Haute École
Groupe ICHEC – ECAM – ISFSC

Enseignement supérieur de type long de niveau universitaire

Data Trusts

L'émergence d'un modèle de gestion de données fondé sur la responsabilité fiduciaire et les considérations de data gouvernance y afférentes

Mémoire présenté par :
Grégory TRUONG

Pour l'obtention du diplôme de :
Master en sciences commerciales
Année académique 2020-2021

Promoteur :
Thierry Van den Berghe

Boulevard Brand Whitlock 6 - 1150 Bruxelles

Haute École
Groupe ICHEC – ECAM – ISFSC

Enseignement supérieur de type long de niveau universitaire

Data Trusts

L'émergence d'un modèle de gestion de données fondé sur la responsabilité fiduciaire et les considérations de data gouvernance y afférentes

Mémoire présenté par :
Grégory TRUONG

Pour l'obtention du diplôme de :
Master en sciences commerciales
Année académique 2020-2021

Promoteur :
Thierry Van den Berghe

Boulevard Brand Whitlock 6 - 1150 Bruxelles

Remerciements

Je tiens à adresser mes remerciements aux personnes qui m'ont apporté leur aide pour la réalisation de ce mémoire :

Anna Artyushina
PhD candidate at York University, Toronto, Canada
Faculty of Graduate Studies
Researcher in data trusts and data governance in smart cities.

Dr. Johanna Walker
Research Fellow / Knowledge Partner, Web and Internet Science Group
University of Southampton, UK - Department of Electronics and Computer Science (ECS).
Co-editor of the special issue of the Journal of Community Informatics on Data Literacy

Dr. Kieron O'Hara
Associate Professor in Electronics and Computer Science
University of Southampton, UK - Department of Electronics and Computer Science (ECS).
Research Fellow for the Centre for Policy Studies
Editor of Foundations and Trends in Web Science

Jack Hardinges
Program Lead at the Open Data Institute

à mon maître de mémoire :

Dr Thierry Van Den Berghe
Professeur d'informatique de gestion
ICHEC – Bruxelles Management School

Et pour leur soutien constant :
à Arthur, Anh Dung, Francine, Marie et Sébastien.

Engagement anti-plagiat du mémoire

« Je soussigné, TRUONG Grégory Tri-Anh, 2020 - 2021, déclare par la présente que le mémoire ci-joint est exempt de tout plagiat et respecte en tous points le règlement des études en matière d'emprunts, de citations et d'exploitation de sources diverses signé lors de mon inscription à l'ICHEC, ainsi que les instructions et consignes concernant le référencement dans le texte respectant la norme APA, la bibliographie respectant la norme APA, etc. mises à ma disposition sur Moodle.

Sur l'honneur, je certifie avoir pris connaissance des documents précités et que le travail présenté est original et exempt de tout emprunt à un tiers non cité correctement. »

Dans le cadre du dépôt en ligne, la signature consiste en l'introduction du mémoire via la plateforme ICHEC-Student.

Table des matières

Liste des figures et tableaux

Introduction

Dans une communication datée du 19 février 2020 et intitulée « *Une stratégie européenne pour les données* », la Commission Européenne dévoile les contours de sa stratégie digitale pour les cinq prochaines années.

On y découvre la vaste ambition affichée de créer un marché des données paneuropéen, une économie nourrie de ces données et au-delà, d'ériger un véritable modèle de *data governance* afin d'encadrer leur collecte et leur utilisation.

Les possibles bénéfices évoqués grâce à une meilleure utilisation des données concernent une vaste gamme de secteurs socio-économiques : le secteur privé (gain de productivité, marchés plus concurrentiels) mais également les domaines de la santé, de l'environnement et de la lutte contre le changement climatique, les services publics (rendus plus transparents et plus efficaces) et les transports…

Pour mener à bien sa stratégie, la Commission Européenne (2020) prévoit un plan d'investissement de plusieurs milliards d'euros. Elle évoque également quelques outils à exploiter, telle la désormais célèbre *Blockchain*. Parmi les dispositifs novateurs mentionnés figure aussi, le méconnu *Data Trust* - la fiducie de donnée -, un mécanisme de *data gouvernance* dont les gestionnaires assurent une responsabilité fiduciaire sur le modèle du Trust anglo-saxon.

Hors d'Europe, ce mécanisme du *Data Trust* fait l'objet de propositions et de débats intellectuels depuis déjà plusieurs années, avec des publications académiques régulières, des projets pilotes, des déploiements réussis et des échecs instructifs.

L'objectif de ce mémoire est de faire la clarté sur ce concept de '*Data Trust'* afin de comprendre les raisons qui suscitent l'intérêt pour ce dispositif, et de le situer dans le contexte plus large du marché de la donnée.

Pour ce faire, nous commençons par procéder à une analyse de la donnée comme actif et passons en revue ses caractéristiques spécifiques. Grâce à la typologie des biens d'Ostrom, nous mettons en lumière les zones de tensions entre acteurs et leurs stratégies respectives.

Nous opérons ensuite une revue des questions juridiques relatives à l'application de la notion de propriété aux données, en conservant un point de vue international.

Nous clôturons la première partie de ce travail par une macroanalyse du marché de la donnée, de sa structure et de ses caractéristiques économiques. Les difficultés particulières et la recherche d'alternatives face à la domination, de facto, d'une poignée de géants technologiques sont également abordées.

La seconde partie de ce travail traite spécifiquement du *Data Trust.*

Nous revenons sur l'origine et les fondements intellectuels du concept de *trust* de tradition anglo-saxonne dont le mécanisme s'inspire.

Nous fournissons, ensuite, un historique détaillé permettant de contextualiser l'émergence du concept de *Data Trust* et la manière dont l'idée d'appliquer la notion de responsabilité fiduciaire aux données s'est progressivement construite. Les principaux courants y sont représentés.

Les caractéristiques du *Data Trust* et le rôle des différents acteurs sont expliqués. Une différence est faite entre les applications littérales et figuratives du concept de *trust* comme mécanisme de gestion de données.

Nous nous arrêtons également sur le très controversé projet d'*Urban Data Trust* mené par Alphabet autour de la *smart city* de Toronto, les oppositions citoyennes qu'il a générées et les causes de son échec.

Afin d'illustrer le propos, quelques cas d'applications privés et civiques du mécanisme sont fournis. Les usages particuliers, tels qu'identifiés par les chercheurs, sont scrutés et nous nous attardons en particulier sur son potentiel disruptif en cas d'asymétrie de pouvoir entre acteurs ou lorsque la protection d'intérêts minoritaires est nécessaire.

En raison de la nature non déterministe du mécanisme, et au vu des formes variées au travers desquelles le concept trouve à s'exprimer, quelques exemples supplémentaires sont rapidement présentés pour refléter adéquatement la variété des dispositifs.

Enfin, nous passons en revue deux classifications complémentaires des différents types de *Data Trust*. Pour chaque catégorie de *Data Trust,* nous listons les acteurs en faisant la promotion, leurs objectifs et les enjeux sous-jacents.

Cette classification est également l'occasion d'évoquer rapidement d'autres dispositifs de data gouvernance (data co-ops, data commons, …), ainsi que la notion centrale de confiance qui donne son nom au dispositif et le concept plus large de « licence sociale ».

Nous concluons en notant la grande diversité de dispositifs désignés par l'appellation de *Data Trusts* et en écartant certains de ces dispositifs qui tiennent plus du marketing que de la réelle fiducie de données.

Partie 1 : Une nouvelle économie basée sur les données

1. Les données comme actif

1.1 Le pétrole de l'économie digitale

Il y a quelques années, le magazine américain *Wired* publiait un article intitulé « *Data is the new oil of the digital economy* ».

Son auteur, Joris Toonders (2014), invitait les entreprises à investir dans les infrastructures data en les considérant non plus comme des centres de coût mais bien comme des postes générateurs de profit. Il enjoignait ses lecteurs à traiter désormais les données comme un actif de premier plan, en les comparant aux gisements inexploités de pétrole du 18ème siècle.

À ceux capables de comprendre et d'extraire la valeur de cet actif, il prédisait d'immenses récompenses grâce à l'économie digitale dont la révolution ne faisait que commencer.

Depuis cet article initial paru dans *Wired*, cette comparaison entre le nouveau pétrole et les données réapparait régulièrement en titre de diverses publications, tant généralistes que spécialisées :

- « *The world's most valuable resource is no longer oil, but data* » (The Economist, 2017).
- « *Mastercard's boss just told a Saudi audience that 'data is the new oil'* » (MSNBC, 2017).
- « *Mégadonnées : le nouveau pétrole* » (L'actualité, 2017).
- « *Why data is the new oil* » (Syracuse University - School of Information Studies, 2018)
- « *Data is the new oil* » (Hacker Noon, 2019).
- « *Data is the new oil and that's a good thing* » (Forbes, 2019).
- « *Data is the new oil : Five reasons that's not necessarily a good thing.* (Data Center Knowledge, 2019)
- « *Is data the new oil ? Competition issues in the digital economy* » (European Parliament Think Thank, 2020)
- …

À tel point que cette qualification des data comme « nouveau pétrole » fait désormais partie intégrante de l'imaginaire économique et technologique.

Or, aussi largement diffusée soit-elle, cette comparaison mérite d'être analysée avec esprit critique avant d'être acceptée en l'état.

1.2 Similitudes entre le pétrole et les données

Il existe effectivement un certain nombre d'éléments qui plaident dans le sens d'un parallèle avec le pétrole et sa place stratégique dans l'économie précédente :

- Les données brutes ('raw data' ou 'machine data') sont désormais commercialisées comme des biens économiques. Pour Zech (2016), elles sont devenues une matière première négociée - une *commodity* - dont les applications de Big Data se nourrissent. Les données sont même devenues des ressources cruciales et de grande valeur. D'après Monnerie (2018), leur valeur est en hausse constante. Pour Szczepanski (2020), elles sont désormais si essentielles qu'elles pourraient être considérées comme l'actif le plus précieux de l'économie moderne.

- Pour Wendy Hall, professeure en Sciences de l'Informatique à l'Université de Southampton et Jérôme Pesenti, Vice-Président de l'Intelligence Artificielle chez Facebook (2017), auteurs d'un rapport officiel sur le développement de l'intelligence artificielle à destination du gouvernement britannique, les enjeux économiques et stratégiques sont colossaux. L'objectif du Big Data et du développement de l'IA est de reproduire ou de surpasser les capacités humaines pour la réalisation de certaines tâches, à l'aide des systèmes informatiques.

  L'estimation des bénéficies économiques qu'ils citent dépasse l'entendeur avec des contributions annuelles de l'intelligence artificielle à l'économie qui pourraient atteindre l'équivalent de la production de la Chine et de l'Inde combinée d'ici 2030. Il est également attendu de l'AI, des impacts majeurs dans les secteurs de la santé, de la finance, de l'automobile, de la logistique, de la technologie, des communications, du divertissement, de la vente de détail, de l'énergie et de l'industrie manufacturière. Avec le plus grand potentiel pour le secteur de la santé.

  Comme le pétrole brut, les données ont besoin d'être « raffinées » avant utilisation : d'après Zech (2016), il est courant que les données ne soient pas initialement captées dans un but particulier. Les décisions de procéder à une analyse survenant en général dans un second temps, postérieur à la collecte. Il considère que c'est même là une caractéristique distinctive du Big Data.

  Schlosser (2018), du World Economic Forum, note qu'à l'heure actuelle 10% seulement des données sont collectées dans un format directement exploitable, une quantité dérisoire.

  Comme pour le pétrole, un « raffinage » (nettoyage, mise en forme, maintenance) avant exploitation s'avère donc essentiel dans l'immense majorité des cas, en particulier lorsque l'information est concaténée à partir de banques de données multiples alimentées par des systèmes capteurs épars et fragmentés.

Monnerie (2018) identifie un cycle en trois étapes:

- La captation de la matière première (big data).
- La transformation en ressource économiquement exploitable (smart data).
- L'exploitation (qu'il s'agisse de revente ou d'utilisation par les moteurs de recherche, la publicité ciblée ou d'algorithmes).

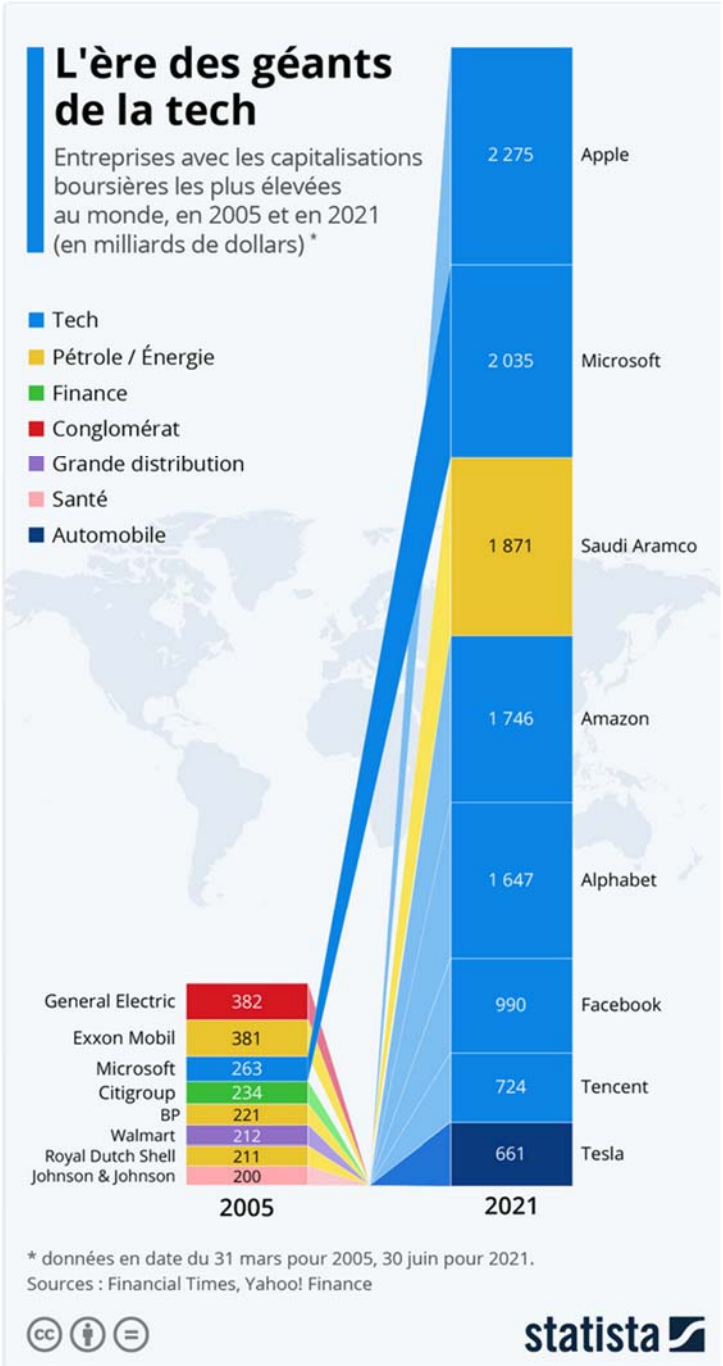

Figure 1 : L'ère des géants de la tech.

Source : Gaudiaut, T. (2021). *Infographie : L'ère des géants de la tech.* Statista Infographies. Consulté le 04 août 2021, à l'adresse https://fr.statista.com/infographie/22656/classement-entreprises-capitalisation-boursiere/

- Aujourd'hui, le sommet du classement mondial des entreprises dotées des plus importantes capitalisations boursières est lourdement dominé par des sociétés actives dans le secteur de la technologie et des données (Apple, Microsoft, Amazon, Alphabet, Facebook et Tencent). Il y a quinze ans, ce palmarès était dominé par un assortiment plus varié d'entreprises et les géants énergétiques (Exxon, General Electric, BP, Royal Dutch Shell) y étaient les plus représentés.

  Cette rotation au sommet et le remplacement des multinationales énergétiques par les GAFA + Microsoft sont cohérents avec un changement de paradigme économique où les données seraient la nouvelle ressource majeure stratégique à la place des anciennes, énergétiques.

- Enfin, les données sont perçues comme un pilier central, un élément fondateur, du prochain cycle techno-industriel.

  Pour la Commission Européenne (2015), il est question de transition vers une industrie 4.0, *smart* et connectée, où les données seraient tout à la fois un facteur de production, un actif et même dans certains cas, une monnaie d'un nouveau type (grâce au développement des cryptomonnaies). Le big data, l'interconnexion des objets (Internet of Things) et les services digitaux (dont le *cloud computing*) étant les piliers de cette nouvelle industrie.

  Pour Qiang (2018), cette révolution 4.0 succédera à l'ère d'Internet et de l'Internet mobile et sera plutôt caractérisée par la combinaison des données de masse (big data) et d'intelligence artificielle.

  Quelle que soit la vision du futur des uns et des autres, on constate que les données de masse y occupent systématiquement une place stratégique et centrale.

« *Just as steam powered much of the First Industrial Revolution, the free flow of data will be fundamental to powering what the World Economic Forum and others are calling the Fourth Industrial Revolution* » (Banga, 2016, p. 3)

On le voit, en esquissant sur base de quelques éléments le portrait d'une ressource centrale nécessaire à l'avènement d'une ère économique nouvelle, il est possible de faire ressortir certains points communs qui semblent corroborer la validité de la comparaison pétrole / data.

Toutefois, aussi séduisante soit cette comparaison, au premier abord, les données et l'économie digitale sont par certains aspects très différentes du circuit économique historique et des matières premières traditionnelles.

Si bien que Schlosser (2018), dans un article pour le World Economic Forum, rejette vigoureusement la comparaison constante entre pétrole et data, et souligne même qu'elle finit par engendrer des traitements inadéquats.

Il estime que le traitement des données comme ressource finie conduit certaines autorités publiques à les amasser dans des silos, à les enfermer derrière des barrières et, ce faisant, à priver les citoyens d'une partie des bénéfices possibles, tant économiques que sociaux, qui pourraient être obtenus par le partage, la diffusion et la mise en commun des données.

C'est que les données disposent également de caractéristiques uniques qui les différencient fondamentalement des autres actifs et sur lesquelles il convient de s'arrêter pour obtenir un aperçu plus nuancé de la situation.

1.3 Caractéristiques spécifiques des données comme actif

Intangibilité : La première et la plus évidente différence entre les matières premières traditionnelles et les données est le caractère intangible de ces dernières.

Pour Szczepanski (2020), la part intangible de l'économie a connu un essor général au cours des 20 dernières années en raison du progrès technologique. Il s'agit ici de la part de l'économie qui est dématérialisée et qui est constituée de flux de données (par exemple du divertissement en streaming, des flux d'information financière entre places boursières, un vaste assortiment de trafic web, etc.). Cette part de l'économie qui ne repose plus des échanges physiques (comme le commerce direct de matières premières, de biens manufacturés ou encore de services en présentiel) est en croissance constante depuis l'avènement des NTIC.

Pour Ciuriak & Wylie (2018), le rôle prédominant de certains actifs à travers l'histoire est le fruit d'une co-évolution en lien avec les technologies cruciales d'une ère. Durant l'époque féodale, lorsque la création de valeur était principalement centrée sur l'agriculture, la terre était le facteur de production essentiel. Avec l'avènement de la révolution industrielle, des machines-outils et des chaînes de production c'est le capital financier et, nous l'avons évoqué, l'énergie qui furent les ressources prépondérantes.

Dans la suite de leur réflexion, Ciuriak & Wylie (2018) estiment que de nos jours, la dématérialisation des actifs va de pair avec l'émergence successive d'une *knowledge-based economy* (dont l'actif majeur est la propriété intellectuelle) puis d'une *data-driven economy* (fondée sur les données et les algorithmes).

L'intangibilité croissante des actifs - dont les données - serait donc la résultante d'une co-évolution des ressources clés en symbiose avec le développement technologique.

De son côté, le *Copenhagen Institute for Futures Studies* (Scharff, 2021), préfère parler de « megatrends » : des vagues de fonds, des tendances de très long terme,

inévitables, qui traversent et transforment durablement l'économie et la société. L'institution en identifie quatorze. Le phénomène de « *dématérialisation* » est l'une d'entre elles, ainsi que l'émergence d'une « *société du savoir* », fondée sur les data, et « *fonctionnant en réseaux* ».

Bien que différentes, ces grilles de lecture partagent la vision globale d'une économie de demain largement dématérialisée et d'une intangibilité croissante des actifs. Une tendance économique de long terme dans laquelle les données, et leur caractère immatériel, occupent une place centrale.

Reproductibilité : Une autre caractéristique particulière des données est le fait qu'elles soient reproductibles à un coût marginal nul ou quasi nul une fois les dépenses initiales effectuées (Mills, 2019 ; Scassa, 2018). La copie des données ne coûte rien (ou presque rien) même si la collecte requiert des investissements de départ qui peuvent être élevés.

Pour Szczepanski (2020), l'innovation dans la plupart des secteurs de l'économie digitale subit de hauts coûts fixes dès le départ mais le fait que la reproduction soit possible à un coût marginal nul, a pour effet de créer des économies d'échelle de plus en plus importantes.

Naturellement, l'existence d'économies d'échelle conséquentes va pousser les entreprises à croître un maximum pour en bénéficier. Si bien que le marché de la donnée finit par prendre la forme d'un oligopole dominé de manière systémique par quelques acteurs historiques ayant atteint une taille critique (Artyushina, 2020 ; Birch et Al., 2020 ; Delacroix & Lawrence, 2019 ; Micheli et Al., 2020 ; Mills, 2019 ; Monnerie, 2018 ; Szczepanski, 2020).

L'accès aux nouveaux entrants est rendu difficile par ces barrières à l'entrée : tant les investissements initiaux que la présence des géants bien en place qui bénéficient d'ores et déjà des économies d'échelle évoquées.

Nous aurons l'occasion de revenir plus en détail sur les caractéristiques économiques du marché de la donnée dans le chapitre correspondant (cf. infra p.37). Pour l'instant, actons simplement cette spécificité supplémentaire de la donnée en tant qu'actif : la reproductibilité à coût marginal quasi nul et son effet amplifiant sur les potentielles économies d'échelle.

Non-déplétion : Contrairement au pétrole qui est une ressource finie, la quantité de données qu'il est possible de générer ne connait pas de limite. Pour Schlosser (2018), il s'agit là d'une différence majeure avec les matières premières traditionnelles qui s'épuisent et dont il existe une quantité limitée. Alors que la valeur du pétrole, exemple type d'une ressource *single use*, provient de sa rareté et des difficultés d'extraction, les données, elles, sont réutilisables, copiables et partageables. Leur exploitation ou leur partage est non-déplétif : ils n'entraînent pas la destruction de la ressource.

Combinables : Pour Mills (2019), des données supplémentaires sont même créées à l'occasion du traitement. En analysant et combinant les données collectées, nous en créons de nouvelles (*implied data*, *derived data*).

Par exemple, grâce à l'assemblage d'une liste de prénoms nous pouvons déduire que le prénom Jean est populaire (*implied data)* et en accolant à cette liste de prénoms des adresses, nous obtenons un jeu de coordonnées complet (*derived data).*

Pour Scassa (2018), ces données dérivées jouent un rôle important dans le contexte du big data et des opérations analytiques. Elles peuvent prendre la forme de profils digitaux complexes ou de données prédictives.

Comme nous le verrons dans la partie consacrée au copyright (cf. infra p.28), cette distinction entre données originales d'un côté et données dérivées ou résultant d'un traitement de l'autre peut être importante en regard du droit de la propriété intellectuelle.

Ces possibilités combinatoires associées aux énormes volumes de données et à la puissance de calcul contemporaine sont la clé qui doit permettre de libérer le plein potentiel économique pressenti de la *data-driven economy* (cf. infra p.32).

Abondance : Le marché contemporain de la donnée est également caractérisé par son abondance. Pour Hall et Pesenti (2017), depuis l'an 2000 nous assistons à une augmentation exponentielle des quantités de données générées globalement. Des volumes de données massifs sont continuellement collectés à un rythme difficile à imaginer et de plus en plus facilement.

En s'appuyant sur des sources provenant de chez IBM, Szczepanski (2020) note qu'en 2016, 90% des données dans le monde avaient été créées lors les deux années précédentes.

La Commission Européenne (2020) estime que la quantité de données globales produites en 2018 était de 33 zettabytes (1 zettabyte = 1 milliard de térabytes), et qu'elle va plus que quintupler d'ici à 2025 pour atteindre un rythme effréné de 175 zettabytes par an.

Pour mettre cette valeur en perspective, il suffit de contempler le fait qu'il aura fallu attendre l'année 2012 pour que la quantité totale cumulée de données électroniques existant dans le monde atteigne 1 zettabyte. Désormais nous produisons cette quantité des dizaines – et bientôt des centaines – de fois, annuellement.

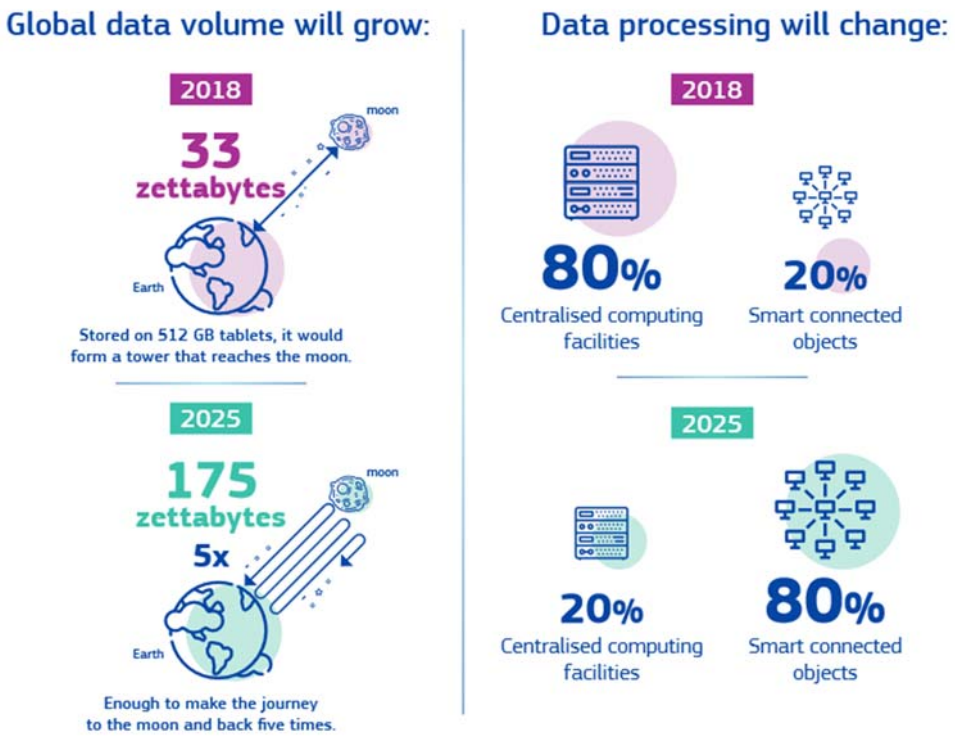

Figure 2 : Évolutions 2018 - 2025 des volumes et de la répartition des données.

Source : European Commission (2021). *European Data Strategy.* Récupéré le 08/12/2020 de https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283

Cette explosion prévue du volume de donnée est à mettre en lien avec la multiplication des appareils connectés et le déploiement progressif de l'Internet of Things (IoT) :

« les modalités de stockage et de traitement des données vont connaître des bouleversements au cours de cinq prochaines années. À l'heure actuelle, 80 % des opérations de traitement et d'analyse des données se déroulent dans des centres de données et des installations informatiques centralisées, et 20 % dans des objets connectés tels que des voitures, des appareils ménagers ou des robots de l'industrie manufacturière, ainsi que des installations informatiques proches de l'utilisateur (*edge computing* ou traitement des données à la périphérie). D'ici à 2025, ces proportions vont probablement s'inverser. » (Commission européenne, 19 février 2020, p.2).

L'augmentation des objets connectés et la décentralisation y afférente vont entraîner une métamorphose profonde du paysage digital.

Paradoxalement, en dépit de l'abondance de la matière première, de nombreuses données sont actuellement inexploitées. C'est qu'en l'absence de structure permettant leur partage, selon des règles qui satisfassent aux intérêts divergents des différentes parties, elles restent bien souvent inaccessibles à ceux qui pourraient les valoriser, dans le cadre de recherches scientifiques ou de projets d'intelligence artificielle.

Pour Reed et Ng (2018), beaucoup de données potentiellement utiles sont ainsi actuellement détenues en silos, tant par les entités publiques que par des organismes privés.

La question centrale du partage :

Reproductibles à coût marginal nul, ne s'épuisant pas lorsqu'elles sont exploitées, créatrices de valeur lorsqu'elles sont combinées à grande échelle, abondantes… par bien des aspects, on se rend compte que les caractéristiques spécifiques des données se prêtent à leur partage et à leur mise en commun. D'autant plus que nous avons, à présent, les moyens de les stocker, de les traiter et de les analyser à échelle industrielle.

Mais pour de multiples raisons (intérêts concurrentiels divergents, crainte de perte d'exclusivité, protection nécessaire des *data subjects*, manque de confiance légitime, absence de *know-how*), elles demeurent trop souvent enfermées en silos et inexploitées.

Or les promesses de création de valeur de cette nouvelle économie basée sur les données dépendent précisément de la capacité à partager les quantités massives de data dont les systèmes de machine learning ont besoin. Et ce, en prenant en considération une multitude d'intérêts divergents, y compris la protection des intérêts du citoyen.

C'est dans ce contexte associé à une défiance croissante envers la poignée de géants technologiques qui dominent de manière systémique les flux de données que l'intérêt pour des modèles alternatifs de data gouvernance a été ravivé.

De nombreux auteurs (Hall et Pesenti, 2017 ; Hardingues et Wells, 2019 ; Micheli et al., 2020 ; Mills, 2019 ; O'Hara, 2019 ; Reed et Ng, 2019) estiment que le *Data Trust* - ce mécanisme de data gouvernance modulable qui offre des possibilités de mutualisation éventuellement assorties d'une responsabilité fiduciaire - pourrait être une réponse à ce problème.

Nous y reviendrons en détail dans la seconde partie de ce travail qui lui est consacrée (cf. infra p.42).

1.4 données dans la typologie des biens

La typologie des biens d'Ostrom

Dans son travail sur la gestion des biens communs, Elinor Ostrom (2010) utilise une typologie considérée comme faisant référence pour catégoriser différents types de biens. Cette classification est particulièrement pertinente pour appréhender les biens économiques dont une composante est commune ou partagée entre différents acteurs.

Deux critères de différenciation sont utilisés :

- Le critère d'exclusion : qui correspond à la facilité ou la difficulté avec laquelle il est possible de priver autrui de l'accès au bien.

- Le critère de rivalité : qui correspond à la mesure dans laquelle la consommation d'un bien par une personne réduit la possibilité d'un tiers à le consommer lui aussi.

Un stylo à bille ou une pomme sont des exemples typiques de biens privés, un type de bien pour lequel les deux critères (exclusion et rivalité) sont élevés. Le propriétaire d'une pomme peut facilement en refuser l'accès à autrui (exclusion). Et s'il la consomme, il n'est plus possible à un concurrent d'en profiter (rivalité).

À l'opposé du spectre, l'éclairage public et la défense nationale sont des exemples de biens publics pour lesquels les deux critères d'exclusion et de rivalité sont faibles. Tous les citoyens peuvent bénéficier d'un bien public, comme la défense nationale, sans pour autant diminuer le bénéfice qu'un tiers en retire (non-rivalité). Ce type de biens est déployé de manière globale au profit d'une collectivité sans qu'il soit aisé d'en exclure un bénéficiaire spécifique (non-exclusion).

Il est possible de représenter sur deux axes, grâce à un tableau, ces critères et les dénominations des types de biens économiques correspondants :

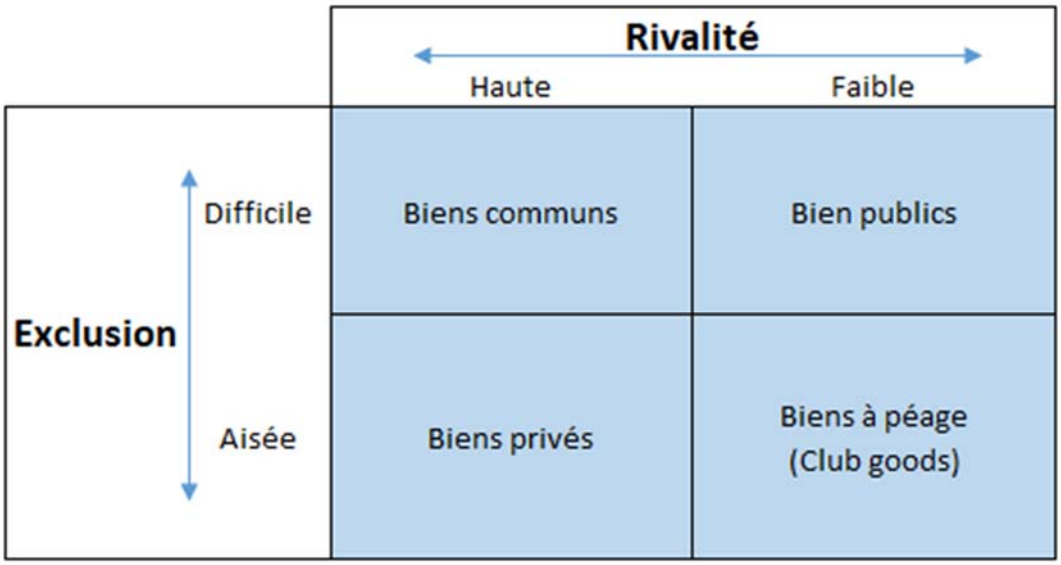

Figure 3 : Classification des biens communs d'après Ostrom (2010).

Les deux derniers types de biens qui présentent des caractéristiques mixtes (basse sur un axe, élevée sur l'autre) sont les biens communs et les biens à péage (parfois appelés bien « club »).

Un exemple souvent utilisé pour illustrer la catégorie des biens communs (exclusion faible, rivalité élevée) est celui du pâturage public, ouvert et accessible à tous les bergers. En l'absence de barrières et de propriété privée, tous peuvent faire paître leurs troupeaux sur ce terrain commun. Il n'y a donc pas d'exclusion. Par contre, le bien est dit « rival » car il existe un risque d'épuisement de la ressource commune si chaque acteur cherche à maximiser son utilité individuelle au détriment du bien commun. C'est ce que Hardin (1968) a appelé *la tragédie des communs* dans un célèbre article économique éponyme.

Un autre exemple de bien commun est celui des eaux internationales et des hautes mers, accessibles à tous les bateaux de pêche. Comme pour le pâturage commun, il n'y a pas d'exclusion, l'accès à la ressource étant ouvert à tous. Le critère de rivalité, lui, est manifeste en raison du risque de surexploitation de la ressource.

Enfin, la télévision câblée est une illustration de bien à péage (ou bien « club »). Il n'existe pas réellement de rivalité (une personne qui regarde la télévision câblée n'en prive pas les autres spectateurs) toutefois les possibilités pratiques d'exclusion sont réelles et elles amènent le distributeur à instaurer un accès conditionné avec péage, ici sur le modèle de l'abonnement (d'autres modèles existent).

Les données dans la typologie des biens

La question qui nous intéresse est de déterminer la place des données, dans cette typologie afin de comprendre plus finement les zones de tension qui existent entre différents acteurs et leurs stratégies respectives.

Nous l'avons vu (cf. supra p.16), les données ont la particularité d'être abondantes, aisément reproductibles et partageables. Si bien qu'il est possible et facile pour plusieurs acteurs d'exploiter simultanément des copies d'un même jeu de données, et ce, sans altérer aucunement la version dont disposent les autres. En ce sens, nous sommes dans le cas de figure parfait d'un bien non-rival dont l'utilité ne décroît pas quand le nombre d'utilisateurs augmente.

Les choses deviennent cependant plus complexes lorsque le critère d'exclusion doit être évalué.

A priori, on pourrait considérer intuitivement que les banques de données sont des biens auxquels s'applique le critère d'exclusion. Après tout, si nous nous mettons un instant dans la position d'un géant technologique, captant quotidiennement de vastes quantités de données par le biais d'un réseau social, de services gratuits ou d'appareils mobiles, on comprend aisément qu'il est facile à cet acteur, s'il le souhaite, de conserver les données collectées sans en partager l'accès ni à un concurrent, ni au public, ni à un quelconque tiers. En ce sens, on pourrait être tenté de considérer qu'il est aisé d'appliquer le critère d'exclusion.

Cette proto-analyse nous placerait dans le cas de figure d'un bien à péage (non-rivalité mais exclusion). Et effectivement, dans leurs relations avec les annonceurs publicitaires, on peut considérer, d'une certaine façon, que Google ou Facebook

mettent à disposition l'accès à leurs données exclusives contre paiement. Ce qui est cohérent avec le modèle économique des biens à péage.

Pour Lau, Penner et Wong (2019), cette grille de lecture n'est cependant pas la plus appropriée et il est préférable de distinguer plusieurs niveaux lorsque l'on considère les caractéristiques des données.

La couche *hardware* : un fichier digital a une existence persistante sur un support de stockage sous la forme de 0 et de 1 qui peuvent être décodés pour présenter l'information de manière lisible. Les supports de stockage physiques peuvent se présenter sous de multiples formes : disques durs d'ordinateur personnel, serveurs dans un data center, mémoire interne de téléphone, mémoires amovibles diverses (cartes SD, micro SD, clés USB), etc.

La couche *content* : l'information a également une existence plus abstraite, plus conceptuelle, qui « flotte librement », indépendamment du support physique et du type d'enregistrement. Une image peut, par exemple, être sauvegardée au format pdf ou au format jpg, avec des niveaux de compression différents, des types d'encodage différents mais conceptuellement nous considérerons qu'il s'agit toujours de la même image. Et ce, malgré l'existence de caractéristiques physiques différentes liées au mode d'enregistrement. Un texte est le même qu'il soit enregistré dans un fichier à l'aide d'un logiciel de traitement de texte ou imprimé dans un livre. Une chanson est la même chanson qu'elle soit enregistrée sur un support numérique, fredonnée en rue ou qu'elle soit diffusée à la radio, grâce aux ondes hertziennes. Une série de chiffres griffonnés manuellement sur un morceau de papier ou encodés dans un ordinateur représentent conceptuellement les mêmes valeurs arithmétiques.

Il existe donc différents niveaux, des couches distinctes, et l'on peut constater que ces niveaux ont des propriétés différentes selon la grille d'analyse des types de biens d'Ostrom.

Comme la plupart des biens économiques, la couche *hardware* présente les caractéristiques tout à fait classiques d'un bien de type privé (exclusion et rivalité).

Considérée séparément, la couche *content* (c'est-à-dire la donnée en elle-même) présente, elle, les caractéristiques d'un bien public (non-rivalité, non-exclusion).

Au niveau de la couche *content*, il n'existe pas de rivalité car les données sont abondantes, aisément reproductibles et partageables comme nous l'avons précédemment souligné.

Le critère d'exclusion est également absent : on estime qu'il n'est pas possible d'empêcher quelqu'un d'utiliser un bien de l'esprit. L'exemple typique est celui de certaines œuvres d'art, au caractère immatériel. Par exemple, tout un chacun peut apprendre une chanson et la fredonner ou apprendre une histoire et la réciter. Il n'est pas possible d'exclure les agents, *au niveau de la couche content*.

Le même raisonnement peut s'appliquer, à l'identique, pour les données digitales. Le critère de non-exclusivité des biens digitaux devient flagrant lorsque l'on pense aux échanges décentralisés, en peer-to-peer et aux vains efforts de l'industrie musicale, pendant des années, pour mettre un terme aux partages illégaux sur

internet. Il est possible de restreindre l'accès au support physique mais à partir du moment où l'information est déjà disponible (en particulier depuis l'avènement d'internet), exclure l'accès à des données qui ont été rendues publiques se révèle extrêmement difficile.

Au niveau de la couche *content*, les données sont donc, fondamentalement, un bien public (non-rivalité, non-exclusion).

Les modèles *Creative Commons* ou l'*Open Data* sont des exemples concrets du traitement des données comme biens publics. La Belgique s'est d'ailleurs dotée d'un projet Open Data à l'initiative du gouvernement (https://data.gov.be). Celui-ci fournit d'innombrables données sur tous les secteurs d'activité de la société. Dans le cadre de la lutte contre la pandémie de COVID-19, Sciensano rend également publiques les données de santé en Open Data (https://epistat.wiv-isp.be/covid/).

Neil Lawrence (2020), professeur *Deepmind* de machine learning à l'Université de Cambridge et membre sénior de l'Alan Turing Institute, estime, lui aussi, que les données sont des biens publics. Il l'affirme explicitement.

Il considère toutefois qu'il faut pousser le raisonnement encore plus loin lorsque l'on exploite des données à caractère personnel (au sens de l'article 4 §1 du GDPR, les données à caractère personnel sont celles qui se rapportent à une personne physique identifiée ou identifiable, directement ou indirectement. Car dans ce cas, aux questions traditionnelles d'exclusion et de rivalité, s'ajoutent les enjeux relatifs à la protection de la vie privée et le respect des droits conférés par le GDPR depuis son entrée en vigueur (ou par des législations similaires, hors d'Europe).

Dans le cas des données à caractère personnel, la série d'intérêts différents qui entrent en considération est donc encore plus complexe : nous sommes en présence d'un bien public, mais auquel s'ajoutent des règles additionnelles qui représentent les droits et intérêts de parties prenantes tierces (les *data subjects*).

Dès lors, dans le cas où l'on exploite des données à caractère personnel, Lawrence (2020) estime que le plus adéquat est d'avoir recours aux méthodes de gouvernance habituellement utilisées pour la gestion des biens communs, parfois appelés *commons* (rivalité, non-exclusion), et ce malgré le fait que les données soient, par essence, des biens publics.

Cette gestion des biens communs, dont Lawrence (2020) recommande d'employer les méthodes est une problématique sur laquelle Ostrom (2010) a longtemps travaillé (ce qui lui a valu le prix Nobel d'économie). Il s'agit de déployer des méthodes de gouvernance démocratique et participative *ad hoc*, de concevoir des règles d'exploitation sur mesure, de créer et gérer une communauté de *stakeholders*, de veiller à ce que les droits de ces *stakeholders* soient respectés et appliqués par les autorités extérieures, de fournir un accès abordable à des méthodes de résolution de conflit et d'utiliser des sanctions graduelles et des exclusions en cas de non-respect des règles communes (Roman, 2021).

Un exemple concret d'utilisation de ces règles de gouvernance, inspirées des *commons*, pour la gestion des données dans un *data trust* est présenté dans la seconde partie de ce travail (cf. infra p. 86).

Conclusion

Les acteurs peuvent avoir des visions différentes de la donnée qui entrent en conflit les unes avec les autres. L'application de la grille de lecture d'Ostrom nous permet toutefois de comprendre que les données sont fondamentalement des biens publics (pour ce qui est de la couche *content*). Les exemples d'exploitation selon ce modèle correspondent à ceux de l'Open Data.

La couche Hardware qui est le support physique d'un enregistrement particulier de la donnée est, elle, un bien privé. Pour les géants technologiques s'assurer du contrôle des supports physiques permet d'avoir la maîtrise des flux de données, de les thésauriser et les commercialiser comme s'il s'agissait d'un bien à péage.

Pour Birch et al. (2020), le but des grandes entreprises du techno-capitalisme est d'acquérir une position dominante et d'obtenir de cette façon, une « rente » financière à partir des données captées. C'est pourquoi ils se concentrent sur le contrôle et la monétarisation des flux de données. Cet objectif prioritaire se fait d'ailleurs parfois aux dépens du droit à la vie privée des citoyens. Pour eux, cette volonté d'accumulation des données personnelles afin d'en obtenir un loyer est la cause structurelle pour laquelle, un scandale comme celui de Cambridge Analytica a pu se produire.

Enfin, bien que les données soient fondamentalement des biens publics, l'existence de droits particuliers lorsqu'il est question de données à caractère personnel fait qu'un mode de gouvernance plus complexe peut être requis.

Plusieurs auteurs (Lawrence, 2020 ; Mills, 2019 ; Paprica et Al., 2021) recommandent d'utiliser des méthodes de gouvernance participatives inspirées de la gestion des biens communs (les *commons*) pour tenir compte de cette complexité additionnelle liée au respect de la vie privée des citoyens.

2 Considérations juridiques : propriété de l'information et droits relatifs

Le développement rapide d'une économie basée sur les données pose la question de savoir qui « possède » ces données et ce que cela signifie réellement (Mills, 2019 ; Scassa, 2018 ; Zech, 2016). En effet, le fait de considérer les données comme des actifs économiques s'accompagne de certaines incertitudes spécifiques sur le plan juridique.

Les questions relatives à la propriété des données, aux droits d'usage liés, au partage de ces données et à la protection des sujets sont celles qui nous intéressent ici (les questions légales spécifiques relatives à la structure du *data trust* sont analysées séparément dans la seconde partie du travail - cf. infra p.43).

C'est, en effet, à la croisée de ces domaines et potentielles zone de tension entre acteurs que le concept de data trust va chercher à proposer des solutions.

Une revue détaillée des multiples régimes légaux nationaux dépasse largement le périmètre du présent travail, aussi parti a été pris d'analyser succinctement les aspects juridiques en lien avec notre sujet sans s'attarder sur un pays particulier.

La manière dont les thématiques sont abordées correspond à celle des publications de droit international économique et technologique consultées.

Les observations présentées ici sont une sélection des points de vue de chercheurs et professeurs de droit universitaires de différentes nationalités (allemand, français, anglais, américain, canadien, hongkongais) qui nous donne un aperçu des grandes questions communes et transversales et de la manière dont elles sont abordées au sein des deux systèmes de droit dominant : common law et droit civil. Mais sans entrer dans les détails au niveau national.

2.1 Le droit de la propriété et l'information

Lau, Penner et Wong (2019) et Lawrence (2020) expliquent que le droit de la propriété n'a pas été appliqué, historiquement, à l'information et que, fondamentalement, celle-ci n'est pas considérée, légalement, comme propriété.

C'est que les informations, y compris les informations personnelles, sont présentes partout. Sans même avoir à les enregistrer, - ni par écrit, ni électroniquement - nous avons constamment à disposition pléthore d'informations sur d'autres personnes et sur le monde. Nous les accumulons en permanence, naturellement, en vivant, en observant notre environnement ou en étant sociables (lorsque nous discutons avec une connaissance d'autres personnes, par exemple).

Scalla (2018) abonde dans le sens d'une absence possible de propriété pour l'information. Elle relève qu'en 1988, la Cour suprême du Canada acquitta, dans le cadre d'un procès criminel, une personne accusée d'avoir volé une liste de noms et d'informations de contact. Dans son jugement, la plus haute cour du Canada

argumenta que l'information ne pouvait être volée car le crime de vol nécessite qu'une personne prive quelqu'un d'autre de quelque chose. Or, dans le cas jugé, la partie accusée avait acquis l'information sans en priver la partie adverse.

Ce jugement illustre certaines des difficultés qu'il faut prendre en considération lorsque l'on pense aux données comme actifs économiques et à la notion de propriété appliquée à l'information.

D'autant plus que le fait que l'information n'ait pas été historiquement soumise au droit de la propriété ne signifie pas pour autant qu'elle ait été dépourvue de valeur dans le passé. Au contraire, l'information a toujours été précieuse et il a, de tout temps, été possible de l'exploiter. Ce n'est en rien un phénomène nouveau.

Lau, Penner et Wong (2019) illustrent la valeur intemporelle de l'information en nous rappelant la célèbre histoire de Nathan Rothschild qui multiplia sa fortune en investissant dans des bonds d'états après avoir été informé de la victoire britannique à Waterloo plus tôt que ses concurrents, grâce à des pigeons voyageurs.

Ce qui a changé, nous expliquent-ils, c'est que nous avons désormais les moyens de traiter les données à une échelle réellement industrielle, à toutes les étapes : extraction, traitement, stockage et analyse.

Cette différence d'échelle et de portée est ce qui différencie fondamentalement la notion de (méga)données telle qu'on l'entend aujourd'hui dans le contexte du *big data*, de la notion d'*information* traditionnelle dont le droit a historiquement estimé qu'elle ne pouvait être considérée comme propriété malgré sa valeur.

Le droit est toutefois mal équipé actuellement pour faire face à l'évolution rapide du secteur de la donnée. Monnerie (2018) estime même que la science du droit est complètement dépassée face aux techniques employées par les géants technologiques. Et les avis divergent sur ce qu'il conviendrait de faire.

Dans une communication intitulée *A Digital Single Market Strategy for Europe*, la Commission Européenne (2015) estimait que quatre grands domaines de la loi devaient évoluer : la protection des données[1], les droits d'auteur, les droits d'usage et la responsabilité.

Concernant les questions de propriété, certains comme Zech (2016) estiment qu'une définition légale des données, comme objet juridique ou comme bien économique, est une étape préalable indispensable pour déterminer les droits de propriété qui doivent s'appliquer.

Pour Scassa (2018), il serait sans doute préférable que l'information reste libre car il n'est pas évident que l'application du droit de propriété aux données soit bénéfique économiquement, politiquement et socialement, pour la collectivité.

De même pour Delacroix et Lawrence (2019) pour qui des droits non exclusifs seraient plus appropriés. Pour eux, les données appartiennent à la catégorie des

[1] Depuis, le Règlement Général sur la Protection des Données est entré en vigueur, le 25 mai 2018.

biens publics (cf. supra p.20) et ce n'est que dans les cas où il existe des intérêts particuliers spécifiques qui méritent d'être protégés (protection de la vie privée, par exemple) qu'il convient de légiférer.

Mills (2019) et Scassa (2018) identifient deux grands groupes qui réclament, de manière croissante la création de droits de propriété s'appliquant aux données :

- Les collecteurs de données : C'est-à-dire les entreprises qui acquièrent ou collectent des données et qui fondent leurs argumentaires sur le capital dépensé (c'est-à-dire sur les coûts engendrés par la collecte de ces données). Leur objectif est d'obtenir des intérêts propriétaires pour commercialiser les données et d'en priver leurs concurrents.

- Les générateurs de données : C'est-à-dire les individus concernés par l'utilisation répétée ou abusive de leurs données personnelles. Leur argumentaire se base sur le fait que les données collectées les concernent individuellement et sont non fongibles. Leur objectif principal est l'obtention d'un intérêt propriétaire d'afin exercer un plus grand contrôle lié au respect de la vie privée. De plus, certains auteurs (Artyushina, 2020 ; Delacroix et Lawrence, 2019 ; Hardinges et Wells, 2019 ; Mills, 2018) mentionnent la possibilité pour le sujet d'obtenir une partie de la valeur créée par l'exploitation commerciale de ses données personnelles. Sous la forme d'un dividende, par exemple.

L'absence de droit de la propriété applicable à l'information ne signifie pas pour autant que ces deux groupes soient dépourvus légalement.

En pratique, les entreprises revendiquent souvent les effets de la propriété, en invoquant soit le droit du copyright, soit les différentes lois relatives portant sur la confidentialité de certaines informations dont la plupart des pays sont dotés. En Europe, la directive portant sur la protection juridique des bases de données (directive européenne du 11 mars 1996) intervient également.

De plus, même si l'information n'est pas soumise au droit de la propriété, Lau, Penner et Wong (2018) nous rappellent que les supports matériels (c'est-à-dire l'équipement informatique : les serveurs, disques durs, supports USB, etc.) le sont, eux, bien entendu. Et qu'il est facile aux entreprises d'en réguler l'accès.

De leur côté, les individus bénéficient des diverses lois de protection des données personnelles. Les citoyens européens sont couverts par le célèbre règlement UE 2016/679 dit Règlement Général sur la Protection des Données (RGPD), qui fait référence mondiale en la matière. D'autres juridictions ont aussi adopté des législations de protection des données personnelles dont certaines s'inspirent parfois partiellement du RGPD.

2.2 Le droit de propriété intellectuelle : copyright et droit d'auteur

Lorsque nous avons précédemment évoqué le caractère intangible des données, nous avons mentionné le point de vue de Ciuriak & Wylie (2018) pour qui il existe une continuité entre la *knowledge-based economy*, dont l'actif majeur est la propriété intellectuelle, et la *data-driven economy*, fondée sur les données et les algorithmes (cf. supra p.15). Chacune représentant une étape successive de la dématérialisation progressive de l'économie.

Le domaine du droit qui s'occupe traditionnellement des actifs correspondant à la *knowledge-based economy* est le droit de la propriété intellectuelle.

Dans les pays de tradition légale anglo-saxonne, aussi appelés pays de la *Common Law*, cette matière est principalement réglée par la loi du *copyright*. Tandis que dans les pays continentaux, dont la tradition est issue du *droit civil*, ces matières sont principalement régies par le *droit d'auteur* et les droits voisins.

L'expression anglaise *copyright* est parfois utilisée abusivement en français pour désigner les droits d'auteurs et les droits voisins mais il s'agit bien de droits distincts. Et ce malgré le fait qu'ils possèdent des points communs et ont tendance à se rejoindre sur la forme en raison des accords internationaux qui ont un effet harmonisant (la convention de Berne pour la protection des œuvres littéraires et artistiques, l'accord de l'OMC sur les aspects des droits de propriété intellectuelle).

Ces lois descendent de traditions philosophiques différentes. Le copyright s'intéresse principalement aux aspects économiques de la propriété intellectuelle tandis que le droit d'auteur est centré sur le créateur et ses œuvres.

Copyright dans les pays de common law :

Pour Scassa (2018), le droit du copyright a spécifiquement été pensé pour s'appliquer aux actifs intangibles, il dès lors logique de se demander s'il pourrait aussi s'appliquer aux données. Et si oui, quand et comment ?

En répondant à cette question, elle commence par souligner une limite fondamentale de cette protection légale : ni les faits ni les idées ne peuvent être visés par un copyright. Ils appartiennent au domaine public. Seule l'*expression originale* des faits et idées peut faire l'objet d'un copyright. Une restriction équivalente existe dans le droit d'auteur.

En raison de l'exclusion des faits du droit de la propriété intellectuelle, les compilations de faits bénéficiaient d'une protection limitée. Il existe toutefois des raisons de penser que ce paradigme est en train d'évoluer avec l'ère du big data.

Les systèmes de collecte de données opèrent désormais en continu et sont omniprésents. Les méthodes de traitement, d'analyse et de stockage des données se sont énormément complexifiées.

Scassa (2018) souligne que traditionnellement, l'application d'une protection basée sur le copyright requérait un auteur humain. En analysant plusieurs décisions de justice récentes, elle remarque cependant que les cours de justice ont commencé à faire évoluer quelque peu leurs décisions face à la réalité changeante.

Dans les pays de *Common Law*, et principalement aux États-Unis, plusieurs avocats inventifs ont argumenté que l'information brute était transformée par des calculs complexes et des algorithmes. D'après eux, les data obtenues en sortie de traitement devaient donc être considérées comme le résultat d'un « travail » et non pas comme un simple assemblage de faits. En conséquence le copyright devait pouvoir s'y appliquer.

Dans un jugement rendu récemment, les données générées par un algorithme classique (par opposition à un algorithme de machine learning) se sont vu reconnaître des qualités originales suffisantes par une cour de justice que pour pouvoir bénéficier de la protection du copyright. De même dans un autre cas, relatif à un processus de collecte de données complexe (collecte de données sismiques sous-marines) où la protection du copyright a pu s'appliquer.

On constate donc une évolution des décisions dans les pays de *common law*. Ce qui semble ouvrir la voie à la protection des compilations de données (mais pas des données individuelles).

Cependant, Scassa (2018) estime que la protection conférée par le copyright demeure très incertaine. Le copyright ne protège qu'une expression, c'est-à-dire ici un arrangement particulier des données. Cela signifie que si quelqu'un effectuait une copie de toutes les données et les ordonnait différemment, cela ne constituerait pas une infraction au copyright.

Droit d'auteur dans les pays de droit civil européens :

En Europe, la situation semble plus claire. D'après Guadamuz (2017), la plupart des pays européens (Allemagne, Espagne …) disposent de lois nationales qui précisent que les œuvres créées par un être humain sont les seules à pouvoir être protégées par le droit d'auteur. La compilation de données automatisée, avec ou sans traitement machine postérieur, apparait donc exclue.

Cette situation est confirmée par un arrêt historique de la CJUE dans l'affaire *Infopaq* (C-5/08 *Infopaq International A/S* c. *Danske Dagbaldes Forening*), et plusieurs décisions européennes allant dans le même sens. Ces arrêts précisent que le droit d'auteur ne concerne que des œuvres originales, ce qui signifie « création intellectuelle propre à son auteur ». La plupart des interprétations légales considèrent que cela signifie que pour bénéficier de la protection du droit d'auteur, une intervention humaine est indispensable.

Directive européenne sur la protection juridique des bases de données :

Outre le droit d'auteur, il existe une directive européenne s'appliquant aux bases de données : la directive 96/9/EC du 11 mars 1996. Toutefois d'après Zech (2016), la protection garantie par ce droit *sui generis* ne concerne pas les données elles-mêmes. Elle s'applique uniquement aux investissements réalisés afin de mettre en place la base de données. De plus, des arrêts rendus par la CJUE (*BHB/Hill* et *Fixtures Marketing I-III*) plafonnent les investissements pris en considération.

Si bien que non seulement le contenu des bases de données n'est pas couvert mais le big data ne peut de toute façon pas bénéficier de la protection de cette directive en raison du plafond.

Exception au droit de la propriété intellectuelle :

Enfin, notons que le droit de la propriété intellectuelle n'offre jamais de protection absolue. Il contient des exceptions qui permettent d'effectuer des citations limitées dans le cadre de la recherche scientifique, de la critique, de la parodie, du journalisme, de l'éducation, etc.

Certaines exceptions sont parfois prévues explicitement par la loi. Battisti et Schöpfel (2017) soulignent ainsi l'importance cruciale pour la recherche scientifique de l'exception dont bénéficie le *Text and Data Mining* par le biais de la *Directive on copyright in the Digital Single Market* (directive EU 2019/790).

Cette exception autorise les universités, les instituts de recherches et les hôpitaux publics à employer des techniques de *data mining*, à des fins de recherche scientifique, y compris lorsque ces techniques sont appliquées à des documents protégés par le droit d'auteur. Pour bénéficier de l'exception, les organisations doivent toutefois disposer initialement d'un accès légal aux data (ce qui inclut tous les documents en libre accès).

Cette exception existe aussi, mais de manière un peu plus limitée, pour les autres entités qui opèrent à des fins non scientifiques (par exemple les sociétés commerciales).

Pour Battisti et Schöpfel (2017), ces exceptions sont importantes car le droit des bases de données et les courtes permissions de citation ordinaires prévues par le droit d'auteur ne sont pas adaptés aux outils de *Text and Data Mining* et constituaient de véritables verrous.

Les outils de *text and data mining* permettent de collecter d'importantes quantités de données. La ratification de ces exceptions par le Parlement Européen en 2019 est donc utile dans le cas de l'intelligence artificielle, bien que cela ne règle pas tous les problèmes du secteur (les exceptions votées ne fournissent, par exemple, pas d'incitants encourageant les détenteurs de data, à ne plus les accumuler en silo).

2.3 Lois relatives à la confidentialité de certaines informations

Signalons rapidement que la plupart des juridictions sont dotées de lois relatives à la confidentialité de l'information. Dans l'Union Européenne, le cadre législatif a été harmonisé par la *directive EU 2016/943 sur la protection des savoir-faire et des informations commerciales non divulguées (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites*.

Il est ici question des secrets professionnels et commerciaux, des recettes et formules, des listes de clients ou encore des interdictions d'enrichissement liées à des accès privilégiés (*insider trading*), etc.

Pour Lau, Penner et Wong (2019), ces interdictions trouvent leur origine dans l'existence d'une relation et dans la nature de cette relation : qu'il s'agisse d'accès privilégié, d'une relation de confiance, de devoir professionnel, etc.

Elles sont donc entièrement contextuelles et ne doivent pas être comprises comme découlant de l'existence d'une propriété de l'information.

Zech (2016) relève que la directive européenne qui couvre ces matières permet que les données puissent être, dans certains cas, considérées comme des secrets commerciaux. Toutefois, cette protection ne confère pas de réels droits exclusifs. En effet, elle nécessite l'existence d'un secret préalable. En ce sens, elle ressemble à la protection que l'on peut obtenir en restreignant l'accès à un support physique. Si les données sont d'ores et déjà accessibles au public, cette protection s'estompe.

De plus, Scassa (2018) souligne qu'il peut y avoir des situations où l'intérêt général nécessite que la confidentialité soit levée. Dans le cas d'une décision automatique générée par l'IA et le big data, par exemple, il pourrait être nécessaire de comprendre les facteurs qui ont amené à la prise de cette décision, voire de mettre en place une supervision par les autorités. Dans l'intérêt public, la loi permet la création de telles exceptions au principe de confidentialité.

3 Le marché des données

3.1 Une ambition européenne

Avant de procéder à l'analyse de la structure et des caractéristiques économiques du marché des données, il est important de prendre un instant pour examiner l'aspect politique du macro-environnement, comme on le ferait dans le cadre d'une analyse PESTEL, car les autorités européennes sont très actives dans ce secteur et leurs décisions auront un impact tangible au niveau européen.

La stratégie digitale européenne pour les 5 prochaines années est décrite dans une communication de la Commission Européenne intitulée « *Une stratégie européenne pour les données* ». Ce document datant de février 2020 dévoile la vaste ambition affichée de créer un marché des données paneuropéen, une économie nourrie de ces données et au-delà, d'ériger un véritable modèle de *data gouvernance* afin d'encadrer leur collecte et leur utilisation.

Trois objectifs sont affichés par la Commission Européenne (2020) pour le futur *Single Data Market* :

- La bonne circulation des données au sein de l'Union Européenne, entre secteurs, dans l'intérêt de tous ;
- Le respect des règles européennes, en particulier concernant la protection de la vie privée et des données, et le droit de la concurrence ;
- La mise en place de régulations équitables, pratiques et claires concernant l'accès et l'utilisation des données.

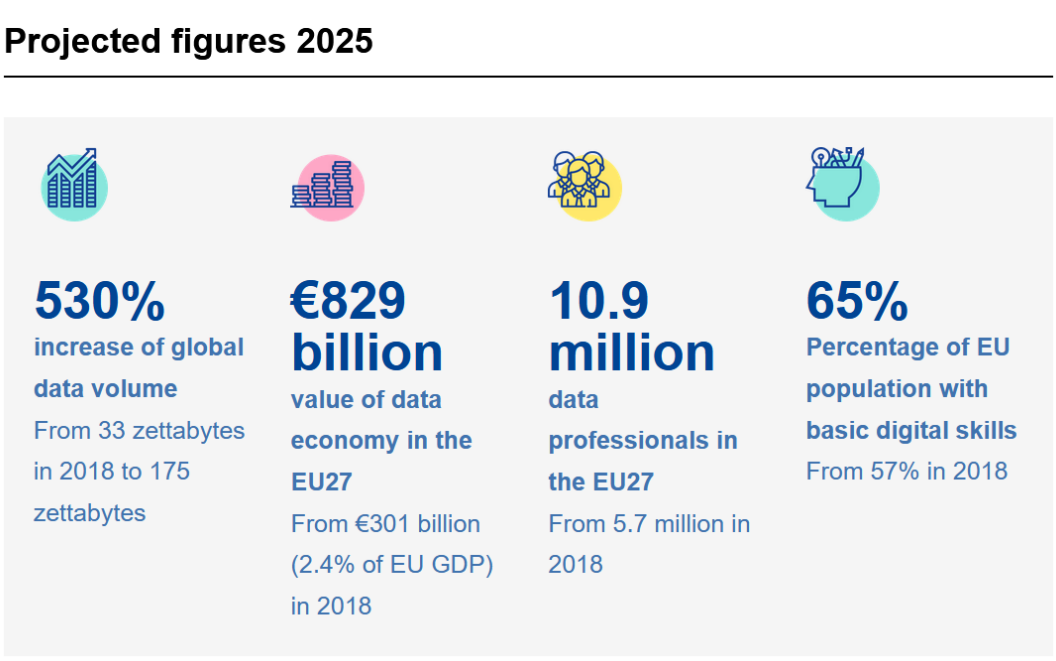

Figure 4 : Projections 2025 de la Commission Européenne pour le *Single Data Market*.

Source : European Commission (2021). *European Data Strategy*. Récupéré le 08/12/2020 de https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy

« À terme, l'Europe vise à tirer parti des avantages d'une meilleure utilisation des données, notamment une productivité accrue et des marchés plus concurrentiels, ainsi que des améliorations dans les domaines de la santé et du bien-être, de l'environnement, une gouvernance transparente et des services publics efficaces. » (Commission Européenne, 2020, p.1).

Quelques exemples d'applications civiles sont cités par la Commission (2020, 2021) pour illustrer le potentiel bénéfique que l'exploitation massive des données pourrait avoir en Europe : l'optimisation de la génération d'électricité par les parcs éoliens, une meilleure allocation des ressources de santé dans la lutte contre les maladies infectieuses, le monitoring en temps réel du trafic ferroviaire et automobile avec notification des retards, prévention des embouteillages et une réduction des impacts écologiques, une réduction du coût des services publics et un support analytique pour la lutte contre le changement climatique...

Avec à la clé, pour chacun de ces projets, des économies estimées par la Commission en centaines de millions. Voire en milliards d'euros, lorsque l'estimation tient compte des gains de productivité (il est question de gains de l'ordre de 20 milliards d'euros en coûts du travail pour la seule optimisation du trafic automobile).

Pour atteindre ses objectifs, le plan de la Commission (2020) prévoit des investissements directs à hauteur de 4 à 6 milliards d'euros cofinancés par l'Europe et les entreprises. Ce faisant, elle souhaite créer un contexte favorable à l'émergence d'outils novateurs qui permettent simultanément le respect des droits des citoyens, dans le sillon du RGPD, et favorise l'émergence d'« un cadre transsectoriel pour l'accès aux données et leur utilisation » (Commission Européenne, 2020, p.15).

Outre les investissements directs, la Commission Européenne entend encourager la mise en place de dispositifs de data gouvernance innovants tel le *Data Trust*, qu'elle nomme spécifiquement dans sa communication de 2020. D'une manière générale, les autorités européennes sont très attentives aux problématiques relatives au partage et à la circulation des données. Le secteur se sachant scruté avec attention, tous les acteurs sont attentifs à l'évolution du cadre réglementaire.

Pour concrétiser ces ambitions stratégiques, la Commission Européenne a adopté, le 25 novembre 2020, la Proposition de règlement du Parlement Européen et du Conseil sur la gouvernance européenne des données, plus communément appelé 'Data Governance Act'.

De plus, à l'occasion de conversations informelles avec des professeurs de droit des universités d'Oxford et de Birmingham, Paul Nemitz, un haut conseiller de la Commission Européenne a confirmé l'intérêt de la Commission pour le *Data Trust* et les dispositifs innovants de data gouvernance (data coopératives, etc.). Avec comme objectif le développement d'un écosystème varié pour les données (Delacroix, McFarlane & Nemitz, 2021). La création d'une structure juridique, sur mesure, pour la gestion des droits des données est également mentionnée.

3.2 Difficultés relatives à la délimitation du marché

Concernant le marché des données, Monnerie (2018), note que sa structure spécifique présente des difficultés uniques majeures qui compliquent sa délimitation. Lorsque l'on cherche à mesurer la puissance d'un opérateur économique pour évaluer l'état de la concurrence sur un marché, c'est traditionnellement la notion de « marché pertinent » qui sert de référent et la « part de marché » qui est l'unité de mesure employée.

Or, il estime que, concernant le traitement de l'information, cette unité n'est pas tout à fait adaptée. Dans la mesure où l'information présente des caractéristiques particulières (cf. supra p.15), dont un coût marginal d'acquisition et de reproduction qui peut être minime, évaluer la puissance de marché en estimant les facilités de production, comme si l'information était un bien traditionnel, donne des résultats qui ne semblent pas refléter adéquatement le réel pouvoir de marché des acteurs en place. Cette mesure est encore complexifiée en raison de la gratuité d'une partie des services lors de la captation des flux de données.

Théoriquement, l'utilisateur peut facilement quitter les services des géants technologiques et il suffit d'un clic pour se rendre sur un site autre que ceux des GAFA. En pratique, on observe toutefois la domination d'une poignée de géants technologiques sur les marchés observés (Artyushina, 2020 ; Birch et Al., 2020 ; Delacroix & Lawrence, 2019 ; Mills, 2019 ; Monnerie, 2018 ; Szczepanski, 2020).

Si bien que dans le cas du marché des données, Monnerie (2018) estime nécessaire de repenser l'unité de mesure, elle-même. Mais, malgré le mérite critique de son argument, il échoue à proposer une mesure alternative à la classique « part de marché » calculé à partir de la valeur monétaire (cf. infra p. 38).

Dans le cadre des études réalisées pour le compte du Parlement Européen (Szczepanski, 2020) ou du côté des communications de la Commission Européenne (2020), on note toutefois qu'outre les classiques mesures de part de marchés par secteurs d'activité, les autorités européennes s'intéressent également au volume de trafic brut comme indicateur. Ceux-ci permettant la prise en compte des volumes liés aux services gratuits.

Monnerie (2018) remarque également que la nature même du marché pose question. L'opérateur d'un moteur de recherche ne se contente pas de livrer une information pertinente à l'utilisateur. Simultanément, il collecte, traite et analyse les données de ses visiteurs qui sont, elles-mêmes, réemployées pour alimenter d'autres services.

Szczepanski (2020) souligne que cette capacité des géants technologiques, à acquérir des données asymétriquement, à travers toute une série de produits et de marchés différents pose des questions légitimes pour la régulation de la concurrence.

En effet, il existe de vastes économies de gamme qui bénéficient aux acteurs dominants. Grâce aux données et aux analyses acquises dans le cadre d'une activité précédente, ils peuvent ensuite se déployer dans de nouveaux secteurs avec un coût réduit et opérer transversalement sur plusieurs marchés. Pour les

nouveaux entrants qui ne disposent pas de ces facilités, ces économies de gamme constituent une barrière à l'entrée.

Les chercheurs en droit de la concurrence s'interrogent, dès lors, pour savoir s'il faut considérer le marché global de la production de données dans son ensemble ou s'il faut le sous-diviser, en fonction de l'usage économique des données.

Nous reproduisons plus loin (cf. infra p.38) quelques données relatives aux parts de marché des grands acteurs. Il convient toutefois de garder à l'esprit que le marché de la donnée, si on le considère dans son ensemble, est plus complexe que la simple somme de ces « sous-marchés » : moteur de recherche, médias sociaux, navigateur internet, publicités en ligne, vente en ligne … et ce, parce que les flux de données et les analyses qu'ils permettent agissent comme des forces transversales qui non seulement opèrent dans ces sous-marchés, mais ont aussi un effet de réseau plus large.

3.3 Structure du marché

Ayant noté les difficultés méthodologiques relatives à la délimitation exacte du marché, nous pouvons toutefois avancer quelques observations simples quant à sa structure.

Mills (2019) propose une représentation schématique des échanges entre acteurs et des flux de données dans le cas où il n'existe pas d'intervention normative (Laissez Faire).

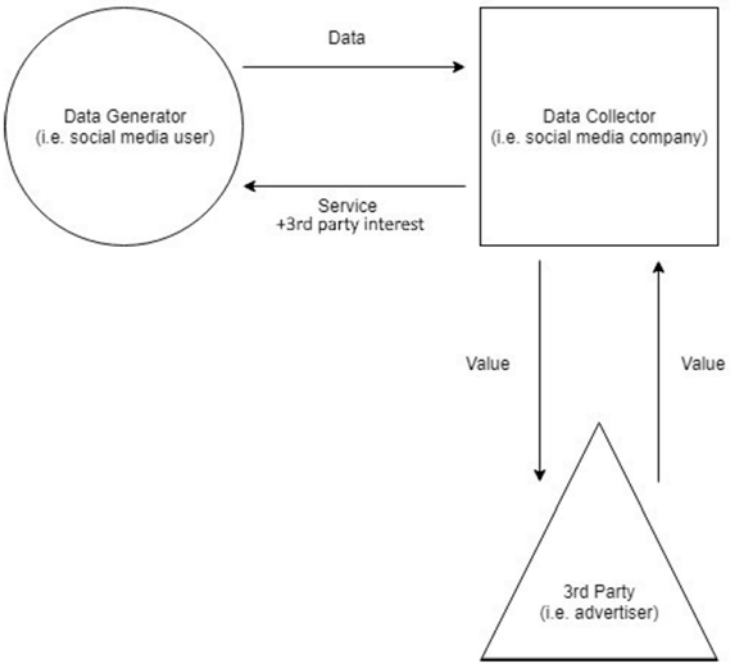

Figure 5 : Data flows in a Laissez-Faire Model.

Source : Mills, S. (2019). Who Owns the Future ? Data Trusts, Data Commons, and the Future of Data Ownership. *SSRN Electronic Journal*, 1. https://doi.org/10.2139/ssrn.3437936

La vision simplifiée des flux est celle d'un usager qui joue le rôle du « fournisseur » de données et celle d'un tiers, ici l'annonceur publicitaire, dans le rôle du « client ».

Plus spécifiquement, l'on parlera de *Data Generator* pour désigner l'utilisateur de service qui génère les données et de *Data Collector* pour désigner l'entreprise qui opère le service, collecte les données et les exploite.

L'usager du service digital, le *Data Generator*, fournit les données qui sont les matières premières. Le *Data Collector* effectue un travail algorithmique et transforme cette matière première. Et un tiers, ici l'annonceur publicitaire, achète un accès au résultat de ce travail algorithmique afin de diffuser sa publicité en ciblant exactement le segment démographique souhaité.

Par rapport au principal flux de données, le *Data Generator* (dans notre exemple l'usager du réseau social) n'est pas - quoi qu'il en pense - le client mais bien le fournisseur. Une situation parfois illustrée trivialement par l'adage : « si c'est gratuit, c'est que vous êtes le produit ».

Aussi convenue soit cette observation, il n'est pas inutile de la réitérer. Car de nombreux utilisateurs ne se considèrent pas comme des fournisseurs dans la relation qu'ils entretiennent avec les services digitaux. Du point de vue de l'économie de la donnée, c'est pourtant le cas. Pour Szczepanski (2020), cette confusion est entretenue par une asymétrie d'information et de pouvoir prononcée si bien que les *Data Generators* ne sont pas conscients de la nature et de la quantité des données qu'ils envoient aux *Data Collectors.*

Cette lecture de la situation, fortement simplifiée, n'est cependant vraie que si on se limite à l'analyse du flux principal d'information. En réalité, les données sont employées pour une multitude d'opérations économiques créant un effet de réseau particulièrement complexe.

Par le biais d'une boucle de rétroaction, ces mêmes données servent à transformer et personnaliser l'expérience utilisateur du *Data Generator* qui les a lui-même émis et vont, en retour, modifier son comportement (personnalisation des recommandations et de l'expérience).

Cet aspect récursif est d'ailleurs l'élément fondateur qui définit le point de départ de la *data economy* pour Ciuriak & Wylie :

« *The point where the data-driven economy truly started is when the use of big data created a feedback that modified the social and economic behaviour that generated the data in the first place* » (Ciuriak & Wylie, 2018).

Ces mêmes données sont aussi utilisées pour alimenter d'information des services tiers qui peuvent être intégrés (comme une Marketplace à l'intérieur d'un réseau social) ou externes (un projet de machine learning séparé). Elles peuvent également servir à lancer de nouveaux services grâce aux économies de gamme et sont parfois même exploitées par des acteurs étrangers, à des fins détournées, comme lors du scandale *Cambridge Analytica*.

Pour ces raisons, Monnerie (2018) considère que nous sommes en présence d'un marché « biface » voire, même, d'un marché « multiface ».

Par marché biface, on entend un marché qui sert simultanément deux clientèles distinctes mais interdépendantes : par exemple, un réseau social offre un service de communauté digitale à ses usagers qui constituent une première clientèle et dans le même temps il sert des annonceurs publicitaires qui constituent une seconde clientèle, au sein d'un même marché avec des liens de dépendances entre acteurs.

L'expression de marché « multiface » couvre une réalité similaire mais avec plus de deux clientèles. On entend par là que les grandes entreprises digitales actives sur le marché de la donnée vont chercher à satisfaire simultanément des demandes en aval et en amont sans qu'il soit toujours possible de les distinguer.

Alphabet/Google est un bon exemple d'écosystème multiface où les clientèles sont à la fois variées et indissociables : diffusion publicitaire, service mail, moteur de recherche, plateforme vidéo, navigateur web, stockage dans le cloud, service cartographique, système d'exploitation pour téléphone mobile et son magasin d'applications liées… Ces services servent des clientèles variées (utilisateur, annonceurs publicitaires, créateurs de contenu) et mouvantes entre ces différents services. Alors qu'un compte unique permet d'accéder à l'ensemble et de centraliser les données.

3.4 Caractéristiques économiques du marché

D'après Szczepanski (2020), le marché des données est caractérisé par des coûts fixes élevés, d'importantes économies d'échelle et de gamme, le développement d'écosystèmes, de puissants effets de réseau directs et indirects, de hautes barrières à l'entrée et une forte concentration oligopolistique.

Coûts fixes et économies d'échelle

L'innovation dans la majorité des secteurs de l'économie digitale requiert de hauts coûts fixes. Toutefois, en raison des possibilités de copie à coût marginal nul inhérentes à la nature des données (cf. supra p.16), les coûts de production sont souvent très bas par rapport au grand nombre d'utilisateurs, ce qui amène d'énormes économies d'échelle. L'existence d'économies d'échelle n'est pas exclusive à l'économie digitale mais, en raison des coûts marginaux quasi nuls, elles y ont un effet plus prononcé.

Économies de gamme

Comme nous l'avons déjà mentionné, les entreprises peuvent également bénéficier d'économies de gamme en réemployant les flux de données d'un service

existant pour créer et alimenter de nouveaux services et ensuite exploiter transversalement les données collectées, sur plusieurs sous-marchés.

Les économies de gamme obéissent au même principe que les économies d'échelle. Si ce n'est qu'elles ne se manifestent pas lors d'une augmentation des quantités produites, mais bien par la possibilité de bénéficier de réduction de coût lors de l'élargissement de la gamme de produits (ici réutilisation gratuite d'une ressource, les données, de manière horizontale pour lancer et exploiter de nouveaux services).

Taille critique et domination oligopolistique

Ces énormes économies d'échelle et de gamme, associées aux hauts coûts fixes, poussent les entreprises à croître de manière importante. Si bien que l'on constate l'émergence de plateformes hégémoniques selon un schéma oligopolistique.

Cette concentration du pouvoir aux mains de quelques firmes qui contrôlent la plus grande part du marché est un phénomène très visible et constaté par de nombreux auteurs (Artyushina, 2020 ; Birch et Al., 2020 ; Delacroix & Lawrence, 2019 ; Micheli et Al., 2020 ; Mills, 2019 ; Monnerie, 2018 ; Szczepanski, 2020).

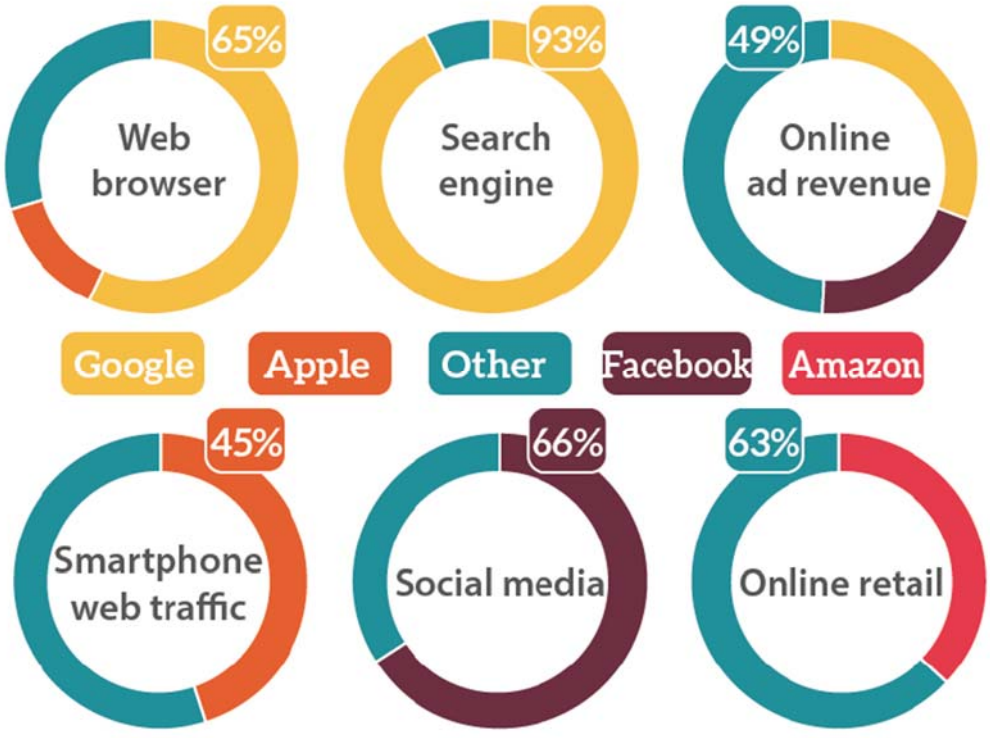

Figure 6 : Global market share by company.

Source : Szczepanski, M. (2020). *Is data the new oil? Competition issues in the digital economy.* EPRS | European Parliamentary Research Service. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI%282020%29646117

Écosystèmes

Pour Szczepanski (2020), il existe d'autres caractéristiques spécifiques du marché de la donnée. L'une d'entre elles est le développement d'écosystèmes qui procurent aux entreprises déjà en place un fort avantage compétitif. Ces écosystèmes s'étendent parfois sur l'ensemble de la chaîne de valeur numérique et sont donc présents simultanément sur de multiples marchés digitaux. Ils permettent aux acteurs dominants de constituer d'immenses jeux de données difficilement duplicables pour un nouvel entrant.

En raison du manque d'interopérabilité, ces écosystèmes créent des barrières à la sortie pour les utilisateurs et à l'entrée pour les compétiteurs qui sont forcés d'entrer en concurrence avec l'ensemble de la gamme des services.

Effets de réseau

Les puissants effets de réseaux sont une autre caractéristique spécifique du marché de la donnée. Pour Szczepanski (2020), il en existe de deux types : directs et indirects.

Les effets de réseau directs sont liés à la quantité d'utilisateurs d'un service. Cet aspect joue un rôle crucial pour en attirer d'autres, typiquement dans le cas des réseaux sociaux qui nécessitent d'atteindre une certaine taille critique.

Les effets de réseau indirects sont eux en lien avec le nombre de personnes qui offrent du contenu sur une plateforme (par exemple le nombre de producteurs de vidéo sur YouTube ou le nombre de vendeurs sur Amazon).

Ces effets de réseau peuvent constituer des barrières à l'entrée si fortes qu'elles sont même capables d'empêcher l'entrée sur le marché des acteurs les plus puissants: l'échec du réseau social de Google, Google+, malgré toutes les ressources à disposition de l'entreprise est une illustration frappante de cet effet.

Barrières à l'entrée

Pour les nouveaux entrants, les barrières à l'entrée sont nombreuses et difficilement franchissables.

Elles sont causées par les caractéristiques économiques du marché que nous venons de décrire : hauts coûts fixes, vastes économies d'échelle et de gamme dont bénéficient les entreprises déjà en place et grâce auxquelles elles ont atteint une taille critique, écosystèmes fermés et non portabilité des données qui empêchent les usagers de quitter un service et qui forcent les nouveaux entrants à faire concurrence à l'ensemble de la chaîne de valeur, effets de réseau liés au nombre d'usagers ou de personnes proposant du contenu…

À celles-ci, il faut encore ajouter les barrières à l'entrée liées à l'application des réglementations, dont la conformité avec le RGPD.

Problèmes de concurrence

« Il est clair que les GAFA disposent d'un flux de données si dense qu'ils peuvent être qualifiés d'opérateurs systémiques en ce qu'ils dominent un réseau critique d'information ». (Monnerie, 2018, p.436).

L'accumulation du pouvoir dans les mains de quelques acteurs dominants et la forme oligopolistique du marché génèrent des inquiétudes croissantes.

L'OCDE (2016) remarque qu'une poignée de firmes « superstar » hautement innovantes et productives sont concentrées dans quelques secteurs, tandis que le reste de l'économie, aux États-Unis et en Europe souffre d'une productivité stagnante. Elle estime que la non-diffusion technologique constitue un frein pour la productivité générale et que les barrières à l'entrée et les coûts élevés de transition vers une *knowledge-based economy*, rendent la concurrence de plus en plus difficile pour les entreprises qui ne sont pas leaders de marché.

Pour Szczepanski (2020), il existe effectivement des signes que l'intensité de la concurrence est peut-être en train de ralentir. Détrôner une entreprise qui contrôle une part significative du marché et a établi une position dominante pourrait s'avérer très compliqué à l'avenir.

Le risque est que les entreprises qui sont déjà en place actuellement soient systématiquement les mieux placées lors de la prochaine révolution technologique, centrée sur le *machine learning* et l'intelligence artificielle. Dans ce cas, les énormes quantités de données nécessaires à ces technologies constitueront des barrières à l'entrée supplémentaires insurmontables pour les nouveaux entrants.

La détention d'une vaste quantité de données est vue comme un avantage compétitif significatif pour les entreprises déjà établies. À tel point que l'OCDE (2016) note que la simple détention de gigantesques banques de données est suffisante pour renforcer la position des dominants et empêcher les concurrents d'attirer des clients. À vaste échelle, la détention de données exclusives aurait donc, en soi, un effet préjudiciable perceptible sur la concurrence.

Pour les observateurs du marché, la question est donc posée ouvertement de savoir s'il ne serait pas nécessaire de réguler le partage de données, et même de le rendre obligatoire dans certains cas spécifiques.

3.5 Le partage de données et le besoin de data gouvernance

Dans ce contexte, l'accumulation et la thésaurisation des données sont vues comme un obstacle alors qu'elles pourraient être exploitées simultanément par

plusieurs entreprises avec des effets bénéfiques pour la compétitivité et l'innovation.

Szczepanski (2020) note, à cet égard, que la dissémination la plus large possible des données et leur utilisation par le plus grand nombre d'entreprises pourraient, théoriquement, augmenter le bien-être économique et social. À condition que le respect de la vie privée et la sécurité soient préservés.

Crémer, de Montjoye et Schweitzer (2019) suggère, dans un rapport préparé à l'attention de la Commission Européenne, que la portabilité des données pourrait être imposée là où les effets de lock-in sont particulièrement forts, grâce à des standards ouverts et des législations ciblant certains secteurs.

Toutefois, l'effet de ces mesures pourrait être limité en raison des économies d'échelle déjà exploitées par les entreprises en place et des puissants effets de réseau que nous avons mentionné (cf. supra p.39)

L'exploration pour trouver de solutions appropriées continue donc. Et il n'existe actuellement, nulle part dans le monde de modèle général qui parvient à tenir compte de la multiplicité des facettes du problème, y compris ceux liés au respect de la vie privée et à la cybersécurité.

En Europe, les enjeux relatifs à la protection de la vie privée sont en général bien compris. Avec l'adoption du Règlement Général sur la Protection des Données, l'Union Européenne fait figure de leader en matière de data gouvernance. Toute politique entendant s'attaquer aux problèmes de concurrence présents sur le marché des données devra impérativement en tenir compte.

Du côté du FMI, Carrière-Swallow et Haksar (2019) préconisent une approche transversale intégrée pour faire face à ces multiples enjeux, dont les objectifs seraient :

- D'encourager le contrôle de leurs données par les usagers ;
- De rendre le partage de données obligatoire, entre firmes, afin d'augmenter la compétitivité et d'affaiblir le pouvoir de marché de celles déjà en place ;
- De clarifier la distribution des bénéfices économiques générés par les données ;
- D'empêcher la fragmentation internationale des marchés de données ;
- Tout en étant robuste sur le plan de la cybersécurité.

Pour répondre à ces défis complexes, la Commission Européenne adapte continuellement son arsenal réglementaire et explore en parallèle plusieurs pistes.

Dans une communication datée de février 2020 et intitulée *Une stratégie européenne pour les données*, quelques outils à exploiter sont évoqués, telle la désormais célèbre *Blockchain*. Parmi les dispositifs novateurs mentionnés figure aussi, le méconnu *Data Trust* - la fiducie de donnée -, un dispositif de data gouvernance dont les gestionnaires peuvent, dans certains cas, assurer une responsabilité fiduciaire sur le modèle du Trust anglo-saxon.

Dans la seconde partie de ce travail, nous analysons et décrivons ce dispositif du *Data Trust* avec pour objectif l'évaluation de son potentiel à faire face aux enjeux complexes évoqués.

Partie 2 : Le Data Trust

4 Le trust, un concept de tradition anglo-saxonne

4.1 L'origine médiévale du trust, en marge du droit commun

Afin de comprendre les fondements intellectuels de la notion de trust (dont le *Data Trust* s'inspire), remontons dans le temps :

En 1066, Guillaume le Conquérant et ses soldats normands profitent d'une crise de succession et envahissent l'Angleterre. Le 14 octobre, ils triomphent à Hastings et s'emparent du trône anglais. Ce faisant, le duc de Normandie emporte une bataille qui aura des conséquences profondes pour l'histoire du royaume britannique.

Au lendemain de la victoire, Guillaume se fait couronner roi d'Angleterre et redistribue les terres confisquées à la noblesse vaincue. Dans la foulée, de substantielles modifications seront progressivement apportées par les vainqueurs au système administratif anglais déjà sophistiqué pour son époque.

Parmi les réformes promues par la nouvelle lignée de rois anglais, celle de la justice figure en bonne place. À la demande de la Couronne, des juges itinérants parcourent le pays et compilent les édits existants. Il en résultera la compilation progressive d'une jurisprudence uniforme : la '*common law*' anglaise[2].

Scott, A. (1966) professeur émérite de droit à l'Université de Harvard et éminent spécialiste du sujet relate l'apparition progressive du *trust* dans ce contexte féodal mouvant :

Dès le départ, le *trust* répond à des besoins pratiques. Il s'agit de transférer la gestion de terres à un tiers tout en en conservant le fruit, pour soi-même ou pour un bénéficiaire désigné. De tels transferts deviennent communs au 12ème et au 13ème siècle.

Il n'existe au départ aucun cadre juridique, les jurisprudences qui constituent la nouvelle *common law*, ne couvrent pas cette pratique. Le système repose donc entièrement sur la parole donnée et l'honneur de celui qui se voit confier les titres des propriétés avec la mission de faire fructifier ces terres, en marge du droit émergent.

Delacroix et Lawrence (2019) relatent que quelques générations plus tard, au 14ème siècle, les croisés qui avaient transféré leurs titres de propriété pendant leur

[2] La '*common law*' issue du droit anglais est un système juridique largement basé sur les précédents et la jurisprudence. Encore aujourd'hui, dans la plupart des anciennes colonies britanniques, il constitue le fondement du système juridique en application. Il est l'un des systèmes juridiques dominants à l'échelle mondiale (l'autre étant le droit civil de tradition romaine, parfois également appelé droit continental).

absence vont se heurter au refus de restitution de leurs terres au retour de croisade. Sans recours légal, ils en appellent au chancelier, le plus haut officier légal, désigné par le Roi. Le *Chancellor* arbitrera en leur faveur, estimant que ceux à qui ces terres avaient été confiées « devaient être contraints *in equity* [c'est-à-dire en actions et en capital] à faire ce que la conscience requérait d'eux » (Scott, 1966, p.177).

Cette histoire des origines contient déjà tous les éléments essentiels qui vont définir le trust dans sa version moderne : la mise en gestion d'un droit ou d'un bien pour le bénéfice d'un tiers, l'importance cruciale de l'honneur et de la confiance, l'obligation de la parole tenue, rendue exécutoire par une autorité officielle si nécessaire.

4.2 Les éléments fondamentaux d'un trust

Un trust est donc un contrat par lequel un gestionnaire (le *trustee*) se voit confier la responsabilité fiduciaire de gérer un droit ou un actif pour le bénéfice d'autrui (le *beneficiary*).

Trois rôles essentiels existent au sein de tout trust :

- Le *settlor* (on parle aussi parfois de *trustor* ou *creator* en anglais. En français, on parle de fiduciant ou de constituant) est la personne physique ou morale qui transfère temporairement la propriété de ses biens ou de ses droits.

- Le *trustee* (nommé en français le fiduciaire ou plus simplement l'administrateur) est la personne physique ou morale qui se voit confier la propriété ou les droits avec pour mission de les gérer dans l'intérêt des bénéficiaires.

- Le *beneficiary* (le bénéficiaire) est la personne physique ou morale désignée par le *settlor* pour recevoir les bénéfices de la gestion effectuée par le *trustee.*

En plus de ces trois rôles, deux autres éléments fondamentaux constituent le trust :

- Les actifs placés en trust. Il peut s'agir de biens, de propriétés mais également d'une gamme assez large de droits.

- L'objet du trust est la raison pour laquelle le trust est constitué, son objectif. L'objet de tout trust doit être valide et reconnu par la loi (il n'est pas possible d'établir un trust dans un but illégal).

Le *trust* est un dispositif très flexible. Chacun des rôles peut être rempli par une ou plusieurs personnes. Ainsi, un groupe constitué de différentes personnes morales et physiques peut très bien occuper l'une des positions.

Certaines personnes peuvent également occuper plusieurs positions simultanément. Lau, Penner et Wong (2019) expliquent que le *trustee* peut, par exemple, également être un bénéficiaire du trust, tout en exécutant son devoir d'administrateur. Par contre, il ne peut pas être le seul bénéficiaire de ce trust. Ainsi, un *trustee T* peut administrer un trust au profit d'un bénéficiaire *B* ou au profit de *T* et de *B*. Mais pas uniquement au profit de *T.*

D'après Penner (2019), le *settlor* d'un trust peut également en être le bénéficiaire ou l'un des bénéficiaires. Un arrangement relativement typique est celui par lequel un *settlor* crée un *trust* et stipule qu'il en recevra les revenus durant son vivant puis, qu'à sa mort, les actifs en gestion reviendront à un autre bénéficiaire désigné. Une autre possibilité est de stipuler que les actifs resteront sous gestion au sein du trust, après le décès du *settlor* et d'attribuer au bénéficiaire désigné les revenus générés par ces actifs.

La loi des *trusts* est relativement permissive. Une large gamme d'actifs de natures différentes peut être détenue par l'intermédiaire d'un trust : des biens mais également des droits (comme des obligations). La règle générale est que tout droit qui dispose d'une valeur économique intrinsèque peut être détenu au sein d'un trust.

Edwards (2004) explique que les biens ou droits transférés par le constituant au trust quittent le patrimoine du *settlor* et sont transférés aux fiduciaires (les *trustees*) mais qu'ils restent soumis à la volonté du constituant par le biais de l'acte de fiducie.

Pour Lau, Penner et Wong (2019), le *trustee* dispose d'une certaine liberté d'action pour remplir son mandat. La manière dont les bénéfices du *trust* parviennent au bénéficiaire peut varier en fonction du type d'actifs détenus par le trust. Par exemple, si un *trustee* estime opportun et conforme à la mission du trust qu'un bénéficiaire dispose d'une nouvelle voiture, ce *trustee* peut directement remettre l'argent destiné à l'achat au bénéficiaire. Il peut également acheter la voiture pour le bénéficiaire en utilisant les fonds du *trust* de manière à ce que le bénéficiaire possède le titre de propriété du véhicule. Enfin, le *trustee* peut encore utiliser les fonds du *trust* pour acheter une voiture en son nom propre (celui du *trustee*) et octroyer au *bénéficiaire* une licence d'utilisation. Dans ce dernier cas, le bénéficiaire aura l'usage du véhicule mais ne pourra pas le revendre (un choix de gestion qui peut être approprié si le bénéficiaire souffre, par exemple, d'une addiction au jeu).

La responsabilité fiduciaire est centrale et tempère le pouvoir discrétionnaire du *trustee* qui doit toujours agir dans l'intérêt supérieur du bénéficiaire et du trust. Dans l'éventualité où un *trustee* ne respecterait pas sa mission, il est possible au bénéficiaire de recourir au tribunal, pour forcer la main de l'administrateur. C'est un point crucial du dispositif.

Le *trustee* ne peut pas laisser ses propres intérêts entrer en conflit avec ceux du bénéficiaire.

Delacroix et Lawrence (2019) soulignent qu'en cas de plainte, la charge de la preuve incombe au *trustee* qui doit démontrer qu'il a cherché à promouvoir l'intérêt du bénéficiaire avec impartialité, prudence, transparence et une loyauté complète.

Il s'agit d'un niveau responsabilité plus élevé que ce qui est exigé par d'autres branches du droit. En droit anglo-saxon, ce qu'il est raisonnable d'attendre d'une personne « normale » est appelé « *duty of care* » (une notion qui correspond approximativement au fait d'agir « en bon père de famille », en droit belge).

La responsabilité fiduciaire, à laquelle est soumise le *trustee*, est juridiquement plus contraignante que ce simple « *duty of care* ».

Pour Hardinges (2018), la responsabilité fiduciaire est même considérée, en droit anglais, comme le plus haut niveau d'obligation qu'une partie puisse avoir envers une autre.

Le trust est, nous l'avons dit, un dispositif extrêmement flexible. Les objectifs d'un trust peuvent varier et cette adaptabilité est l'un des avantages souvent cités. Un trust peut être établi pour bien des raisons : pour exécuter un testament, pour protéger des actifs contre des créditeurs ou des saisies, pour gérer des terres et des propriétés immobilières (land trust), pour des raisons fiscales, pour prendre soin d'une personne handicapée ou dépendante, pour transférer un patrimoine à des petits enfants en « sautant » la génération des parents, et bien d'autres raisons…

Certains trusts sont révocables par le *settlor*, d'autres sont irrévocables - même par le *settlor* - à moins d'obtenir la permission préalable des bénéficiaires. Certains trusts peuvent être anonymes, le bénéficiaire ignorant alors qui est son bienfaiteur (on parle de *blind trust*).

Un type particulier de trust est le trust « charitable » ou *public trust*. Penner (2019) explique que, contrairement aux autres types de trusts, un trust charitable n'opère pas au bénéfice de personnes spécifiques. Il est établi au profit d'une cause qui appartient à l'un des quatre grands domaines suivants : social, éducatif, religieux ou d'intérêt général.

Dans ce cadre, des personnes physiques peuvent tout à fait bénéficier du trust (par exemple des étudiants peuvent bénéficier d'un trust public à visée éducative) mais ces personnes ne sont pas les *bénéficiaires* au sens légal. Il n'existe pas de bénéficiaires au sens habituel. Le trust public ou charitable est au service de la cause désignée, pas des personnes qui pourraient en bénéficier. Et c'est donc à cette cause supérieure, que le ou les *trustee(s)* doivent allégeance. Différentes juridictions octroient à ce type de trust, sous conditions, des avantages fiscaux. Un tel trust ne peut poursuivre un but de lucre.

Exemples de trust

Exemple de trust privé, établi au profit d'une personne bénéficiaire :

- Mike Ilitch était un homme d'affaires américain ayant bâti sa fortune en vendant des pizzas. Il était également propriétaire d'équipes de baseball et de hockey, à Détroit. Un jour, il décida de constituer un trust pour venir en aide à Rosa Parks, une militante afro-américaine célèbre pour son engagement contre la ségrégation raciale aux États-Unis dans le cadre du mouvement des droits civiques. Rosa Parks avait besoin d'être relogée suite à une agression à son domicile et Ilitch utilisa le trust comme outil légal pour l'aider à couvrir les frais de son déménagement et ensuite à payer son loyer tous les mois pendant des années.

Exemple de trust public, établi au profit d'une cause d'intérêt général :

- En Angleterre et au Pays de Galles, un trust nommé le « canal and river trust » (https://canalrivertrust.org.uk/) est responsable de la gestion et de l'entretien de milliers de kilomètres de canaux et rivières. Outre un financement public, il collecte les droits d'usage (pour la navigation fluviale, etc.) et les affecte à la maintenance et à la valorisation des cours d'eau. Il agit en partenariat avec les autorités publiques locales (des membres élus ou désignés par les autorités locales constituent le conseil d'administration qui assume le rôle de *trustee*). Conformément à l'exception régissant le trust « charitable », aucune personne bénéficiaire n'est désignée, le bénéficiaire est ici la cause d'intérêt général.

Comprendre le trust dans le cadre du droit continental

Enfin, soulignons que le trust est un concept peu connu et généralement mal appréhendé dans les pays européens, que ce soit en droit belge ou dans les autres pays de droit continental.

Dans un article publié par la revue de la Fédération Royale du Notariat Belge, qui consacre au trust l'une de ses études, on peut lire :

« *Le trust désoriente le juriste de droit continental. Ce dernier tente de l'approcher – « ce n'est pas une sinécure ! » – en empruntant des voies qui lui sont connues : le démembrement du droit de propriété, la stipulation pour autrui, la donation, la fondation, l'association de fait... Mais ce faisant, il fait fausse route. Le trust n'est rien de ces institutions. Au mieux en présente-t-il certains traits. Mais cette ressemblance n'est qu'apparente. Le trust est une institution distincte – un concept sui generis* » (Van Boxtael et Fonteyn, 2015, p.46).

De la même manière, Scott (1966) notait déjà, il y a soixante ans que les académiques allemands, français, néerlandais et même ceux d'Amérique centrale ou latine avaient la plus grande difficulté à insérer le mécanisme du trust dans leurs systèmes jurisprudentiels respectifs.

Pour Van Boxtael et Fonteyn (2015), il est pourtant possible de comprendre, dans le cadre de notre système de droit civil, le trust. À condition de garder en tête trois particularités essentielles :

1. Le trust est dénué de personnalité juridique. Il n'a pas de personnalité morale. On ne peut donc pas le comparer à une société, une association ou une fondation. Il ne peut pas posséder de biens ou agir en justice. Par contre, il est reconnu par le droit international.

2. L'essence du trust est une relation juridique qualifiée de « fiduciaire ». Celle-ci joue un rôle central. Il s'agit principalement d'une relation entre le *trustee* et le bénéficiaire. En effet, dès que le *settlor* a mis des biens « en trust », il se retire. Le *trustee* se voit alors investi d'un certain nombre de prérogatives principalement liées à la détention et l'administration des biens du trust dans le cadre de cette relation. Mais il reçoit également des devoirs et des obligations qui encadrent ces pouvoirs d'administration des biens. Et les conditions auxquelles le *trustee* est soumis sont réelles et fortes. À tel point que le droit anglo-saxon parle de « *dual ownership* », comme si en pratique les biens mis en *trust* appartenaient déjà simultanément au bénéficiaire et au *trustee*. Cette notion de « *dual ownership »* désoriente les juristes continentaux qui n'y sont pas habitués.

3. Finalement, le trust crée un patrimoine distinct dans le patrimoine du *trustee*. Ce patrimoine est séparé des autres biens propres, il est « à côté ». Et les actifs détenus en gestion ne le sont qu'au titre du trust.

5 Émergence progressive du concept de *Data Trust*.

L'idée d'appliquer les concepts du trust et de la responsabilité fiduciaire à la gestion des données est apparue progressivement.

Le concept a connu récemment un engouement soutenu, en particulier depuis la publication en 2017 du rapport «*Growing the artificial intelligence industry in the UK*» rédigé par Hall et Pesenti à la demande du gouvernement britannique.

Mais il existe également toute une série d'autres travaux préalables ou contemporains sur lesquels il est intéressant de s'attarder afin d'obtenir un bon aperçu des réflexions, et des enjeux qui gravitent autour du concept de *Data Trust*.

Nous présentons ci-dessous une sélection de quelques étapes clés qui ont jalonné l'émergence progressive du concept. Cette sélection, forcément imparfaite, ne prétend pas être exhaustive. Elle permet, par contre, nous semble-t-il, de se faire une idée du contexte et des différents objectifs - parfois contradictoires - des acteurs qui s'intéressent au dispositif du *Data Trust*.

5.1 Kenneth Laudon (début des années 90)

Parmi les travaux les plus précoces figurent ceux du professeur Laudon, au début des années 1990.

D'après Khan et Pozen (2019), Laudon serait le premier à avoir employé le terme de « fiducie d'informations ». Ils rapportent que Laudon proposa en 1993 la mise en place d'un marché de l'information, alors qu'Internet était encore balbutiant. Il considérait que quand ce marché de l'information serait en place, des « fiducies d'informations » émergeraient naturellement.

D'après lui, ces fiducies accepteraient des dépôts d'information et chercheraient à en maximiser la rentabilité en les commercialisant ou par d'autres moyens.

5.2 Les frères Winickoff (2003)

En 2003, David Winickoff, et Richard Winickoff, respectivement docteur en droit et docteur en médecine, font le constat que les avancées dans le domaine de la génétique et de la bio-informatique, rendent les collections d'informations médicales et d'échantillons biologiques (sang et tissus) particulièrement précieuses pour la recherche pharmacologique.

Ils remarquent que dans plusieurs pays, des entreprises de biotechnologie constituent des banques génomiques à grande échelle (constituées de millions d'échantillons dans certains cas) et en commercialisent l'accès. Ce qui génère d'importantes questions éthiques, légales, médicales et sociales, y compris concernant les problématiques liées au respect de la vie privée.

Dans un article publié dans le *New England Journal of Medicine*, les frères Winickoff (2003) relèvent toute une série de problèmes posés par la pratique des *biobanks* privées à caractère commercial aux États-Unis :

- Les biobanks commerciales demandent la plupart du temps aux donneurs de leur remettre, d'avance, un consentement ouvert valable pendant une durée indéterminée. Ce qui ne peut être considéré comme un consentement éclairé ou un choix informé lié à une expérience spécifique, le sujet ne sachant rien des projets de recherche futurs pour lesquels ses échantillons vont être utilisés.

- Les biobanks commerciales présentent souvent leurs projets de manière trompeuse :

    - La collecte des informations et des échantillons par des médecins et des infirmières communique l'impression trompeuse d'un contexte scientifique ou éducatif non lucratif.

    - Les formulaires employés donnent l'impression que les échantillons de sang et de tissu sont sans valeur marchande et qu'ils seraient jetés, si le donneur ne consentait pas à leur utilisation à des fins de recherche. Or, en réalité, ils sont parfois conservés pendant des années et acquièrent une valeur de marché.

    - Les échantillons sont prélevés par les hôpitaux et les formulaires mentionnent des « protocoles de recherche en milieu hospitalier » alors qu'en réalité les hôpitaux commercialisent ces échantillons à des biobanks privées qui elles même les utilisent pour générer un profit.

- De plus en plus fréquemment, les comités d'éthique censés superviser les activités de recherche approuvent des protocoles à durée indéterminée. Ils renoncent également à la supervision éthique de certains projets de recherche. Censés être indépendants, ils sont parfois sous pression pour favoriser les intérêts des institutions qui les emploient.

Pour Winickoff et Winickoff (2003), les dons médicaux ont une signification morale profonde qui requière un meilleur respect du consentement du donneur, notamment son droit d'information et de retrait. De plus, si les échantillons médicaux sont d'ores et déjà commercialisés, se pose la question des droits de propriété et des bénéfices qu'en tire la communauté. L'intérêt général et les bénéfices pour la population étant des critères déterminants pour évaluer la justification éthique d'une recherche.

Pour régler ces problèmes, les deux frères américains proposent de recourir au *trust charitable*. Un modèle de *trust* où les bénéficiaires sont remplacés par une cause d'intérêt général (cf. supra p.15).

Dans le modèle qu'ils proposent, lorsqu'une personne marque son accord pour faire un don d'informations médicales, de sang ou de tissus, le formulaire de consentement précise que l'organisation recevant la contribution est une biobank mais d'un type différent, car organisée sous la forme d'un *trust charitable*. Celle-ci a donc l'obligation fiduciaire d'agir en *trustee* et d'assurer la protection de la contribution. Les bénéficiaires sont remplacés par l'intérêt général du public.

Ce modèle présente plusieurs avantages par rapport aux biobanks privées à caractère commercial. Par essence, l'objectif d'un *trust charitable* servant l'intérêt général est plus en accord avec le caractère altruiste d'un don médical que le but de lucre d'une entreprise commerciale. De plus, les hôpitaux qui sollicitent des dons endossent ici un rôle de gardiens et non plus de revendeurs. Enfin, l'architecture du trust peut être créée de manière à ce que les donneurs participent à la gestion du trust. Et le trust à l'avantage de la longévité, contrairement aux sociétés commerciales qui pourraient être amenées à revendre leur inventaire en cas de faillite.

Les possibilités de participation démocratique au processus décisionnel sont un point important de la proposition. L'établissement d'une biobank utilisant la structure du *trust charitable* permet, en effet, que les donneurs participent au comité de *trustees* qui assure la gouvernance de l'organisme.

Pour les frères Winickoff (2003), plusieurs formes de participation sont envisageables : par exemple, la mise en place d'un comité de donneurs qui évalue les demandes de recherche et dispose d'un droit de véto sur des projets spécifiques. Ou encore l'élection d'un ou plusieurs donneurs pour siéger directement au sein du comité de *trustees* qui administre la biobank.

Les possibilités sont nombreuses pour favoriser l'implication démocratique des patients : il est par exemple envisagé que tous les projets de recherche puissent être présentés sur un site web avec un délai suffisant pour permettre aux donneurs de se retirer d'un projet de recherche qu'ils n'approuveraient pas (selon un système d'opt out).

David et Richard Winickoff (2003) sont convaincus qu'un modèle coopératif basé sur le *trust charitable*, s'il est conçu avec soin, permettrait aux donneurs de se sentir en confiance lorsqu'ils soumettent à la biobank leurs données médicales lors des visites hospitalières. Ils ont la conviction que la participation démocratique des patients, accompagnée par le leadership de l'institution médicale créerait un sens communautaire permettant la défense de l'intérêt général, avec succès.

La proposition décrite par les deux frères américains dans leur article de 2003 ne constitue pas encore un *Data Trust* au sens où les chercheurs qui travaillent sur le numérique l'entendent aujourd'hui (c'est-à-dire entièrement digital). En effet, les actifs gérés par une biobank sont, en partie, physiques (des échantillons de tissus, d'organe, de sang…).

Toutefois, l'autre moitié des actifs détenus en gestion par une biobank sont déjà des informations pures (des données médicales) et elles aussi prennent de la valeur lorsqu'elles sont agrégées, exactement comme les data générées par le

secteur digital. Nous sommes donc déjà face un modèle hybride (une partie des actifs gérés par le trust sont tangibles, une autre intangible).

Force est de constater que ce modèle basé sur le trust charitable, tel que proposé par les frères Winickoff pourrait théoriquement s'appliquer aux données sans grand changement structurel.

Certains auteurs comme Milne et Al. (2021) ou Hardinges (2020) de l'Open Data Institute reconnaissent explicitement l'influence intellectuelle des frères Winickoff sur les réflexions qui ont mené à l'émergence du *Data Trust*.

Chez d'autres chercheurs qui ont travaillé, plus tard, la question des *Data Trust* strictement digitaux (Artyushina, 2019 ; Delacroix et Lawrence, 2019 ; McDonald, 2015 et 2021 ; Mills, 2019) ont voit émerger des préoccupations de gouvernance démocratique très proches et dont les arguments et les solutions sont forts similaires.

Pour ces raisons, les travaux des frères Winickoff ont un statut précurseur, au moins sur le plan intellectuel et ce bien qu'ils appartiennent au domaine médical et non digital. L'utilisation originale du mécanisme légal du trust afin de faire respecter les intérêts d'un groupe épars de sujets, est au cœur des propositions que l'on retrouvera plus tard, concernant le *Data Trust*.

5.3 Lilian Edwards (2004)

Un an plus tard, Lilian Edwards, une académique écossaise spécialisée dans les questions juridiques en lien avec Internet, la propriété intellectuelle et l'intelligence artificielle publiera un article intitulé « *The problem with privacy* ».

Plusieurs auteurs (Delacroix et Lawrence, 2019 ; O'Hara, 2019 ; Hardingues et Wells, 2019) considèrent qu'il s'agit de la première référence à un proto-concept de *Data Trust* tel qu'on l'entend aujourd'hui.

Dans son article, Edwards (2004) aborde plusieurs points qui restent d'actualité presque vingt ans plus tard : le déséquilibre de pouvoir qui existent entre les consommateurs et les sites web, la relative ignorance des consommateurs quant aux procédés techniques et aux méthodes de collecte de données, l'absence de réels choix concernant les politiques de protection de la vie privée, le manque de contrôle une fois les données collectées.

En raison de cette asymétrie de pouvoir entre individus et sites web, Edwards (2004) rejette l'idée selon laquelle un consommateur serait en position crédible lui permettant de négocier une compensation appropriée pour la cession consentie de ses données personnelles. En particulier parce qu'il est impossible au consommateur de connaître la réelle valeur de ses données numériques avant que celles-ci ne soient agrégées avec les données qui ont été collectées sur tous les autres utilisateurs du site web et des divers services digitaux de l'entreprise.

Ce sont les processus d'agrégation, le data mining, et la création de profils digitaux analytiques postérieurs à la collecte qui multiplient la valeur des données précédemment cédées gratuitement par les consommateurs.

Les *data collectors* et *data processors* accumulent donc des bases de données de grande valeur remplies de préférences et d'information personnelles que les utilisateurs de sites web ont laissées derrière eux, presque sans s'en rendre compte. Alors que du côté des utilisateurs, ces données sont pensées séparément et qu'individuellement elles sont perçues comme n'ayant presque pas de valeur.

Or, Edwards (2004) remarque qu'historiquement l'institution légale utilisée pour gérer de façon équitable un bien qui est créé à partir de dons de multiples petits contributeurs, et qui ne représente un élément de valeur substantielle que pris dans son ensemble, est le trust. Du moins, dans les pays de la *common law*.

« Si nous nous tournons vers un modèle inspiré du *trust*, nous pouvons imaginer un régime au sein duquel les données sont remises gratuitement, mais où les *data collectors* et les *data processors* ont une obligation continue d'honorer l'intérêt supérieur et la confiance des *data subjects* ». (Edwards, 2004, p.326).

Delacroix et Lawrence (2019) soulignent qu'Edwards fut la première à discerner le potentiel du *trust* comme mécanisme légal pour gérer une ressource digitale – les données personnelles – qui s'accumulent par petites contributions successives (et dont l'agrégat finit avoir de la valeur.

Pour concrétiser cette vision, elle propose ouvertement que les *data subjects* soient vus à la fois comme *settlors* et bénéficiaires d'un trust tandis que le *data controller* est considéré comme *trustee*. Ce qui signifierait que le *data controller* serait redevable d'une obligation fiduciaire vis-à-vis du *data subject*.

Alors que le contexte, le pays et le secteur d'activité sont tout à fait différents, Edwards arrive donc à une conclusion relativement proche de la proposition des frères Winickoff, un an avant elle. Avec le trust comme instrument idéal pour gérer un agrégat d'informations.

5.4 L'UK Biobank (2006)

En 2006, une gigantesque base de données biomédicale nommée *UK Biobank* (https://www.ukbiobank.ac.uk/) est établie au Royaume-Uni pour collecter à échéance régulière les données génétiques et médicales d'un demi-million de personnes recrutées à cet effet.

Conformément aux recommandations des frères Winickoff, trois ans plus tôt, l'*UK Biobank* s'éloigne des précédents modèles de Biobanks commerciales privées et adopte le modèle du *trust charitable* au service de l'intérêt général, piloté par un *board* de *trustees*.

Le modèle est proche de celui promu par les deux frères américains, en 2003, mais il existe toutefois une différence : le projet de l'UK Biobank se concentre sur l'agrégation des données de santé et non sur la collecte d'échantillons physiques. Le *trust* collecte principalement de l'information génétique et médicale. En ce sens, le projet est beaucoup plus proche de ce que sera un *Data Trust* du secteur numérique (fondamentalement, il s'agit d'une base de données). Il ne s'agit plus d'un modèle hybride.

Aujourd'hui, l'UK Biobank (2021) se décrit elle-même comme la databank la plus fournie de son genre, au monde. Elle est utilisée pour mener à bien de nombreuses recherches scientifiques sur une multitude de maladies (cancer, maladies cardiaques, etc.). Les données des participants sont anonymisées et mises à disposition de chercheurs scientifiques, à l'international.

Milne, Sorbie et Dixon-Woods (2020), expliquent que la recherche dans le secteur de la santé dépend de plus en plus de ces jeux de données à grande échelle qui sont soit collectés volontairement (à l'occasion d'études épistémologiques) ou sont générés à partir des dossiers médicaux électroniques.

La bonne volonté des donneurs est un facteur déterminant pour une partie substantielle de la recherche génétique et médicale. Dans un contexte d'innovation technologique rapide et d'inquiétudes liées au respect de la vie privée, la question de la confiance du public est donc centrale.

D'après Holm, Kristiansen et Ploug (2020), les citoyens ont une attitude positive vis-à-vis des recherches scientifiques qui utilisent les données de patients et, d'une manière générale, il existe une confiance envers les praticiens, les hôpitaux et les chercheurs qui traitent leurs données. Mais malgré cette confiance, une proportion significative de la population souhaite une forme de contrôle sur la manière dont leurs données sont utilisées et échangées. Et le fait de ne pas répondre à cette attente est un élément susceptible de diminuer la confiance des patients et leur volonté de collaborer à des collectes de données.

L'un des défis relevés par l'UK Biobank est de parvenir à concevoir un modèle digne de confiance et qui balance les différents intérêts (ceux du donneur et des autres parties prenantes) tout en restant pertinent si l'environnement (technologique, régulatoire, etc.) évolue dans le futur. Ce qui est sans cesse le cas.

La structure légale du trust charitable permet de présenter au donneur une entité durable avec des buts clairs, des obligations de respecter les objectifs énoncés, des règles de gouvernance transparentes, une charte, etc. Et comme l'institution est gérée activement, elle peut s'adapter en respectant l'ensemble des intérêts (en cas de changement législatif, par exemple).

Au cours des années qui suivirent l'établissement de l'UK Biobank, les réflexions issues du monde médical et celles issues du secteur digital vont continuer à converger lentement. Si bien qu'aujourd'hui, l'Open Data Institute (2021), une institution uniquement dédiée à la gestion des données numériques, n'hésite pas à considérer l'UK Biobank comme un *Data Trust* à part entière et à le citer en exemple sur son site web.

C’est également vrai dans l’autre direction : quinze ans après la mise en place de l’UK Biobank, certains chercheurs issus du secteur de la santé ont embrassé le mécanisme du *Data Trust*, tel que compris par les acteurs du secteur digital, et le revendiquent explicitement pour la gestion des données médicales (Milne et Al., 2021 ; Paprica et Al ; 2020).

5.5 Jack Balkin (2014 et années suivantes)

Mais nous n’en sommes pas encore là. Alors que des projets pratiques se mettent en place dans le secteur de la santé, le débat académique se poursuit dans le secteur digital. Car, ce qui est évident et établi de longue tradition dans le secteur médical (l’obligation fiduciaire des médecins envers leurs patients) ne coule pas de source dans le secteur des données digitales. Et il faudra plusieurs années avant que ces réflexions n’émergent.

Khan et Pozen (2019) soulignent l’importance d’une série d’articles rédigés et publiés, à partir de 2014 et au cours des années suivantes, par Jack Balkin, un éminent professeur de droit constitutionnel à l’université de Yale, aux États-Unis.

Comme Edwards plus de dix ans avant lui, Balkin (2017) observe que les utilisateurs ordinaires sont profondément dépendants et vulnérables face aux géants technologiques qui collectent, traitent, analysent et valorisent leurs données. Il estime que la technologie est constitutive de relations de pouvoir déséquilibrées entre les utilisateurs et les entreprises. Pour contrebalancer ce rapport de force et s’assurer que les géants technologiques ne trahissent pas la confiance des citoyens, Belkin propose de s’inspirer de l’obligation fiduciaire.

Il prend l’exemple des docteurs, des avocats, des comptables, et autres métiers similaires qui collectent de nombreuses informations personnelles et les traitent d’une manière qui n’est pas toujours comprise par leurs clients. Une relation de confiance est nécessaire en raison de cette asymétrie de connaissances. Pour cette raison, ces professionnels ont l’obligation légale d’agir de bonne foi, de servir les intérêts de leurs clients et patients, de ne pas créer de conflits d’intérêts et de ne pas divulguer certaines informations les concernant.

Pour Balkin (2017), les entreprises qui collectent de vastes quantités d’informations sur leurs utilisateurs (comme Facebook, Google, Microsoft, Uber, etc.) sont dans une position similaire caractérisée par son asymétrie : les utilisateurs sont complètement transparents pour ces organisations alors qu’à l’inverse les opérations de ces entreprises sont tout à fait opaques pour leurs utilisateurs.

Il estime donc que les fournisseurs d’accès à Internet, les moteurs de recherche, les réseaux sociaux, etc. devraient être soumis à une forme d’obligation fiduciaire similaire à celle des médecins et avocats, bien que plus limitée. Il propose de traiter ces entreprises comme des *fiducies d’information* avec pour objectif explicite qu’il leur soit interdit de trahir la confiance de leurs usagers après avoir collecté leurs données et qu’elles ne puissent pas travailler contre leurs intérêts.

Il est intéressant de noter que, pour Balkin (2017), ce sont les entreprises (comme Google ou Facebook, par exemple), qui doivent directement assumer une responsabilité fiduciaire et doivent être traitées, elles-mêmes, comme des *fiducies d'information*. Contrairement au point de vue d'Edwards (2004), dix ans plus tôt, pour qui cette responsabilité fiduciaire devait être endossée par une entité tierce (un trust constitué à cet effet).

D'après Khan et Pozen (2019), les propositions de Balkin eurent un certain retentissement dans le monde académique aux États-Unis et de nombreux juristes exprimèrent publiquement leur approbation. Des législateurs issus des deux grands partis américains manifestèrent également leur intérêt. Un groupe de sénateurs démocrates allant même jusqu'à introduire une proposition de loi (le *Data Care Act*) basé sur les propositions de Balkin, une première fois en 2018 puis une seconde fois en mars 2021 (cette dernière proposition suit toujours son cours à travers le long parcours législatif). Bien qu'il soit difficile d'estimer son degré de sincérité, Mark Zuckerberg, le CEO de Facebook, a également marqué son support publiquement, à deux reprises, envers la proposition de Balkin (Zittrain, 2019).

5.6 Sean McDonald (2015 et années suivantes) : le trust civique

Parallèlement aux réflexions de Balkin, Sean McDonald (2015) fait, lui aussi, le constat que nous laissons de plus en plus de traces digitales derrière nous, que les entreprises qui les collectent se les approprient à notre insu et sont, *de facto,* libres d'en faire ce qu'elles veulent.

Il considère que, malgré les incroyables progrès permis par la digitalisation, celle-ci est négativement affectée par des limites inhérentes aux organisations sous-jacentes. Les entreprises privées sont contraintes par leur stabilité financière et par leur obligation de maximiser le profit de leurs actionnaires. Tandis que les institutions non lucratives sont, elles, dépendantes de décisions discrétionnaires émanant de leurs directions ou de ceux qui les financent.

Pour McDonald (2015), malgré les bienfaits que la technologie procure, aucun modèle ne permet une participation publique satisfaisante pour la gestion des actifs digitaux.

Afin de pallier ce manque, il propose un modèle qu'il appelle le *trust civique*.

En tant qu'avocat chevronné (il est, entre autres, membre du barreau de New York, enseignant invité à Stanford et Harvard et conseiller auprès du comité 'éthique et intelligence artificielle' de l'*Institute of Electrical and Electronics Engineers*), il utilise son expertise légale pour donner une existence concrète à sa vision.

Bâtissant sur les idées d'Edwards (2004), il rédige les documents légaux nécessaires à la création d'un *trust* dont les règles de gestion sont spécifiquement pensées et adaptées pour répondre aux besoins spécifiques de la gestion d'actifs numériques et pour contrebalancer les asymétries de pouvoirs déjà évoquées.

À la différence des réflexions d'Edwards, qui restaient théoriques, McDonald va donc donner une existence concrète au modèle de *trust* qu'il préconise (et pendant plusieurs années, ensuite, il continuera à mettre des *data trust* en place par le biais de l'entreprise qu'il a cofondée et en partenariat avec le CIGI - le Centre for International Governance Innovation, dont il est un *senior fellow*).

Dans le modèle proposé et mis en place par McDonald (2015), une organisation endossant le rôle de *trustee* est créée pour détenir le code informatique et les données qui peuvent être générées par une technologie. Ce *trust civique* qui détient le code et les données accorde ensuite une licence d'exploitation sur ces actifs digitaux à une société commerciale contre rémunération.

Ce qui distingue ces *Trusts Civiques* des trusts normaux, c'est l'inscription d'une obligation fiduciaire tant pour le trust (et donc ses *trustees*) que pour la société commerciale qui bénéficie de la licence d'exploitation du code et des données.

Les entités ont toutes les deux l'obligation de mettre en place des processus de gouvernance participative. Et cette obligation est rédigée de manière à constituer une obligation fiduciaire au sens légal du terme (avec recours près les tribunaux en cas de non-respect, donc).

On a donc une structure dont les actes juridiques sont explicitement écrits pour qu'il existe des contre-pouvoirs entre acteurs. Avec l'obligation légale que soit mise en place une gouvernance démocratique participative.

McDonald (2015) explique que, comme dans le cas d'un trust classique, les cinq éléments fondamentaux sont présents :

- Le **Settlor** (ou Trustor) est l'entité ou la personne qui détient initialement les données ou le code (par exemple, parce qu'il est *Data Collector*). Il peut soit créer un nouveau *civic trust,* soit apporter ses données à un *civic trust* déjà existant.

  Une fois l'apport réalisé, les données ne lui appartiennent plus. Concrètement, il est toutefois commun que le *settlor* continue à bénéficier d'une licence d'exploitation (parfois même une licence exclusive).

  Point notable : En raison de la séparation des patrimoines entre le *settlor* et le *trust*, si le *settlor* vient à faire faillite ou est racheté, les actifs digitaux peuvent très bien continuer à être utilisés par le *trust*.

- Le **bénéficiaire** est la personne, le groupe de personne ou la cause au profit de qui les actifs sont exploités. Comme dans un trust classique, les bénéficiaires disposent d'un droit de recours contre le *trustee* si celui-ci n'effectue pas correctement son travail. Ils sont les bénéficiaires finaux, et à ce titre, ils « possèdent », d'une certaine façon, les actifs digitaux.

  McDonald (2015) recommande que les bénéficiaires d'un trust civique soient un groupe de personnes largement défini. Et ce, afin que les données

et technologies détenues par le trust (en particulier celles financées par de l'argent public ou émanant d'institutions publiques) servent le plus grand nombre.

Une large base de bénéficiaires permet également à un nombre important de citoyens ordinaires de bénéficier d'une base légale pour effectuer un recours, si nécessaire. C'est une façon pour le *Civic Digital Trust* de reconnaître que la technologie peut avoir un impact potentiel non seulement sur ses utilisateurs directs mais également sur la communauté.

- Les **actifs** détenus peuvent théoriquement être tout ce qui a de la valeur (comme dans un trust classique). Dans le cadre d'un *Civic Data Trust*, il s'agira typiquement de code au sens large (les bases de données, les standards, les process, l'interface, etc.) et de toutes les données le traversant.

  D'après McDonald (2015), le *Civic Data Trust* peut très bien détenir uniquement l'un ou l'autre mais la détention des deux (code et données) est l'option qui offre le plus de possibilités pour protéger l'objectif du *Trust Civic*. En effet, le fait de posséder les actifs digitaux constitue un levier si l'entreprise commerciale tierce à qui une licence d'exploitation est accordée ne remplit pas sa part du marché (il suffit de révoquer la licence accordée).

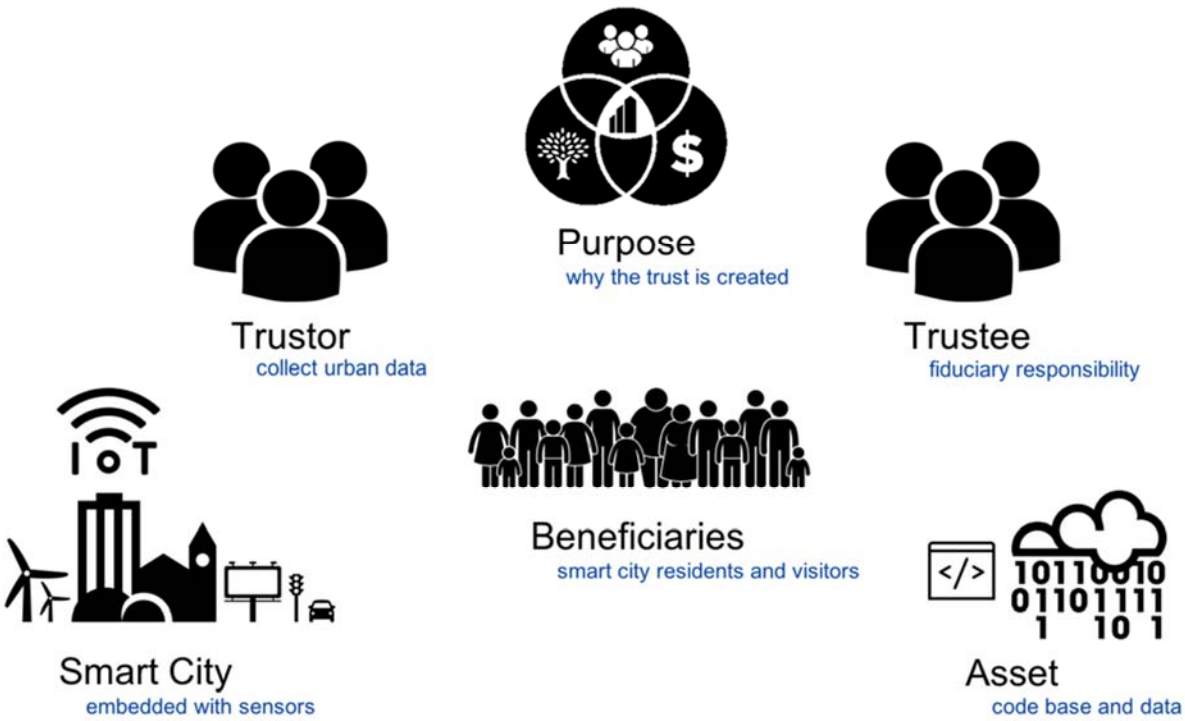

Figure 7 : Emploi d'un Civic Trust pour la gestion des données d'une smart city.

Source : MaRS (2018). *What is a Civic Digital Trust? Récupéré de https://marsdd.gitbook.io/datatrust/trusts/what-is-a-civic-digital-trust*

- L'**objet** (*purpose*) du *trust* est tout à la fois sa raison d'exister, sa mission et ses principes de gouvernance. Cet objectif officiel crée une obligation fiduciaire qui contraint légalement le *trustee* à respecter un certain nombre de valeurs et de procédures. Les bénéficiaires peuvent utiliser l'objet déclaré du trust pour forcer le *trustee* à rendre des comptes si nécessaire.

  Dans un *Civic Trust*, l'objet intègre explicitement des principes de gouvernance participative (et ceux-ci définissent les droits des usagers de la plateforme technologique). Comme tous les *trusts,* le *Civic Trust* est par nature un instrument modulable lors de sa conception mais l'objectif déclaré de McDonald (2015) étant de créer une entité juridique qui protège l'intérêt public face à l'asymétrie de pouvoir constatée, les statuts et les règles établies représentent donc cette volonté (Cf infra p. 86, Paprica et Al. - pour un exemple concret de règles de gouvernance participatives au sein d'un *data trust*).

  Lors de sa rédaction, l'objet du *Civic Data Trust* peut définir des modalités de gouvernance participatives qui correspondent aux besoins spécifiques du *trust* et sont adaptées sur mesure en fonction des buts poursuivis : élection représentative, représentation égalitaire, démocratie directe, partenariats avec des experts, pouvoir décisionnel pondéré par l'expertise ou par la participation…

- Le **Trustee** est la personne, l'organisation ou le groupe de personnes et d'organisations qui gère les actifs confiés au trust et poursuit le but officiel dans l'intérêt des bénéficiaires.

  D'après McDonald (2015), idéalement le *trustee* devrait disposer d'une stabilité financière suffisante que pour ne pas subir de pressions externes. Pour lui, dans le cadre de la vision démocratique et participative qui est promue, il est plus logique de prévoir dans les statuts fondateurs qu'un groupe ou une organisation endosse le rôle central de *trustee* plutôt qu'une personne unique qui concentrerait le pouvoir exécutif.

  McDonald (2015) encourage à expérimenter avec la participation des bénéficiaires (différents modes de participation, élection des *trustees*, etc.) et défend les bienfaits d'une pluralité de modèles de gestion différents (pondérés par l'expertise, démocratiques). Il estime également nécessaire de définir des règles de contrôle interne obligeant les organes décisionnels (au premier rang desquels le *trustee*) à rendre des comptes de manière transparente sur sa gestion. Ces règles permettant la révocation du *trustee* et son remplacement si nécessaire.

  O'Hara (2020) estime cependant que trouver des *trustees* talentueux peut s'avérer difficile car la position, elle-même, représente un risque inhérent

en raison de la responsabilité légale associée (en cas d'échec lors de la poursuite de l'objet du *trust*).

Enfin, McDonald (2015) souligne qu'en cas d'échec du trust, les utilisateurs ou le public peuvent décider de ce qu'il adviendra des actifs détenus en gestion (le code et les données). En cas de liquidation, ceux-ci peuvent être transférés à un autre *trust civique*, à un organisme public, au gouvernement et même à une entreprise privée. C'est un élément qui modère le risque et qui permet d'expérimenter avec la structure.

Le fait que cette décision soit prise démocratiquement et non par le biais d'une enchère au plus offrant constitue, pour McDonald, une étape essentielle vers la construction d'un patrimoine numérique commun et durable.

5.7 Hall & Pesenti (2017) : L'industrie de l'IA

Au mois d'octobre 2017, une publication vient bousculer les réflexions et les travaux en cours. Cette parution attire soudainement l'attention. Elle braque les projecteurs sur le concept de *Data Trust* et le pousse sur le devant de la scène internationale (du moins dans le secteur professionnel du big data) et, simultanément, engendre de la confusion sur ce que couvre exactement l'appellation *Data Trust*.

À la demande du gouvernement britannique, Wendy Hall, professeure en Sciences de l'Informatique à l'Université de Southampton et Jérôme Pesenti, Vice-Président de l'Intelligence Artificielle chez Facebook, rédigent ensemble un rapport intitulé « *Growing the artificial intelligence industry in the UK* ».

Ce rapport, auquel il est parfois fait référence sous la simple appellation '*The AI Report*' (Hardinges et Wells, 2019 ; Rinik, 2019), explore les énormes bénéfices économiques potentiels liés au développement de l'intelligence artificielle (cf. supra p.12).

Pour O'hara (2019), qui enseigne à l'Université de Southampton et qui est donc collègue direct de l'une des auteurs, le contexte tacite de ce rapport est la supposition que les États-Unis et la Chine ont l'avantage d'être des marchés plus larges que le Royaume-Uni et moins fragmentés que l'Union Européenne. L'enjeu est donc la recherche d'un avantage compétitif alternatif dans un secteur hautement stratégique et prometteur économiquement.

L'objectif de ce rapport, tel que décrit par Hall et Pesenti (2017) est de faire en sorte que le Royaume-Uni devienne l'endroit au monde le plus favorable pour que les entreprises actives dans le secteur de l'IA démarrent, croissent, prospèrent et réalisent tous les bénéfices que la technologie promet.

Après un rappel historique de la longue tradition britannique dans le domaine informatique et la remémoration des travaux d'Alan Turing, le rapport contient, comme on peut s'y attendre, une vaste quantité d'informations sur le secteur :

- Une carte géographique recensant la localisation des entreprises actives dans le domaine de l'IA ;
- La liste des acquisitions majeures du secteur ces dernières années ;
- Le détail des énormes impacts économiques escomptés grâce à l'IA (630 milliards de livres sterling pour le seul Royaume-Uni d'ici à 2035 et une augmentation de la croissance attendue de 2,5 à 3,9% en VAB - Valeur Ajoutée Brute) ;
- Des cas d'application et exemples d'entreprises novatrices ;
- Les applications dans le secteur public et privé ;
- La liste des investissements et soutiens financiers effectués par les pays à la pointe dans le secteur ;
- etc.

Mais c'est plus loin, dans la seconde partie du rapport (celle qui contient les recommandations des auteurs), que l'on trouve la section qui nous intéresse dans le cadre de ce travail.

Hall et Pesenti (2017) recommandent, pour développer le secteur de l'IA, de prendre des mesures qui facilitent l'accès aux données dans une vaste gamme de secteurs.

Et pour ce faire, ils recommandent plus spécifiquement, de :

- Développer les ***Data Trusts**,* afin d'améliorer la confiance et la facilité lors du partage de données ;
- Rendre une plus grande quantité de données de recherche lisibles par ordinateur (‘*machine readable*’) ;
- Encourager le *text and data mining* comme outils essentiels standards pour la recherche.

Dans ce cadre, le rapport assigne des rôles à certaines institutions (Universités, organismes spécialisés comme l'Alan Turing Institute ou l'Open Data Institute, etc.).

18 recommandations détaillées sont classées en quatre grandes catégories. La toute première de ces recommandations, consiste à :

- Créer un programme de partenariat public / privé pour développer les ***Data Trusts*** dans le but de faciliter l'accès aux data.

Dans cette partie de leur rapport, Hall et Pesenti (2017) décrivent ensuite les *Data Trusts* comme des dispositifs cadre d'accords éprouvés et fiables qui garantissent des échanges sûrs et mutuellement bénéfiques.

Les auteurs envisagent qu'une organisation de soutien (qu'ils appellent le DTSO – *Data Trust Support Organisation*) soit mise en place pour diriger le développement d'outils, de modèles et de conseils. Grâce à cette organisation centrale, ceux qui détiennent des données et qui souhaitent les partager et ceux qui en consomment et qui souhaitent y avoir accès pourraient se rencontrer facilement pour former des *Data Trusts* au gré de leurs besoins et seraient soutenus lors de leurs démarches.

Hall et Pesenti (2017) proposent que le rôle d'organisation de support (le DTSO) soit rempli par une institution neutre et experte. Compte tenu de l'importance de la confiance pour le succès du dispositif, ils estiment qu'il convient de sélectionner une institution dont l'efficacité opérationnelle en matière de cybersécurité et de gestion de données est déjà reconnue. *The Royal Academy of Engineering*, *the Open Data Institute*, *the Digital Catapult* ou encore *the Royal Society* sont cités.

Il est également proposé que le DTSO, cette organisation support, puisse agir comme *trustee*. En tant que tiers, elle pourrait ainsi aider à gérer un *Data Trust.*

Les fonctions principales du DTSO telles que décrites par Hall et Pesenti (2017) seraient de :

- Fournir un cadre qui définisse clairement les données, ou les flux de données, que les parties acceptent d'échanger ;
- Négocier le but du partage de données (l'objet) et les utilisations de celles-ci (y compris analytiques) ;
- Convenir des mécanismes techniques de transfert et de stockage de données ;
- Déterminer les modalités de distribution des bénéfices commerciaux.

Le rapport prévoit encore un plan de déploiement en trois étapes et propose le secteur du transport comme candidat potentiel pour les phases de test (car les organisations du secteur pourraient tirer d'importants bénéfices de l'IA, tout en encourant des risques minimes).

Le rapport contient également quelques recommandations pour augmenter la quantité disponible de travailleurs hautement qualifiés dont le secteur des données a besoin et améliorer leur niveau de compétence ainsi que diverses autres recommandations, dont certaines directement adressées aux pouvoirs publics.

Évolution de la notion de Data Trust

Dans leur rapport, Hall et Pesenti (2017) promeuvent le *Data Trust* comme moyen de faciliter le partage des données et de soutenir l'accès des entreprises, du gouvernement et du monde académique aux data qui sont la matière première nécessaire au développement de l'IA.

Il est donc principalement question de *data sharing*, de mise en commun de banque de données et du développement économique de l'industrie de l'Intelligence Artificielle.

Les bénéfices recherchés par Hall et Pesenti (2017) sont :

- La création d'un cadre de confiance pour l'utilisation des données sensibles.
- La baisse des coûts de transaction liés au partage de données.

Il y a donc une évolution de la notion de *Data Trust* par rapport aux précédents travaux. Les considérations relatives à l'asymétrie de pouvoirs entre acteurs, soulignées par les auteurs précédents (Edwards, 2004 ; McDonald, 2015 ; Balkin, 2017) sont complètement évacuées.

Même le cadre de confiance dont il est ici question ne vise plus à rassurer les citoyens quant au respect de leur vie privée mais bien à encourager les collecteurs de données (en particulier les entreprises trop petites pour rivaliser directement avec les GAFA) à mettre leurs données en commun. La recherche d'un cadre de confiance sert donc ici à surmonter certaines réticences entre acteurs économiques (intérêts concurrentiels perçus comme divergents, crainte de perte de contrôle, souhait de protéger une forme d'exclusivité, secrets professionnels, absence de *know how,* etc.) et non le droit à la vie privée des *data subjects*.

Malgré ce focus sur le *data sharing* et le développement de l'IA, il est bien question de *Data Trust*, qui a une place centrale dans le rapport. Mais le dispositif est avant tout décrit ici comme un outil de partage de données. Et la définition fournie dans le rapport est volontairement vague et ouverte :

« Ces trusts ne sont pas des entités ou des institutions légales mais plutôt un ensemble de relations sous-tendues par un cadre reproductible, conforme aux obligations des parties, pour partager les données d'une manière juste, sûre et équitable. » (Hall et Pesenti, 2017, p.46).

D'après cette définition, le *Data Trust* qu'Hall et Pesenti (2017) appellent de leurs vœux n'est donc plus un véritable *trust* au sens légal. Il ne fait que s'inspirer librement du concept. Et ce que signifie un partage des données « juste, sûr et équitable » est laissé à l'appréciation de chacun.

Plusieurs auteurs (Delacroix et Lawrence, 2019 ; Mills, 2019 ; O'Hara, 2019 ; Lau, Penner et Wong, 2019) soulignent l'importance de cette différence majeure. En effet, le choix d'Hall et Pesenti de ne pas adhérer à la définition légale du trust a pour effet de soustraire les parties à la responsabilité fiduciaire (« Ces trusts ne sont pas des entités ou des institutions légales »). Ce qui entraîne l'impossibilité de recourir au tribunal pour rendre exécutoire l'obligation fiduciaire, si nécessaire. L'abandon de la structure légale du trust constitue un virage fondamental par rapport aux travaux des auteurs précédents.

Certains auteurs comme Delacroix et Lawrence (2019) se montrent très critiques vis-à-vis de ce choix, allant jusqu'à estimer que le nom de *Data Trust*, dans ce contexte, n'est rien d'autre que du marketing.

Un point de vue avec lequel Rinik (2019), conférencière à l'Université de Winchester sur le droit de la propriété intellectuelle et le droit du *trust*, est en accord. Très critique, elle estime même que, les vagues notions « d'intérêt public » mentionnées dans le rapport ne sont qu'une façade derrière laquelle se dissimule l'appétit vorace des entreprises et des gouvernements. Et que le seul objectif de ce *Data Trust* allégé est de mettre les acteurs en confiance grâce à l'appellation de *trust* pour s'approprier ensuite leurs données personnelles et alimenter une croissance économique continue, tout en bafouant le droit à la vie privée.

D’autres comme Reed et Ng (2018) se montrent plus réceptifs et estiment que cette vision d’un *Data Trust* qui ne serait qu’inspiré du *trust* légal pourrait se concrétiser au sein d’un cadre contractuel. Ils estiment que ce modèle pourrait être utile pour développer la mise en commun de données et le développement de l’IA, en particulier s’il était déployé par secteurs (de manière à ce que toutes les données d’un même secteur soient soumises aux mêmes traitements). L’enjeu étant de permettre à des entreprises de petite ou moyenne taille (par rapport aux GAFA) de se grouper et de mettre en commun leurs ressources dans un cadre de confiance.

D’autres universitaires comme Reed et Ng (2018), se montrent plus réceptifs et estiment que cette vision d’un *Data Trust* qui ne serait qu’inspiré du *trust*, sans en adopter la structure légale et les contraintes, pourrait se concrétiser au sein d’un cadre contractuel. Ils estiment que ce modèle pourrait être utile pour développer la mise en commun des données et le développement de l’IA, en particulier s’il était déployé par secteurs (de manière à ce que toutes les données d’un même secteur soient soumises à des traitements standardisés). L’enjeu étant de permettre à des entreprises de petite ou moyenne taille (par rapport aux GAFA) de se grouper et de mettre en commun leurs actifs digitaux dans un cadre de confiance alors qu’actuellement, isolées, elles ne sont pas en mesure d’entrer en concurrence avec les géants technologiques qui occupent une position oligopolistique (cf. supa p.38).

Reed et Ng (2018) reconnaissent toutefois que, de manière réaliste, il ne serait pas possible aux *data subjects* d’être une partie prenante d’un tel contrat - contrairement aux modèles jusqu’à présent défendus par Edwards (2004) et McDonald (2015). Le *Data Trust* tel que proposé par Hall et Pesenti dans leur rapport ne serait donc, d’après eux, qu’un arrangement contractuel entre seules entreprises et institutions.

En résumé, le modèle proposé par Hall et Pesenti (2017) diffère fondamentalement sur deux points clés, des travaux précédents :

- L’objectif principal est le partage de données pour faciliter l’exploitation économique dans le secteur de l’IA, et non plus la protection des acteurs vulnérables.
- Le mécanisme s’inspire librement du *trust* mais sans en reprendre la structure légale ni l’obligation fiduciaire, centrale au dispositif.

Le rapport « *Growing the artificial intelligence industry in the UK »* eut un très large retentissement lors de sa sortie.

McDonald (2019) relève que le gouvernement britannique décida de l’utiliser comme cadre de référence pour effectuer un investissement d'un milliard de livres sterling dans le secteur de l'intelligence artificielle.

À partir de cette publication, d’autres acteurs majeurs comme Alphabet et la Commission Européenne commencèrent également à s’intéresser très fortement au dispositif du *Data Trust*.

5.8 Sidewalk Toronto (2017 – 2020) : La smart city

La smart city

La gestion des données urbaines au sein d'une *smart city* est un autre secteur pour lequel la collecte massive de données peut, théoriquement, être à l'origine d'importants bénéfices sociaux pour les habitants de la ville mais où les asymétries de pouvoir et les défis relatifs au respect de la vie privée sont également aigus.

Avec la multiplication des objets connectés, l'environnement numérique va connaître de profonds bouleversements dans les années à venir. Comme nous l'avons précédemment évoqué, la Commission Européenne (2020) prévoit que d'ici à 2025, 80% du traitement des données aura lieu de manière décentralisée, dans des *smart devices* connectés (cf. supra p.18) et non plus dans des *data centers* et des installations centralisées.

Cette tendance de fond risque d'avoir un effet profond sur la gestion des villes de demain. Pour Picon (2018), spécialiste de l'histoire de l'architecture et de la technologie qui enseigne à Harvard et à l'école polytechnique de Lausanne, la juxtaposition d'un imaginaire cybernétique à la cité n'est pas un phénomène nouveau. Mais les développements technologiques récents, en particulier la combinaison de capteurs, d'infrastructures connectées et de géolocalisation en temps réel marque un tournant et va permettre la collecte d'une quantité gigantesque de données urbaines.

Non seulement la masse d'informations brute générée dans les villes va s'accroître considérablement mais surtout, elle va permettre de cerner les phénomènes, les comportements et les flux de façon beaucoup plus granulaire et instantanée, voire prédictive.

La ville de demain, bardée de capteurs, de compteurs, de caméras et de portails magnétiques offrira un état des déplacements et des consommations en temps réel plutôt que sous la forme de mesures de débits agrégées comme c'était le cas autrefois. Avec des applications dans le secteur de la santé, de la lutte contre la criminalité ou pour alléger la congestion des infrastructures : voirie, transport, réseau électrique, eau et assainissement … pour ne citer que les plus évidentes.

Non seulement il sera possible de suivre simultanément des milliards d'occurrences mais ces traces digitales vont également s'accumuler, ce qui va permettre d'avoir accès à des représentations beaucoup plus fidèles de la vie quotidienne et des consommations au sein de la cité, à la fois en temps réel et sur des périodes étendues.

Or, le déploiement des technologies du big data, et potentiellement de l'IA, dans la cité pose les mêmes problèmes de gestion, d'asymétrie de pouvoir entre acteurs et suscite les mêmes tensions en matière de droit à la vie privée qu'ailleurs. Et, peut-être, y sont-ils même plus aigus encore, s'agissant de l'environnement urbain, lieu de vie des citoyens où il sera quasi impossible de s'y soustraire.

Pour Picon (2018), trois groupes d'acteurs dont il faut tenir compte coexistent :

- Le premier groupe est composé d'entreprises du numérique qui se tournent vers les *smart cities* dans le but de développer de nouveaux marchés. Initialement, les entreprises faisant partie de cette catégorie étaient les grands équipementiers (IBM, Cisco), rejoints, avec le développement du big data, par les entreprises du secteur de la donnée.

- Le second groupe d'acteurs est constitué de politiques (particulièrement les maires) animés par des soucis de gestion efficace ou désirant cultiver leur image, à la pointe du progrès.

- Le troisième et dernier groupe est constitué des habitants de territoire urbain, équipés de smartphones, sans qui rien ne serait possible. Ils génèrent les données numériques qui alimentent les dispositifs de la *smart city* par le biais de la géolocalisation ou en fournissant des inputs directs.

Des développements et des expérimentations sont en cours dans plusieurs pays, souvent à l'échelle d'un quartier lorsqu'il s'agit d'une grande métropole ou parfois sur l'ensemble du territoire dans le cas de villes de taille plus modeste.

Soupizet (2020) liste quelques projets pilotes de *smart cities* en France (à Dijon et à Angers) ou plus loin, à Songdo en Corée du Sud, à Singapour ou encore dans le quartier de *Quayside* à Toronto, au Canada.

L'un de ces projets de *smart city*, celui de Toronto, mené par une filiale d'Alphabet, a été brusquement arrêté en mars 2020, dans un contexte de tensions avec les habitants et les autorités publiques. Malgré cet échec et son arrêt soudain, ce projet nous intéresse particulièrement car il est considéré, par de nombreux auteurs comme une bonne illustration des tensions avec les géants technologiques et des difficultés à surmonter pour mettre en place un *Data Trust* en milieu urbain. (Artyushina, 2020 ; O'Hara, 2020 ; Rinik, 2019).

Il constitue, en quelque sorte, un parfait « mauvais exemple » dont bien des enseignements peuvent être tirés. Arrêtons-nous-y un instant.

Alphabet / Sidewalk Toronto

En mai 2017, *Waterfront Toronto* (une agence gouvernementale canadienne en charge de la revitalisation des rives du lac Ontario) lança un appel à projets pour le développement du quartier de *Quayside,* après un siècle d'échecs urbanistiques.

Quelque mois plus tard, en octobre 2017, un accord est conclu avec *Sidewalk Labs*, une filiale d'Alphabet, et le développement de la première *smart city* s'appuyant sur les technologies de Google est annoncé publiquement sous le nom de *Sidewalk Toronto*.

D’après Artyushina (2020) une doctorante de Toronto, spécialisée dans les questions de gouvernance digitale et ayant étudié localement le déroulement des événements dans le quartier de *Quayside*, le projet rencontra des problèmes et fut l’objet de vives controverses dès le départ :

- Sous les pressions du cabinet du Premier ministre canadien, favorable au projet, les responsables de l’organisme public *Waterfront Toronto* ne disposèrent que de quelques jours, un délai largement insuffisant, pour étudier l’accord proposé. Ce qui fit immédiatement scandale.

- *Waterfront Toronto* et *Sidewalk Labs*, la filiale d’Alphabet, gardèrent les détails de l’accord secret et n’en rendirent public qu’un bref résumé.

- Les manœuvres secrètes et la rétention d’informations perdurèrent pendant les deux ans et demi que dura le projet, avant qu’il échoue finalement en mars 2020.

- Durant les 9 premiers mois, *Sidewalk Labs* refusa simplement de fournir toute documentation sur le projet, même aux fonctionnaires dont le travail était d’évaluer la proposition.

- Quand la pression publique devint trop forte, *Sidewalk Labs* publia des douzaines de longs documents qui détaillaient les aspects les moins controversés (les logements sociaux, les constructions en bois, etc.) mais ne divulguaient rien sur le business model de la smart city ni sur la collecte de données.

- Un an après le lancement du projet, les experts en protection de la vie privée, engagés comme consultants, démissionnèrent et rompirent le silence en prenant la parole dans la presse pour dénoncer les plans de gestion de données secrets de *Sidewalk Labs*. Ils révélèrent que la filiale d’Alphabet avait décidé de rompre sa promesse d’anonymiser les données collectées au sein de la *smart city* pour satisfaire des acheteurs potentiels et réclamait la possession de l’ensemble de la propriété intellectuelle générée par la *smart city*.

- En juin 2019, *Sidewalk Labs* publia un master plan censé dissiper les inquiétudes des habitants concernant les craintes d’une surveillance digitale généralisée et la privatisation de la gouvernance urbaine. Contrairement au résultat escompté, le document de 1500 pages alarma les citoyens, les experts en protection de la vie privée et les autorités publiques de la ville tant il rendait claire l’intention de l’entreprise de se positionner en monopole local. D’après ce document, *Sidewalk Labs* avait l’intention de créer et de contrôler non seulement une vaste gamme de services digitaux mais également sa propre infrastructure, l’ensemble du parc immobilier et des services publics. Malgré des affirmations contraires de l’entreprise, le document faisait clairement ressortir l’intention de la

société de se positionner comme fournisseur central et inévitable à tout niveau de la cité numérique.

- Plus tard, un document interne, surnommé le *Yellow Book*, contenant le détail des plans de *Sidewalk Labs* concernant la gestion des données digitales, fuita et révéla que l'entreprise comptait implémenter un système de « *crédit social »*. Sur base de leur réputation et de leur volonté à partager leurs données personnelles, les résidents et les entreprises de la *smart city* gagneraient ou perdraient du « *crédit social »* et auraient accès à plus ou moins de services.

- À la lecture du *Yellow Book*, il devint aussi évident que *Sidewalk Labs* estimait que son projet devait bénéficier d'exceptions tout à fait extraordinaires, y compris le pouvoir de lever l'impôt dans la *smart city* et d'être considéré comme l'autorité de référence en matière de taxes et de finance. Le plan de l'entreprise prévoyait notamment d'obtenir le pouvoir de lever les taxes sur la propriété, de les collecter et de les dépenser.

- Enfin, des communications internes, entre l'entreprise et ses partenaires, ont démontré l'intention de *Sidewalk Labs* de monétiser les données captées dans la *smart city* alors que la filiale d'Alphabet n'avait de cesse d'affirmer publiquement le contraire pendant toute la durée du projet.

Pour Artyushina (2020), qui a étudié localement plus de 3000 pages de documentation relatives au projet, il ne fait aucun doute que la filiale d'Alphabet poursuivait une stratégie délibérée typique des grandes plateformes digitales (Alphabet, Amazon, Facebook).

Dissimulés derrière une communication publique conciliante et rassurante vis-à-vis des citoyens, ses objectifs cachés étaient :

- d'obtenir l'ascendant sur ses partenaires (y compris les autorités publiques) ;
- de briguer un monopole sur les infrastructures physiques et, ce faisant, de parvenir à contrôler - seule - l'ensemble des actifs digitaux ;
- de profiter économiquement de tous les acteurs ;
- et de se positionner légalement en plateforme intermédiaire pour éviter l'effet des réglementations.

Les résultats recherchés étaient de collecter, de manière illimitée les données en confinant les *data subjects* dans un rôle de donneurs passifs, sans égard pour leur participation citoyenne à la gouvernance de la cité numérique.

Lorsque nous avions étudié la typologie des biens appliquée aux données dans la première partie du présent travail, nous avions constaté que les géants numériques tentaient de traiter les données comme des biens à péage (cf. supra p.21) et que, les données étant des biens publics, pour en obtenir une rente économique, il leur était nécessaire de parvenir à exclure les autres utilisateurs (critère d'exclusion) en obtenant le contrôle de la couche physique (le support matériel).

Ici nous pouvons constater que la stratégie déployée par la filiale d'Alphabet revient effectivement à chercher à contrôler l'infrastructure de la ville pour monétiser les données pour pouvoir les traiter comme un bien à péage. Cela correspond à notre précédente analyse.

Les données étant fondamentalement des biens publics (auxquelles s'ajoutent les problématiques additionnelles liées au respect de la vie privée, pour les données à caractère personnel, cf. supra p.23), nous constatons également, sans surprise, qu'une telle manœuvre crée d'importantes tensions vis-à-vis des autres acteurs qui ont des conceptions différentes de la nature des données et qui, donc, estiment qu'elles doivent être traitées différemment.

La proposition de Data Trust pour gérer les données de la smart city de Quayside à Toronto.

En 2018, dans le contexte conflictuel que nous avons décrit, alors que les tensions sont vives entre *Sidewalk Labs* et la communauté urbaine, la filiale d'Alphabet propose d'établir un *data trust* pour répondre aux inquiétudes citoyennes concernant la collecte de leurs données personnelles.

La récente parution du rapport d'Hall et Pesenti, il y a peu, eut un retentissement international si bien que le concept de *Data Trust* est, en quelque sorte, « sur le radar » des professionnels du secteur de la donnée de masse qui se tiennent au courant des innovations en matière de *data gouvernance* et d'IA. Il n'est donc pas surprenant qu'à ce moment, *Sidewalk Labs* ait pensé au *Data Trust* pour alléger les tensions autour de son projet à Quayside. L'idée étant « dans l'air du temps » quelques mois après la parution du « *AI Report* ».

D'autant plus qu'un *Data Trust,* à condition d'être bien conçu et bien appliqué, peut objectivement être une réponse tout à fait adaptée pour rééquilibrer les rapports de force et alléger les tensions que connait le projet de *smart city*. L'idée de recourir à un *Data Trust*, n'est donc pas mauvaise, en soi.

Cependant, comme le fait remarquer McDonald (2019), un *Data Trust* ne crée pas automatiquement un modèle de gouvernance juste et démocratique, de par sa simple présence. Comme le *trust* « classique », le *Data Trust* est fondamentalement un instrument modulable et non déterministe.

Il peut être conçu pour favoriser un fonctionnement démocratique, mais également autocratique ou ploutocratique. Son design peut être pensé pour générer, ou pas, des revenus. Il est un outil adéquat pour rééquilibrer les asymétries de pouvoir, mais il ne les règle pas automatiquement de par sa simple existence. Pour qu'il le fasse, il reste nécessaire de développer volontairement une stratégie et un plan en ce sens. Bref, pour livrer le résultat escompté, le *Data Trust* doit incorporer un objet et des règles de fonctionnement qui soient cohérents avec les objectifs promis.

Lorsque *Sidewalk Labs* (2018) dévoile les premiers documents de présentation, on peut lire qu'un *Civic Data Trust* sera créé. Et qu'il :

- s’agira d’une entité indépendante dotée d’un comité de supervision incluant une représentation des citoyens, au sens large ;

- endossera une responsabilité fiduciaire dans le but de servir l’intérêt des *data subjects* et l’intérêt public ;

- s’assurera que la valeur créée à partir des données revienne aux personnes, aux communautés, au gouvernement et aux entreprises ayant généré ces données, et ce dans un cadre qui respecte le droit à la vie privée et qui soit sécurisé.

Les documents rendus publics ne précisent pas explicitement l’approche conceptuelle qui a inspiré *Sidewalk Labs*, ni la structure légale proposée mais différents éléments du projet semblent présenter des similitudes avec le modèle du *Civic Data Trust,* de McDonald (cf. supra p. 56) :

- L’emploi du terme ‘*Civic’ Data Trust* ;
- L’indépendance du *Trust ;*
- L’existence d’une responsabilité fiduciaire ;
- La participation des habitants, aux décisions ;
- L’existence d’une mission d’intérêt général.

Les documents de présentation précisent encore que le *trust* « contrôle, gère et rend accessible au public toutes les données qui peuvent raisonnablement être considérés comme un actif public et constitue un ensemble de règles qui s’appliqueront à tous, y compris *Sidewalk Labs*» (Sidewalk Labs, 2018, p.10)

De plus, l’entreprise fait la promesse de partager les profits résultant de l’exploitation des données avec le public ou les parties tierces disposant d’un intérêt légitime.

Elle explique même clairement que le but de la mise en place de ce *Data Trust*, est de s’éloigner d’un modèle où *Sidewalk Labs* possède et contrôle seul l’ensemble des actifs digitaux.

En observant la suite des actions de l’entreprise, Artyushina (2020) constate toutefois que ces promesses sont des coquilles vides et qu’en pratique, le trust ne met pas un terme aux plans de *Sidewalk Labs* d’acquérir la propriété et le contrôle des données.

Pour la chercheuse canadienne qui a étudié localement le déroulement du projet, la mise en place du *Data Trust* constitue une manœuvre de plus pour s’accaparer les données des citoyens et les transformer en actifs dans le but de se créer une position de rentier.

Concrètement, Artyushina (2020) constate que *Sidewalk Labs* va tenter de modifier la définition légale de ce qu’est une donnée.

L’entreprise invente le terme de « urban data » pour désigner les données anonymisées qui sont collectées dans l’espace public ainsi que dans les endroits

semi-privés : rues, restaurants, halls de buildings, etc. (Digital Governance , 2018, p.14).

Il est fait une distinction entre ces « urban data » et les données à caractère personnel (désignées sous l'appellation de « conventional data »).

Sidewalk Labs tentera d'argumenter que seules les « *urban data* » (c'est-à-dire les données publiques ou semi-publiques) peuvent être considérées comme un actif public et mises en gestion au sein du trust et que les autres données, les « *conventional data* » (c'est-à-dire les données à caractère personnel), doivent rester un actif privé appartenant à la filiale d'Alphabet et à ses partenaires.

En procédant de la sorte, *Sidewalk Labs* tente de mettre la main sur les données personnelles des habitants, malgré la mise en place du *Data Trust* et de sécuriser sur celles-ci des droits de propriété pour ensuite les monétiser.

Lorsque cette tactique échoua, l'entreprise tenta à nouveau de créer différentes catégories légales distinctes pour les données (quatre cette fois) dans le but de contourner la loi canadienne en utilisant d'obscures règles de territorialités en lien avec l'existence d'un accord de libre-échange.

En fondant notre analyse sur la typologie des biens détaillée en première partie (cf. supra p.19), nous constatons que les diverses stratégies de l'entreprise, avant ou après la mise en place du *Data Trust* ont toujours, fondamentalement, le même objectif : Elles consistent à obtenir le contrôle sur les données personnelles, à en limiter l'accès (critère d'exclusion) afin de les monétiser c'est-à-dire de les exploiter comme un bien à péage, exactement comme nous l'avions vu.

Cet exemple confirme également qu'il existe pour l'entreprise un marché « biface », tel que Monnerie (2018) l'avait décrit (cf. supra p.36) : d'un côté *Sidewalk Labs* cherche à fournir des services digitaux à la ville et aux habitants qui sont ses premiers clients, mais d'un autre, elle cherche simultanément à vendre les données personnelles collectées à des acheteurs externes, qui sont un second groupe de clients distincts. Et ce malgré, le fait que les intérêts de ces deux catégories de clients soient en opposition.

D'après Artyushina (2020), *Sidewalk Labs* apporta ensuite des changements fondamentaux à sa proposition de *trust* :

- Le *trust* fut renommé 'Urban Data Trust' et non plus 'Civic Data Trust' abandonnant toute référence à la gouvernance civique (et au modèle de McDonald).

- *Sidewalk Labs* annonça que le trust n'en serait pas un au sens légal. La structure légale du *trust* avec ses rôles et ses obligations ne serait finalement pas utilisée et contrairement à ce qui avait été précédemment annoncé, il n'y aurait donc pas d'obligation fiduciaire (ni par conséquent de recours possible devant les tribunaux).

Après cette annonce, *Sidewalk Labs* ne prit jamais la peine de définir la structure du *trust* en détail, ni même ses relations avec les agences gouvernementales.

On comprend néanmoins que l'abandon de la responsabilité fiduciaire et le fait que *Sidewalk Labs* cherche à soustraire les données personnelles du dispositif réduisent à néant l'utilité du *trust* du point de vue de l'intérêt des citoyens.

Artyushina (2020) relate qu'en novembre 2019, l'agence publique *Waterfront Toronto* finit par rejeter cette proposition de *Data Trust* complètement dénaturée, au point de n'être plus que cosmétique. Quelques mois plus tard, en mars 2020, *Sidewalk Labs* annula complètement le projet de *smart city,* citant des raisons financières liées à la crise du COVID.

La question du consentement dans la smart city

Wylie (2021), une critique proéminente du projet, relève que la proposition de gouvernance digitale de *Sidewalk Labs* fait tout à fait l'impasse sur la question essentielle de savoir, si les résidents de la cité doivent être traqués ou pas.

Pour Rinik (2019), la collecte de données dans l'espace public, l'emploi de caméras, de capteurs et d'éventuels logiciels de reconnaissance faciale est un sujet hautement controversé qui pose la question éthique de la surveillance constante des citoyens. Particulièrement lorsqu'elle est opérée par de larges entreprises privées.

Dans le cadre du projet de smart city à Quayside, l'un des points qui suscita rapidement l'inquiétude des citoyens est l'absence de mécanisme d'opt-out pour les résidents.

Sidewalk Labs répondit à ces critiques en introduisant un principe de « consentement par panneau d'affichage ».

Figure 8 : Consentement par panneau d'affichage à *Sidewalk Toronto*.

Source : Lu, J. (2019). *How can web ring transparency to urban tech ? These icons are a first step.* Sidewalk Talks.
Récupéré de https://medium.com/sidewalk-talk/how-can-we-make-urban-tech-transparent-these-icons-are-a-first-step-f03f237f8ff0

D'après Lu (2019), l'objectif de ce système de signalisation était d'aider à expliquer le rôle des différents capteurs invisibles présents dans la ville numérique : les icônes noires indiquant le but de la collecte de données, les icônes bleues et jaunes renseignant, elles, le type de données collectées (bleu pour anonymisées, jaune pour identifiables). De plus, les icônes blanches contenaient des QR codes et des URL redirigeant vers des canaux numériques pour obtenir des d'informations supplémentaires.

Artyushina (2020) explique qu'il était donc prévu que les passants soient avertis des enregistrements vidéo ou audio dans l'espace public, par l'intermédiaire de ces panneaux. Mais une fois informés, ils n'ont d'autre choix que de consentir à la collecte de données ou de quitter l'espace pour retirer leur consentement.

Pour Wylie (2018), ce système est inacceptable car il n'est fonctionnellement pas possible de s'exclure de l'espace public. Et il ne s'agit donc pas d'une option d'opt out raisonnable.

Carr et Hesse (2019) mettent en garde. Pour eux, la *smart city* est vendue comme une technologie neutre, comme une initiative supplémentaire pour étendre la *data-driven economy*. Mais ils estiment que le projet d'Alphabet à Toronto révèle que les infrastructures digitales sont des potentiels chevaux de Troie et que les pouvoirs locaux peuvent se révéler vulnérables face aux manœuvres des grandes entreprises.

Dans le cadre du projet de *smart city* mené par *Sidewalk Labs* à Toronto, force est de constater que le *Digital Trust* proposé présentait, lui aussi, des caractéristiques plus proches de celles d'un cheval de Troie, destiné à servir des intérêts cachés que de celles d'un dispositif destiné à servir la communauté de façon transparente.

Sans remettre en cause l'utilité générale des *Data Trusts*, y compris leur potentiel dans le cadre des *smart cities*, Artyushina (2020) nous invite à considérer le déroulé des événements, à Quayside, comme un avertissement.

5.9 Bottom-up Data Trusts – Delacroix et Lawrence (2019)

Après le dévoiement du concept de Data Trust par *Sidewalk* à Toronto, en 2018, et suite à la publication du « rapport AI » d'Hall et Pesenti (2017) qui s'éloignait substantiellement des réflexions précédentes, en abandonnant la forme légale du trust et en se concentrant sur de nouveaux objectifs (partage des données, développement de l'intelligence artificielle), plusieurs voix s'élevèrent pour réclamer un retour aux fondements originels du *Data Trust :* l'*« empowerment »* des *data subjects* et la prise en compte des asymétries de pouvoir.

Parmi ces voix réclamant un retour aux sources, celle de McDonald (2019) qui, dans un article intitulé « *reclaiming Data Trusts* », invite à se réapproprier le concept de *Data Trust*.

Celles également de Sylvie Delacroix, professeur de droit et d'éthique à l'université de Birmingham et de Neil Lawrence, professeur '*DeepMind'* de machine learning à l'université de Cambridge. Tous deux membres de l'Alan Turing Institute.

Dans un article qui fait aujourd'hui référence pour de nombreux académiques, intitulé « *Bottom-up data Trusts : disturbing the 'one size fits all' approach to data governance* », Delacroix et Lawrence (2019) estiment que la question de la gestion des données personnelles doit, fondamentalement, être appréhendée comme une problématique appartenant au domaine des droits de l'homme et nom au registre commercial.

Ils remarquent que la plupart de nos actes quotidiens, comme nos habitudes d'achat ou les amitiés que nous formons, sont désormais transformés en données, lisibles par des machines. Que la collecte systématique de nos données permet de disséquer nos vies comme jamais auparavant. Et que, si l'économie digitale contient, larvée, la promesse que nos ordinateurs seront d'ici peu capables de prédire nos comportements et de satisfaire nos besoins en les anticipant, cela suppose un degré de surveillance qui peut rapidement basculer de la bienveillance à la limitation maligne de nos libertés.

Ces traces digitales que nous laissons dans notre sillage peuvent être utilisées pour créer de la richesse et même des bénéfices sociétaux, mais elles peuvent également être exploitées pour manipuler nos opinions, comme l'a démontré le scandale Cambridge Analytica. Pour les *data subjects* que nous sommes tous, elles constituent donc une vulnérabilité.

Face à cette vulnérabilité, Delacroix et Lawrence (2019) estiment que les cadres réglementaires 'top-down' (comme le RGPD) sont absolument nécessaires mais qu'ils ne sont pas suffisants.

La quantité de temps et d'efforts requis pour faire valoir les droits conférés par de telles réglementations (GDPR en Europe, lois similaires dans d'autres régions du monde) est telle que les citoyens ordinaires ne les utilisent que dans les situations les plus importantes. Et bien que la situation soit meilleure en Europe que dans le reste du monde, l'exercice de ces droits continue à requérir un niveau considérable de connaissances et d'investissement personnel.

De plus, et c'est l'aspect le plus insidieux du problème, nous fournissons nos données de manière cumulative, progressive, par petites doses quotidiennes. Celles-ci s'accumulent jusqu'à former un agrégat exploitable qui dresse de nous un portrait digital complet. Mais ces petites fuites journalières ne sont pas perçues comme problématiques prises individuellement. Les cadres légaux actuels ne résolvent pas ce cumul progressif d'incidences mineures. Si bien que nous n'avons jamais tant été exposés et définis par notre passé qu'aujourd'hui.

Pour Delacroix et Lawrence (2019), des droits de propriété appliqués aux données (cf. supra p. 25) ne sont pas une solution susceptible de résoudre ce problème. Car ils ne modifient pas le fait qu'isolé, l'individu ne dispose d'aucun pouvoir dans ses relations avec les géants technologiques.

C’est pourquoi, en s’appuyant sur les travaux précédents de Balkin (cf. supra p. 55) ils proposent un modèle de *Data Trust* qui adopte strictement la forme juridique du *trust*, y compris les obligations légales inhérentes et remette la responsabilité fiduciaire au centre du dispositif.

Dans leur proposition, le *data subject* est à la fois le *settlor,* qui apportent les données et le *bénéficiaire* à qui le *trust* et le *trustee* doivent une indivisible loyauté.

Mais surtout, Delacroix et Lawrence (2019) imaginent un écosystème avec de multiples *Data Trusts* différents (certains plus favorables à la commercialisation des données qui verserait un dividende à ses affiliés, ou d’autres plus attachés à la protection de la vie privée, par exemple). Et la possibilité pour le *Data Subject* de s’affilier à celui qui lui corresponde le mieux.

Il existerait donc toute une série de *Data Trusts,* en concurrence les uns avec les autres, proposant des politiques de gestion variées. Et le citoyen serait libre de rejoindre le collectif qu’il estime le meilleur pour ses intérêts.

Delacroix et Lawrence (2019) envisagent que cet écosystème de *Data Trusts* mélange des dispositifs publics et privés représentant des valeurs différentes (avec un *Data Trust* public auquel seraient inscrits par défaut les citoyens n’ayant pas effectué de choix). D’une certaine façon le système qu’ils promeuvent n’est pas sans rappeler celui des mutualités de soin de santé belge (de grands collectifs incarnant des traditions et des valeurs différentes et proposant des services différant légèrement les uns des autres).

Ces *Data Trusts* pourraient, par exemple, proposer des options différentes face à la recherche médicale. Ce qui aiderait à surmonter les obstacles auxquels fait face, actuellement, la recherche scientifique (en constituant de grands jeux de données collectifs), tout en offrant aux citoyens le moyen de poser des choix qui représentent leurs convictions personnelles. Chaque *Data Trust* proposant une position différente.

Un autre exemple d’application est la négociation de termes et conditions. Pour un individu isolé, effectuer des choix spécifiques concernant la manière dont les médias sociaux gèrent ses données personnelles requiert la lecture de conditions générales longues et obscures, et une implication personnelle extrêmement chronophage. Par le biais de la mutualisation, ces *Data Trusts* seraient à même de rééquilibrer quelque peu l’asymétrie de pouvoir et de négocier collectivement de meilleures conditions.

Pour Delacroix et Lawrence (2019), il existe également des applications dans le secteur financier (conditions financières, crédit, etc.) ou du commerce (concernant les données collectées par les supermarchés par le biais des cartes de fidélité, par exemple).

Mais surtout, et c’est le cœur de leur argument, construire ces *Data Trusts*, à partir des *data subjects* et pour les *data subjects*, en utilisant pleinement les mécanismes du *trust* et de l’obligation fiduciaire, permet de remettre dans les mains des citoyens les leviers pour exercer un pouvoir décisionnel les concernant.

Pour Hardinges (2020), cette proposition de Delacroix et Lawrence (2019) permet aux citoyens de mettre en commun les droits qu'ils ont sur leurs données personnelles dans le cadre juridique du *trust,* et ce, d'une façon qui reflète leurs préférences et leurs besoins. Mais surtout, contrairement à la majorité des droits qui se concentrent sur la possibilité pour la personne de décider par elle-même, l'originalité de cette proposition réside dans le fait de reconnaître que la plupart des approches individualistes concernant la gestion des données sont lacunaires. En traitant les données comme un actif communautaire, le *data subject* voit son pouvoir personnel multiplié.

La gestion collective des données dans le cadre d'un modèle « bottom-up » est complémentaire au cadre réglementaire « top-bottom » des lois sur la protection de la vie privée.

Ayant étudié de manière extensive la compatibilité de leur proposition avec le GDPR, Delacroix et Lawrence (2019) concluent qu'un aménagement réglementaire serait peut-être nécessaire pour que leur proposition soit applicable sur le territoire européen. En effet, ils estiment qu'il existe, pour le moment, un doute quant au fait de savoir s'il est possible de mandater un tiers (un trustee) pour exercer les droits de portabilité, d'accès et d'effacement conférés par le GDPR à la place du *data subjetc*[3]. Ce qui est nécessaire pour implémenter leur proposition.

Sur ce point, Paul Nemitz, un haut conseiller de la Commission Européenne, se montre rassurant lors de conversations informelles avec Sylvie Delacroix et Ben McFarlane, professeur de droit à Oxford. Il explique que, dans le cadre du *Data Governance Act*, la commission étudie actuellement des mécanismes de certification qui puissent garantir que les intermédiaires traitant les données respectent certaines normes qualitatives. Et que le point clé, concernant les *Data Trusts* sera plutôt que le processus qui permette aux personnes d'effectuer des choix reste simple et facilement compréhensible par les personnes que le *Data Trust* doit servir (Delacroix, McFarlane, & Nemitz, 2021).

[3] Il existe un débat entre professeurs universitaires de droit pour savoir quelle interprétation exacte il faut donner à l'article 80(1) du GDPR. En particulier, la question est de savoir si une loi nationale permettant de donner un mandat, à un trustee, pour exercer les droits de portabilité, d'accès ou d'effacement conférés par le GDPR serait légale ou pas.

6. Les différentes catégories de Data Trusts

6.1 Quelques exemples supplémentaires

Comme nous venons de le voir, la réflexion intellectuelle autour des *Data Trusts* s'est construite par étapes successives : avec des avancées, des reculs, des expérimentations et des divergences d'opinions.

Malgré ce trajet sinueux, on peut distinguer un sillon clair, une progression régulière de la réflexion et un intérêt continu pour le dispositif.

Outre les étapes clés que nous avons sélectionnées pour illustrer l'historique du chapitre précédent, plusieurs projets concrets ont été menés et de nombreuses publications supplémentaires ont paru durant les dernières années.

Hardingues & Wells (2019) relatent ainsi qu'en 2018, OpenCorporates, la plus grande base de données d'entreprises au monde, annonça son passage à une structure de gouvernance basée sur le trust.

De son côté, l'Open Data Institute (2021) mena trois projets pilotes dans différents contextes :

- À Londres, un data trust fut créé (en partenariat avec l'organe de gouvernance régional de la région de Londres et 'Le borough royal de Greenwich') pour mener à bien deux projets. Le premier visait à collecter et analyser les données de chauffage dans les logements résidentiels. Le second a analysé des informations relatives à la disponibilité des places de parking disponibles pour les véhicules électriques.

- Des producteurs et des distributeurs de boisson et de nourriture ont exploré les possibilités offertes par les data trusts pour réduire le gaspillage alimentaire et analyser leurs données de vente, mises en commun.

- Le WILDLABS Tech Hub a utilisé un Data Trust pour traiter les images, les sons et les données acquises par les fonctionnaires douaniers dans le cadre de la lutte contre le commerce international illégal d'espèces sauvages, aux frontières.

On le voir, les cas d'application sont variés.

Parmi les publications notables, un rapport conjoint d'Element AI et de Nesta (2021), deux firmes actives dans le domaine de l'intelligence artificielle, mérite attention. Il continue sur la lignée ouverte par Hall et Pesenti (2017) et explore les applications du *Data Trust* dans le domaine de l'IA.

Le rapport de l'Ada Lovelace Institute (2021) offre la synthèse la plus récente et la plus à jour.

Enfin, d'autres projets concrets ont encore été mis en place au cours des dernières années. Le lecteur curieux trouvera en annexe (*annexe 1* et *annexe 2*) du présent travail, un listing d'une quinzaine de *Data Trusts* et *Data Coopératives* existants compilés par Gomer et Simperl (2020).

6.2 La classification d'O'Hara

Sean McDonald (2019) constate que l'intérêt en provenance des gouvernements et des grandes entreprises s'est considérablement accru autour du concept de *Data Trust* à partir de la parution du rapport de Hall et Pesenti en 2017 (cf. supra p. 60). Et que depuis, on assiste à la multiplication des investissements financiers publics et privés, accompagnés d'une hausse des manifestations de bonne volonté de la part du monde politique.

Le secteur privé en a pris note, lui aussi, ce qui a entrainé l'émergence rapide d'un marché de la consultance centré sur le *trust*. Pour McDonald (2019), il en résulte que le discours public sur la question du *Data Trust* est actuellement dominé par des groupes bien financés dont l'objectif est de maximiser le partage des données.

Mais, pour lui, le partage des données à grande échelle sous l'impulsion du secteur privé n'est pas la seule façon d'utiliser les *Data Trusts,* ni même la plus importante. Il estime qu'il existe d'autres usages du *Data Trust,* notamment ceux au service de l'intérêt public et trouve important de tenir compte de la variété d'usages différents lorsque l'on modèle l'environnement politique de la gouvernance fiduciaire des données.

McDonald (2019) n'est pas le seul à remarquer cette multitude d'acteurs aux objectifs différents. O'Hara (2020) relève lui aussi cette coexistence de différents groupes autour de la question du *trust* digital.

Pour y voir un peu plus clair, il propose de les classer en trois grandes catégories avec chacune des intérêts propres et des visions différentes du *Data Trust* :

- ceux qui cherchent à augmenter la quantité de données *collectées*,
- ceux qui cherchent à augmenter la quantité de données *partagées*
- et enfin ceux qui cherchent à *diminuer* la quantité de données partagées.

6.3 Les Data Trusts pour les agrégateurs de données

Pour O'Hara (2020), cette première catégorie des agrégateurs de données contient les géants très connus du secteur technologique (Alphabet, Amazon, Facebook) et certains acteurs de taille légèrement inférieure (Netflix, Palantir, …). Ces entreprises disposent d'énormes moyens financiers et ont la ferme intention de maintenir leur richesse. Leur but est d'augmenter la quantité de données collectées.

Comme nous l'avons vu, ces entreprises dominent le marché et tendent à constituer un oligopole (cf. supra p. 38). Le risque auquel ces grandes entreprises doivent faire face est le fait que la thésaurisation de données à grande échelle pourrait déclencher des enquêtes antitrust (avec la possibilité que leurs activités soient divisées d'autorité par les régulateurs). Il est donc important pour eux que leurs pratiques continuent à être considérées comme étant socialement acceptables.

Pour ces entreprises, le *Data Trust* est avant tout une appellation, une « marque », avec un certain potentiel pour mettre en confiance une partie sceptique du public. Pour ces grands agrégateurs, l'usage principal du *Data Trust* est de rassurer. Nous sommes dans le registre de la communication et des relations publiques.

La proposition de *Sidewalk Labs*, filiale d'Alphabet, d'avoir recours à un « *Urban Data Trust* » pour gérer les données sensibles générées par le projet de smart city à Toronto, en est un exemple manifeste.

L'*Urban Data Trust* tel que proposé par *Sidewalk Labs* n'est pas un trust, au sens légal. Il ne contient pas d'obligation fiduciaire, trop contraignante pour une simple mission de communication. Il n'adopte pas non plus de structure alternative (organisme public, *data commons,* data coopérative, …) à même de favoriser des pratiques de gouvernance démocratique.

Ce « *trust* » ne forme pas un tout cohérent. Il s'agit plutôt d'un amalgame de mesures *ad hoc* assemblées chaotiquement pour faire face aux inquiétudes exprimées par les habitants sans apporter de modification substantielle à la gestion des données.

Bien que *Sidewalk Labs* (2018) ait cherché à promouvoir, avec enthousiasme, son projet en le décrivant comme une initiative de data gouvernance innovatrice et sans précédent, l'objectif unique semble avoir été de conserver un monopole sur la collecte et l'exploitation des données personnelles des citoyens (cf. supra p. 71).

Les auteurs qui se sont exprimés sur le projet de *Sidewalk Labs* (Artyushina, 2020 ; Carr et Hesse, 2020, McDonald, 2019 ; Delacroix et Lawrence, 2019 ; Goodman et Powles, 2019) se montrent extrêmement critiques. Et les opinions sont, à l'exception de celles émanant directement des travailleurs de *Sidewalk Labs,* unanimement négatives.

L'*Urban Data Trust* déployé à Toronto avait pour mission de rassurer le public concernant la collecte de données. Mais, dans le cadre de ce projet de smart city, force est de constater que ce n'est pas seulement la gestion des données qui faisait l'objet d'une profonde méfiance de la part des citoyens. C'est l'ensemble du projet et l'entreprise, elle-même.

Pour qu'il existe un climat de confiance et que les citoyens acceptent de participer à un projet de partage de données, ajuster superficiellement le dispositif ne suffit pas. Le *Data Collector* doit bénéficier de ce que O'Hara (2020) et Milne et Al. (2021) appellent la « licence sociale ». C'est-à-dire la réelle confiance de la population qui se manifeste par une permission tacite de procéder à la collecte et au traitement des données (cf. infra p. 86).

Un autre projet susceptible d'appartenir à cette première catégorie de *Data Trusts* créés à des fins de communication par les grands agrégateurs de data est l'*oversight board* (https://oversightboard.com/) mis en place par Facebook pour superviser les décisions de modération de sa plateforme. Ce comité de surveillance est censé endosser le rôle de « Cour Suprême » en matière de modération de contenu.

On peut remarquer que sa création fut annoncée en novembre 2018, soit quelques mois à peine après les premières révélations du scandale de *Cambridge Analytica* (début 2018). On peut donc légitimement se demander si la motivation première n'est pas, comme dans le cas de *Sidewalk Labs,* avant tout de rassurer le public.

La création de ce *trust*, qui prend la forme d'un comité décisionnel, survient d'ailleurs au même moment (fin 2018) que la création du *Urban Data Trust* à Toronto. Dans un climat où le concept de *Data Trust* est en vogue.

Il existe toutefois une différence de taille avec le projet de *Sidewalk Labs* : l'*oversight board* de Facebook adopte bien, lui, la forme légale d'un trust de droit du Delaware, avec l'ensemble des obligations légales en découlant. L'avenir dira si ce projet est un succès ou un échec.

6.4 Les Data Trusts fonctionnels

Pour O'Hara (2020), le second groupe d'acteurs qui s'intéresse aux *Data Trusts* est composé entreprises et d'institutions qui souhaiteraient mettre en commun et agréger des données mais qui n'y parviennent pas, pour diverses raisons.

Le rapport d'Hall et Pesenti (2017) sur la croissance de l'industrie de l'intelligence artificielle au Royaume-Uni est la publication à l'origine de l'intérêt de cette seconde catégorie d'acteurs pour les *Data Trusts.*

Comme décrit dans le rapport d'Hall et Pesenti (2017), l'industrie de l'intelligence artificielle et du *machine learning* requiert d'énormes quantités de données. Or, de nombreuses entreprises, même de grande taille, ne collectent pas des quantités de données comparables à celles dont disposent les GAFA. Elles ont donc intérêt à se grouper si elles souhaitent concurrencer l'oligopole formé par les géants technologiques (cf. supra p.38) et récolter une part des gigantesques bénéfices promis par la *data-driven economy* dans de nombreux secteurs (cf. supra p.12).

Malgré cet intérêt à mettre en commun leurs données, il existe plusieurs obstacles pour les entreprises et les institutions qui souhaitent partager leurs data (et ce, même lorsqu'il s'agit de données ne revêtant pas un caractère personnel).

O'Hara (2020) liste quelques-unes de ces difficultés : les éventuels soucis relatifs à la qualité des données reçues, l'inquiétude qu'un tiers profite de sa propriété intellectuelle, les potentiels dommages à la réputation de l'entreprise découlant d'usages inappropriés des données par un tiers, le fait que des partenaires potentiels ne soient tout simplement pas au courant de l'existence des données ou des entreprises avec qui il pourrait être bénéfique de collaborer, le caractère délicat

de certaines données qui peut pousser les managers à ne pas prendre de risques ou encore d'éventuelles complications liées aux différences de régimes légaux entre pays.

L'utilisation du *trust* ou d'un mécanisme similaire vise ici à surmonter ces difficultés entre fournisseurs de données et receveurs. Pour les membres de ce groupe, l'objectif est d'augmenter la quantité de données partagées, dans un cadre de confiance, et ce soit dans un but de lucre ou pour favoriser la recherche scientifique.

Il est donc question de mettre en lien ceux qui détiennent ou qui génèrent les données et ceux qui en consomment de grandes quantités, par exemple les start-up actives dans le secteur analytique et prévisionnel.

Pour O'Hara (2020), il est extrêmement rare que ce type de *Data Trust* adopte la forme juridique traditionnelle d'un *trust* et l'obligation fiduciaire ne fait donc, généralement, pas partie d'un tel dispositif.

Il explique que le mécanisme légal du trust avec son obligation fiduciaire est globalement trop contraignant pour ces acteurs. Et que le terme de *Data Trust* doit, ici, être compris « métaphoriquement ». Le *trust* étant simplement une source d'inspiration pour ce modèle. L'obligation fiduciaire et le rôle du *trustee* y sont symboliques.

Pour lui, il est d'ailleurs plus approprié de définir les *Data Trusts* de cette catégorie de manière fonctionnelle, plutôt que juridique. D'où l' appellation de « *Data Trust fonctionnels »* qu'il leur attribue*.*

En avant-propos du rapport « *Exploring legal mechanisms for data stewardship* », publié par l'Ada Lovelace Institute, Wendy Hall (2021), éminente professeur de Sciences informatiques de l'Université de Southampton et co-auteur du rapport de 2017 « *Growing the artificial intelligence industry in the UK* » confirme cette explication d'O'Hara.

Elle revient sur la rédaction du « rapport IA » de 2017, avec Jérôme Pesenti et explique que le terme *Data Trust* y est utilisé largement et librement. Le type de *Data Trust* qu'ils décrivent, et dont le but est de faciliter le développement de l'intelligence artificielle, n'a pas nécessairement vocation à se manifester sous la forme juridique d'un *trust* légal.

Au contraire, pour elle, les acteurs de cette catégorie ont tout intérêt à s'inspirer de manière très large de divers mécanismes de data gouvernance afin de créer des modèles qui correspondent à leurs besoins propres : *data trusts, data commons, data coopératives,* mécanismes contractuels, …

Si ce type de *Data Trust* endosse la forme d'un dispositif légal, il s'agira la plupart du temps d'un contrat. Le dispositif peut également prendre la forme d'un autre type d'accord (comme un code de conduite ou une charte) voire d'une entreprise fondée spécifiquement pour jouer ce rôle ou même, simplement, d'une architecture informatique partagée.

Approach	Distinguishing feature
Data trusts	Takes what has been learned from the use of legal trusts. Trustees of a data trust will take on responsibility (with some liabilities) to steward data for an agreed purpose.
Data cooperatives	Takes what has been learned from cooperatives. A mutual organisation owned and democratically controlled by members, who delegate control over data about them.
Data commons	Takes what has been learned from managing common pool resources – such as forests and fisheries – and applies the principles to data.
Personal data stores[12]	Stores data provided by a single individual on their behalf and provides access to that data to third-parties when directed to by the individual.
Research partnerships[13]	When data holders provide access to data to universities and other research organisations.

Figure 9 : Data Trusts compared to other approaches.

Source : Hardinges, J., Wells, P., Blandford, A., Tennison, J. & Scott, A., (2019, avril). *Data trusts : Lessons from three pilots.* Open Data Institute. https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit

Hardinges (2020), de l'Open Data Institute, considère que la société HiLo (https://hilomrm.com/) est un bon exemple de ce type de *Data Trusts fonctionnels*.

Cette start-up collecte des jeux de données extrêmement complets en provenance de multiples organisations maritimes. Son objectif est la réduction des accidents de mer et l'amélioration de la sécurité navale, grâce à l'analyse et à l'exploitation des données mises en commun.

Avant d'être une société, HiLo était un projet de trois grandes entreprises du secteur du transport maritime, de l'énergie et de la gestion du risque en mer, qui décidèrent de mettre en commun et de partager leurs données (Shell shipping & Maritime, Maersk Tankers et Lloyd's Register Consulting).

Lors de la conception de leur prototype analytique et de leur modèle prédictif, HiLo put bénéficier de l'expertise d'académiques de l'Imperial College London. L'Alan Turing Institute vérifia également leur modèle statistique (HiLo, 2021).

Face au succès du projet, une société indépendante fut ensuite fondée et commença à accepter des clients externes. Aujourd'hui, cette start-up collecte et gère les données maritimes de 55 organisations. Lorsqu'elle est rejointe par un nouveau client, celui-ci lui remet ses données internes existantes, qui sont ensuite anonymisées. Des experts passent en revue les données et les intègrent dans le modèle prédictif (HiLo, 2021).

Pour HiLo, le problème qu'il s'agit ici de régler est la mise en commun de vastes quantités de données pour nourrir le modèle prévisionnel.

Fonctionnellement, on est relativement proche du modèle d'une *data coopérative*

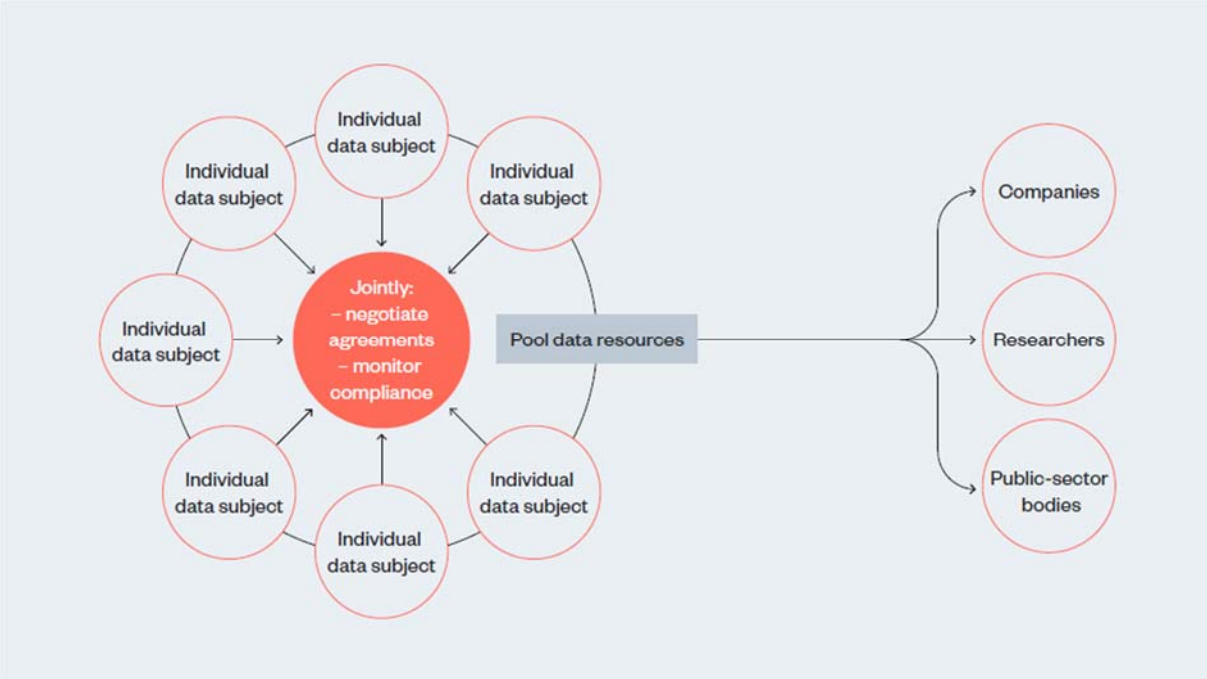

Figure 10 : Exemple de Data Coopérative d'usagers.

Source : Ada Lovelace Institute. (2021, mars 04). *Exploring legal mechanisms for data stewardship.* Consulté 14 août 2021, à l'adresse https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

Dans une *data coopérative*, des membres (traditionnellement des *data subjects* mais ici, des entreprises) mettent en commun leurs ressources et les gèrent collectivement ou en déléguant la gestion à un représentant.

Comme nous l'avons mentionné, le fait de s'inspirer librement de différents modèles de data gouvernance est un élément constitutif de ces *Data Trusts fonctionnels.* Pour O'Hara (2019) ce type de *Data Trusts* peut être modelé en fonction des besoins spécifiques des *data scientists* et des différentes parties prenantes.

Les fonctions de ce modèle sont tant le *data processing,* en lui-même, que le fait de servir d'interface transparente, et conçue sur mesure, entre les différents membres.

Les citoyens peuvent être associés au dispositif en tant que *data subjects* mais ils ne sont pas toujours présents car le *Data Trust fonctionnel* est, dans bien des cas, mis en place par un groupe d'institutions ou d'entreprises, sans interaction directe avec les citoyens (comme c'est le cas, par exemple, d'HiLo).

S'agissant d'un dispositif créé sur mesure. L'addition d'organes de gouvernance ou de contrôle n'est pas rare. La littérature décrivant ce modèle inclut parfois des comités de supervision, par exemple.

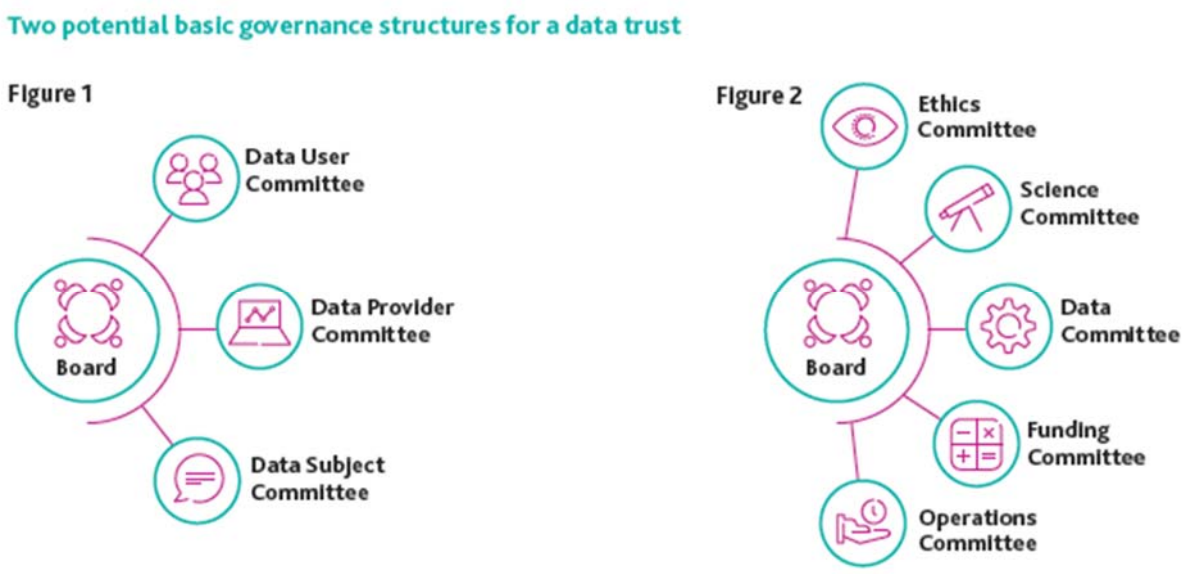

Figure 11 : Exemple de comités de supervision dans un *Data Trust fonctionnel*.

Source : BPE Sollicitors, Pinsent Masons, & Queen Mary University of London. (2019, avril). *Data trusts : legal and governance considerations*. https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf

6.5 Bottom-up Data Trusts et Civic Data Trusts

Lorsque nous avons étudié les liens entre l'information et le droit de la propriété, dans la première partie de ce travail (cf. supra p. 27), nous avions noté que Mills (2019) et Scassa (2018) identifiaient deux groupes qui, tous deux, réclamaient la création de droits de propriété s'appliquant aux données mais pour des raisons opposées :

- D'un côté, les entreprises réclamaient la création de droits de propriété dans le but de commercialiser les données ;

- Tandis que de l'autre, les citoyens cherchaient, eux, à obtenir des droits de propriété sur leurs données mais pour mettre un terme à l'utilisation répétée ou abusive de leurs données personnelles.

Il est intéressant de constater à quel point la classification des *Data Trusts* proposée par O'Hara (2020) reflète des lignes de fracture similaires.

Les deux premières catégories de *Data Trusts*, que nous venons de décrire concernent des entreprises dont l'objectif fondamental est la commercialisation des données.

À l'opposé, les membres de cette troisième catégorie sont des *data subjects* dont l'objectif est de réduire la quantité de données partagées. Ou, plus exactement, de

s'assurer que le partage de leurs données personnelles se fasse dans le respect de règles tacites et formelles avec lesquelles ils sont en accord.

Les *Data Trusts* promus par les membres de cette troisième catégorie sont les *Bottom-up Data Trusts,* de Lawrence et Delacroix (2019). Au sein desquels, les *data subjects* mutualisent leurs données afin de, collectivement, constituer un contre-pouvoir aux entreprises du secteur digital face auxquelles, ils sont autrement isolés. Ainsi que les *Data Trusts civiques* issus du modèle de McDonald (2015) (cf. supra p. 46).

Ces types de *Data Trust,* quels que soient leurs objets, adoptent toujours la structure juridique d'un *trust* classique et ils respectent l'ensemble des obligations légales liées, car l'obligation fiduciaire est ici centrale au dispositif. Dans ce modèle le *trustee* a donc une obligation de loyauté indivisible envers les bénéficiaires (les data subjects ou la cause d'intérêt général), qui disposent d'un droit de recours devant les tribunaux, le cas échéant.

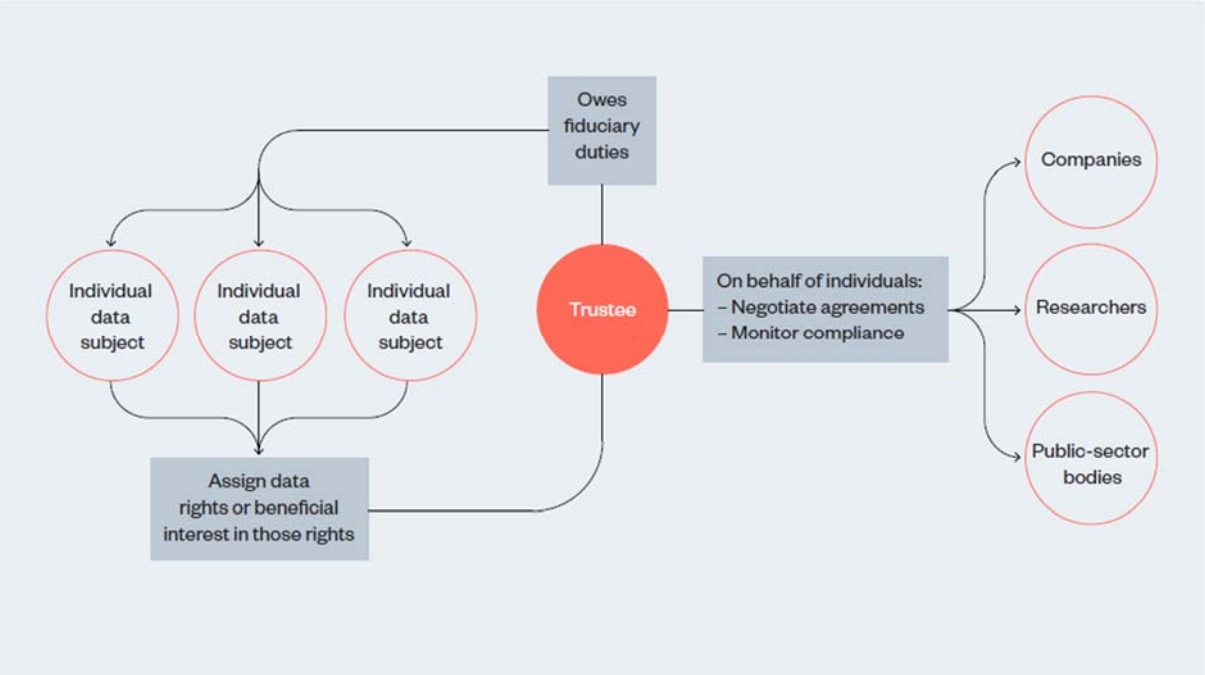

Figure 12 : Exemple de Data Trust adoptant la forme légale d'un trust.

Source : Ada Lovelace Institute. (2021, mars 04). *Exploring legal mechanisms for data stewardship.* Consulté 14 août 2021, à l'adresse https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

Dans le modèle de Delacroix et Lawrence (2019), les data subjects sont simultanément les *settlors* (ils font apport des droits qu'ils possèdent sur leurs données personnelles) et les bénéficiaires du *trust*.

Tandis que dans le cas du *Data Trust civique* préconisé par McDonald (2015, 2019), le bénéficiaire peut être une cause d'intérêt général, éducative ou sociale selon le modèle du *trust charitable* (cf. supra p. 46).

Ces deux modèles de *trust* ont en commun le souci de contrebalancer les asymétries de pouvoir et de proposer des modes de gouvernance démocratique et participative. Contrairement au *Data Trust fonctionnel,* ici l'intérêt des *data subjects* occupe une place cruciale.

Cette catégorie de *Data Trust* est également celle qui a les racines intellectuelles les plus anciennes comme nous l'avons vu en retraçant le long historique des publications et des réflexions sur le sujet (cf. supra p. 49).

Il nous semble également approprié de classer ici certains projets de *data trusts* dont l'objet est en lien avec la recherche biomédicale.

Ceux-ci partagent, en effet, les mêmes caractéristiques fondamentales que les *Data Trusts* de cette catégorie : l'adoption de la forme légale du *trust,* en ce compris l'obligation fiduciaire au sens strict, la promotion de modes de gouvernance participatifs, le respect du consentement éclairé des *data subjects* et un objet social en lien avec l'intérêt général.

Les plus récentes publications issues du secteur de la recherche médicale concernant les *data trusts* (Milne, Sorbie et Dixon-Woods, 2020 ; Paprica et Al, 2020) citent d'ailleurs ouvertement le modèle *bottom-up* de Delacroix et Lawrence (2019).

Pour Milne, Sorbie et Dixon-Woods (2020), la mise en place d'une gouvernance participative dans les biobanks de grande taille est un élément essentiel car l'enjeu du secteur n'est pas seulement de respecter le seul cadre réglementaire et légal mais bien d'aller au-delà et de s'assurer d'obtenir la « licence sociale » des donneurs. C'est-à-dire leur confiance et leur autorisation tacite, au sens large.

Dans ce contexte, divers *trusts et data trusts* ont été mis en place dans les biobanks, avec une emphase sur la transparence et l'engagement constant avec les parties prenantes (y compris les donneurs et les personnes affectées par les recherches effectuées).

Les donneurs participent de diverses manières : dans certains cas, ils font partie d'un comité consultatif, dans d'autres ils sont directement représentés au sein des organes décisionnels.

Milne, Sorbie et Dixon-Woods (2020) relatent que dans certains cas, les parties prenantes ont été consultées en amont, avant la mise en place du *data trust.* Ce fût le cas, par le biais de consultations populaires, avant la mise en place de l'UK Biobank, dont nous avons parlé (cf. supra p. 53) : un comité de 25 patients et donneurs fut constitué, et avec l'appui d'experts, il leur a été demandé de se prononcer sur les valeurs qui devraient être celles d'une biobank. Ils parvinrent à un fort consensus.

Dans la plupart des cas, les comités d'éthique, de gouvernance et d'accès aux données restent néanmoins l'apanage d'experts même quand une participation des patients ou du public est incluse. D'après Milne, Sorbie et Dixon-Woods (2020) il existe cependant des exceptions, comme celle du METADAC, un comité éthico-légal ayant la responsabilité de l'accès aux données dans le cadre d'études longitudinales. La plupart des membres de ce comité sont des experts en droit, en

épidémiologie, en bioéthique ou en sociologie mais la présence d'un ou deux représentants des participants aux études scientifiques est obligatoire.

Un autre exemple de gouvernance participative dans le secteur médical est celui du *Michigan Biotrust* (MBT). Ce biotrust ne collecte pas des data mais bien des échantillons sanguins. Il nous semble intéressant de le citer, malgré tout, car il est l'un des rares qui endosse la forme légale d'un *trust charitable* conforme au modèle décrit par les frères Winickoff, en 2003 (cf. supra p. 49).

Milne, Sorbie et Dixon-Woods (2020) notent que les gestionnaires de ce *Biotrust* ont une responsabilité fiduciaire et l'obligation légale d'utiliser les échantillons sanguins pour le bénéfice de la santé publique et de la population. Leur travail est supervisé par trois comités : un scientifique, un traitant des questions éthiques et un troisième dont la mission est de s'assurer que le *biotrust* respecte les valeurs de la communauté. Des citoyens siègent au sein de ce dernier comité.

On le voit, la participation démocratique dans les *bottom-up trust* et les *trusts civiques* peut théoriquement prendre plusieurs formes : participation en amont, en aval, dans des comités aux côtés d'experts, directement au sein des organes décisionnels, etc.

L'ensemble de ces modes de gouvernance peut s'appliquer aussi bien aux *trusts* classiques qu'aux *data trusts*.

Dans certains cas, cette gouvernance participative s'inspire directement des travaux d'Ostrom (Roman, 2021) relatifs à la gestion des biens communs dont nous avons parlé en première partie (cf. supra p. 19).

Un exemple concret est décrit dans les travaux de Paprica et Al. (2020).

Paprica et Al. (2020) rapportent qu'au Canada, 19 professionnels travaillant pour 15 organisations différentes du secteur des soins des santé (hôpitaux, universités, initiatives de gestion de données du secteur) ont fait le constat qu'il existait un besoin de guidance pratique pour l'installation et la gestion d'infrastructures permettant le partage et l'accès aux données médicales tout en assurant leur protection.

En s'appuyant simultanément sur leurs expériences de 1ère ligne dans la gestion d'infrastructures data et sur une synthèse de la littérature scientifique connexe, ils ont produit une série de recommandations pour la gestion des données informatiques à caractère médical par le biais d'un *Data Trust* participatif.

Box 1: Min specs for data trust establishment and operations

1. Legal: The data trust must fulfill all legal requirements, including the authority to collect, share and hold data
2. Governance
   a) The data trust must have a stated purpose
   b) The data trust must be transparent in its activities
   c) The data trust must have an accountable governing body
   d) Governance must be adaptive
3. Management
   a) There must be well-defined policies and processes for the collection, storage, use and disclosure of data
   b) Policies and processes must include data protection safeguards which are reviewed and updated regularly
   c) There must be an ongoing process to identify, assess and manage risks
4. Data user requirements
   a) All data users must complete training before they access data
   b) All data users must agree to a data user agreement that acknowledges that data use will be monitored and includes consequences for non-compliance
5. Public and stakeholder engagement
   a) There must be early and ongoing engagement with stakeholders including members of the public
   b) Where there is a reasonable expectation that specific subpopulations or groups would have a particular interest in, or would be affected by, an activity of the data trust, there must be direct engagement tailored for that subpopulation/group

Figure 13 : Spécifications minimales pour l'établissement et l'opération d'un Data Trust

Source : Paprica et Al. (2020). Essential Requirements for Establishing and Operating Data Trusts. *International Journal of Population Data Science*, *5*. https://doi.org/10.23889/ijpds.v5i1.1353

Cette synthèse est tout à fait représentative des règles de gestion participatives et transparentes que l'on rencontre au sein des *Data Trusts* de la troisième catégorie. Elle est aussi fortement inspirée des méthodes de gestion des biens communs, décrites par Ostrom (Roman, 2021).

Pour ces *Data Trust,* il est important de disposer de la « licence sociale » du public et pour ce faire la mise en place d'un cadre de confiance est nécessaire. La défense des intérêts du public est la raison d'être de ces dispositifs. Les règles de gouvernance, les modalités de participation et la structure légale reflètent cet état de fait.

6.6 Classification de Mills

En complément de la classification d'O'Hara, nous proposons de nous arrêter rapidement sur une autre classification des *Data Trusts*. Celle de Mills, un économiste de l'université de Manchester.

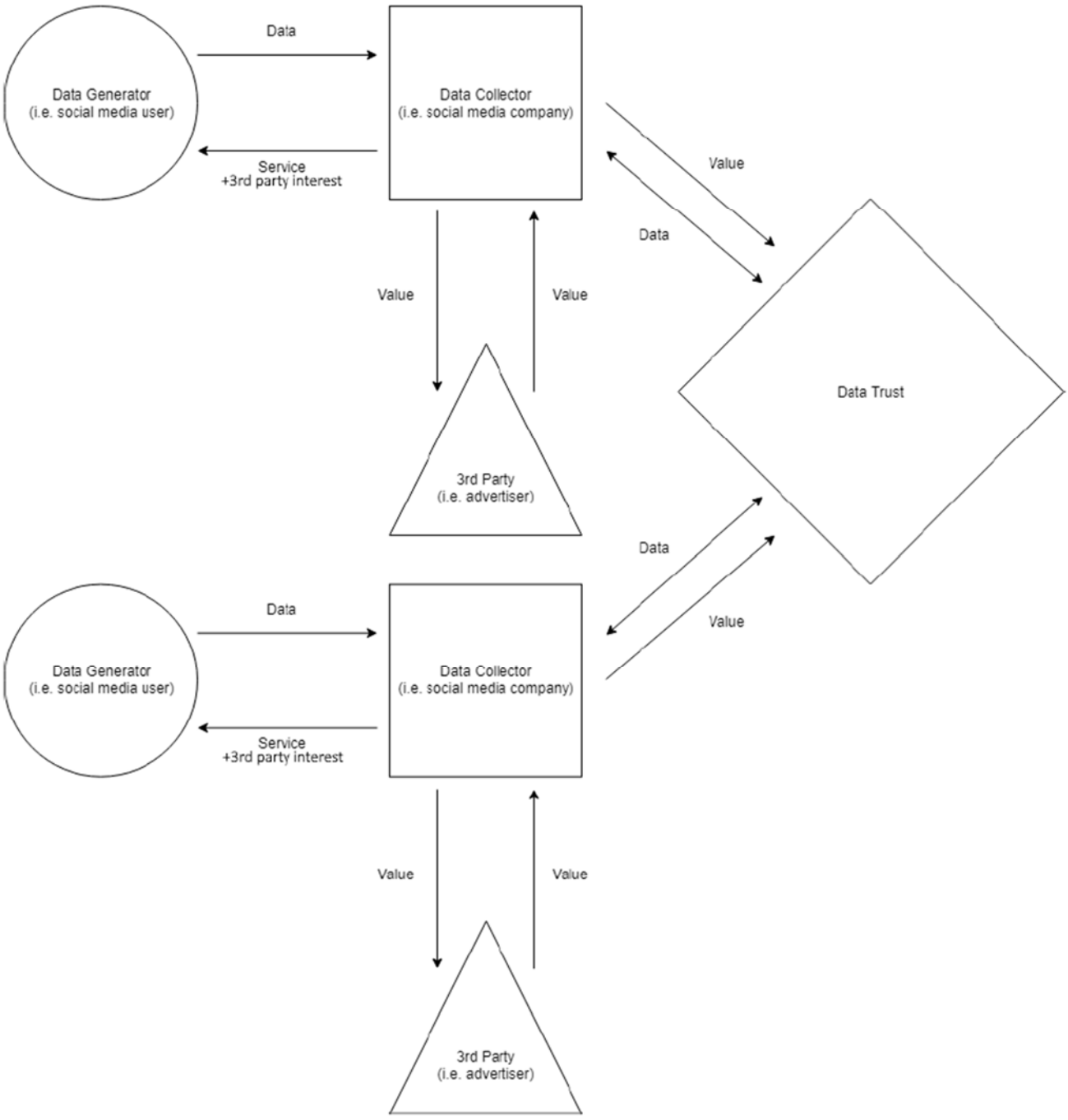

Figure 14 : Data flows in a Collector Centric Data Trust.

Source : Mills, S. (2019). Who Owns the Future ? Data Trusts, Data Commons, and the Future of Data Ownership. *SSRN Electronic Journal*, 1. https://doi.org/10.2139/ssrn.3437936

En empruntant au registre de l'économie politique, Mills (2019) se concentre sur la structure des Data Trusts pour les catégoriser. Car il estime que la structure sous-jacente d'un *Data Trust* influence et révèle les rapports de force entre parties.

Cette classification structurelle est intéressante à utiliser, en complément de celle de O'Hara. Pour évaluer si la structure proposée pour le *Data Trust* est cohérente avec l'objectif promu officiellement. Ou s'il existe des incohérences.

Pour Mills (2019), la première catégorie est composée des *Data Trusts* centrés sur les *data collectors* (cf. figure 14, page précédente).

Dans cette catégorie, ce sont les *Data Collectors* qui établissent et contrôlent le *Data Trust.* Leur objectif est de partager et de mettre en commun leurs données, dans un cadre de confiance. Comme on peut le constater sur le schéma, dans ce modèle les *data subjects*, sont relégués en périphérie et ils n'ont pas de réellement de prise sur le *Data Trust*.

Cette catégorie correspond, *grosso modo,* aux datas trusts de la seconde catégorie de O'Hara (celle des *data trusts fonctionnels*) utilisés pour le développement de l'IA et des modèles prédictifs (cf. supra p. 80).

La seconde catégorie de *Data Trusts* identifiés par Mills (2019) présente des structures centrées sur la donnée. Cet intéressant modèle n'a pas réellement d'équivalent dans la classification d'O'Hara.

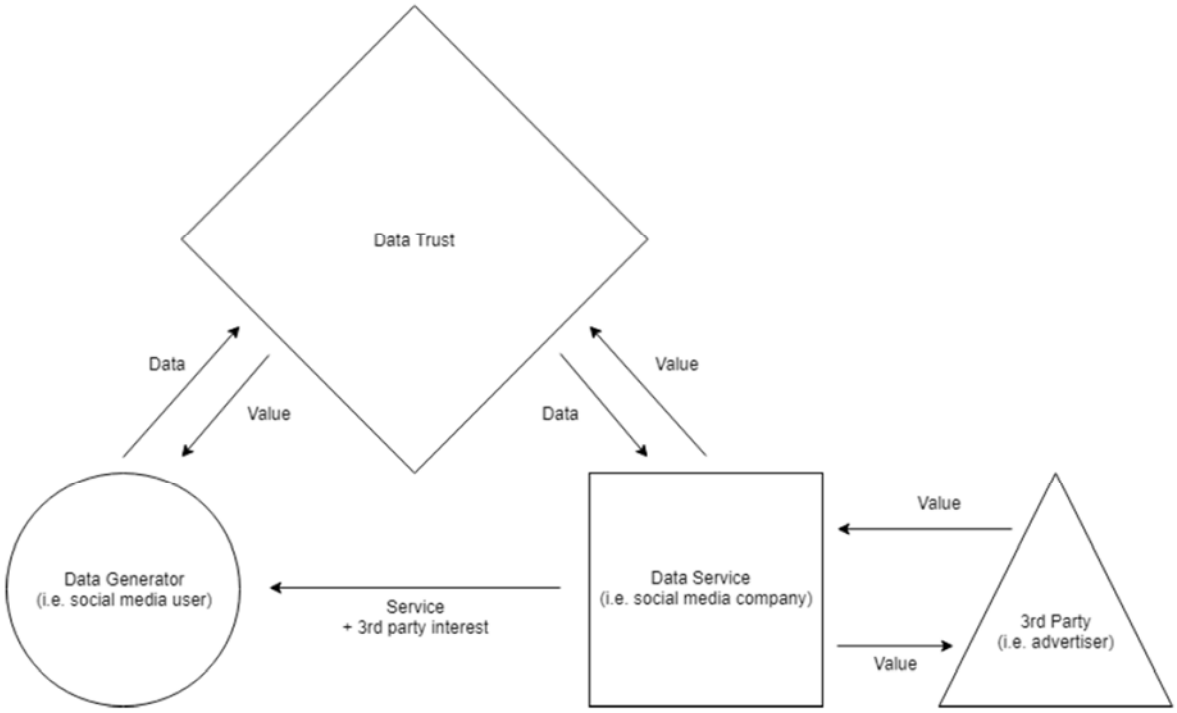

Figure 15 : Data flows in a Data Centric Data Trust.

Source : Mills, S. (2019). Who Owns the Future ? Data Trusts, Data Commons, and the Future of Data Ownership. *SSRN Electronic Journal*, 1. https://doi.org/10.2139/ssrn.3437936

Il s'agit d'un modèle équilibré où le *Data Trust* est équidistant du *data generator* (le *data subject*) et du *data collector*. Le processus décisionnel est quasi équilibré. Dans un tel modèle, la définition des rôles et de l'objet du *trust* sera primordiale pour déterminer les rapports de force entre parties.

Ce modèle offre les meilleures opportunités de relations équilibrées entre participants. Mais il n'est pas évident de savoir ce qui pousserait un *Data Collector* (par exemple, une entreprise animant un réseau social) à collaborer aussi égalitairement avec des *data subjects*, à moins d'y être contraint.

Dans un tel modèle, le rôle des pouvoirs publics pourrait être de rendre une telle collaboration obligatoire par voie réglementaire.

La troisième et dernière catégorie de *Data Trusts* identifiée par Mills (2019) présente des structures centrées sur les générateurs de données (*data subjects*).

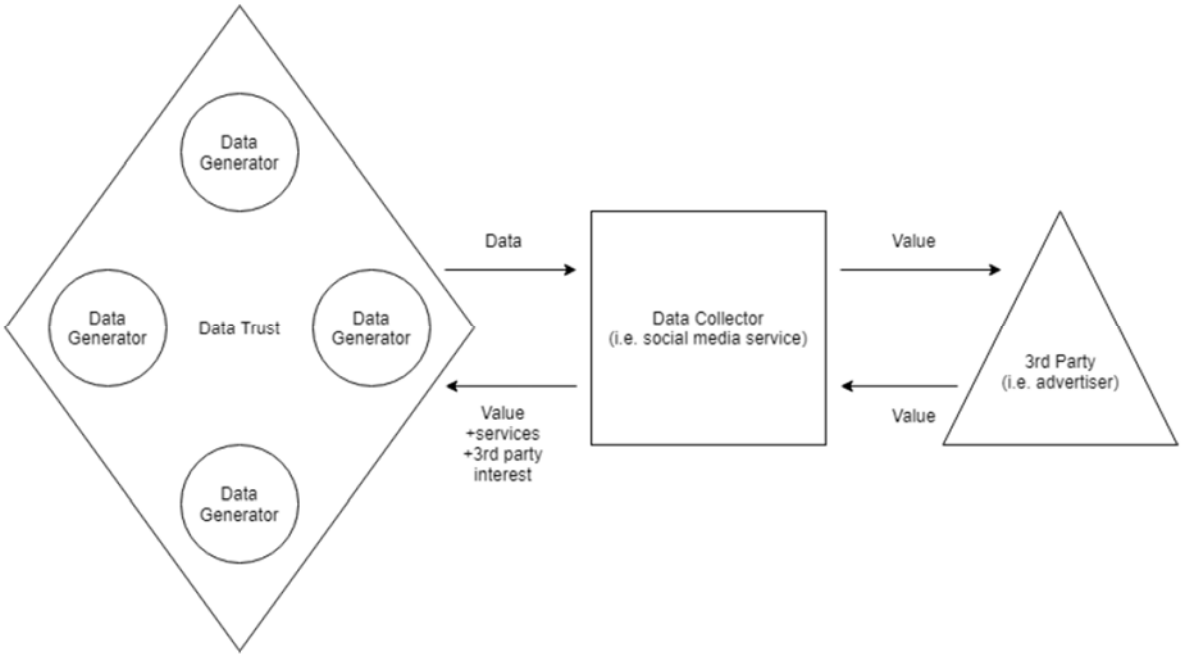

Figure 16 : Data flows in a Generator Centric Data Trust.

Source : Mills, S. (2019). Who Owns the Future ? Data Trusts, Data Commons, and the Future of Data Ownership. *SSRN Electronic Journal*, 1. https://doi.org/10.2139/ssrn.3437936

Ce modèle correspond à la troisième catégorie identifiée par O'Hara (2020) et au modèle *bottom-up* préconisé par Delacroix et Lawrence (2019).

Ce type de *Data Trust* est l'émanation directe des *data subjects* qui mutualisent leurs données et exercent souvent le contrôle sur la structure. Le *Data Trust* sert ici directement les intérêts des *data subjects* ou une cause d'intérêt général qui a leur approbation.

C'est la structure la plus démocratique et la plus respectueuse des droits citoyens. La « licence sociale » y est de la plus grande importance pour opérer légitimement. Par

contre, il n'est pas certain que ce type de structure soit exploitable dans les cadres commerciaux.

Pour les *data trusts* de ce type qui refuseraient de commercialiser les données collectées (par exemple par qu'ils estiment prioritaire de se concentrer sur la protection de la vie privée), se pose la question du financement du *trustee* et de l'infrastructure.

Pour O'Hara (2020), il pourrait en effet être difficile pour ce type de *Data Trust* d'attirer un *trustee* talentueux. Comme ce type de *trust* adopte toujours la structure légale du *trust*, le *trustee* y endosse une responsabilité fiduciaire. Ce qui signifie l'existence d'un certain risque pour lui et d'importants devoirs.

Dans un tel modèle, le rôle des pouvoirs publics pourrait être de subsidier le *trustee* de manière à régler le problème du financement.

Cette grille d'analyse de Mills (2019) nous permet de jeter un regard additionnel intéressant sur l'aspect structurel du *trust* et par conséquent sur les rapports de force sous-jacents. Elle doit, à notre avis, être utilisée en complément des autres grilles de lecture et peut permettre d'identifier d'éventuelles incohérences entre le discours, l'objet revendiqué et l'implémentation du *Data Trust*.

Dans le cas de l'*Urban Data Trust* de Toronto, par exemple, cette grille d'analyse aurait permis de mettre aisément en lumière le fait que la structure proposée par *Sidewalk Labs* était plus proche de la première catégorie de structures (où les *data subjects* sont relégués en périphérie) alors qu'elle se présentait officiellement comme un projet de la 2ème ou de la 3ème catégorie, censé être beaucoup plus démocratique.

Conclusion

À l'issue de ce travail nous ne pouvons que constater la très grande diversité des dispositifs se présentant sous l'appellation de *Data Trusts.*

Par essence, le *trust* est un instrument flexible et modulable qui peut être conçu pour répondre à de multiples besoins. Il n'est donc pas illogique qu'il en existe une certaine variété.

Il apparait toutefois nécessaire d'écarter les propositions qui tiennent plus du marketing que de la réelle fiducie de données. Comme le souligne Artyushina (2020), le cas de l'*Urban Data Trust* de Toronto doit ici nous servir d'avertissement. Il nous semble donc que les dispositifs de la première catégorie de O'Hara (cf. supra p. 78) ne peuvent qu'inviter à la prudence. Voire à la méfiance. Et qu'il est préférable de les écarter.

Comme nous l'avons vu, les *Data Trusts* de la seconde catégorie sont ceux qui attirent sur eux tous les projecteurs. Actifs dans le domaine de l'intelligence artificielle, où les enjeux financiers et stratégiques sont colossaux, ils concentrent logiquement la majorité des investissements et de l'attention.

Plus informels, la structure légale du trust est, pour eux, moins importante. Le dispositif fait surtout partie d'un écosystème de modes de gouvernance variés et flexibles s'inspirant librement les uns des autres. Il existe peu de doutes sur le succès futur de ces dispositifs qu'ils conservent le nom de *Data Trust* ou qu'ils continuent à évoluer.

À l'occasion de discussions informelles avec des professeurs de droit des universités d'Oxford et de Birmingham, Paul Nemitz, un haut conseiller de la Commission Européenne, a expliqué que le Commission entendait favoriser l'émergence d'un écosystème de modes de data gouvernance variés. Et envisageait pour ceux-ci de recourir à un système de certification qualitative. (Delacroix, McFarlane & Nemitz, 2021).

Au vu des importantes ambitions européennes pour le marché de la donnée (Commission, 2020), les *Data Trusts* de la seconde catégorie sont bien positionnés pour occuper une place importante dans ces futurs écosystèmes composés de dispositifs de data gouvernance certifiés.

Il est par contre moins certain qu'ils soient à même de répondre aux inquiétudes des citoyens en matière de vie privée ou qu'ils parviennent à s'articuler harmonieusement avec le RGPD.

La troisième catégorie de *Data Trust* bénéficie de la plus longue tradition académique et endosse fidèlement la structure légale du *trust.* Sa légitimité peut difficilement être remise en question.

Soucieux des asymétries de pouvoir, ces *Data Trusts* se positionnent sur les lignes de tension que nous avons mises en évidence grâce à la typologie des biens d'Ostrom (cf. supra p. 21).

Mc Donald (2019) souligne que comme tous les instruments susceptibles d'avoir un impact, ils occupent un espace contesté politiquement. C'est ce qui fait tout leur intérêt.

Comme ces *Data Trusts* adopte la structure légale du *trust* de tradition anglo-saxonne, la question de savoir quelle forme juridique il pourrait adopter dans un contexte européen devra être tranchée. Paul Nemitz (Delacroix, McFarlane & Nemitz, 2021), forcément très au fait des projets de la Commission se montre rassurant sur ce point.

Les documents publiés autour de l'adoption du *Data Governance Act* (Commission Européenne, 2021) font également allusion à l'éventuelle possibilité de créer une nouvelle forme juridique, en droit européen. D'ici là, cette question reste ouverte.

Quel que soit leur avenir, ces *Data Trusts* de la troisième catégorie (qui peuvent, eux, être considérés comme des fiducies de donnés au sens le plus strict du terme) ont l'immense mérite de poser des questions essentielles pour chaque citoyen numérique :

Celles du devenir de nos traces digitales, de la techno-surveillance, de la nécessité de penser des modes de data gouvernance démocratique.

Ils proposent également une gamme de solutions flexibles qui peuvent compléter judicieusement les approches verticales, de type réglementaire. Et ils puisent dans la longue tradition de mode de gouvernance participative.

Pour toutes ces raisons, ils nous semblent des instruments de data gouvernance précieux.

Comme le disent Hardinges et Wells (2019), les *Data Trusts* ne seront sans doute pas le chapitre final en matière de partage de données. Par contre, ils peuvent certainement aider.

Bibliographie

Ouvrages et Monographies

- O'hara, K. (2019, février 13). *Data Trusts : Ethics, Architecture and Governance for Trustworthy Data Stewardship* [Monograph]. University of Southampton. https://doi.org/10.5258/SOTON/WSI-WP001

- Penner, J. E. (2019). The Law of Trusts. *Paperback, eleventh edition. Oxford University Press*. *ISBN: 9780198795827*.

Articles scientifiques

- Alsaad, A., O'Hara, K., & Carr, L. (2019). Institutional Repositories as a Data Trust Infrastructure. *Companion Publication of the 10th ACM Conference on Web Science*, 1-4 https://doi.org/10.1145/3328413.3329402

- 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. https://doi.org/10.1016/j.tele.2020.101456

- Balkin, J. M. (2017). Free Speech in the Algorithmic Society : Big Data, Private Governance, and New School Speech Regulation*. Yale Law School, Public Law Research Paper No. 615.* https://doi.org/10.2139/ssrn.3038939

- Battisti, M. & Schöpfel, J. (2017). Quel paysage juridique pour l'exploration de données ? *I2D - Information, données & documents*, 54(2), 25-26. https://ezproxy.ichec.be:2098/10.3917/i2d.172.0025

- Birch, K., Chiappetta, M., & Artyushina, A. (2020). The problem of innovation in technoscientific capitalism : Data rentiership and the policy implications of turning personal digital data into a private asset. *Policy Studies*, *41*(5), 468-487. https://doi.org/10.1080/01442872.2020.1748264

- Carr, C., & Hesse, M. (2020). When Alphabet Inc. Plans Toronto's Waterfront : New Post-Political Modes of Urban Governance. *Urban Planning*, *5*, 69. https://doi.org/10.17645/up.v5i1.2519

- Ciuriak, D., & Wylie, B. (2018). Data and Digital Rights : More Questions Than Answers - But Enumerating the Questions is Essential *(SSRN Scholarly Paper ID 3300263). Social Science Research Network* https://doi.org/10.2139/ssrn.3300263

- Delacroix, S., & Lawrence, N. D. (2019). Bottom-up data trusts : Disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law*, *9*(4), 236-252. https://doi.org/10.1093/idpl/ipz014

- Delacroix, S., & Montgomery, J. (2020). From Research Data Ethics Principles to Practice : Data Trusts as a Governance *Tool (SSRN Scholarly Paper ID 3736090). Social Science Research Network.* https://dx.doi.org/10.2139/ssrn.3736090

- Edwards, L. (2004). The Problem with Privacy. *International Review of Law Computers & Technology*, Vol. 18 (No. 3), pp. 263-294, https://papers.ssrn.com/abstract=1857536

- Gomer, R. C., & Simperl, E. (2020). Trusts, co-ops, and crowd workers : Could we include crowd data workers as stakeholders in data trust design? *Data & Policy*, 2. https://doi.org/10.1017/dap.2020.21

- Goodman, E. P., & Powles, J. (2019, novembre). *Urbanism Under Google : Lessons from Sidewalk Toronto* (SSRN Scholarly Paper ID 3390610). 88 Fordham Law Review. 457 (2019), http://dx.doi.org/10.2139/ssrn.3390610

- Hardin, G. (1968). The Tragedy of the Commons. *Science*, *162*(3859), 1243-1248. https://doi.org/10.1126/science.162.3859.1243

- Hardinges, J., & Wells, P. (2019). Data trusts will not be the final word on data sharing, but they might help. *Public Money & Management, 39* (5), 320-321. https://doi.org/10.1080/09540962.2019.1611232

- Holm, S., Kristiansen, T., & Ploug, T. (2020). Control, trust and the sharing of health information : The limits of trust. *Journal of Medical Ethics*, medethics-2019. https://doi.org/10.1136/medethics-2019-105887

- Khan, L., & Pozen, D. (2019). A Skeptical View of Information Fiduciaries. *Harvard Law Review, Vol. 133, pp. 497-541, 2019* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3341661

- Lau Jia Jun, J., Penner, J. E., & Wong, B. (2019). The Basics of Private and Public Data Trusts. *NUS Law Working Paper No. 2019/019, EW Barker Centre for Law & Business* Working Paper 19/03. https://doi.org/10.2139/ssrn.3458192

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

- Mills, S. (2019). Who Owns the Future ? Data Trusts, Data Commons, and the Future of Data Ownership. *SSRN Electronic Journal*, 1. https://doi.org/10.2139/ssrn.3437936

- Milne, R., Sorbie, A., & Dixon-Woods, M. (2021). What can data trusts for health research learn from participatory governance in biobanks ? *Journal of Medical Ethics*. https://doi.org/10.1136/medethics-2020-107020

- Monnerie, N. (2018). Les défis de la commercialisation des données après le RGPD : aspects concurrentiels d'un marché en développement. *Revue*

internationale de droit économique, t. xxxii(4), 431-452. https://doi.org/10.3917/ride.324.0431

- Ostrom, E. (2010). Beyond Markets and States : Polycentric Governance of Complex Economic Systems. *The American Economic Review*, *100*(3), 641-672. https://www.aeaweb.org/articles?id=10.1257/aer.100.3.641

- O'Hara, K. (2019, février 13). *Data Trusts : Ethics, Architecture and Governance for Trustworthy Data Stewardship* [Monograph]. University of Southampton. https://doi.org/10.5258/SOTON/WSI-WP001

- O'Hara, K. (2020). Data Trusts. *European Data Protection Law Review*, *6*(4), 484-491. https://doi.org/10.21552/edpl/2020/4/4

- Paprica, P. A., Sutherland, E., Smith, A., Brudno, M., Cartagena, R. G., Crichlow, M., Courtney, B., Loken, C., McGrail, K. M., Ryan, A., Schull, M. J., Thorogood, A., Virtanen, C., & Yang, K. (2020). Essential Requirements for Establishing and Operating Data Trusts. *International Journal of Population Data Science*, *5*. https://doi.org/10.23889/ijpds.v5i1.1353

- Picon, A. (2018). Villes et systèmes d'information : de la naissance de l'urbanisme moderne à l'émergence de la *smart city*. *Flux*, 111-112(1), 80-93. https://doi.org/10.3917/flux1.111.0080

- Rosenbaum, S. (2010). Data Governance and Stewardship : Designing Data Stewardship Entities and Advancing Data Access. *Health Services Research*, *45*(5p2), 1442-1455. https://doi.org/10.1111/j.1475-6773.2010.01140.x

- Scassa, T. (2018, septembre 04). Data Ownership. *CIGI Papers No. 187, Ottawa Faculty of Law Working Paper No. 2018-26*. https://doi.org/10.2139/ssrn.3251542

- Scott, A. W. (1966). The Importance of the Trust. *University of Colorado Law Review*, *39*, 177.

- Stalla-Bourdillon, S., Thuermer, G., Walker, J., Carmichael, L., & Simperl, E. (2020). Data protection by design : Building the foundations of trustworthy data sharing. *Data & Policy*, *2*. https://doi.org/10.1017/dap.2020.1

- Soupizet, J.-F. (2020). La smart city : Mythe et réalité. *Futuribles*, *434*(1), 49-65. Cairn.info. https://doi.org/10.3917/futur.434.0049

- Van Boxstael, J.-L., & Fonteyn, J. (2015). Le trust et l'acquisition ou la vente d'un immeuble situé en Belgique. *Notamus*, *2015*(1), 46-50. https://dial.uclouvain.be/pr/boreal/object/boreal:165204

- Walker, J., Simperl, E., Stalla-Bourdillon, S., & O'Hara, K. (2019, septembre 16). Decision-making processes for data sharing : A framework for data trusts. *ACM WomENcourage 2019: Celebration of Women in Computing (16/09/19 - 18/09/19)*. https://eprints.soton.ac.uk/434736/

- Winickoff, D., & Winickoff, R. (2003). The Charitable Trust as a Model for Genomic Biobanks. *The New England journal of medicine*, *349*, 1180-1184. https://www.nejm.org/doi/full/10.1056/NEJMsb030036

Rapports et documents officiels

- Ada Lovelace Institute. (2021, mars 04). *Exploring legal mechanisms for data stewardship.* Consulté 9 avril 2021, à l'adresse https://www.adalovelaceinstitute.org/report/legal-mechanisms-data-stewardship/

- BPE Solicitors, Pinsent Masons, & Queen Mary University of London. (2019, avril). *Data trusts : legal and governance considerations*. https://theodi.org/wp-content/uploads/2019/04/General-legal-report-on-data-trust.pdf

- Carrière-Swallow, Y., & Haksar, V. (2019, septembre 23). *The Economics and Implications of Data : An Integrated Perspective*. IMF. https://www.imf.org/en/Publications/Departmental-Papers-Policy-Papers/Issues/2019/09/20/The-Economics-and-Implications-of-Data-An-Integrated-Perspective-48596

- Commission Européenne. (2015, mai 06). *A Digital Single Market Strategy for Europe - Analysis and Evidence. Commission staff working document. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52015SC0100*

- Commission Européenne. (2020, février). *Communication de la Commission au Parlement Européen, au Conseil, au Comité Economique et Social Européen et au Comité des Régions - une stratégie européenne pour les données*. https://eur-lex.europa.eu/legal-content/FR/TXT/PDF/?uri=CELEX:52020DC0066&from=EN

- Commission Européenne (2021, mars 9). *Strategy for Data | Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/strategy-data*

- Crémer, J., de Montjoye, Y-A. & Schweitzer, H., (2019) *Competition policy for the digital era.* Directorate General for Competition - Commission Européenne. https://data.europa.eu/doi/10.2763/407537

- Element AI, & Nesta. (2019, mars). *Community Solutions Portal | Resources | Data Trusts : A New Tool for Data Governance*. Consulté 9 avril 2021, à l'adresse https://portal.futurecitiescanada.ca/resources/data-trusts-a-new-tool-for-data-governance/

- Hardinges, J., Wells, P., Blandford, A., Tennison, J. & Scott, A., (2019, avril). *Data trusts : Lessons from three pilots.* Open Data Institute. https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit

- OECD. (2016, octobre 27). *Big data : Bringing competition policy to the digital era.* https://www.oecd.org/competition/big-data-bringing-competition-policy-to-the-digital-era.htm

- Parlement Européen. (2016, juin 08). *Directive (UE) 2016/943 du Parlement Européen et du Conseil du 8 juin 2016 sur la protection des savoir-faire et des informations commerciales non divulgués (secrets d'affaires) contre l'obtention, l'utilisation et la divulgation illicites*. https://eur-lex.europa.eu/eli/dir/2016/943/oj

- Information Commissioner's Office (2020, octobre 08). *Guide to the General Data Protection Regulation (GDPR)*. ICO. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/

- Hall, W. & Pesenti, J. (2017, octobre 15). *Growing the artificial intelligence industry in the UK*. Londres : Department for Digital, Culture, Media & Sport and Department for Business, Energy & Industrial Strategy. https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk

- Sidewalk Labs. (2018). *Digital Governance Proposals for DSAP Consultation. https://fr.scribd.com/document/390927208/Sidewalk-Toronto-Digital-Governance-Proposals-for-DSAP-Consultation*

- Szczepanski, M. (2020). *Is data the new oil? Competition issues in the digital economy.* EPRS | European Parliamentary Research Service. https://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI%282020%29646117

Sites web et pages web

- Artyushina, A. (2020, août 11). *The EU is launching a market for personal data. Here's what that means for privacy.* MIT Technology Review. https://www.technologyreview.com/2020/08/11/1006555/eu-data-trust-trusts-project-privacy-policy-opinion/

- Banga, A. (2016, janvier 27). *A global economy powered by data*. World Economic Forum. https://www.weforum.org/agenda/2016/01/a-global-economy-powered-by-data

- Delacroix, S., McFarlane, B., & Nemitz, P. (2021, janvier 26). *Understanding the Data Governance Act : In conversation with Sylvie Delacroix, Ben McFarlane and Paul Nemitz*. Data Trusts Initiative. https://datatrusts.uk/blogs/understanding-the-data-governance-act-in-conversation-with-sylvie-delacroix-ben-mcfarlane-and-paul-nemitz

- Gaudiaut, T. (2021, juillet 01). *Infographie : L'ère des géants de la tech*. Statista Infographies. Consulté le 04 août 2021, à l'adresse https://fr.statista.com/infographie/22656/classement-entreprises-capitalisation-

boursiere/

- Guadamuz, A. (2017, octobre) *L'intelligence artificielle et le droit d'auteur*. Organisation Mondiale de la Propriété Intellectuelle. Consulté le 27 juillet 2021, à l'adresse https://www.wipo.int/wipo_magazine/fr/2017/05/article_0003.html

- Hardinges, J. (2018, juillet 10). *What is a data trust?* The Open Data Institute. Consulté le 01 août 2021, à l'adresse https://theodi.org/article/what-is-a-data-trust/

- Hardinges, J. (2018, octobre 19). *Defining a 'data trust'.* The Open Data Institute. Consulté le 01 août 2021, à l'adresse https://theodi.org/article/defining-a-data-trust/

- Hardinges, J. (2020, mai 17). *Data trusts in 2020*. The Open Data Institute. Consulté le 07 août 2021 à l'adresse https://theodi.org/article/data-trusts-in-2020/

- HiLo. (2020, décembre 24). *HiLo Predictive Analysis for shipping & knowledge database center.* Consulté le 13 août 2021 à l'adresse *https://hilomrm.com/about-us/*

- Lawrence, N. D. (2020, octobre 20). *Data Sharing and Data Trusts*. Neil Lawrence's Talks. Consulté 10 juillet 2021, à l'adresse http://inverseprobability.com/talks/notes/data-sharing-and-data-trusts.html

- McDonald, S. & Porcaro, K. (2015, août 04). *The Civic Trust*. Medium. Consulté le 02 août 2021 à l'adresse https://medium.com/@digitalpublic/the-civic-trust-e674f9aeab43

- McDonald, S. & Wylie, B. (2018, octobre 09). *What Is a Data Trust?* Centre for International Governance Innovation. Consulté 4 avril 2021, à l'adresse https://www.cigionline.org/articles/what-data-trust

- McDonald, S. (2018, octobre 17). *Toronto, Civic Data, and Trust*. Medium. Consulté le 20 juillet 2021, à l'adresse https://medium.com/@digitalpublic/toronto-civic-data-and-trust-ee7ab928fb68

- McDonald, S. (2019, mars 05). *Reclaiming Data Trusts*. Centre for International Governance Innovation. Consulté le 4 avril 2021, à l'adresse https://www.cigionline.org/articles/reclaiming-data-trusts/

- Qiang, Y. (2018, juin 25). *La quatrième révolution*. UNESCO. https://fr.unesco.org/courier/2018-3/quatrieme-revolution

- Schlosser, A. (2018, janvier 10). *You may have heard data is the new oil. It's not*. World Economic Forum. Consulté 10 juillet 2021, à l'adresse https://www.weforum.org/agenda/2018/01/data-is-not-the-new-oil/

- UK Biobank (2021). *About us.* UK Biobank. Consulté le 04 août 2021 à l'adresse https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us

- Wylie, B. (2018, mai 04). *You can't opt out of public space, says critic of Toronto's proposed « smart neighbourhood »* | *CBC Radio*. CBC. https://www.cbc.ca/radio/thecurrent/the-current-for-may-4-2018-1.4647190/you-can-t-opt-out-of-public-space-says-critic-of-toronto-s-proposed-smart-neighbourhood-1.4648088

- Zittrain, J. (2019, février 20). *At Harvard Law, Zittrain and Zuckerberg discuss encryption, 'information fiduciaries' and targeted advertisements*. Harvard Law Today. Consulté le 07 août 2021 https://today.law.harvard.edu/at-harvard-law-zittrain-and-zuckerberg-discuss-encryption-information-fiduciaries-and-targeted-advertisements/

Syllabus

- Ejzyn A. (2020). *Stratégie Digitale des entreprises*. Syllabus. ICHEC, Bruxelles.

- Paquet, G., Schrooten, V. et Simon, S. (2018). *Réaliser et rédiger son mémoire en gestion.* Syllabus. ICHEC, Bruxelles.

- Paquet, G., Bawin, I., Schrooten, V. et Wattier, S. (2017). *Séminaire de méthodologie et d'initiation à la démarche scientifique*. Syllabus. ICHEC, Bruxelles.

- Roman, P. (2021). *Topics in Economics and Environment.* Syllabus. ICHEC, Bruxelles.

- Scharff, P-A. (2021). *Management stratégique : Stratégie d'entreprise.* Syllabus. ICHEC, Bruxelles