'appliquent à leurs données ainsi que les mesures de protection à mettre en œuvre pour assurer la confidentialité des données. Les organisations peuvent avoir besoin de soutien juridique pour les aider à clarifier ces distinctions.

De plus, les participants aux entretiens ont reconnu que le partage des données dans le cadre d'un partenariat soulève des questions et des préoccupations pour les organisations, qui peuvent dépasser les préoccupations liées aux violations de la confidentialité des données. Par exemple, en discutant de la participation de son organisation à un projet de mutualisation de données, Anastasia Vaillancourt, directrice du développement, Culture pour tous, a indiqué que son organisation devait relever de nombreux défis. Elle a dit :

> *« De notre côté, les questions sont nombreuses ; par exemple, si nous partageons nos données, même*

pour la cause la plus noble, n'y a-t-il pas toujours un risque de défaillance, de fuite de données ? Quels sont les garanties ou recours ? Et à partir du moment où l'on se fait voler des données, d'usage de surcroît, qu'advient-il ? Ce sont des enjeux très importants qui sont soulevés. Surtout si l'on considère que les données d'usage ne nous appartiennent pas à la base. Techniquement, une donnée de contact appartient à la personne qui a fourni cette information. À partir de là, pouvons-nous réellement partager ces données ? »

Elle souligne que les risques concernant la confidentialité et la sécurité des données deviennent plus complexes lorsque la propriété des données n'est pas claire. Ce manque de clarté peut constituer un obstacle à la poursuite des partenariats des données, car il rend difficile l'attribution des responsabilités concernant les données et les résultats qui en découlent.

La détermination du propriétaire des données est une tâche complexe, car elle exige non seulement l'interprétation de diverses lois, mais aussi la compréhension de l'attitude des individus à l'égard des données. Notre entretien avec Frédéric Julien, directeur, recherche et développement, Association canadienne des organismes artistiques (CAPACOA), a mis en évidence ces aspects attitudinaux de la propriété des données :

« Même s'il s'agit d'une information qui est autrement disponible de façon libre sur le Web, du moment où il y a eu un effort de codification, la personne qui a saisi sa donnée dans un système a l'impression que sa donnée lui appartient, et on pourra débattre si ça lui appartient oui ou non [...] La codification (de l'information) entraîne un sentiment de propriété sur la donnée. Et la propriété tend à refermer les gens, puis à bloquer des initiatives qui sinon faciliteraient une réutilisation de la donnée. Donc il y a finalement un élément attitudinal là-dedans qui est non négligeable. »

Même s'il s'agit d'une information qui est autrement disponible de façon libre sur le Web, du moment où il y a eu un effort de codification, la personne qui a saisi sa donnée dans un système a l'impression que sa donnée lui appartient, et on pourra débattre si ça lui appartient oui ou non [...] La codification (de l'information) entraîne un sentiment de propriété sur la donnée. Et la propriété tend à refermer les gens, puis à bloquer des initiatives qui sinon faciliteraient une réutilisation de la donnée. Donc il y a finalement un élément attitudinal là-dedans qui est non négligeable.

- **Frédéric Julien,** directeur, recherche et développement, Association canadienne des organismes artistiques (CAPACOA)

Coût et valeur des données sous-estimés

En général, avant de pouvoir partager et mutualiser des données, il peut être nécessaire de procéder à un certain nombre d'étapes de collectes et de nettoyages de données. Les participants ont souligné que ces étapes et ces processus de gestion des données à l'interne, qui visent à assurer une qualité élevée des données, peuvent exiger beaucoup de temps et un investissement considérable en ressources humaines et techniques.

Par exemple, pour Patrick Joly, directeur général, Société de gestion de la Banque de titres de langue française (BTLF), les « coûts » engendrés par la collecte, le nettoyage et la maintenance de données et de métadonnées de bonne qualité ont tendance à être sous-estimés. Son organisation a partiellement assumé ces coûts et la responsabilité correspondante dans le cadre de son rôle d'agrégateur de données dans le secteur du livre.

Reconnaissant des incohérences dans le formatage des données et des « silos » dans les modes de production des données et des métadonnées dans le secteur, BTLF a récemment établi une politique visant à guider les organisations dans la structuration et le formatage des données commerciales et des livres. À terme, BTLF espère améliorer la qualité des données que tous les acteurs de la chaîne de valeur peuvent utiliser pour soutenir une meilleure veille stratégique.

Pour certains participants comme Frédéric Julien, dans un environnement concurrentiel, les coûts engendrés au cours de la création et de la maintenance des données peuvent devenir un obstacle à la participation aux partenariats de données numériques. Autrement dit, lorsqu'une organisation investit beaucoup de temps et de ressources dans la maintenance de ses données, elle peut être moins disposée à les partager librement.

Données liées et interopérabilité sémantique

Ensuite, les sujets des données liées et de l'interopérabilité sémantique ont été les fils conducteurs de plusieurs entretiens. Comme discuté au chapitre 3, l'interopérabilité sémantique traite « des vocabulaires partagés et d'un langage commun au moyen de modèles, de définitions et d'attributs communs, avec des résultats comme des registres, des taxonomies, des vocabulaires et des ontologies [traduction libre] » (Open Data Institute, 2018).

Les données liées constituent un outil concret pour réaliser l'interopérabilité sémantique, car elles permettent aux éditeurs de données de soutenir des applications de découverte et d'intégration de données (Schmachtenberg, Bizer et Paulheim, 2014). Pour ce faire, ils relient des éléments de données à un vocabulaire contrôlé et partagé, qui établit des liens entre les contenus Web lisibles par l'homme et les métadonnées lisibles par la machine.

Dans le secteur des arts et de la culture, CAPACOA conduit actuellement un ensemble d'initiatives concernant les données ouvertes liées pour lancer la découverte numérique dans son domaine d'activité. CAPACOA et le Conseil québécois du théâtre conduisent une initiative visant à accroître la présence des arts de la scène au moyen de Wikidata — un outil en ligne qui peut servir de source commune de données liées concernant les gens et les lieux, les événements historiques, les conditions socio-économiques et la culture (Marino et Neto Costa, 2020).

Les données ouvertes liées peuvent également être un outil pertinent pour *Montréal en commun*. Par exemple, FabMob QC, une organisation qui conduit des efforts dans le domaine des données sur la mobilité, assure depuis longtemps la documentation et le partage de ses projets au moyen des outils

La ville intelligente, ça ne dépend pas que de senseurs vidéo, capteurs placés partout à travers la ville. Pour moi, au-delà de l'acquisition technique de la donnée, pour qu'une ville soit vraiment intelligente, il faut que cette donnée-là soit décloisonnée. [...] Ça nécessite une interopérabilité sémantique et aussi une interopérabilité technique pour que l'usager qui a besoin de la donnée puisse y avoir accès en temps opportun. La ville intelligente, pour être vraiment intelligente, devra être décentralisée. Ça ne peut pas se faire autrement.

- Frédéric Julien, directeur, recherche et développement, Association canadienne des organismes artistiques (CAPACOA)

du Web sémantique. L'organisation gère également un site Wikidata, qui vise à « capitaliser tous les projets, retours d'expériences et les erreurs, pour faire émerger une culture commune de l'innovation dans l'action ». Elsa Bruyère, co-fondatrice, FabMob QC, a souligné le potentiel de ces outils sémantiques pour aider à diffuser les activités et les résultats de *Montréal en commun* tout en évitant les silos de données :

> *« Si on a quand même une sémantique et un format de sémantique identique, par exemple en appliquant le Resource Description Framework (RDF), quelqu'un qui ferait une requête pourrait retrouver tous nos résultats dans le cadre du défi, sans avoir à se soucier de savoir sur quelle plateforme il va chercher, sur quel site web il va chercher. À partir d'un des sites Web, il pourrait remonter à d'autres choses. Donc ça nous permettrait d'avoir un croisement beaucoup plus large que ce qu'on a aujourd'hui. Parce que sinon, on risque d'être dans des silos de sites. »*

L'importance des données liées ne se mesure pas seulement à leur rôle dans la diffusion des résultats de *Montréal en commun*. Des liens plus larges pourraient être établis entre l'interopérabilité sémantique et les villes intelligentes. Frédéric Julien a déclaré ceci à ce sujet :

> *« La ville intelligente, ça ne dépend pas que de senseurs vidéo, capteurs placés partout à travers la ville. Pour moi, au-delà de l'acquisition technique de la donnée, pour qu'une ville soit vraiment intelligente, il faut que cette donnée-là soit décloisonnée. [...] Ça nécessite une interopérabilité sémantique et aussi une interopérabilité technique pour que l'usager qui a besoin de la donnée puisse y avoir accès en temps opportun. La ville intelligente, pour être vraiment intelligente, devra être décentralisée. Ça ne peut pas se faire autrement. »*

Intérêt pour les partenariats de données numériques

Bien que les partenariats de données numériques soulèvent des questions juridiques, éthiques et opérationnelles complexes, les participants ont exprimé un intérêt soutenu pour l'exploration de diverses formes de partenariats de données numériques qui sont fondés sur la collaboration, qui génèrent de la valeur et qui s'inscrivent dans la recherche de l'intérêt général.

En particulier, la mutualisation de données et les approches décentralisées du partage de données (par exemple, les données ouvertes liées) ont été reconnues comme des pistes prometteuses, en particulier pour les personnes interrogées du secteur des arts et de la culture. En effet, les participants ont sélectionné un certain nombre d'initiatives différentes comme sources d'inspiration. Il s'agit, par exemple, de projets pilotes de mutualisation de données de Synapse C ou de l'idée d'une « fiducie d'utilité sociale », qui est actuellement à l'étude par Territoires innovants en économie sociale et solidaire (Marchand, 2019).

Montréal en commun conduit aussi à explorer de nouveaux partenariats de données numériques dans l'intérêt du public. Dans les centres de données du système alimentaire et de mobilité, divers projets de collaboration émergent et exigeront des modèles adaptés de gouvernance des données. Par exemple, Récolte lance actuellement un projet de création d'une infrastructure de systèmes alimentaires partagés, qui pourrait finalement lier des données de sources diverses pour assurer le suivi des équipements et les produits alimentaires. Par ailleurs, le Pôle de données sociales est actuellement en train d'explorer la création d'un environnement sécurisé de partage des données entre des agences publiques. Pendant que les partenaires de *Montréal en commun* testent et expérimentent diverses approches, les participants expriment un intérêt et une préoccupation marqués quant à la manière d'assurer la longévité de leurs initiatives, au-delà du financement offert par le programme. Les participants ont aussi exprimé leur intérêt pour en apprendre plus sur le potentiel de différents modèles d'affaires de gouvernance des données et des données ouvertes, qui peuvent offrir des avantages d'intérêt public.

Autres facteurs de succès des partenariats de données numériques

En plus d'avoir permis de confirmer l'intérêt envers les partenariats de données numériques, les entretiens avec les participants ont permis d'identifier des facteurs de succès potentiels de ces initiatives.

- Il est essentiel de s'assurer que toutes les parties intéressées soient alignées et aient une **vision commune** du partenariat de données numériques et de ses objectifs. Reconnaissant que les organisations qui participent à de tels partenariats peuvent avoir des cultures de données différentes, les participants ont estimé que l'adoption d'un **vocabulaire commun par les parties** était une des clés du succès de l'initiative.
- Les parties doivent également développer une **compréhension commune** de ce qu'implique leur participation au partenariat de données numériques ainsi que le partage des **bénéfices** qui peuvent être générés. Des participants, en particulier ceux qui sont engagés dans des projets de mutualisation de données conduits par Synapse C, ont mentionné que les bénéfices potentiels comprennent la production de perspectives sectorielles qui permettent aux organisations de prendre de meilleures décisions et fondées sur des données probantes.

- Les organisations doivent acquérir la **capacité interne** de gérer et de comprendre leurs données avant de participer à des partenariats de données numériques, tandis que la présence d'un **champion des données** au sein de l'organisation peut aider à consolider la gouvernance et la culture des données dans l'organisation et favoriser son engagement dans de tels partenariats.
- Plusieurs participants ont fait remarquer qu'actuellement, le public a une confiance très limitée dans la capacité de nos institutions à gérer et à protéger correctement nos données. Ce faible niveau de confiance du public est probablement exacerbé par certains événements récents et les débats publics en cours, dont la fuite de données de Desjardins (Benessaieh, 2020), et la promotion actuelle d'applications de suivi des contacts pour freiner la propagation de COVID-19, qui suscite des inquiétudes concernant la surveillance gouvernementale (La Presse canadienne, 2020).

Il apparaît donc que **la confiance du public** demeurera un facteur de succès essentiel pour encourager les partenariats de données numériques. La participation citoyenne au sein des initiatives, ainsi que la sensibilisation et l'éducation du public en matière de données, tout comme la poursuite de la transparence au moyen de rapports publics et d'activités d'évaluation seront des éléments clés de toutes futures initiatives.

- Finalement, il est préférable que les partenariats de données numériques soient **dirigés et soutenus par un organe de gouvernance dédié** (par exemple, un groupe de travail ou un comité de gouvernance) et bénéficient du soutien de **ressources expertes** en matière technique et juridique, en particulier quand les partenariats regroupent des acteurs ayant un faible niveau de capacité interne. Les participants aux projets les plus avancés réussissent néanmoins à adopter des politiques de référence et à établir des ententes entre les parties pour définir des normes et des conditions d'utilisation des données.

Au final, les entretiens ont couvert un éventail de sujets, de la compréhension de la gouvernance des données à l'interopérabilité sémantique en passant par les données liées. Bien que nous n'ayons pas pu utiliser les résultats des entretiens pour valider les conclusions individuelles de notre analyse documentaire, ce que nous avons découvert au cours des entretiens a montré comment les questions complexes ainsi que les risques et les défis associés à la gouvernance des données se répercutent sur les gens, dans leur travail quotidien et dans leurs organisations à Montréal.

CONCLUSION

Il est évident que la mise en œuvre de la gouvernance des données n'est pas une tâche facile. De la part des partenaires impliqués, elle exige une prise de décision dans des domaines variés incluant la protection des renseignements personnels, la gestion des accès, l'évaluation des risques, la qualité des données, et plus encore. La gouvernance de données dans le cadre d'un partenariat des données numériques est un processus complexe et continu qui exige des négociations et des compromis pour harmoniser des objectifs variés, favoriser la prise de décision collective et la collaboration, tout en attribuant adéquatement les responsabilités dans des domaines qui comportent des éléments à la fois humains et techniques.

Nos recherches montrent que les partenariats de données numériques sont des véhicules utiles pour mettre les technologies et les données au service du public. Ils ont toutefois émergé dans un contexte en rapide évolution de sorte qu'il n'existe pas de méthode unique pour organiser leur gouvernance. La combinaison de structures, de processus et de mécanismes relationnels nécessaire à une gouvernance des données responsable et efficace dépend en fait de multiples facteurs et des conditions existantes dans lesquelles ils s'inscrivent. Les décisions relatives aux mécanismes les plus appropriés et à leur mise en œuvre doivent donc être prises par les parties concernées en fonction de leur contexte, leurs besoins et leurs objectifs. Nos recherches ont démontré qu'un partenariat caractérisé par une dynamique collaborative adoptera des approches de co-construction et d'expérimentation pour y parvenir.

Nous constatons que de plus en plus d'initiatives cherchent à générer des bénéfices sociétaux par la mise en commun des données, tout en reconnaissant leur valeur publique et la possibilité de les gérer de façon collective. Les partenariats de données dans l'intérêt public sont des initiatives qui se démarquent par le fait que les partenaires s'engagent à créer des bénéfices tangibles pour le public, déploient des stratégies de participation citoyenne et adhèrent à des principes forts en matière de gouvernance des données.

À cet égard, les partenariats de données numériques s'assurent de préserver le bien commun en s'appuyant sur trois principes : la responsabilité, l'efficacité et l'imputabilité. D'abord, pour assurer la confiance du public et la légitimité de leurs initiatives, les partenariats de données numériques doivent mettre en place toutes les mesures nécessaires pour traiter les données de manière éthique et responsable. Ensuite, les partenariats de données numériques doivent mettre en place des mécanismes de gouvernance qui favorisent une gestion efficace et cohérente des données. Finalement, nous caractérisons l'imputabilité comme une variété de mécanismes par lesquels une reddition de compte claire et transparente peut avoir lieu à destination des parties prenantes et du public.

À travers une série d'entrevues éclairantes, nous avons pu valider la manière dont la gouvernance des données s'intègre à la gestion quotidienne dans les organisations montréalaises. Nos interlocuteurs ont mis en évidence les obstacles et les défis relatifs aux données qui se présentent à eux, notamment en ce qui concerne la protection de la vie privée, la propriété des données, leur qualité et leur interopérabilité. Ils ont témoigné que pour réussir, les partenariats de données numériques ont besoin d'un soutien et d'une expertise spécifiques, ainsi que d'une direction forte et d'un organe décisionnel dédié.

Dans l'ensemble, nos recherches démontrent que, malgré les difficultés, l'établissement de partenariats de données numériques pour servir des objectifs d'intérêt public présente un intérêt considérable. Alors que les gouvernements, le secteur privé, les universités et les organisations à but non lucratif et caritatives continuent d'explorer de nouveaux

partenariats de données numériques à Montréal et ailleurs, nous concluons ce rapport par quelques apprentissages clés tirés de notre recherche, qui nous l'espérons contribueront à leur succès futur.

1. Reconnaître que l'intérêt public est défini et négocié par les citoyens

L'intérêt public est une chose vivante qui existe dans un état de constante délibération et de négociation. En tant que tel, il peut être difficile à cerner. Dans certains cas, un consensus clair sur la nature de l'intérêt public peut émerger. Toutefois, lorsque de nouvelles questions sociales ou technologiques sont concernées — par exemple, l'utilisation de l'intelligence artificielle — l'intérêt public peut être moins clair.

Un partenariat de données numériques qui intègre la participation publique dans sa gouvernance de manière soutenue et continue a plus de chances de percevoir des changements dans l'opinion publique et de garantir que les objectifs du partenariat de données numériques s'alignent sur les besoins des citoyens. De plus, un partenariat de données numériques sera plus crédible et aura plus de chances de créer des avantages clairs pour le public s'il développe ses objectifs en collaboration avec un large éventail de voix : femmes, communautés indigènes, immigrants récents, résidents à faible revenu et autres groupes.

2. Investir du temps dans votre processus collaboratif et d'expérimentation

L'établissement d'un partenariat solide en matière de données exige du temps et des efforts soutenus de la part de tous les partenaires. Ceux-ci doivent faire un investissement initial pour bâtir la confiance, démontrer une volonté de collaborer, développer une compréhension commune des problèmes auxquels l'initiative s'attaque et se regrouper autour d'objectifs communs.

Ce cycle de collaboration encourage également la volonté d'expérimenter, de développer de nouveaux projets pilotes et d'exploiter leur impact pour créer de nouveaux moyens de valoriser les données. L'élaboration de cas d'usage est un moyen utile pour commencer à explorer le potentiel des données qui font l'objet d'une potentielle mise en commun pour déterminer quels avantages publics peuvent être générés.

3. Créer une gouvernance des données adaptée à vos besoins

On retrouve beaucoup de publicité autour de certains « modèles » de gouvernance des données qui promettent d'être la solution pour une utilisation éthique, efficace et responsable des données, tels que les fiducies des données ou encore les collectifs de données. Mais la réalité est qu'il n'existe pas de modèle standard pour gérer un partenariat de données. Même si un partenariat est intéressé par l'un de ces modèles, il faut du temps et des efforts pour mettre en place les structures, les procédures et les relations qui seront les plus adaptées à son contexte. La détermination des mécanismes appropriés pour un partenariat de données numériques dépend de la compréhension de plusieurs facteurs, tels que le niveau de gouvernance, les caractéristiques des données, et le champ d'application concerné. La bonne nouvelle, c'est que cet effort sera récompensé par un cadre de gouvernance des données adapté aux besoins des partenaires.

Malgré l'absence de modèle clé en main, il existe des cadres juridiques prometteurs dans le contexte des partenariats de données numériques. Par exemple, la fiducie d'utilité sociale (FUS) est un véhicule juridique

particulier au droit civil québécois qui répond à une volonté d'appropriation commune et de gouvernance collective (Marchand, 2019). Ce cadre juridique attire l'intérêt de nombreux chercheurs actuellement et pourrait faire l'objet d'expérimentations futures, malgré les nombreuses questions qu'il continue de soulever, à savoir notamment si les données peuvent constituer une forme de propriété qui pourrait bénéficier de la désignation de FUS.

Cet exemple démontre que les partenariats de données numériques ne peuvent se développer en vase clos. Plusieurs exigences en matière de gouvernance des données sont déjà définies dans les lois, règlements ou normes existants. Les lois québécoises et canadiennes sur la protection de la vie privée notamment établissent un cadre auquel la gouvernance des données ne peut déroger. Pour de nombreux commentateurs, ces lois sont dépassées et devront être révisées par les législateurs. Cette situation offre néanmoins un espace d'innovation pour les organisations ambitieuses.

4. Documenter votre impact et partager vos succès

Il est très utile de savoir dans quelle mesure le partenariat de données numériques a atteint son objectif. Tout d'abord, le suivi et la communication des progrès permettront aux parties prenantes locales et au public de rester engagés dans l'initiative. Deuxièmement, le fait de documenter les conditions de départ, les mesures prises, les défis rencontrés et les leçons apprises aidera les partenaires à s'adapter et à améliorer leur initiative au fil du temps. Troisièmement, en racontant l'histoire de l'initiative, les partenaires apporteront une contribution inestimable aux communautés de pratique locales, régionales et mondiales, qui pourront en tirer des enseignements et s'inspirer. La recherche sur ce qui fonctionne le mieux pour les partenariats de données numériques en est encore à ses débuts, ce qui démontre l'importance de documenter les cas d'utilisation, les succès, les échecs et les impacts de ces initiatives, afin de garantir que nous puissions continuer à nous appuyer sur les succès.

ANNEXE

Liste de personnes interrogées

- Sophie Tremblay, avocate et cheffe de l'exploitation, Novalex
- Jean-Sébastien Bélanger, chef du Service aux membres et à la clientèle, Musée des beaux-arts de Montréal
- Patrick Joly, directeur général, Société de gestion de la Banque de titres de langue française
- Anastasia Vaillancourt, directrice du développement, Culture pour tous
- Frédéric Julien, directeur, recherche et développement, Association canadienne des organismes artistiques
- Elsa Bruyère, co-fondatrice, FabMob QC
- Audray Fontaine, coordonnatrice en transfert des savoirs, Centre de recherches interdisciplinaires en études montréalaises
- Lorenzo Daïeff, chargé de projet, SALIM Défi des villes intelligentes, Récolte

BIBLIOGRAPHIE

Abiteboul, S., Miklau, G., Stoyanovich, J., & Weikum, G. (2016). Data, Responsibly. *Dagstuhl Reports, 6*(7), 73. http://www.dagstuhl.de/dagpub/2192-5283

Abraham, R., Schneider, J. et vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424-438. 10.1016/j.ijinfomgt.2019.07.008

Abrams, M. (2014). The Origins of Personal Data and its Implications for Governance. *SSRN Electronic Journal*. 10.2139/ssrn.2510927

Allen, M. et Cervo, D. (2015). *Multi-domain master data management: advanced MDM and data governance in practice*. Morgan Kaufmann.

Ana Brandusescu, Michael Canares et Silvana Fumega. (2020, 21 août). Open data standards design behind closed doors? *ILDA*. https://idatosabiertos.org/diseno-de-estandares-de-datos-abiertos-a-puertas-cerradas/

Ansell, C. et Gash, A. (2007). Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory, 18*(4), 543-571. 10.1093/jopart/mum032

Artyushina, A. (2020). Is civic data governance the key to democratic smart cities? The role of the urban data trust in Sidewalk Toronto. *Telematics and Informatics, 55*, 101456. 10.1016/j.tele.2020.101456

Bass, T., et Old, R. (2020). *Common Knowledge : Citizen-led data governance for better cities (DECODE)*. Nesta. https://www.nesta.org.uk/report/common-knowledge-citizen-led-data-governance-better-cities/

Bass, T., Sutherland, E. et Symons, T. (2018). *Reclaiming the Smart City: Personal data, trust and the new commons*. https://www.nesta.org.uk/report/reclaiming-smart-city-personal-data-trust-and-new-commons/

Battiste, M. (2008). Research Ethics for Protecting Indigenous Knowledge and Heritage: Institutional and Researcher Responsibilities. Dans *Handbook of Critical and Indigenous Methodologies*. SAGE Publications, Inc. 10.4135/9781483385686

Benessaieh, K. (2020, 19 juin). *Vol de données chez Desjardins: la catastrophe, un an plus tard*. La Presse. https://www.lapresse.ca/affaires/entreprises/2020-06-19/vol-de-donnees-chez-desjardins-la-catastrophe-un-an-plus-tard

Bennett, C. J. et Raab, C. D. (2018). Revisiting the governance of privacy: Contemporary policy instruments in global perspective: Revisiting the governance of privacy. *Regulation & Governance*. 10.1111/rego.12222

Bertino, E., Merrill, S., Nesen, A. et Utz, C. (2019). Redefining Data Transparency: A Multidimensional Approach. *Computer, 52*(1), 16-26. 10.1109/MC.2018.2890190

Bhargava, R. et D'Ignazio, C. (2015, 30 juin). *Designing Tools and Activities for Data Literacy Learners*. Data Literacy Workshop at ACM Web Science Conference, Oxford, Royaume-Uni. https://www.media.mit.edu/publications/designing-tools-and-activities-for-data-literacy-learners/

Bolychevsky, I., Ruhaak, A., Bunting, M., McMillan, A., Cameron, S., Voznick, M. et Pasquarelli, W. (2019). *Exploring the potential of data trusts in reducing food waste*. Open Data Institute. https://docs.google.com/document/d/1v9O-3exRdZFu6h-xqo11Ej3Ul4cyePfRCYYBDrpUoNBk/edit

Budin-Ljøsne, I., Teare, H. J. A., Kaye, J., Beck, S., Bentzen, H. B., Caenazzo, L., Collett, C., D'Abramo, F., Felzmann, H., Finlay, T., Javaid, M. K., Jones, E., Katić, V., Simpson, A. et Mascalzoni, D. (2017). Dynamic Consent: a potential solution to some of the challenges of modern biomedical research. *BMC Medical Ethics, 18*(1), 4. 10.1186/s12910-016-0162-9

Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X. et Greene, C. S. (2020). Responsible, practical genomic data sharing that accelerates research. *Nature Reviews Genetics.* 10.1038/s41576-020-0257-5

Calzada Prado, J. et Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri, 63*(2). 10.1515/libri-2013-0010

Cavoukian, A. (2009). Privacy by Design : The 7 Foundational Principles. https://www.ipc.on.ca/wp-content/uploads/Resources/7foundationalprinciples.pdf

Cavoukian, A. et Dixon, M. (2013). *Privacy and Security by Design: An Enterprise Architecture Approach.* Information and Privacy Commissioner of Ontario. https://www.ipc.on.ca/wp-content/uploads/Resources/pbd-privacy-and-security-by-design-oracle.pdf

Centre for Information Policy Leadership. (2014). *A Risk-based Approach to Privacy: Improving Effectiveness in Practice.* https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/white_paper_1-a_risk_based_approach_to_privacy_improving_effectiveness_in_practice.pdf

Commissaire à l'information et à la protection de la vie privée de l'Ontario. (2018, juillet). *Privacy Fact Sheet: General Data Protection Regulation.* https://www.ipc.on.ca/wp-content/uploads/2018/07/fs-privacy-gdpr.pdf

Commissariat à la protection de la vie privée du Canada. (2017). *Aperçu des lois sur la protection des renseignements personnels au Canada.* https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/02_05_d_15/

Commissariat à la protection de la vie privée du Canada. (2018, 24 mai). *Lignes directrices pour l'obtention d'un consentement valable.* https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/collecte-de-renseignements-personnels/consentement/gl_omc_201805/

Commissariat à la protection de la vie privée du Canada. (2019a). *Principes relatifs à l'équité dans le traitement de l'information de la LPRPDE.* https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/la-loi-sur-la-protection-des-renseignements-personnels-et-les-documents-electroniques-lprpde/p_principle/

Commissariat à la protection de la vie privée du Canada. (2019b). *Survol de la LPRPDE.* https://www.priv.gc.ca/fr/sujets-lies-a-la-protection-de-la-vie-privee/lois-sur-la-protection-des-renseignements-personnels-au-canada/la-loi-sur-la-protection-des-renseignements-personnels-et-les-documents-electroniques-lprpde/lprpde_survol/

Commission d'accès à l'information du Québec. (s. d.). *Un renseignement personnel, c'est quoi ?* https://www.cai.gouv.qc.ca/entreprises/un-renseignement-personnel-cest-quoi/

Commission européenne. (2020). Communication de la commission au parlement européen, au conseil, au comité économique et social européen et au comité des régions : Une stratégie européenne pour les données. https://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX%3A52010DC0245

Conrad, E., Misenar, S. et Feldman, J. (2016). *CISSP study guide* (Third edition). Elsevier, Syngress.

Consumer Data Research Centre. (2020). *Protecting Data.* https://data.cdrc.ac.uk/protecting-data

Le Contrôleur européen de la protection des données (s.d.). *Délégué à la protection des données.* https://edps.europa.eu/data-protection/data-protection/reference-library/data-protection-officer-dpo_fr

Coutts, S. et Gagnon-Turcotte, S. (2020). *Data Governance and Digital Infrastructure: Analysis and Key Considerations for the City of Toronto.* Open North. https://www.toronto.ca/wp-content/uploads/2020/08/95fb-2020-07-10-Open-North-Data-Governance-Report-Main-report-WEB.pdf

Curty, R. G. et Qin, J. (2014). Towards a model for research data reuse behavior. *Proceedings of the American Society for Information Science and Technology, 51*(1), 1-4. 10.1002/meet.2014.14505101072

D'Addario, J., Dodds, L., Brown, W. et Maddison, J. (2020). *Sharing data to create value in the private sector.* Open Data Institute. https://theodi.org/article/report-sharing-data-to-create-value-in-the-private-sector/

de Lusignan, S., Liaw, S.-T., Krause, P., Curcin, V., Vicente, M. T., Michalakidis, G., Agreus, L., Leysen, P., Shaw, N. et Mendis, K. (2011). Key concepts to assess the readiness of data for international research: data quality, lineage and provenance, extraction and processing errors, traceability, and curation. Contribution of the IMIA Primary Health Care Informatics Working Group. *Yearbook of Medical Informatics, 6*, 112-120. http://epubs.surrey.ac.uk/188358/

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M. et Blondel, V. D. (2013). Unique in the Crowd: The privacy bounds of human mobility. *Scientific Reports, 3*(1), 1376. 10.1038/srep01376

Du Perron, S. (2020a, 17 juin). Projet de loi 64 : une réforme à l'Européenne du droit à la protection des renseignements personnels. *Laboratoire de cyberjustice.* https://www.cyberjustice.ca/2020/06/17/projet-de-loi-64-une-reforme-a-leuropeenne-du-droit-a-la-protection-des-renseignements-personnels/

Du Perron, S. (2020b, 24 novembre). Le temps des réformes : cinq comparaisons entre le projet de loi n° 64 et le projet de loi C-11. *Autonomisation des acteurs judiciaires par la cyberjustice.* https://ajcact.openum.ca/2020/11/24/le-temps-des-reformes-cinq-comparaisons-entre-le-projet-de-loi-n-64-et-le-projet-de-loi-c-11/

Earley, S., Henderson, D. et Data Management Association (dir.). (2017). *DAMA-DMBOK: data management body of knowledge* (2nd edition). Technics Publications.

Elliot, M., Mackey, E. et O'Hara, K. (2020). *The Anonymisation Decision Making Framework 2nd Edition: European Practitioners' Guide.* UKAN. https://msrbcel.wordpress.com/framework/

Faniel, I. M. et Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW), 19*(3-4), 355-375. 10.1007/s10606-010-9117-8

Faundeen, J. L., Burley, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., Hutchison, V. B., Martín, E., Montgomery, E. T., Ladino, C. C., Tessler, S. et Zolly, L. S. (2014). *The United States Geological Survey Science Data Lifecycle Model* ([U.S. Geological Survey Open-File Report] no 2013–1265). http://dx.doi.org/10.3133/ofr20131265

Gal, M. S. et Rubinfeld, D. L. (2019). Data Standardization. *New York University Law Review, 94*, 737-770. https://www.nyulawreview.org/issues/volume-94-number-4/data-standardization/

Gasser, U. et Almeida, V. A. F. (2017). A Layered Model for AI Governance. *IEEE Internet Computing, 21*(6), 58-62. 10.1109/MIC.2017.4180835

Gellert, R. (2016). We Have Always Managed Risks in Data Protection Law: Understanding the Similarities and Differences Between the Rights-Based and the Risk-Based Approaches to Data Protection. *European Data Protection Law Review, 2*(4), 481-492. 10.21552/EDPL/2016/4/7

Girard, M. (2018). *Canada Needs Standards to Support Big Data Analytics* ([Policy Brief] no 145). Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/documents/PB%20no.145web.pdf

Groth, P., Munroe, S., Miles, S. et Moreau, L. (2008). Applying the Provenance Data Model to a Bioinformatics Case. Dans L. Grandinetti (dir.), *High Performance Computing and Grids in Action* (p. 250-264). IOS Press.

Groupe des politiques et de la recherche du Commissariat à la protection de la vie privée du Canada. (2016). *Consentement et protection de la vie privée : Document de discussion sur les améliorations possibles au consentement sous le régime de la Loi sur la protection des renseignements personnels et les documents électroniques.* Commissariat à la protection de la vie privée du Canada. https://www.priv.gc.ca/media/1807/consent_201605_f.pdf

Guidoin, S., Marczak, P., Pane, J. et McKinney, J. (2016). *Identifying recommended standards and best practices for open data.* Open North & ILDA. http://geothink.ca/wp-content/uploads/2016/02/Identifying-Recommended-Standards-Open-Data-Open-North.pdf

Hardinges, J., Wells, P., Blandford, A., Tennison, J. et Scott, A. (2019). *Data trusts: Lessons from three pilots.* Open Data Institute. https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLU-T3oOobnYJGibEhspr2v87jg/

IEEE. (1990). Interoperability. Dans *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*. IEEE.

Infocomm Media Development Authority of Singapore et Personal Data Protection Commission. (2019). *Trusted Data Sharing Framework.* Government of Singapore. https://www.imda.gov.sg/-/media/Imda/Files/Programme/AI-Data-Innovation/Trusted-Data-Sharing-Framework.pdf

Involve, Understanding Patient Data and the Carnegie UK Trust et Scott, K. (2018). *Data for Public Benefit: Balancing the Risks and Benefits of Data Sharing.* Involve, The Carnegie UK Trust, & Understanding Patient Data. https://www.involve.org.uk/sites/default/files/field/attachemnt/Data%20for%20Public%20Benefit%20Report_0.pdf

Jacobson, I., Spence, I. et Bittner, K. (2011). USE-CASE 2.0: The Guide to Succeeding with Use Cases. https://www.ivarjacobson.com/sites/default/files/field_iji_file/article/use-case_2_0_jan11.pdf

Janssen, M., Estevez, E. et Janowski, T. (2014). Interoperability in Big, Open, and Linked Data--Organizational Maturity, Capabilities, and Data Portfolios. *Computer, 47*(10), 44-49. 10.1109/MC.2014.290

Järvinen, T. L. N., Sihvonen, R., Bhandari, M., Sprague, S., Malmivaara, A., Paavola, M., Schünemann, H. J. et Guyatt, G. H. (2014). Blinded interpretation of study results can feasibly and effectively diminish interpretation bias. *Journal of Clinical Epidemiology, 67*(7), 769-772. 10.1016/j.jclinepi.2013.11.011

Johnson, K. (2020, 28 septembre). Amsterdam and Helsinki launch algorithm registries to bring transparency to public deployments of AI. *VentureBeat.* https://venturebeat.com/2020/09/28/amsterdam-and-helsinki-launch-algorithm-registries-to-bring-transparency-to-public-deployments-of-ai/

Jones, K. H., Laurie, G., Stevens, L., Dobbs, C., Ford, D. V. et Lea, N. (2017a). The other side of the coin: Harm due to the non-use of health-related data. *International Journal of Medical Informatics, 97*, 43-51. 10.1016/j.ijmedinf.2016.09.010

Jones, K. H., Ford, D. V. et Lyons, R. A. (2017b). *The SAIL Databank: 10 years of spearheading data privacy and research utility, 2007-2017.* Swansea University. https://saildatabank.com/wp-content/uploads/

SAIL_10_year_anniversary_brochure.pdf

Kaye, J., Whitley, E. A., Lund, D., Morrison, M., Teare, H. et Melham, K. (2015). Dynamic consent: a patient interface for twenty-first century research networks. *European Journal of Human Genetics, 23*(2), 141-146. 10.1038/ejhg.2014.71

Khatri, V. et Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148-152. 10.1145/1629175.1629210

Klievink, B., van der Voort, H. et Veeneman, W. (2018). Creating value through data collaboratives: Balancing innovation and control. *Information Polity, 23*(4), 379-397. 10.3233/IP-180070

Kum, H.-C. et Ahalt, S. (2013). Privacy-by-Design: Understanding Data Access Models for Secondary Data. AMIA Joint Summits on Translational Science Proceedings. *AMIA Joint Summits on Translational Science, 2013*, 126-130. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3845756/

Loi sur la protection des renseignements personnels et les documents électroniques, LC 2000, c 5. https://laws-lois.justice.gc.ca/fra/lois/p-8.6/

Marchand, M.-A. (2019). *Les fiducies d'utilité sociale : synthèse de connaissances.* Territoires innovants en économie sociale et solidaire (TIESS). https://bit.ly/FUS-synthese_pdf

Marino, V. et Joana Neto Costa. (2020, 22 juin). The Wikidata project for the performing arts is on! – LDFI. https://linkeddigitalfuture.ca/2020/06/22/the-wikidata-project-for-the-performing-arts-is-on/

Micheli, M., Ponti, M., Craglia, M. et Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society, 7*(2), 205395172094808. 10.1177/2053951720948087

Montenegro, M. (2019). Subverting the universality of metadata standards: The TK labels as a tool to promote Indigenous data sovereignty. *Journal of Documentation, 75*(4), 731-749. 10.1108/JD-08-2018-0124

Mozilla Insights. (2020). *Shifting Power Through Data Governance.* Mozilla Insights. https://foundation.mozilla.org/en/initiatives/data-futures/data-for-empowerment/

Office québécois de la langue française. (1999). Protection de la confidentialité. Dans *Le grand dictionnaire terminologique (GDT).* http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2074344

Office québécois de la langue française. (2000). Propriétaire de fichier. Dans *Le grand dictionnaire terminologique (GDT).* http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=2074764

Office québécois de la langue française. (2004). Donnée. Dans *Le grand dictionnaire terminologique (GDT).* http://gdt.oqlf.gouv.qc.ca/ficheOqlf.aspx?Id_Fiche=8358482

Open Data Institute. (2018). *Types of open standards for data.* Open Standards for Data Guidebook. https://standards.theodi.org/introduction/types-of-open-standards-for-data/

Open Data Institute. (s. d.). *The Data Spectrum.* https://theodi.org/about-the-odi/the-data-spectrum/

Open Data Institute. (2019). *Data trusts: Lessons from three pilots.* https://theodi.org/article/odi-data-trusts-report/

Opendatasoft. (2020). *Choisir et décrire vos métadonnées : nos conseils pour rendre vos données découvrables, réutilisables et interopérables.* https://www.opendatasoft.com/fr/livre-blanc-metadonnees

Organisation de coopération et de développement économiques (OCDE). (2019). *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies.* OECD Publishing.

https://doi.org/10.1787/276aaca8-en

Ostrom, E. (1990). *Governing the commons: the evolution of institutions for collective action.* Cambridge University Press.

Otto, B. (2011). *A Morphology of the Organisation of Data Governance.* European Conference on Information Systems (p. 272). https://aisel.aisnet.org/ecis2011/272

Le Parlement européen et le Conseil européen. (2016). Le Règlement (UE) 2016/679 [le Règlement général sur la protection des données]. https://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=EN

Pearson, S. et Casassa-Mont, M. (2011). Sticky Policies: An Approach for Managing Privacy across Multiple Parties. *Computer,* 44(9), 60-68. 10.1109/MC.2011.225

Peer, L., Green, A. et Stephenson, E. (2014). Committing to Data Quality Review. *International Journal of Digital Curation, 9*(1), 263-291. 10.2218/ijdc.v9i1.317

Peña Gangadharan, S. et Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society, 22*(7), 882-899. 10.1080/1369118X.2019.1593484

Porcaro, K. (2020). *Failure Modes for Data Stewardship.* Mozilla Insights. https://mzl.la/2Zu34tH

La Presse canadienne. (2020, 17 juin). *COVID-19 : les applications de traçage, alliées ou ennemies?* Radio-Canada.ca. https://ici.radio-canada.ca/nouvelle/1712841/applications-tracage-coronavirus-debat-experts-canada-pour-contre-vie-privee-sante

Projet de loi n° 64, Loi modernisant des dispositions législatives en matière de protection des renseignements personnels, 1e sess, 42e lég, Québec, 2020. http://www.assnat.qc.ca/Media/Process.aspx?MediaId=ANQ.Vigie.Bll.DocumentGenerique_159567&process=Default&token=ZyMoxNwUn8ikQ+TRKYw-PCjWrKwg+vIv9rjij7p3xLGTZDmLVSmJLoqe/vG7/YWzz

R v Jarvis, 2019 SCC 10, [2019] 1 SCR 488 https://scc-csc.lexum.com/scc-csc/scc-csc/en/item/17515/index.do

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B. et Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report.* SSHRC. http://rgdoi.net/10.13140/RG.2.1.1922.5044

Riley, J. et National Information Standards Organization (U.S.). (2017). *Understanding Metadata: What Is Metadata, and What Is It For?* National Information Standards Organization (U.S.). http://www.niso.org/publications/understanding-metadata-riley

Ritchie, F. (2017). *The 'Five Safes': a framework for planning, designing and evaluating data access solutions.* Data for Policy 2017, Londres. http://dx.doi.org/10.5281/zenodo.897821

Rocher, L., Hendrickx, J. M. et de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications, 10*(1), 3069. 10.1038/s41467-019-10933-3

Royal Academy of Engineering. (2019). *Towards trusted data sharing : guidance and case studies - Data sharing checklist.* https://www.raeng.org.uk/policy/publications-(1)/interactives/data-sharing

Royaume Uni. Information Commissioner's Office. (2019). *Data sharing code of practice.* https://ico.org.uk/media/about-the-ico/consultations/2615361/data-sharing-code-for-public-consultation.pdf

Royaume Uni. Information Commissioner's Office. (s. d.). *Right to data portability.* ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-pro-

tection-regulation-gdpr/individual-rights/right-to-data-portability/

Samarati, P. et de Vimercati, S. C. (2001). *Access Control: Policies, Models, and Mechanisms.* R. Focardi et R. Gorrieri (dir.), Berlin, Heidelberg (p. 137-196). https://doi.org/10.1007/3-540-45608-2_3

Scassa, T. (2018a). *Data Ownership* (no 187). Centre for International Governance Innovation. https://www.cigionline.org/sites/default/files/documents/Paper%20no.187_2.pdf

Scassa, T. (2018b, 7 juin). *Enforcement powers key to PIPEDA reform.* Policy Options. https://policyoptions.irpp.org/magazines/june-2018/enforcement-powers-key-pipeda-reform/

Schmachtenberg, M., Bizer, C. et Paulheim, H. (2014). Adoption of the Linked Data Best Practices in Different Topical Domains. Dans P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. Knoblock, D. Vrandečić, P. Groth, N. Noy, K. Janowicz et C. Goble (dir.), *The Semantic Web – ISWC 2014* (vol. 8796, p. 245-260). Springer International Publishing. 10.1007/978-3-319-11964-9_16

Sebastian-Coleman, L. (2018). *Navigating the labyrinth: an executive guide to data management* (1st edition). Technics Publications.

Shamsi, J. A. et Khojaye, M. A. (2018). Understanding Privacy Violations in Big Data Systems. *IT Professional, 20*(3), 73-81. 10.1109/MITP.2018.032501750

Sheehan, M., Thompson, R., Fistein, J., Davies, J., Dunn, M., Parker, M., Savulescu, J. et Woods, K. (2019). Authority and the Future of Consent in Population-Level Biomedical Research. *Public Health Ethics*, phz015. 10.1093/phe/phz015

Smart Dubai et Nesta. (2020, mars). Data sharing toolkit: approaches, guidance and resources to unlock the value of data. https://www.nesta.org.uk/documents/1832/Data_Sharing_Toolkit_1.pdf

Solove, D. J. (2006). A Taxonomy of Privacy. *University of Pennsylvania Law Review, 154*(3), 477. 10.2307/40041279

Sposito, F. A. (2017). *What do data curators care about? Data quality, user trust, and the data reuse plan.* 2017 IFLA World Library and Information Congress, Wroclaw, Poland (p. 7). http://library.ifla.org/1797/1/S06-2017-sposito-en.pdf

The British Academy, techUK et The Royal Society. (2018). *Data ownership, rights and controls: Reaching a common understanding* ([Discussions at a British Academy, Royal Society and techUK seminar on 3 October 2018]). https://royalsociety.org/~/media/policy/projects/data-governance/data-ownership-rights-and-controls-October-2018.pdf

The British Academy et The Royal Society. (2017). *Data management and use: governance in the 21st century.* https://royalsociety.org/-/media/policy/projects/data-governance/data-management-governance.pdf

The Engine Room. (2016). *The Hand-Book of the Modern Development Specialist: Being a Complete Illustrated Guide to Responsible Data Usage, Manners & General Deportment.* https://the-engine-room.github.io/responsible-data-handbook/assets/pdf/responsible-data-handbook.pdf

Thuermer, G., Walker, J. et Simperl, E. (2019). *Data sharing toolkit: Lessons learned, resources and recommendations for sharing data.* Data Pitch. www.datapitch.eu

Vangen, S. et Huxham, C. (2012). The Tangled Web: Unraveling the Principle of Common Goals in Collaborations. *Journal of Public Administration Research and Theory, 22*(4), 731-760. 10.1093/jopart/mur065

Verhulst, S. G. et Sangokoya, D. (2015, 22 avril). *Data*

Collaboratives: Exchanging Data to Improve People's Lives. Medium. https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

Verhulst, S. G., Young, A., Winowatan, M. et Zahuranec, A. J. (2019). *Leveraging Private Data for Public Good* (p. 57). https://datacollaboratives.org/static/files/existing-practices-report.pdf

Ville de Montréal. (2020, octobre). Charte des données numériques. https://laburbain.montreal.ca/sites/default/files/charte_donnees_numeriques_1_0.pdf

Visceral Visions. (2020). *Culturebrew.art.* https://www.visceralvisions.com/culturebrewart

Wade, M. et Hulland, J. (2004). Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly, 28*(1), 107-142. 10.2307/25148626

Weill, P. (2004). Don't Just Lead, Govern: How Top-Performing Firms Govern IT. *MIS Quarterly Executive, 3*(1), 1-17. https://aisel.aisnet.org/misqe/vol3/iss1/3

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U. et Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society, 18. http://ci-journal.net/index.php/ciej/article/view/1286

Yoon, A. (2017). Data reusers' trust development. *Journal of the Association for Information Science and Technology, 68*(4), 946-956. 10.1002/asi.23730

Zook, M., Barocas, S., Boyd, D., Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A. et Pasquale, F. (2017). Ten simple rules for responsible big data research. *PLOS Computational Biology, 13*(3), e1005399. 10.1371/journal.pcbi.1005399