

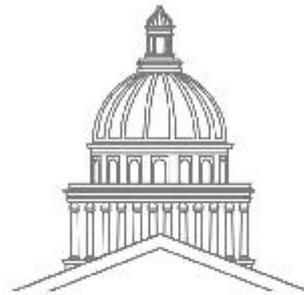
Université Panthéon-Assas

Institut Français de Presse (IFP)

Mémoire de master 2/ septembre 2018

Mémoire de Master 2 de recherche Médias,
Publics et Cultures Numériques
dirigé par Mme Cécile MEADEL

**Le marché des données n'existe pas :
étude des différentes formes de
valorisation des données.**



UNIVERSITÉ PARIS II
PANTHÉON-ASSAS

Aude ROUILLOT

Sous la direction de Cécile MEADEL

Date de dépôt : 03/09/2018

Avertissement

La Faculté n'entend donner aucune approbation ni improbation aux opinions émises dans ce mémoire ; ces opinions doivent être considérées comme propres à leur auteur.

REMERCIEMENTS :

Je remercie Mme Cécile Méadel, ma directrice de mémoire, pour son accompagnement et ses conseils lors de mon travail de recherche.

Je remercie tous les professionnels cités dans ce mémoire, qui m'ont accordé de leur temps, et qui ont nourri ma réflexion.

Je remercie Lydie Bouzenard et Laurent Rouillot, mes parents, pour leur soutien, leur relecture, leurs conseils et leurs avis.

Je remercie Emile Petitperrin pour son soutien au quotidien et l'intérêt qu'il a porté à mon travail.

Je remercie Alice, Savannah, Marine, Victoria et Sandra pour les conseils, le soutien, et les séances de travail en groupe.

Résumé :

Mots clés : Marché des données, Data, Big Data, Open Data, Smart City, économie circulaire, données personnelles, données privées

L'expression de « marché des données », très répandue dans les médias et dans l'imaginaire collectif, suscite aujourd'hui un grand nombre de fantasmes. Entre les énormes profits réalisés par les GAFAs, et les scandales à propos du commerce des données personnelles, c'est toujours du même « marché » dont il est question. Seulement, dès que l'on tente de comprendre ce qu'il y a vraiment derrière ce « marché des données » (les acteurs, l'organisation des échanges), on découvre une réalité complexe et difficile à saisir. En déconstruisant cette expression, on découvre un écosystème riche, composé d'une grande diversité d'acteurs qui s'approprient et valorisent les données selon des modèles différents.

Les données ne forment pas un tout homogène. Données personnelles, données ouvertes (Open Data), ou encore données privées des entreprises (confidentielles ou non) ne sont pas soumises aux mêmes régulations, et répondent à des usages différents.

Derrière le terme de « marché », des échanges variés, pas forcément marchands, entre différents acteurs (internauts, GAFAs, Data Brokers, annonceurs etc.) se basent sur l'exploitation intensive des données personnelles (presque exclusivement). Permettant d'offrir des services publicitaires et marketing, la donnée n'est pas échangée pour elle-même, mais bien pour ce qu'elle permet de faire. Elle obtient de la valeur à travers les services qu'elle permet de produire (ciblage publicitaire, marketing analytique etc.) Plus qu'un « marché des données », il faudrait parler d'une économie capitaliste irriguée par la donnée.

Enfin, au-delà de cette économie fondée sur le profit, d'autres types de valorisation des données émergent, reposant sur l'ouverture de celles-ci (Open Data, Smart City). En effet, ouvrir ne veut pas dire détruire, et donner accès à ses données librement peut générer de la valeur, parallèlement à l'exploitation capitaliste des données.

Introduction

La notion de « marché des données », voire « marché des données personnelles » est souvent utilisée dans les médias, afin de parler de la valorisation des données. On peut par exemple lire dans un article du *Monde* de mai 2015 :

« Selon le cabinet IDC, le marché des données des citoyens de l'Europe des vingt-huit s'élevait à 60 milliards d'euros en 2016 et devrait atteindre 80 milliards en 2020. »¹

Dans ce mémoire, je propose de montrer en quoi l'utilisation de ce terme de « marché des données » est problématique. Utilisé à tout va, il recouvre en fait des réalités variées et hétérogènes, que je vais tenter d'éclaircir. Les sujets comme l'utilisation des données ou la protection des données personnelles sont d'une grande actualité aujourd'hui. À l'heure du Règlement Général pour la Protection des Données² ou du scandale Facebook-Cambridge Analytica³, il me semble indispensable de chercher à comprendre les spécificités de la collecte, la réutilisation et la marchandisation des données.

Si la collecte des données est un phénomène qui date déjà des années 1970 – 1980 (Rochelandet, 2010) sa massification, leur traitement et réutilisation, c'est-à-dire la datafication, sont des phénomènes récents. Ils vont de paire avec le développement des technologies du numérique, dès la deuxième moitié du XXème siècle. Celles-ci connaissent un tournant important dans les années 2000. Contrairement aux autres supports médiatiques comme la télévision, la presse ou la radio, Internet se pose comme support participatif. Les utilisateurs peuvent créer du contenu, laisser leur trace. Ils sont de plus en plus incités à prendre ce rôle actif avec le lancement des grands réseaux sociaux : Facebook en 2004, Twitter en 2006... Ces plateformes poussent les internautes à *liker*, commenter, publier des photos, des avis, partager des articles etc. La commercialisation de

¹ https://www.lemonde.fr/economie/article/2017/05/28/des-donnees-personnelles-tres-convoitees_5135092_3234.html

² Le RGPD est un règlement européen entré en vigueur le 25 mai 2018. Il définit la nouvelle manière dont les entreprises utilisent et gèrent les données personnelles qu'elles collectent. Il permet une adaptation aux évolutions technologiques comme le Cloud, ou les objets connectés. Dans les grandes lignes, le RGPD impose : de la transparence de la part des entreprises quant à l'utilisation des données personnelles, un consentement des personnes concernées, la légalité et sécurité du traitement.

³ Ce scandale fait référence au fait que Facebook a laissé la compagnie britannique de conseil Cambridge Analytica recueillir, depuis 2014, des informations sur ses utilisateurs, sans les en informer pour ensuite s'en servir afin d'influencer les électeurs américains lors de la campagne présidentielle de 2016. Tout cela a été fait sans en informer les électeurs américains, et donc sans leur consentement.

l'iPhone au même moment¹, suivie par la popularisation du smartphone, permet à de plus en plus d'individus d'accéder à du contenu sur Internet tout en étant mobiles. Pour illustrer cette idée, on peut par exemple noter qu'en 2017, 65% des Français possèdent un smartphone, et 60% accèdent à Internet grâce à leur mobile, via un navigateur ou des applications². Chaque minute dans le monde sept millions de Snaps sont envoyés via Snapchat, deux-cent-seize millions de photos sont aimées sur Facebook et deux millions et demi sur Instagram³. La popularisation du smartphone a également permis un traçage physique, géographique des individus via les données de géolocalisation (Rochelandet, 2010).

Cette utilisation intensive, personnelle et massive d'Internet génère un grand nombre de données. Une recherche sur Google, un trajet sur Google Map, le visionnage d'une série sur un service de VOD (vidéo à la demande) ... Toutes ces actions variées, mais désormais si courantes, sont des sources inépuisables de données. Je présenterai plus en détail ce point dans la suite du mémoire, en précisant quels sont les procédés qui permettent de récolter ces données. Mais il faut retenir que l'utilisation quotidienne de ces services permet aux entreprises qui en sont à l'origine de recueillir des informations de plus en plus précises sur les utilisateurs.

Cette ère du numérique place les données au cœur de nos sociétés et de nos économies. Celles-ci ne sont pas seulement des traces laissées par les internautes. Elles peuvent aussi provenir de bien d'autres sources, comme des capteurs mesurant les précipitations, la pollution de l'air, le bon fonctionnement d'un véhicule etc. Depuis une vingtaine d'années, nous avons affaire à une véritable explosion quantitative des données numériques. Chaque jour, environ deux trillions et demi de données sont produites⁴. D'après IBM, ces deux dernières années ont suffi pour que soient produites 90% des données jamais produites dans le monde⁵. Face à une telle augmentation, les techniques de stockage, de gestion de base de données et de traitement de l'information ont dû s'adapter. Ces évolutions sont rassemblées sous le concept de Big Data⁶. Il est difficile de trouver une définition précise du Big Data tant il s'apparente à des réalités diverses en fonction de la personne qui en parle. On peut néanmoins s'appuyer sur celle de la CNIL :

¹ L'iPhone est commercialisé à partir du 29 juin 2007 aux Etats-Unis, et à partir du 28 novembre 2007 en France.

² <https://www.blogdumoderateur.com/chiffres-internet/>

³ <https://www.blogdumoderateur.com/chiffres-internet/>

⁴ <https://www.lebigdata.fr/definition-big-data>

⁵ <https://www.lebigdata.fr/chiffres-big-data>

⁶ <https://www.lebigdata.fr/definition-big-data>

*« On parle depuis quelques années du phénomène de Big Data, que l'on traduit souvent par « données massives ». Avec le développement des nouvelles technologies, d'internet et des réseaux sociaux ces vingt dernières années, la production de données numériques a été de plus en plus [importante] : textes, photos, vidéos, etc. Le gigantesque volume de données numériques produites combiné aux capacités sans cesse accrues de stockage et à des outils d'analyse en temps réel de plus en plus sophistiqués offre aujourd'hui des possibilités inégalées d'exploitation des informations. Les ensembles de données traités correspondant à la définition du Big Data répondent à trois caractéristiques principales : volume, vitesse et variété ».*¹

Je souhaiterais approfondir cette définition en me référant à un entretien avec Laurent Rouillot, ingénieur en génie logiciel, que j'ai mené dans le cadre de mon mémoire, que je présente plus loin dans cette introduction. Il insistait surtout sur la dimension non structurée caractéristique des données du Big Data :

*« Le Big Data pour moi c'est une base de données qui se différencie des bases de données traditionnelles, des bases de données relationnelles qu'on voit dans le monde de l'entreprise, par le fait qu'on intègre dans cette base de données des données non structurées. Une base de données relationnelle d'entreprise c'est une base où on a, au préalable, défini le contenant de façon très précise, et tout ce qu'on insère dans la base de données doit, en termes de format, satisfaire au contenant. Alors que dans le cadre des données de type Big Data, on va prendre ce qu'on trouve et après, derrière, il va falloir construire les algorithmes nécessaires pour réussir à exploiter et extirper l'information qui nous intéresse. »*¹

Donc, selon lui, la donnée structurée ne pose aucun problème, même si elle est massive. En effet, à partir du moment où elle est structurée, elle est facilement utilisable. Une base de données classique peut contenir un grand nombre de données sous le même format, classés les unes par rapport aux autres, il n'est alors pas difficile de s'y retrouver. Or, la particularité du Big Data est que toutes les données récupérées ne sont pas dans des formats similaires : on a affaire à des vidéos, des textes, des images, etc. Les bases qui accueillent ces données ne sont pas structurées, il faut donc la mise en place d'outils spéciaux, l'adaptation des technologies à cette nouveauté.

Une autre spécificité du Big Data qu'il n'avait pas souligné dans cet entretien, mais dont il m'a parlé par la suite est la suivante, et c'est peut-être celle qui nous intéresse le

¹ <https://www.cnil.fr/fr/definition/big-data>

plus pour ce mémoire. Pour reprendre ses mots, il existe une sélection au moment du stockage des données structurées en raison du contenant rigide de la base de données relationnelle. Il parle même plutôt de « *censure* » que de sélection. Cette sélection n'existe pas dans le Big Data. Par conséquent, les données structurées n'exprimeront que ce que l'on a prévu dans le contenant, alors que l'interprétation des données de Big Data est illimitée. Cette interprétation repose alors sur l'algorithme d'exploitation.

Les GAFA (Google, Amazon, Facebook, Apple)² sont les premiers à s'être approprié ces nouvelles techniques permettant de gérer et d'utiliser des volumes massifs de données. Ils répondent à une nouvelle stratégie, de plus en plus importante aujourd'hui, qui est de récolter un maximum de données afin de pouvoir les valoriser.

En effet, en 2006, Clive Humby, un mathématicien britannique, actuellement Chief Data Scientist chez Starcount, co-fondateur de Dunnhumby³, faisait cette remarque :

*« Data is the new oil. It's valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc. to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value »*⁴.

Cette idée de « nouvel or noir » est très reprise dans les médias français. Les données irriguent l'économie numérique¹. Posséder des données peut rapporter beaucoup, mais à l'état brut, elles ne servent à rien. Il devient donc indispensable pour les entreprises de se doter des moyens de les récolter, ainsi que de les traiter.

Cette valorisation peut prendre plusieurs formes, comme nous le verrons ensuite. À titre d'exemple, on peut parler d'une des plus évidentes qui est l'utilisation des données à des fins marketing. Des entreprises qui fournissent des biens ou des services vont avoir tout intérêt aujourd'hui, dans un monde où les consommateurs sont submergés d'informations, à se différencier, à mieux cibler leur public. Les données récupérées et traitées vont alors permettre de créer des publicités personnalisées, plus efficaces et donc plus rentables. Cet exemple sera développé dans le II du mémoire.

¹ <https://www.cnil.fr/fr/definition/big-data>

² L'expression GAFA désigne à l'origine les quatre grandes entreprises du web, Google, Amazon, Facebook et Apple. Aujourd'hui, en plus de celles-ci, elle désigne une dizaine d'autres entreprises comme Microsoft, IBM, Airbnb... On peut également utiliser l'expression « Géants du web » pour les désigner (CF définition Wikipédia).

³ Starcount et Dunnhumby sont deux entreprises de Data Science

⁴ Traduction : « *La donnée est le nouveau pétrole. Comme le pétrole, elle a de la valeur, mais si elle n'est pas raffinée, elle ne peut pas vraiment être utilisée. Comme le pétrole qui doit être changé en gaz, plastique, produits chimiques etc., la donnée doit être traitée, analysée pour devenir valorisable et rentable,* », source : <https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>

On comprend donc également pourquoi il devient intéressant de pouvoir recouper un maximum de données, de sources diverses et variées. Elles permettent alors de fournir des informations plus précises sur les utilisateurs, les consommateurs². Les différents acteurs qui possèdent des données vont donc avoir intérêt à se les échanger. C'est dans ce contexte qu'on entend parler de « marché des données ».

Le terme de « marché des données » revient souvent dans les articles de presse que j'ai pu lire au cours de mes recherches. Si certains tentent de donner une image assez précise de cette réalité³, la plupart du temps, on retrouve tout de même souvent deux schémas : « le marché des données » vaut plusieurs milliards de dollars, et les entreprises ont grand intérêt à se saisir de ce « *nouvel or noir* »⁴ ; mais en même temps, ce n'est pas éthique, il y a un manque de confiance de la part des utilisateurs qui ne savent pas où partent leurs données, de ce que les entreprises font avec⁵. Ce sont les deux thèmes principaux que l'on retrouve dans les articles de presse lorsque l'on recherche des informations sur le marché des données.

Pour réussir à trouver une information qui diffère un peu, par exemple sur les types d'acteurs précis que l'on retrouve sur ce « marché des données », sur les espaces dédiés à la revente ou à l'achat, cela s'avère être très difficile. Cette expression récurrente est, pour moi problématique car elle a tendance à lisser une réalité qui s'avère beaucoup plus complexe qu'elle n'en a l'air.

Ces articles ont une visée informative, non pas scientifique. On peut donc penser qu'il est en quelque sorte normal de passer par des raccourcis. Néanmoins, cela est pour moi paradoxal, dans le sens où si l'inquiétude des utilisateurs est vraiment importante, le fait de lisser une réalité complexe derrière une expression floue sans l'explicitement correctement est un écueil. Je tiens à préciser ce que je trouve si problématique dans cette expression de « marché des données ».

Le premier problème qui se pose est que le bien échangé auquel cette expression fait référence n'est pas uniforme. Dans un entretien mené avec Jean-Marc Lazard, co fondateur

¹ On peut par exemple citer *Les Echos* qui titrent un article du 11 septembre 2017 « Data, le nouvel or noir », ou *Franceinfo*, avec son article du 12 mars 2015 : « Le Big Data, le nouvel or noir de l'économie numérique », ou encore celui du *Journal du net* du 22 août 2017 : « L'or noir du numérique, le Big Data nous livre ses secrets ».

² Pour continuer sur l'exemple de l'utilisation marketing, mais ça n'est pas la seule utilisation, comme je vais le montrer dans la suite du mémoire.

³ https://www.lesechos.fr/07/03/2013/lesechos.fr/0202599460114_la-ruce-vers-l-or-des-donnees-personnelles.htm#

⁴ <https://www.lesechos.fr/partenaire/le-comptoir-mm-de-la-nouvelle-entreprise/partenaire-1185-data-le-nouvel-or-noir-2113238.php>

⁵ <https://www.lesechos.fr/idees-debats/cercle/cercle-183763-le-siecle-de-la-donnee-2182542.php>

d'OpenDataSoft, que je présenterai dans la suite du mémoire, celui-ci me disait : « *le mot data exprime des réalités totalement différentes. Ce qui est d'ailleurs un challenge pour nous, quand on s'adresse à nos clients* »

« Les données » sont en effet très hétérogènes, elles ne sont pas toutes collectées de la même manière, ne sont pas accessibles de façon égale. Certaines sont plus sensibles que d'autres. Certaines peuvent avoir énormément de valeur monétaire alors que d'autres sont accessibles librement et gratuitement.

L'utilisation du terme « marché » est également problématique. Il renvoie à des réalités économiques très précises. Parler de marché c'est parler d'une articulation de l'offre et de la demande autour de biens ou de services. Or, ici, les vendeurs et acheteurs se mélangent, ne sont pas clairement définis. De même, ce qui est échangé, la donnée, n'est ni un bien, ni un service, ne peut être commercialisée à l'état brut, doit être traitée, modifiée en fonction des attentes des acheteurs, mais également en fonction des capacités et des opportunités des vendeurs.

Finalement, l'utilisation de la notion de « marché des données » est problématique parce qu'elle fait référence à une réalité complexe, hétérogène, parfois même obscure. La marchandisation des données correspond en fait à différents types d'échanges, entre différents acteurs, qui coexistent. Il est également important de souligner que la valorisation des données ne passe pas que par des échanges marchands. Il existe d'autres modèles permettant de mettre en valeur ces données, qu'il est important d'avoir en tête lorsque l'on parle de « marché des données ».

À travers mes recherches, j'ai pu trouver un certain nombre d'articles scientifiques ou d'ouvrages qui traitent, en partie, de la valorisation ou marchandisation des données. Je souhaite ici présenter trois travaux qui m'ont particulièrement aidée au début de mes recherches. Bien que mon mémoire ne repose pas exclusivement sur eux, ils ont eu une place importante dans ma réflexion et dans l'orientation de mes recherches suivantes.

Pour mieux comprendre ce qu'on appelle « données personnelles », j'ai trouvé l'ouvrage de Fabrice Rochelandet, *Économie des données personnelles et de la vie privée* (Rochelandet, 2010) très instructif. Le premier chapitre « Définitions : vie privée et données personnelles » est particulièrement intéressant puisqu'une partie explique ce qui fait la particularité des données personnelles et leur mode de collecte et d'exploitation. Le reste du texte est trop tourné vers la question de l'éthique, à savoir, est-il acceptable

d'utiliser les données personnelles ? Comme je ne souhaite pas me poser cette question dans mon mémoire, j'ai accordé moins d'intérêt aux autres chapitres.

Le livre *Datanomics, les nouveaux business models des données* (Chignard et Benyayer, 2015) a été très intéressant pour m'aider à mieux comprendre les spécificités économiques des données. Cet ouvrage a été rédigé par Simon Chignard Data Editor de data.gouv.fr, dans le cadre de la mission Etalab¹, et Louis-David Benyayer, membre du réseau d'experts de la mission Etalab (Chignard et Benyayer, 2015, 4^{ème} de couverture). Les cinq premiers chapitres m'ont principalement intéressés pour comprendre les spécificités économiques des données, et ainsi pouvoir commencer à déconstruire l'idée d'un « marché des données ».

Pour ce qui est de l'Open Data, j'ai trouvé l'ouvrage *Open Models: Les business modèles de l'économie ouverte* (Benyayer, 2015), de Louis-David Benyayer très intéressant. C'est un livre qui rassemble des articles assez concis, de plusieurs auteurs, et qui traitent des approches ouvertes remettant en cause les organisations traditionnelles. Un chapitre est consacré à l'Open Data : « Les Open Models de la Data », coordonné par Chloé Bonnet (Benyayer, 2015, p66-102), co-fondatrice de l'agence de conseil en innovation Five by Five. Quatre articles ont particulièrement attiré mon attention dans ce chapitre, puisqu'ils parlent de la création de valeur à partir de l'Open Data. Ils donnent des exemples concrets et remettent en cause les croyances liées à la valorisation des données ouvertes. Complétés par les entretiens que j'ai pu faire, ces courts articles étaient très enrichissants.

Au-delà de ces trois travaux clés pour mon mémoire, j'ai pu remarquer, à travers mes lectures, que le sujet de la collecte et de la valorisation des données est, bien souvent, traité sous un angle éthique. Par exemple, est-il bien ou non de monétiser des données, de les utiliser à des fins marketing ? Les articles ou ouvrages qui montraient la valorisation des données sous un angle plus neutre et économique étaient plus rares. Trop souvent, les questions de droit et d'éthique prenaient le dessus, et alors je pouvais retrouver ce fameux terme de « marché des données » qui, à défaut d'être explicité, était souvent critiqué. Je ne souhaite pas donner de point de vue éthique, ni approfondir la partie légale dans ce mémoire. Je souhaite montrer quels sont les acteurs qui collectent les données, et par quels moyens ils les valorisent, pour enfin comprendre de quoi on parle lorsqu'on utilise le terme vague de « marché des données ».

¹ Voir Partie I.

Je présenterai plus en détail les autres travaux scientifiques qui m'ont été utiles dans le développement de mon mémoire, aux moments pertinents.

Afin d'en apprendre davantage sur les différents acteurs qui collectent et utilisent les données, j'ai réalisé un certain nombre d'entretiens. Pour rencontrer des professionnels du milieu de la Data, je me suis rendue à plusieurs événements, comme des salons : DATA JOB 2017 (« rendez-vous majeur des professionnels de la Data »)¹, Big Data Paris 2018². J'ai aussi tenté de contacter plusieurs dizaines de personnes via le réseaux LinkedIn. Il n'a pas été évident d'obtenir des réponses positives. J'ai néanmoins réussi à mener des entretiens avec plusieurs personnes intéressantes aux profils variés. Mes questions touchaient toujours tant à la profession des individus que j'interrogeais, à leur domaine d'expertise, qu'à leur point de vue plus général sur la valorisation et la marchandisation des données.

Le premier entretien que j'ai mené a été avec Laurent Rouillot, mon père, ingénieur en génie logiciel. Il travaille depuis bientôt six ans à la création d'un logiciel de gestion de base de données. S'il ne portait pas vraiment sur l'utilisation et la valorisation des données, cet entretien m'a beaucoup servi pour comprendre en détail ce qu'était le Big Data. Il m'a servi de base pour approfondir mon sujet.

J'ai ensuite rencontré Théo Roux, jeune journaliste à l'Atelier BNP Paribas³. J'avais lu un de ses articles sur les Data Brokers⁴ au début de mes recherches, que j'avais trouvé intéressant et que j'avais envie d'approfondir. L'entretien a été très enrichissant, un bon début pour avancer dans mes recherches.

Après avoir découvert un article qui parlait du projet de Smart City Rennaise⁵, je suis ensuite entrée en contact avec Norbert Friant, responsable du service numérique au sein de Rennes Métropole. Si au départ, je ne m'intéressais qu'à la marchandisation des données, je me suis rendu compte, au fil de mes recherches que d'autres formes de valorisations (non marchandes) existaient. C'est dans ce contexte que j'ai commencé à m'intéresser à la Smart City, à l'ouverture contrôlée et la mise en commun des données.

¹ <http://datasciencepost.com/fr/datajob-2017-rendez-vous-majeur-des-professionnels-de-la-data/>

² <https://www.bigdataparis.com/>

³ « L'Atelier BNP Paribas est une société de prospective et d'innovation qui conseille et accompagne le Groupe BNP Paribas et ses clients dans leur transformation digitale ». <https://atelier.bnpparibas/about-us>

⁴ Les Data Brokers sont des courtiers en données. Ils récupèrent et rachètent des données, les rassemblent pour leur donner plus de valeur et les revendent aux entreprises qui le veulent. CF Partie 2.

⁵ La Smart City, ou « ville intelligente », est caractérisée par une meilleure gestion de ses services et infrastructures grâce à l'utilisation des NTIC (Nouvelles technologies de l'information et de la communication). L'article dans lequel j'ai entendu parler de la Smart City Rennaise pour la première fois vient du *Journal du net* : <https://www.journaldunet.com/economie/services/1206939-rennes-metropole-echange-donnes-privées/>.

Dans cet entretien, Norbert Friant a notamment insisté sur la dimension d'intérêt public d'ouverture des données, mais aussi sur le retard tant culturel que factuel en France, qui empêche d'exploiter l'ouverture des données à son maximum. Cela a été pour moi très intéressant, surtout pour en apprendre davantage sur la valorisation non marchande des données.

J'ai également réalisé un entretien avec Cédric Guilmin, auteur du mémoire « Approche Open Data au Bureau Economique de la Province de Namur », qui avait accepté de me parler de son travail de recherche. J'étais entrée en contact avec lui en lisant un de ses commentaires sur le site <https://data.rennesmetropole.fr/>, dans lequel il disait travailler sur l'Open Data. Cet entretien fut particulièrement technique, et ne m'a pas apporté beaucoup d'informations pertinentes pour mon mémoire. Il m'a néanmoins fait découvrir l'entreprise OpenDataSoft, dont il parle dans son mémoire, que j'ai pu contacter par la suite.

J'ai en effet eu la chance de pouvoir rencontrer Jean-Marc Lazard, co-fondateur et président d'OpenDataSoft¹. Dans cet entretien, il m'a présenté le logiciel que propose son entreprise, de sa vision personnelle de l'Open Data et de la Smart City, ainsi que de la marchandisation des données. Pour lui finalement, la tentative de délimitation d'un « marché des données » n'est pas forcément pertinente. Les données sont partout, et le challenge aujourd'hui, serait, selon lui, une meilleure exploitation, circulation et utilisation des informations et des données.

Dans le contexte de Smart City Rennaise, j'avais pu lire un article sur le projet de mise en place d'un « troc de données » entre la ville et différentes entreprises². J'avais tenté de contacter ces entreprises, et avait obtenu une réponse positive de la part d'Enedis³. J'ai donc rencontré Jean-François Montagne, responsable de la gouvernance des données. Dans cet entretien nous avons évoqué les différentes possibilités de mise à disposition et d'échanges (non marchands) de données, hypothétiques ou factuelles. C'était très intéressant de découvrir comment une entreprise importante chargée de mission de service public se positionne par rapport à ces questions. Si la question du « troc de données » n'en est qu'à l'état de projet, il a été très enrichissant de connaître le point de vue de Jean-François Montagne à ce sujet.

¹ OpenDataSoft est une entreprise française fondée en 2011. Elle propose une solution logicielle sous forme de SaaS (*solution as a service* c'est-à-dire que le logiciel n'est pas installé sur l'ordinateur de l'utilisateur mais est accessible via internet) qui permet d'héberger et de diffuser les données d'entreprises, de collectivités etc. CF Partie 3.

² <https://www.journaldunet.com/economie/services/1206939-rennes-metropole-echange-donnees-privées/>

³ Enedis est l'entreprise chargée de mission de service public, responsable de la gestion et de l'aménagement de la majorité du réseau de distribution d'électricité en France.

Finalement, le dernier entretien que j'ai mené a été avec Michael Chevalier, auteur du mémoire « Les enjeux juridiques concernant les nouveaux modèles d'affaires basés sur la commercialisation des données » (Chevalier, 2015). Je l'ai contacté après avoir trouvé son mémoire en ligne. Sa partie sur les Data Brokers m'avait particulièrement intéressée, et je souhaitais approfondir certains points qui me semblaient obscurs, et que je n'arrivais pas à éclaircir moi-même. Cet entretien s'est donc déroulé sous la forme de questions/réponses précises à propos du fonctionnement des Data Brokers. Il a ainsi pu m'aider à y voir plus clair à ce sujet assez peu traité dans la littérature scientifique.

Ces recherches et rencontres m'ont permis de découvrir que le terme de « marché des données » reflétait une réalité bien plus complexe qu'on ne le pense, et que la valorisation des données ne passe pas uniquement par ce « marché ».

Comment s'articulent les différentes mises en valeur, marchandes ou non, des données, selon les spécificités de ces dernières ?

Afin de répondre à cette problématique, je vais commencer par préciser de quelles données on parle. Je vais montrer que les données peuvent provenir de nombreuses sources différentes, qu'elles ne sont pas toutes sensibles, que même si certaines sont qualifiées de « personnelles », elles restent très hétérogènes (I).

Une fois que le paysage des données sera plus clair, je vais m'atteler à déconstruire l'idée d'un « marché des données ». Cette expression fait en effet référence à une réalité bien plus complexe qu'on ne peut le penser. Je montrerai qui sont les acteurs qui composent ce « marché », comment s'organisent les échanges, jusqu'à remettre en cause l'utilisation de cette expression, qui est pour moi, problématique (II).

Finalement, je présenterai d'autres formes de valorisation des données, non marchandes, comme l'Open Data, l'échange de données dans le contexte de la Smart City. Mon but est de montrer que d'autres types d'échanges sont possibles, motivés par d'autres valeurs que la recherche de profit, et qui pourraient véritablement mettre l'individu au cœur de l'utilisation des données (III).

Partie I : Déconstruction de l'image homogène des données.

L'expression « marché des données » est pour moi problématique car elle lisse une réalité très vaste. Les données ne sont pas toutes les mêmes. Elles ne sont pas collectées de la même manière, ni par les mêmes acteurs. Elles n'ont pas toutes la même importance, leur valeur ne se mesure pas forcément de la même manière, et elles ne portent pas toutes les mêmes enjeux.

Certaines données sont considérées comme privées, voire sensibles. D'autres peuvent être mise à disposition gratuitement au grand public, ou à une audience restreinte... Leur récolte n'est pas homogène non plus. Certaines méthodes sont plus transparentes que d'autres par exemple.

Je souhaite donc, dans cette partie, faire un point précis sur les différents types de données. Pour ce faire je vais m'inspirer de l'article de Henri Isaac, « La donnée, une marchandise comme les autres ? » (Isaac, 2018), dans lequel il dit que les données sont soumises à trois régimes juridiques différents, en fonction de leur type. On a, dans un premier temps, les données à caractère personnel, régies par la Loi 78-17 du 6 janvier 1978, « Internet et liberté ». Il sera très intéressant de revoir, précisément, ce qu'elles sont, comment elles sont collectées, et de questionner la dimension « privée » de ces données, dans un contexte de méfiance et d'inquiétude quant à la réutilisation des données personnelles.

Dans un second temps, il existe les données publiques. Depuis 2016, elles sont soumises à un régime d'ouverture, avec la loi pour une République numérique. De même, ici, il sera intéressant de voir les motivations qui poussent les différents acteurs à ouvrir leurs données, ainsi que le processus d'ouverture par lequel elles doivent passer.

Et enfin, il y a les « autres » données. Elles ne sont soumises à aucune régulation, et sont régies par une base contractuelle permettant leur échange entre les différents acteurs.

Cette division en trois des types de données me semble très pertinente par rapport à mes autres recherches, c'est pourquoi j'ai décidé de diviser cette partie en trois axes, qui reprendront respectivement ces trois sujets.

Pour la réflexion sur les données personnelles, je m'appuierai sur l'ouvrage de Fabrice Rochelandet, *Economie des données personnelles et de la vie privée*, que j'ai déjà présenté en introduction (Rochelandet, 2010). Je vais plus particulièrement m'intéresser au premier chapitre dans lequel on retrouve une classification des données personnelles, ainsi que des informations précises quant aux différentes façons dont elles sont récoltées. Dans le même thème, je vais utiliser l'article « Les modèles d'affaires numériques sont-ils trop indiscrets ? Une analyse empirique » de Cecera Grazia, Fabrice Le Guel et Fabrice Rochelandet (Grazia, Le Guel, Rochelandet, 2015), qui présente assez précisément les méthodes de récoltes des données.

Pour développer la sous-partie sur l'Open Data, je vais m'appuyer principalement sur trois articles. Le premier, « L'Open Data comme nouvelle forme de gouvernance numérique : enjeux, marchés, modèles, idéologies », d'Alain Van Cuyck (Van Cuyck, 2013) présente le mouvement d'ouverture des données et ses enjeux. Le second, « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique » de Valérie Peugeot, (Peugeot, 2010) montrent comment trois secteurs distincts (scientifique, public et privé) s'insèrent dans le mouvement d'Open Data, en prenant l'exemple concret de la Grande Bretagne. Enfin, je reprendrai les notions développées par Jérôme Denis et Samuel Goëta dans « La fabrique des données brutes. Le travail en coulisse de l'Open Data » (Denis, Goëta, 2013) pour montrer que l'ouverture des données n'est pas un processus naturel, simple et évident, mais qu'il nécessite un grand nombre d'outils et des ressources importantes.

Enfin, dans ma troisième partie, plus courte que les deux premières, je présenterai les « autres » données, en m'appuyant plutôt sur des articles de presse, sur l'actualité, et des exemples concrets. Je montrerai deux aspects des données privées d'entreprises : les données sensibles et confidentielles, et les données privées, qui pourraient être ouvertes, mais qui sont plutôt partagées à des acteurs précis.

1. PRESENTATION DES DONNEES PERSONNELLES ET DE LEURS MODES DE COLLECTE.

Les préoccupations autour des données personnelles remontent en France aux années 1970. A cette époque, l'INSEE lance le projet Système Automatisé pour les Fichiers Administratifs et le Répertoire des Individus (SAFARI). Il s'agit de connecter tous les fichiers administratifs concernant les Français, par le biais de leur numéro INSEE, créant un répertoire général permettant d'identifier les individus, et d'ensuite transférer ce répertoire sur des bandes magnétiques¹. C'est *Le Monde* qui dévoile ce projet, qui fait rapidement scandale. L'article intitulé « Safari, la chasse aux Français »² dénonce le danger de l'interconnexion des fichiers et des informations (sans réellement les définir précisément néanmoins). Suite aux réactions que cette annonce provoque, le projet est abandonné.

La loi 78-17 du 6 janvier 1978, aussi appelée « Informatique et Liberté », et la création de la CNIL découlent de cette affaire. Elle est présentée comme un organe de « *surveillance de nos libertés* »³ par les médias de l'époque. Déjà à la fin des années 1970 et dans les années 1980, l'idée d'interconnexion des fichiers des individus inquiète. Dans le journal de 20 heures du 17 janvier 1980 sur Antenne 2, l'exemple de la commande par téléphone montre que cette idée de dresser un profil précis du consommateur en recoupant ses informations personnelles (nom, âge etc) aux pratiques de consommation, ainsi qu'à d'autres informations de sources diverses. Comme le dit la voix off, « *celui qui peut consulter cette fiche a du coup un énorme pouvoir sur vous* »⁴. Le consommateur qui découvre l'existence d'un fichier le concernant peut alors faire appel à la CNIL pour demander un droit d'accès à cette fiche.

Les préoccupations qui concernent les données personnelles et ce qui peut en être fait remontent donc au début de la démocratisation de l'informatique. Dès les années 1970, dans l'imaginaire commun, l'utilisation des données personnelles est souvent considérée comme louche, pouvant mettre en danger les libertés individuelles au niveau politique

¹ Article issu du Blog de Simon Chignard, auteur de *Datanomics. Les nouveaux business models des données* : , <https://donneesouvertes.info/2018/01/26/safari-la-chasse-aux-francais-40-ans-apres/>.

² Philippe Boucher, « SAFARI, la chasse aux sorcières », *Le Monde*, 21 mars 1974.

³ Journal de 20h du 17 janvier 1980 sur Antenne 2, présenté par David Pujadas, <https://sites.ina.fr/cnil-40-ans/focus/chapitre/4>

⁴ <https://sites.ina.fr/cnil-40-ans/focus/chapitre/4>, vidéo à 1'04

(dans le cadre du projet SAFARI), comme au niveau publicitaire ou marketing (avec l'idée du ciblage, qui peut s'apparenter à de la manipulation).

Depuis, la réutilisation des données personnelles est généralement vue comme problématique, et peut mettre à mal la confiance des individus.

Afin de bien comprendre les enjeux derrière l'utilisation des données personnelles il me semble important de réfléchir en trois temps. Premièrement, il faut réussir à comprendre ce que sont vraiment les données personnelles. Je partirai de la définition officielle de la CNIL, et de celle du RGPD, et les approfondirai afin d'avoir une vision englobante de ce qu'on appelle « données personnelles »¹.

Deuxièmement, il est nécessaire de comprendre comment ces informations sont collectées, avant d'étudier une quelconque réutilisation. A partir des articles qui traitent de la collecte des données personnelles (Rochelandet, 2010 ; Cecere, Le Guel, Rochelandet, 2015), j'ai dégagé trois types de collecte des données personnelles que je détaillerai.

Finalement, je questionnerai la dimension « privée » des données personnelles, en m'appuyant notamment sur la politique d'utilisation des données de Facebook² pour montrer que la collecte et la réutilisation des données sont prévues, dans les conditions générales d'utilisation des plateformes.

a. Quelle définition des données à caractère personnel ?

On retrouve plusieurs définitions officielles de la donnée personnelle. Pour la loi de 1978 « Informatique et Libertés »³, c'est « *toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à plusieurs éléments qui lui sont propres* ». Sur le site internet de la CNIL⁴ on retrouve la même idée, formulée de cette manière : une donnée personnelle est « *toute information identifiant directement ou indirectement une personne physique (ex. nom, no d'immatriculation, n° de téléphone, photographie, date de naissance, commune de résidence, empreinte digitale...)* ». L'article 4 du RGPD donne

¹ Il s'agit ici de montrer que ce n'est pas seulement la photo de profil, ou le numéro de téléphone, mais bien toutes les traces qu'on laisse derrière nous.

² <https://fr-fr.facebook.com/about/privacy/>

³ Article 3 de la Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

⁴ <https://www.cnil.fr/fr/definition/donnee-personnelle>

cette définition : « *toute information se rapportant à une personne physique identifiée ou identifiable (ci-après dénommée « personne concernée») ; est réputée être une « personne physique identifiable» une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale* »¹.

Globalement, les définitions insistent sur le fait que ces données permettent de reconnaître, d'identifier une personne, soit sur une photo, soit par un nom, par une adresse etc. La définition du RGPD est peut-être la plus complète, bien qu'assez floue à mon avis, surtout à la fin. Ces différentes facettes de l'identité de l'individu sont justement ce qui fait la particularité des données personnelles à l'ère du numérique. Pour moi, cela fait référence aux différentes actions des individus, aux traces laissées lors de la navigation sur internet, mais aussi au fait que les données sont croisées entre elles. Un commentaire laissé par un internaute sur YouTube, sous un pseudonyme, peut-être recoupé avec l'adresse Gmail², à tous les emails envoyés par cette personne³, aux vidéos qu'elle a visionnées sur la plateforme etc. Tout ce recoupement permet d'avoir une vision globale de l'individu bien plus inquiétante que la simple photo ou adresse qui permet de le reconnaître.

Les données des individus ne sont pas utilisées que dans un simple contexte de reconnaissance. Elles peuvent servir à donner des profils d'individus très orientés. Fabrice Rochelandet donne une classification des données personnelles. Il les divise en deux parties : les données objectives et les données subjectives (Rochelandet, 2010). Les premières sont celles qui permettent une reconnaissance et une identification de l'individu sans biais interprétatif. Ce sont donc les données d'identification (comme le numéro de téléphone), les données physiques (comme le poids, la taille), les données sociodémographiques, juridiques, financières et foncières. Les secondes sont les données qui concernent les préférences, centres d'intérêts, les opinions et activités politiques, religieuses et syndicales, le comportement, les relations, les déplacements... Elles

¹ Article 4, Règlement 2016/679 du Parlement européen et du Conseil *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (règlement général sur la protection des données)*.

² Le compte Gmail est relié au compte YouTube.

³ <http://www.lefigaro.fr/secteur/high-tech/2018/07/03/32001-20180703ARTFIG00135-gmail-permet-a-des-developpeurs-de-lire-les-e-mails-de-ses-utilisateurs.php>

permettent de donner une image des habitudes des individus, de leur style de vie, et donc de créer des profils qui peuvent être très orientés.

Enfin, pour compléter ces différentes sources, j'aimerais m'appuyer sur un article du Journal Du Net qui, utilisant l'article 4 du RGPD¹, donne une définition claire et simple de la donnée personnelle. C'est une donnée qui « *peut être fournie volontairement par la personne mais également être générée par la simple utilisation d'un service. C'est notamment le cas des clics sur une page web ou encore d'une donnée de localisation. Ces informations collectées sont qualifiables de données personnelles à partir du moment où, associées à d'autres données telles qu'une adresse IP, elles peuvent identifier une personne. Au quotidien, l'utilisation des logiciels communicants d'une voiture ou encore son numéro de série, peut rattacher la voiture à une personne physique spécifique* »². La donnée est personnelle dès qu'elle peut être individualisée, dès qu'elle se rapporte à une personne précise, et qu'elle permet de l'identifier, d'identifier un comportement.

On voit bien que les données personnelles ne se résument pas qu'à un simple numéro de téléphone, ou une photo de l'individu. Elles sont plus vastes, et, recoupées entre elles, permettent de mieux cerner des individus, ou groupes d'individus. Il devient intéressant pour différents acteurs de récupérer ces données, de les croiser afin d'avoir une vision plus précise sur les utilisateurs (potentiels ou effectifs) de certains services. J'expliquerai en deuxième partie quelles sont les volontés et les motivations derrière la collecte de ces données. Je vais tout d'abord m'atteler à montrer quels sont les procédés de récolte de ces informations.

b. Quelles techniques de collecte des données à caractère personnel ?

Avant d'étudier ces techniques, il est intéressant d'avoir en tête l'encadrement de la collecte des données par la loi 78-17 du 6 janvier 1978, dite « Internet et Liberté ».

¹ Article 4, Règlement 2016/679 du Parlement Européen et du Conseil *relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données (règlement général sur la protection des données)*.

² <https://www.journaldunet.com/ebusiness/le-net/1206397-definition-donnee-personnelle/>

L'article 2 de la loi 78-17 du 6 janvier 1978¹ définit le traitement (automatisé ou non) des données personnelles : « *Constitue un traitement de données à caractère personnel toute opération ou tout ensemble d'opérations portant sur de telles données, quel que soit le procédé utilisé, et notamment la collecte, l'enregistrement, l'organisation, la conservation, l'adaptation ou la modification, l'extraction, la consultation, l'utilisation, la communication par transmission, diffusion ou toute autre forme de mise à disposition, le rapprochement ou l'interconnexion, ainsi que le verrouillage, l'effacement ou la destruction* ». La collecte des données fait donc partie du traitement, et doit répondre aux attentes des articles suivants. Tout d'abord, il est nécessaire de définir un responsable du traitement. L'article 3 dit que « *Le responsable d'un traitement de données à caractère personnel est, sauf désignation expresse par les dispositions législatives ou réglementaires relatives à ce traitement, la personne, l'autorité publique, le service ou l'organisme qui détermine ses finalités et ses moyens* »². Ce Responsable du traitement est soumis à des obligations et des principes. L'article 6 précise les conditions auxquelles doivent satisfaire les données personnelles sujettes au traitement : elles doivent être « *collectées (...) de manière loyale et licite* », pour des « *finalités déterminées* ». Elles doivent être « *adéquates [et] pertinentes* » par rapport aux finalités pour lesquelles elles sont collectées, mais aussi « *exactes, complètes et [...] mises à jour* ». Elles ne peuvent être conservées pendant une durée excédant « *la durée nécessaire aux finalités pour lesquelles elles sont collectées* »³. Ainsi, en France, l'utilisation des données personnelles ne peut, dans la loi, se faire qu'à condition qu'une finalité la motive, et qu'elles ne soient conservées que pendant un temps fixé correspondant à cette finalité.

Je ne vais pas chercher à montrer si cela est respecté par les acteurs qui récoltent des données. Je souhaitais simplement rappeler en premier lieu les fondements juridiques sur lesquels s'appuient, en théorie, les méthodes de collectes de données. Maintenant que cela est fait, je vais présenter ces méthodes.

Dans leur article « *Entreprise et respect de la vie privée du consommateur. De l'usage autorisé à l'utilisation souhaitable des données personnelles* », Régis Dumoulin et Caroline Lancelot Miltgen expliquent que la manière dont les données personnelles sont collectées a beaucoup évolué depuis les années 1990 (Dumoulin, Lancelot Miltgen, 2012).

¹ Article 2, Loi n° 78-17 du 6 janvier 1978, *relative à l'informatique, aux fichiers et aux libertés*.

² Article 3, Loi n° 78-17 du 6 janvier 1978, *relative à l'informatique, aux fichiers et aux libertés*.

³ Article 6, Loi n° 78-17 du 6 janvier 1978, *relative à l'informatique, aux fichiers et aux libertés*.

Ils soulignent plusieurs choses. Dans un premier temps, si, avant, la collecte par les entreprises était ciblée sur des informations liées aux consommateurs d'un marché précis, aujourd'hui, il est plus intéressant pour elles de récolter des données qui permettent de dresser un profil non pas du consommateur mais de l'individu (habitudes, préférences etc) (*Ibid.*, 2012). Dans un second temps, là où on avait une récolte d'information directe, par téléphone, ou par courrier, l'arrivée des Nouvelles Technologies de l'Information et de la Communication (NTIC)¹ et l'évolution des pratiques des utilisateurs ont permis à la collecte de données de s'étendre et de l'intensifier.

Ces techniques de collecte sont présentées par Fabrice Rochelandet dans son ouvrage cité plus haut (Rochelandet, 2010). Comme pour la différence entre les données subjectives et objectives, il présente les méthodes de collecte des données sous la forme d'un tableau². On retrouve également des précisions sur ces méthodes de « *tracking* » dans un article qu'il a co-écrit en 2015 : « Les modèles d'affaire numériques sont-ils trop indiscrets ? » (Cecere, Le Guel, Rochelandet, 2015, p 80). Je vais m'appuyer sur ces deux travaux pour présenter les techniques de collecte de données personnelles.

On l'a vu plus haut, à la base, les méthodes de collecte d'information avaient pour but d'offrir un service personnalisé et de meilleure qualité à l'internaute. Petit à petit ces techniques dites de « *tracking* » deviennent de plus en plus intrusives. Elles sont en évolution constante, une sorte de « *jeu du chat et de la souris* » (*Ibid.*, 2015, p 81). Les limites techniques de certaines méthodes de « *tracking* » permettent aux internautes de la contrer, obligeant un renouvellement et une complexification des méthodes de collecte.

En m'appuyant sur les différents travaux cités ci-dessus, j'en suis arrivée à dégager trois types de collecte des données personnelles, s'appuyant sur des techniques différentes.

¹ Cette notion fait référence aux techniques permettant le traitement et la transmission des informations. On parlait de TIC, aujourd'hui, du fait des évolutions technologiques liées au numérique on parle des NTIC.

² Voir en Annexe 1.

Les techniques s'apparentant au « *face-à-face* » (Dumoulin, Lancelot Miltgen, 2012).

Ce sont les données issues de formulaires remplis sur internet comme des formulaires d'inscription à des services en ligne (newsletter, réseau social, site de vente en ligne etc), les données liées aux transactions au moment d'un achat en ligne (adresse, données bancaires etc) (Rochelandet, 2010)¹.

Je reprends l'idée d'une collecte "*face-à-face*", prépondérante avant les années 1990, faite alors par téléphone (Dumoulin, Lancelot Miltgen, 2012), afin de qualifier cette méthode de récupération de données objectives qui passe aujourd'hui par des formulaires ou des transactions. L'individu transmet les données qui lui sont explicitement demandées afin de pouvoir accéder à un service. C'est la méthode la plus traditionnelle de collecte d'informations personnelles, mais qui ne permet pas forcément d'obtenir des données très précises et très intéressantes pour les entreprises.

Les traces collectées à l'insu de la personne.

Elles sont à la fois les plus problématiques du côté des utilisateurs, et les plus intéressantes du côté des entreprises. Elles évoluent avec la technologie et s'adaptent aux usages. Ainsi, on va retrouver par exemple les données de géolocalisation des utilisateurs de smartphones. On a également les données enregistrées associées à un identifiant, qui sont en fait les données de navigation. C'est par exemple le cas du compte Gmail qui centralise les données de navigation sur Google ou sur YouTube, ou même de Facebook Connect. C'est une fonctionnalité qui permet à un site internet, qui délivre un service sous condition d'inscription, de donner la possibilité aux utilisateurs de se connecter en utilisant leur compte Facebook. C'est par exemple le cas de Spotify. Cela permet à Facebook de récupérer les listes de musiques écoutées chaque semaine par les utilisateurs.

On va ensuite retrouver toutes les techniques qu'on pourrait qualifier de « cachées », dont l'objectif est de récolter les traces laissées par l'utilisateur lors de sa navigation. Cecere, Le Guel et Rochelandet (2015) appellent ces techniques les « *traceurs* ». Ils décrivent ces traceurs en annexe de leur article.

¹ Voir en Annexe 1.

Parmi ces traceurs on retrouve d'abord les cookies, des « *fichiers stockés sur l'ordinateur de l'internaute lorsqu'il consulte des sites* ». Ils sont controversés, car ils peuvent contenir des données personnelles que les individus transmettent sans le vouloir ni même savoir qui les récupère. Plusieurs types de cookies existent. Les cookies classiques correspondent à des fichiers qui peuvent être supprimés facilement de l'ordinateur de l'utilisateur, ou qui s'effacent automatiquement au bout d'un certain temps. Ils peuvent venir du site web visité, ou bien d'un site tiers, potentiellement inconnu de l'internaute. Mais dans les deux cas, les navigateurs permettent d'effacer ces cookies. Il existe cependant des evercookies, qui sont une forme de cookies quasiment impossible à effacer. Le document rendu public en 2013 par Edward Snowden¹ montrait notamment que la NSA utilise ces evercookies pour tracer les utilisateurs de TOR, le réseau informatique superposé mondial et décentralisé².

Dans le même type de technologie que les cookies, on retrouve le DOM Storage, qui est un genre de « *cookie en mieux* »³. Il permet globalement de stocker une plus grande quantité d'informations et de manière plus sécurisée que les cookies.

L'article cite également le « *Web bug* », en français, le pixel espion⁴, est une petite image numérique, d'un pixel généralement, utilisée par des sites internet. Ils permettent de collecter des données sur l'activité des internautes.

L'adresse IP est également une manière de récolter des informations de navigation sur un internaute, car elle permet « *d'identifier de façon unique un ordinateur sur le réseau* ». Cette technique pourrait être détrônée par le fait qu'avec le développement de l'IoT (Internet of Things⁵), désormais, chaque appareil mobile doit avoir un identifiant unique et permanent.

Les données issues de l'exposition de soi.

Les informations partagées sur les réseaux sociaux (photos, vidéos, *likes*, commentaires, informations sur les goûts, sur la relation amoureuse etc.) sont récupérées par la plateforme, et peuvent ensuite être réutilisées¹.

¹ Dans le cadre du scandale des écoutes de la NSA.

² <https://en.wikipedia.org/wiki/Evercookie> et [https://en.wikipedia.org/wiki/Tor_\(anonymity_network\)](https://en.wikipedia.org/wiki/Tor_(anonymity_network))

³ <https://openclassrooms.com/fr/courses/245710-ajax-et-lechange-de-donnees-en-javascript/245519-dom-storage-et-persistence>

⁴ https://fr.wikipedia.org/wiki/Pixel_espion

⁵ En français : l'internet des objets. Cette expression fait référence aux objets connectés qui produisent des données quotidiennement.

Ce point soulève une question intéressante. Quand on parle de données à caractère personnel, on pense rapidement “données privées”. On peut se dire qu’en effet, nos navigations sur internet, notre adresse postale, ou nos coordonnées bancaires sont des informations qui nous concernent personnellement et que l’on aimerait garder, dans une certaine mesure, privées. Les informations partagées sur les réseaux sociaux ont cependant cette particularité d’être très personnelles, mais en même temps, ouvertes à une large audience (si ce n’est pour dire, publiques). J’aimerais approfondir cette dimension, ce paradoxe qui fait que les données les plus personnelles ne sont pas forcément privées, dans le point suivant.

c. Les données personnelles sont-elles privées ?

Pour terminer cet axe de questionnement autour des données personnelles, j’aimerais questionner la dimension « privée » de ces données. Si, dans l’imaginaire commun, elles relèvent plutôt de la sphère privée, dans les faits, il en est autrement.

Les réseaux sociaux incitent les individus à partager du contenu personnel, voire privé. Tout d’abord, on peut voir que l’utilisation massive, croissante de plateformes comme Facebook engendre des effets de réseaux importants². Être inscrit et actif sur Facebook devient aujourd’hui indispensable. « Un monde de verre : Facebook ou les paradoxes de la vie privée (sur)exposée » est un article de Yannick Estienne qui, basé sur des entretiens d’utilisateurs de Facebook, montre en quoi il est trop discriminant de ne pas être actif sur le réseau social. « *Les discours des usagers font (...) état d’une forme de décrochage relationnel et de discrimination passive des non-usagers de Facebook* » (Estienne, 2011). Les individus qui décident de ne pas s’inscrire ne reçoivent pas les invitations aux événements, ne sont pas tenus au courant de ce qu’il se passe etc. Ainsi donc, les réseaux sociaux massifs comme Facebook attirent un maximum d’utilisateurs, et les incitent à partager des informations personnelles.

¹ <https://fr-fr.facebook.com/about/privacy/>

² L’effet de réseau est le fait que l’utilité d’un bien dépend du nombre de personnes qui l’utilisent. L’exemple classique est celui du téléphone, mais ça fonctionne aussi avec Facebook : plus il y a de personnes qui l’utilisent, plus il est intéressant de l’utiliser aussi.

La frontière se brouille entre public et privé (Dumoulin, Lancelot Miltgen, 2015). Les informations partagées sont dès lors accessibles aux autres utilisateurs (en fonction des réglages de confidentialité) mais aussi au réseau social lui-même, qui récolte toutes les traces laissées par ses utilisateurs. L'utilisation massive des réseaux sociaux, l'hyperconnexion des individus font que des données très personnelles, relevant même de l'intime sont divulguées, mises à disposition d'un large public. On peut qualifier cela de « paradoxe de la vie privée », et le définir ainsi : « *la contradiction entre, d'une part, une méfiance et une inquiétude bien réelle face à des menaces également bien réelles en matière de protection de sa vie privée, et, d'autre part, un désir très fort de s'exposer en livrant volontairement des informations sur soi susceptibles de porter atteinte à sa vie privée* » (Estienne, 2011). Les internautes sont inquiets face aux scandales liés à l'utilisation des données personnelles, mais décident néanmoins de continuer à s'afficher publiquement sur internet. Ils continuent à poster des informations, parfois très intimes, en ayant « explicitement » accepté de les partager avec la plateforme sur laquelle elles sont postées.

Dans la loi 78-17 du 6 janvier 1978, ainsi que dans le RGPD, les données personnelles ne sont pas considérées comme étant la propriété des individus. Elles sont un droit attaché à la personne (Isaac, 2018). La collecte de ces données est donc autorisée, si l'individu concerné l'accepte « explicitement ». En acceptant les conditions générales d'utilisation de Facebook, l'internaute autorise la plateforme à récolter ses données, et à les utiliser de différentes manières, ces manières étant exposées explicitement par la plateforme¹. Le texte exposant ces utilisations et les contextes de collecte des données est très long, et n'est probablement pas lu la plupart du temps, lorsqu'un nouvel individu s'inscrit.

Néanmoins, il me semble indispensable de prendre connaissance de ce que Facebook s'autorise à faire avec nos données personnelles. Dans sa politique d'utilisation des données, on peut par exemple découvrir le fait que Facebook ne récolte pas uniquement nos données de navigation sur sa plateforme, mais également nos données de géolocalisation, observe l'utilisation que l'on fait de notre smartphone (batterie, application en arrière-plan etc), accède à notre bibliothèque de photographie, récupère nos

¹ <https://fr-fr.facebook.com/about/privacy/>

conversations sur Messenger etc. Il s'autorise à les partager à des entreprises partenaires, dont les noms, les activités et les usages de ces données ne sont pas précisés.

L'objectif ici n'est pas de dénoncer la politique de Facebook, mais bien de mettre en évidence ce « paradoxe de la vie privée », ainsi que de mettre à mal l'idée assez répandue qui sous-entend que les données personnelles sont privées. Facebook donne accès à tous ses utilisateurs à la manière dont il va collecter et utiliser les données. En cliquant sur « j'accepte les conditions générales d'utilisation » lors de l'inscription, l'internaute autorise explicitement Facebook à faire tout ce qui est cité dans ces textes.

Des données les plus intimes, aux données objectives, les individus partagent tout et donnent leur consentement explicite aux réseaux sociaux pour que ceux-ci les réutilisent. On ne peut donc définitivement pas dire que les données personnelles sont privées.

d. Pour conclure sur les données personnelles ...

Les données à caractère personnel ne sont pas seulement les informations se référant directement à un individu, mais également les traces qu'il laisse malgré lui, par exemple dans le cadre de l'utilisation des réseaux sociaux. Recoupées avec d'autres informations, elles permettent d'obtenir un profil précis de l'individu, de ses pratiques, habitudes, préférences.

Les données personnelles peuvent être collectées de différentes manières, certaines plus obscures que d'autres. La collecte peut passer par des formulaires que l'internaute doit remplir afin de s'inscrire à une plateforme par exemple. Les « *traceurs* » (Cecere, Le Guel, Rochelandet, 2015) permettent de récolter des informations sur les pratiques de navigations des individus, sans que ceux-ci ne soient au courant. Les informations personnelles publiées sur les réseaux sociaux sont également récoltées.

Si on s'imagine souvent que les données personnelles vont de pair avec la sphère privée, on se rend compte que les pratiques des internautes et les politiques concernant la collecte et l'utilisation des données des plateformes faussent cette idée. Les données à caractère personnel ne sont pas privées, elles peuvent être ré-utilisées à partir du moment où l'individu donne son consentement explicite, ce qu'il fait dans la majeure partie des cas sans forcément le savoir.

2. PRESENTATION DU MOUVEMENT D'OPEN DATA ET DES DONNEES SUJETTES A L'OUVERTURE.

Si le mouvement d'ouverture des données commence à vraiment faire parler de lui à la fin des années 2000, il est déjà question d'Open Data dans les années 1990. Cette volonté d'ouverture apparaît dans le monde scientifique. En 1995, le Comité sur les Données géophysiques et environnementales du Conseil national de la Recherche aux Etats-Unis publie un document : *De l'échange complet et ouvert des données scientifiques*¹. Il souligne l'importance d'une recherche coopérative, de dimension internationale, sur les changements environnementaux (atmosphère, biosphère, océans), afin de pouvoir croiser les informations entre chercheurs (Peugeot, 2010).

Concrètement, la création de Wikipédia en 2001 est un exemple d'ouverture de l'information. La plateforme permet aux individus de mettre en ligne leur savoir, d'alimenter et d'améliorer certains sujets. Le contenu est mis à disposition gratuitement à tous les internautes qui peuvent le lire, le copier, le réutiliser.

L'idée d'ouvrir les données se répand petit à petit dans d'autres secteurs, notamment le secteur public dès les années 2000. Un mouvement de libération des données est lancé dans plusieurs pays. Les États deviennent de véritables « *data providers* », (Van Cuyck, 2013) produisant un très grand nombre de données publiques. En 2009 les États-Unis lancent leurs plateforme data.gov, suivis un an plus tard par le Royaume-Uni. Ce sont les premiers sites qui regroupent et donnent accès gratuitement à des données gouvernementales. En 2013, ce sont déjà quarante-deux pays, dont la France, et quatre institutions internationales (Union Européenne, OCDE, Nations Unies, Banque Mondiale) qui ouvrent leurs données (Van Cuyck, 2013).

Mais qu'est-ce que l'Open Data concrètement ? Sur le site internet gouvernement.fr on peut lire la définition suivante : « *L'open data désigne l'effort que font les institutions, notamment gouvernementales, qui partagent les données dont elles disposent. Ce partage doit être gratuit, dans des formats ouverts, et permettre la réutilisation des données. La politique du Gouvernement s'appuie notamment sur le droit d'accès aux documents administratifs (loi Cada 1978), qui considère que les données produites ou détenues par les administrations, dans le cadre de leurs missions de service public, doivent être mises à*

¹ *On the Full and Open Exchange of Scientific Data* (A publication of the Committee on Geophysical and Environmental Data - National Research Council).

disposition du public. Cela ne concerne ni les informations personnelles, ni celles touchant à la sécurité nationale, ni celles couvertes par les différents secrets légaux »¹. L'Open Data fait donc référence à la fois au mouvement d'ouverture des données, mais aussi aux données ouvertes.

Pour approfondir cette idée, je vais aborder le sujet de l'Open Data en répondant à trois questions. Pourquoi ouvrir ses données ? Qui ouvre ses données ? Comment ouvrir ses données ? Pour y répondre je m'appuierai sur plusieurs travaux.

Je vais utiliser l'article d'Alain Van Cuyck (2013), « L'Open Data comme nouvelle forme de gouvernance numérique : enjeux, marchés, modèles, idéologies » (Van Cuyck, 2013), qui présente le mouvement d'ouverture des données en France et dans le monde, tout en parlant des enjeux économiques et politiques.

Celui de Valérie Peugeot (2010), « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique » est également très intéressant puisqu'il parle de l'ouverture des données dans trois secteurs bien distincts : les secteurs scientifique, public et privé, ce qui permet de voir comment les différents acteurs s'insèrent dans le mouvement.

L'article « La fabrique des données brutes. Le travail en coulisses de l'Open Data » de Jérôme Denis et Samuel Goëta (2013) vient remettre en cause l'idée d'une libération automatique de données brutes et non structurées. Les auteurs montrent, grâce à une série d'entretiens ethnographiques menés dans plusieurs administrations, que l'ouverture des données répond à un processus complexe et assez lourd, nécessitant l'intervention d'équipe spécialisées en plus des individus qui produisent ces données.

a. Pourquoi ouvrir ses données ?

L'ouverture des données des administrations et collectivités s'inscrit dans le principe fondamental de transparence démocratique. Ce principe s'appuie sur l'idée qu'une démocratie épanouie repose sur la circulation sans entrave de l'information, devenant accessible à tous. En France, ce principe se manifeste à travers la loi 78-753 du 17 juillet 1978, dont l'article 10 stipulait (avant d'être abrogé en 2016) : « *Les informations figurant dans des documents produits ou reçus par les administrations [...], quel que soit le*

¹ <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>

support, peuvent être utilisées par toute personne qui le souhaite à d'autres fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus ». C'est donc l'article qui pose la liberté d'accès de chacun aux documents administratifs.

Les initiatives étatiques pour l'ouverture des données numériques apparaissent à la fin des années 2000. En 2009, est mise en place l'« *Open Government Initiative* », aux États-Unis, porté par Barack Obama. Ce projet est basé sur la transparence du gouvernement, la collaboration des citoyens, l'ouverture des données et des informations, pour plus d'efficacité dans l'action gouvernementale (Van Cuyck, 2013). C'est dans ce contexte qu'a été lancée la première plateforme d'Open Data étatique data.gov, en mai 2009, aux États-Unis, incitant les autres pays à en faire de même.

Dès 2011, la France s'intéresse à l'ouverture de ses données numériques. La mission Etalab est créée, elle lance la politique d'ouverture et de partage des données publiques¹. Elle coordonne et appuie la diffusion et la réutilisation des informations publiques des administrations de l'État. Elle développe et anime (depuis 2011) la plateforme data.gouv.fr, qui met à dispositions les informations en question. Elle promeut l'ouverture des données et les Data Sciences². Elle donne la possibilité aux services publics de publier les données, et aux individus de les modifier, de les enrichir. L'objectif est de pouvoir « *coproduire des informations d'intérêt général* »³

La France s'est également dotée d'une nouvelle loi en octobre 2016 : la Loi pour une République Numérique⁴. Elle « *visé à favoriser l'ouverture et la circulation des données et du savoir, à garantir un environnement numérique ouvert et respectueux de la vie privée des internautes et à faciliter l'accès des citoyens au numérique* »⁵. Elle prévoit, entre autres, l'ouverture « *par défaut* » des données publiques et d'intérêt général. Elle oblige les organismes publics à mettre à disposition gratuitement leurs informations sur internet, en passant par une phase d'anonymisation et de protection, les rendant réutilisables. On passe d'une incitation tant nationale qu'internationale, à une obligation légale d'ouverture.

¹ <https://www.etalab.gouv.fr/qui-sommes-nous>

² Wikipédia donne la définition des Data Sciences suivante : « *extraction de connaissance d'ensembles de données. Elle emploie des techniques et des théories tirées de plusieurs autres domaines plus larges des mathématiques* ». https://fr.wikipedia.org/wiki/Science_des_donn%C3%A9es

³ <https://www.gouvernement.fr/action/l-ouverture-des-donnees-publiques>

⁴ Loi n°2016-1321 du 7 octobre 2016 pour une République numérique.

⁵ <https://www.gouvernement.fr/action/pour-une-republique-numerique>

Mais quelles sont les motivations derrière ces mises en place ? Elles sont de deux sortes : démocratiques et économiques.

Comme on l'a vu plus haut, le mouvement d'ouverture des données va de pair avec le principe de transparence démocratique. Une des premières motivations de l'ouverture des données pour les États, c'est bien cette idée de circulation de l'information, d'*empowerment*. En théorie, le fait de donner accès aux données de l'administration permettrait des actions de militantisme mieux documentées et argumentées (Goëta, Mabi, 2014). Cela permettrait en effet de réduire l'asymétrie d'information entre les citoyens et les administrations, ce qui va en effet dans le sens d'une plus grande prise de pouvoir des individus.

Cet argument démocratique fonctionne en théorie. Dans les faits, on voit bien qu'il en va différemment. Dans un premier temps, comme je l'ai montré dans l'introduction de cette sous-partie, les données sélectionnées et ouvertes sont inoffensives. Les informations sensibles ne sont pas diffusées. De plus, tous les citoyens ne sont pas capables de se saisir et d'utiliser ces données (Goëta, Mabi, 2014).

Bien que cet *empowerment* ne soit que théorique, la promotion de la transparence démocratique me semble une raison suffisante pour motiver l'ouverture de ses données.

La deuxième motivation est économique. Les données sont considérées comme le « *nouvel or noir* », et ce n'est pas parce que certaines sont mises à disposition librement et gratuitement que les retombées économiques sont inexistantes. Sans entrer dans les détails pour l'instant¹, on peut dire qu'en mettant à la disposition des entreprises des jeux de données, les différents acteurs permettent l'innovation, à travers de nouveaux services, des applications qui doivent simplifier des actions quotidiennes. L'application Citymapper s'appuie par exemple sur des jeux de données fournis par les compagnies de transports de certaines grandes villes en France, comme à Lyon par exemple. De plus l'Open Data génère de l'activité et est aussi créateur d'emploi (Peugeot, 2010).

En 2015, Capgemini Consulting, l'entité de conseil en stratégie et transformation du groupe Capgemini, publie deux rapports sur l'état des lieux de l'Open Data en Europe. Le deuxième, *Creating Value through Open Data : Study on the Impact of Re-use of Public Data Resources* traite de la création de valeur à travers l'Open Data. D'après ce rapport,

¹ La valorisation de l'Open Data sera détaillée en III.

le marché européen des données ouvertes serait estimé à 325 milliards d'euros sur la période 2016-2020. On y apprend également qu'avec une utilisation efficace de ces données ouvertes, il serait possible d'économiser six-cent-vingt-neuf millions d'heures d'attente sur les routes de l'Union Européenne, ainsi que de réduire la consommation d'énergie de 16% (ce qui n'est pas encore le cas, du fait d'une utilisation des données ouvertes qui n'est pas totalement efficace). Les retombées économiques positives de l'Open Data semblent donc bien réelles.

b. Qui ouvre ses données ?

Les premiers acteurs concernés par l'ouverture des données sont donc les administrations publiques, étatiques et les collectivités territoriales. Depuis 2016, dans un souci de transparence démocratique, il devient obligatoire pour ces acteurs d'ouvrir certains jeux de données. En plus de la plateforme data.gouv.fr, on peut voir apparaître d'autres initiatives de la part des collectivités territoriales par exemple. Ainsi la métropole de Rennes rend accessible des jeux de données sur son site data.rennesmetropole.fr¹, par exemple.

Certaines entreprises aussi peuvent ouvrir leurs données, comme celles chargées de mission de service public. C'est le cas par exemple de la RATP, ou de Enedis, qui ont mis en place des plateformes pour partager librement une partie de leurs données². On peut retrouver sur la plateforme de la RATP des jeux de données concernant la qualité de l'air dans les différentes stations, les horaires en temps réel, etc. Sur celle d'Enedis, on retrouve des données plus techniques concernant le raccordement au réseau d'électricité, la consommation d'énergie etc.

Les entreprises privées sautent difficilement le pas, néanmoins on peut voir quelques initiatives se mettre en place. Sur son site web, Airbus donne accès à un grand nombre d'informations concernant les produits qu'il fabrique (avions, hélicoptères ...), ainsi qu'à des documents s'adressant aux fournisseurs, détaillant les différents *process*, les codes de bonne conduite etc.³ Cela reste néanmoins assez limité.

¹<https://data.rennesmetropole.fr/page/home/>

²<https://data.ratp.fr/explore/?sort=modified> pour la plateforme de la RATP, <https://data.enedis.fr/explore/?sort=modified>, pour la plateforme de Enedis.

³<https://www.airbus.com/>

Les données partagées ne sont pas des données sensibles ou des données personnelles. Alexandre Léchenet, journaliste au *Monde*, confirme cette idée : « *aujourd'hui les données publiées sont souvent les plus inoffensives* » (Goëta, Mabi, 2014). Les collectivités peuvent poster des données concernant la gestion de l'eau, des déchets, ainsi que des informations concernant des événements, etc. Les entreprises chargées de mission de service public, on l'a dit, ouvrent des données à propos de leur secteur d'activité, qui ne présentent pas de risque. Jean-François Montagne, responsable de la gouvernance des données chez Enedis, lors d'un entretien, insiste sur cette idée : « *on a concentré nos efforts sur de la transparence des données qu'on peut afficher, qui ne font pas de distorsion de concurrence sur le secteur de l'énergie (informations commercialement sensibles), qui ne sont pas des données à caractère perso, et qui ne mettent pas en péril la sûreté du réseau* ».

c. Comment ouvrir ses données ?

Pour répondre à cette question je vais m'appuyer sur l'article « La fabrique des données brutes. Le travail en coulisses de l'Open Data » (Denis, Goëta, 2013), qui apporte un discours très intéressant et assez unique par rapport aux autres travaux que j'avais pu lire avant. Alors que l'ouverture des données est présentée en général comme un processus naturel, cette idée est remise en cause. L'article montre les coulisses de la mise en ligne de jeux de données brutes, qui relève finalement d'une mise en œuvre beaucoup plus importante qu'on ne nous le laisse penser.

D'après la définition que l'on peut lire sur Wikipédia, les données brutes sont « *les données non interprétées émanant d'une source primaire, ayant des caractéristiques liées à celle-ci et qui n'ont été soumises à aucun traitement ou toute autre manipulation* »¹. C'est ce genre de données dont il est question dans le contexte de l'Open Data. Il faut mettre à disposition des données non biaisées, non traitées, afin qu'elles puissent être réutilisées par le plus grand nombre. Trouver, sélectionner et partager des données brutes résulte d'un processus plus complexe qu'on ne le pense.

En réalité, d'après les entretiens ethnographiques qu'ils ont réalisés, Jérôme Denis et Samuel Goëta soulignent trois étapes du processus d'ouverture des données : « *l'identification, l'extraction et la "brutification"* » (Denis, Goëta, 2013). J'en ajouterai

une dernière à la fin qui correspond aux formes choisies pour rendre accessibles les données et pour leur réutilisation.

La phase d'identification des données à ouvrir.

La phase d'identification est la première. Elle est indispensable. En effet, les jeux de données à ouvrir n'existent pas *a priori*. Ils n'attendent pas, sagement, leur ouverture, déjà prêts, dans une base de données bien délimitée. « *Avant la question "comment allons-nous ouvrir tel ou tel jeu de données ?"*, de nombreux acteurs se confrontent à une interrogation plus abyssale encore : *"de quelles données disposons-nous ?"* » (Denis, Goëta, 2013). Cette première phase doit donc permettre de découvrir l'existence, la nature et l'emplacement des données.

Ce ne sont pas les producteurs de ces données qui réalisent ce travail, mais des équipes externes aux services en question. Ce sont par exemple des chargés de projet Open Data qui vont à la rencontre des individus des différents services d'une entreprise, d'une administration ou d'une collectivité. Ces individus sont alors sollicités afin de venir en aide à l'équipe Open Data. Ils doivent permettre une bonne compréhension des types de données qui peuvent être trouvées.

L'identification et l'exploration se font progressivement. Il doit y avoir des échanges réguliers avec les services concernés, permettant l'avancement. L'équipe Open Data doit, à la fin de cette phase exploratoire, déboucher sur une sorte d'inventaire qui répond à des logiques particulières. Cet inventaire peut être orienté de différentes façons, de sorte que les données retenues ne seront pas les mêmes à chaque fois. Parmi ces logiques on retrouve par exemple la volonté de s'aligner avec les pratiques existantes dans d'autres entreprises, administrations ou collectivités, ou bien alors, au contraire, la volonté de se distinguer en sélectionnant un jeu de données inédit. Surtout, un des points importants pour faire cet inventaire est de connaître le degré de sensibilité des données, pour savoir lesquelles peuvent être ouvertes. Les discussions entre les services internes et l'équipe Open Data sont donc indispensables dans cette première phase (Denis, Goëta, 2013).

¹ https://fr.wikipedia.org/wiki/Donn%C3%A9es_brutes

La phase d'extraction des données sélectionnées.

Une fois la phase exploratoire terminée, les données ne sont pas pour autant disponibles et prêtes à être ouverte. Enfermées dans leurs bases de données, elles doivent en être extraites. C'est la deuxième phase : l'extraction.

Le fait d'accéder aux données dépend en général des logiciels qui permettent de les lire et de les ordonner. Ces logiciels ne sont pas unifiés au sein de la structure. En général, il existe plusieurs logiciels, voire plusieurs versions de ces logiciels au sein des différents services. Les équipes d'Open Data ont alors besoin d'autant d'outils sur-mesure d'extraction de données qu'il existe de types et de versions de logiciels. D'après les entretiens menés par Jérôme Denis et Samuel Goëta, cette phase nécessite donc un grand nombre d'instruments, et se révèle être très coûteuse.

La donnée est souvent présentée comme le nouvel or noir, en anglais, on parle de « *commodity* ». Cela correspond à l'idée que la donnée, à l'état brut, dort dans sa base, et qu'elle n'attend qu'une chose : être libérée pour être exploitée, ré-utilisée. Or on voit bien que ces deux premières phases coûteuses, qui nécessitent de nombreux outils, remettent en cause cette vision de l'ouverture des données. Elles ne sont pas directement accessibles, ou « prêtes à l'emploi ». Tout un processus, complexe et onéreux doit être mis en œuvre afin de pouvoir les sélectionner puis les extraire. Une fois extraites, les données ne sont néanmoins pas encore prêtes à être libérées (Denis, Goëta, 2013).

La phase de « brutification » des données extraites.

Cette troisième partie vient contredire l'idée que l'on se fait de la donnée brute, ainsi que la définition que j'ai pu en donner plus haut. Les données brutes, dans le contexte de l'Open Data doivent en fait correspondre à des standards spécifiques. Il est donc très courant que ces données soient traitées afin de les « rendre brutes », c'est ce que l'article présente comme la « *brutification* ».

Ce processus se déroule selon trois types d'interventions. Tout d'abord les données sont soumises à un reformatage. Extraites, elles n'ont pas forcément toutes le même format, et ne sont donc pas forcément toutes lisibles par les outils des développeurs ou le grand public. Il est indispensable de résoudre ce problème en mettant les données sous un format classique, assurant leur lisibilité par le plus grand nombre. A travers leurs

entretiens, Jérôme Denis et Samuel Goëta ont pu constater que le format CSV était souvent considéré comme la référence.

Les données sont également soumises à un nettoyage. Il s'agit là de corriger les erreurs qui peuvent être présentes dans les jeux de données : les données aberrantes (trop ou trop peu élevées), les données manquantes dans les fichiers. Il faut en fait harmoniser tout ce qui a été extrait, afin d'en assurer la lisibilité et la bonne utilisation. Cette harmonisation pourrait s'apparenter à une sorte de premier pas vers l'universalité de la présentation des jeux de données.

Finalement, les données doivent également passer par une « *désindexicalisation* ». Les traces de potentiels usages précédents, de localisation doivent être effacées. Les données non intelligibles doivent être rendues compréhensibles (en remplaçant par exemple les acronymes et abréviations par les vrais termes). Un ensemble de métadonnées (dictionnaire, commentaires) sont également créées permettant la bonne réalisation du projet de partage de données, et la facilitation de leur usage (Denis, Goëta, 2013).

On voit donc, à travers cet article, que les données brutes, « prêtes à l'emploi », n'existe pas *a priori*, mais sont le résultat d'un processus long, coûteux, mobilisant de nombreux outils et individus.

Les formes que peuvent prendre les données ouvertes pour leur utilisation.

Les données mises en ligne peuvent prendre plusieurs formes. Elles peuvent simplement apparaître sur le site internet. Elles sont ainsi consultées, mais leur utilisation et leur circulation ne sont pas assurées, ni simplifiées.

Il existe également des jeux ou bases de données mises à disposition et téléchargeables. Il est alors possible pour ceux qui les récupèrent de les traiter, les réutiliser. Leur circulation est beaucoup plus libre. Néanmoins, cette option comporte certaines limites, comme notamment, la difficulté d'actualisation des informations en temps réel. Si les données évoluent rapidement, il devient très lourd, voire impossible, de les mettre à jour à temps, puis de les reposer sur la plateforme. Les informations contenues dans ces bases risquent de devenir rapidement obsolètes.

Finalement, la solution la plus adaptée, et qui se développe le plus actuellement, est l'ouverture d'une API (*Application Programming Interface*). D'après, le Journal Du Net,

qui fournit une définition simple et compréhensible (ce qui n'est pas évident), une API est « *une solution informatique qui permet à des applications de communiquer entre elles et de s'échanger mutuellement des services ou des données. Il s'agit en réalité d'un ensemble de fonctions qui facilitent, via un langage de programmation, l'accès aux services d'une application* »¹. Dans le contexte de l'Open Data, les API permettent, globalement, d'avoir accès à des jeux de données qui peuvent être mis à jour, et avec lesquels il est possible d'interagir simplement (en tant que développeur).

d. Pour conclure sur les données d'Open Data ...

Ainsi, les motivations derrière l'ouverture des données sont, à la base, politiques et idéologiques. L'Open Data suit le principe de transparence démocratique. Depuis 2016, une loi oblige l'administration à ouvrir une partie de leurs données. Les motivations sont également économiques, l'ouverture des données permettrait de créer de la valeur, aurait des retombées économiques.

Ainsi, les administrations, collectivités et certaines entreprises sont incitées à ouvrir certains jeux de données. Elles ne doivent pas être trop sensibles, car leur ouverture pourrait présenter des risques. Ce sont donc des données « *inoffensives* »², qui peuvent être réutilisées par des individus sans problème.

L'ouverture de ses données, contrairement à l'idée qu'on peut s'en faire, est un processus assez lourd, qui nécessite une équipe et des outils nombreux et coûteux. Une fois que les données à ouvrir sont repérées, sélectionnées, extraites et « *brutifiées* » (Denis, Goëta, 2013), elles sont mises en ligne le plus souvent sous forme d'API permettant un plus grand contrôle de la part de ceux qui les mettent en ligne, mais aussi une réutilisation plus simple et complète pour ceux qui les récupèrent.

¹<https://www.journaldunet.fr/web-tech/dictionnaire-du-webmastering/1203559-api-application-programming-interface-definition-traduction/>

² Voir la citation d'Alexandre Léchenet (Denis, Goëta, 2013).

3. LES DONNEES PRIVEES D'ENTREPRISES : QUELS ENJEUX ET QUELLES UTILISATIONS ?

Les deux premiers axes présentaient les données à caractère personnel, et les données d'Open Data, toutes les deux encadrées par des dispositifs juridiques. Une troisième catégorie existe, qui englobe les autres types de données, c'est-à-dire les données privées produites par les entreprises.

Cette partie sera assez courte, du fait de l'hétérogénéité de ces données. Par élimination, on peut néanmoins dire qu'elles ne correspondent ni aux données personnelles ni aux données d'Open Data. Elles sont trop sensibles pour être diffusées au grand public, et sont donc conservées par les acteurs qui les génèrent, c'est-à-dire les entreprises privées. Elles peuvent être aussi échangées, ou partagées à d'autres acteurs ciblés. Leur nature varie en fonction du secteur d'activité.

Dans un premier temps, j'aimerais voir quelles sont les données qui sont conservées par les entreprises, les données les plus sensibles, qui peuvent présenter un risque si elles sont ouvertes. Ensuite je m'intéresserai aux données qui peuvent être partagées à des acteurs précis.

Mon objectif ici est de montrer, sans trop entrer dans les détails, qu'il existe d'autres types de données que celles dont on entend le plus souvent parler, et que ces données entraînent des problématiques particulières.

a. Les données d'entreprises sensibles et confidentielles.

La première cyberattaque visant une structure étatique remonte à 2007. Elle prend pour cible le système bancaire estonien, le paralysant pendant plusieurs semaines (D'Elia, 2014). Depuis, la cybercriminalité ne fait que se renforcer, utilise des techniques de plus en plus difficiles à contrer. Alors que les entreprises produisent et récoltent de plus en plus de données numériques, il devient indispensable de mettre en place une protection adaptée.

Toutes ces données ne sont pas forcément sensibles et ne doivent pas forcément faire l'objet d'une sécurité particulière. Une majorité des données d'entreprises est de

moindre importance et pourrait être ouverte¹. Parmi celles qui restent, la plupart d'entre elles sont des informations sensibles, qui ne peuvent être ouvertes à tous, mais qui pourrait être partagées à d'autres acteurs ciblés². Finalement, une minorité de données serait confidentielle et nécessiterait une haute protection³.

Les entreprises doivent réussir à catégoriser, trier leurs données afin de savoir lesquelles nécessitent une protection particulière. Il est alors important de décrire les différentes données que l'entreprise possède, par exemple : les informations techniques (méthodes de conception, prototypes, solutions innovantes etc.), commerciales (fichiers clients et fournisseurs, méthodes de distribution etc.), économiques et financières (contrats, assurances, montages juridiques etc.), stratégiques et organisationnelles (projets de recrutement, résultats de veilles stratégiques etc.)⁴. Parmi toutes ses données d'entreprises il faut bien sélectionner celles qui sont confidentielles et celles qui ne le sont pas afin de bien cibler la démarche de protection.

Une protection absolue des données n'existe pas, mais la solution du *Cloud Computing* est de plus en plus fiable. Sans trop entrer dans les détails, cette technique consiste en la location d'un serveur informatique à distance, en passant par l'intermédiaire d'un réseau (Internet)⁵. De cette manière, l'entreprise peut externaliser la protection de ses données vers une autre entreprise spécialisée dans ces questions. Elle propose des solutions innovantes, possède plus de moyens et de ressources pour assurer la sécurité des informations. C'est une solution à la fois économique, la première entreprise n'a pas besoin d'investir dans un système de protection en interne, et bénéfique, qui allège sa responsabilité dans la sécurisation des données confidentielles et sensibles. Les données restent, bien entendu, la propriété de l'entreprise qui fait appel à ce service, mais leur gestion est déléguée à l'entreprise qui propose ce service. Les risques liés à la maintenance, par exemple, sont ainsi diminués⁶.

A titre d'exemple, on pourrait citer l'entreprise d'aéronautique européenne, qui, depuis mars 2018, a choisi d'utiliser le service Cloud de Google (Google G Suite) pour héberger ses données. Dans un contexte de concurrence frontale avec Boeing, Airbus doit

¹ Environ 80% d'après <https://www.sekurigi.com/2015/09/la-protection-des-donnees-sensibles-dentreprise/>

² Environ 15% d'après la même source.

³ Environ 5% d'après la même source.

⁴ https://www.entreprises.gouv.fr/files/files/directions_services/information-strategique-sisse/Guide_LA_PROTECTION_DES_INFORMATIONS_SENSIBLES_DES_ENTREPRISES.pdf

⁵ https://fr.wikipedia.org/wiki/Cloud_computing

⁶ <https://www.sekurigi.com/2015/09/la-protection-des-donnees-sensibles-dentreprise/>

s'assurer de la sécurité de ses données. On sait que Google lit les mails des utilisateurs, récupèrent leurs traces de navigation, etc. ce qui n'est pas le cas pour les entreprises. Ce n'est cependant pas le cas pour les services dédiés aux entreprises, qui profitent néanmoins du côté intrusif de Google. En effet, en analysant les pratiques d'un milliard d'utilisateurs, le géant du web peut traiter la menace plus rapidement que n'importe quel autre éditeur de solution de sécurité. En souscrivant à G Suite, Airbus profite de cette compétence inédite tout en sachant que ses données ne vont pas être analysées par le géant du web¹.

b. Les données privées moins sensibles et leur utilisation.

On l'a vu, les données sensibles et confidentielles ne représentent qu'une minorité des données produites et récoltées par les entreprises. En majorités, elles possèdent des données qui peuvent soit être ouvertes au grand public (mais ce n'est pas mis en place), ou bien qui peuvent être utilisées d'une autre manière. Ces données privées peuvent être partagées, échangées ou vendues à d'autres acteurs précisément ciblés. Elles ne sont en pas soumises à une régulation particulière, et reposent sur une base contractuelle (Isaac, 2018). Les entreprises décident donc ce qu'elles veulent en faire.

Il peut être intéressant de partager ses données avec des acteurs précis. Une fois encore, sans trop entrer dans les détails, on peut évoquer le cas de la SNCF, la RATP, BlablaCar et TransDev. En 2017, ils prévoyaient la mise en commun de certaines de leurs données afin de centraliser les informations concernant les horaires et les prix des transports. Dans l'idée, cela aurait permis une plus grande efficacité et des trajets optimisés. Il n'est pas certain que ce projet se réalise, néanmoins, c'est un bon exemple pour montrer les enjeux qui animent ce genre d'initiatives. Les données ne sont pas sensibles ou confidentielles puisqu'elles vont être mises en commun sur une plateforme. Néanmoins, cette plateforme sera fermée au grand public. L'idée ici est de pouvoir contrer les géants du web comme Google qui propose déjà des services permettant d'optimiser les voyages en prenant en compte les horaires et durées des trains ou des transports en commun par exemple. Ici, les quatre entreprises de transports souhaitent garder le contrôle sur leurs données pour contrer la puissance de Google. Il y a en effet la crainte de voir le portail de recherche recouper les différents voyages possibles, reléguant les entreprises de

¹<https://www.solutions-numeriques.com/deploiement-g-suite-airbus-repond-aux-critiques-quant-a-la-securite-de-ses-donnees/>

transport en bout de chaîne¹. Il est donc plus question de business que de sécurité dans ce cas.

Il existe également des plateformes, récentes, permettant aux entreprises d'échanger, de vendre, d'acheter leurs données privées. C'est par exemple le cas de Dawex. C'est une entreprise française, lancée en 2015 par Fabrice Tocco et Laurent Lafaye. Le *Journal du Net* parle d'une « bourse » qui permet d'échanger des données d'entreprises issues notamment de l'IoT. Il s'agit en effet d'une plateforme ayant pour but de centraliser les échanges de données. Dawex répond également aux problématiques techniques, contractuelles ou financières des entreprises². « Pour aider les acheteurs à avoir une idée des données qu'ils vont acquérir, nous avons créé des outils d'échantillonnage et de data visualisation. On aide aussi à paramétrer le contrat de licence - durée, territoire, usage, capacité à sous-licencier... », explique Laurent Lafaye¹.

Je n'ai pas assez d'information sur ce type de plateforme pour savoir si elles sont le futur du « marché des données » (je n'en parlerai donc pas en deuxième partie du mémoire), mais je trouve néanmoins l'exemple très évocateur pour ce qui est de la présentation des données privées d'entreprises et de ce qui peut en être fait.

c. Pour conclure sur les données privées d'entreprises ...

Les données personnelles, ou celles ouvertes dans le cadre de l'Open Data sont celles dont on entend le plus parler dans les médias. Néanmoins, il existe un autre type de données, que l'on pourrait qualifier de données privées d'entreprises, qui ne sont ni soumises à des régulations particulières. Ces informations présentent des enjeux spécifiques.

Certaines d'entre elles sont sensibles et confidentielles. Les entreprises qui les produisent doivent alors avoir recours à des modes de stockage et de gestion sécurisés, passant de plus en plus par le *Cloud Computing*, en externalisant cette activité.

D'autres ne présentent pas de risque de sécurité, mais sont plutôt des données stratégiques d'un point de vue économique et business. Elles pourraient être partagées,

¹https://www.nouvelobs.com/rue89/rue89-nos-vies-connectees/20141211_RUE6955/mais-pourquoi-la-sncf-planque-t-elle-les-horaires-des-tgv.html et <https://www.usinenouvelle.com/article/sncf-ratp-transdev-et-blablacar-souhaitent-mettre-en-commun-leurs-donnees.N581278>

²<https://www.journaldunet.com/economie/services/1197005-dawex-wall-street-donnees-iot/>

mais les acteurs qui les possèdent préfèrent les conserver, ou les partager avec des acteurs ciblés.

Dans cette première partie, j'ai voulu donner une vision plus claire des données, de leur variété et des enjeux qui s'y rattachent. Pour mieux comprendre ce qu'est le « marché des données » il me semblait important de savoir de quelles données on parle.

Suite à cette présentation, je vais questionner la notion de marché. Depuis le début de mes recherches, je la trouve problématique. En effet, j'ai rencontré beaucoup de difficultés dans la compréhension, et la visualisation du fameux « marché des données ». Les acteurs qui y participent, son organisation etc., ne sont pas des choses évidentes à saisir.

Je vais donc, dans une seconde partie présenter ma vision du « marché des données » (sans aspirer à l'universalité) en m'appuyant sur mes recherches, sur les travaux que j'ai pu découvrir, mais aussi sur ma réflexion personnelle.

¹<https://www.usine-digitale.fr/article/le-francais-dawex-veut-creer-une-bourse-mondiale-de-la-donnee.N577983>

Partie II : « Le marché des données » : une expression problématique pour désigner l'exploitation économique des données.

J'ai décidé de m'intéresser à la notion de « marché des données », centrale aujourd'hui, car les médias en parlent de manière récurrente sans prendre le temps d'expliquer ce qui se trouve derrière. Or ce « marché des données » n'est pas évident à délimiter. Il est même difficile de savoir s'il existe vraiment, tant les travaux à son sujet en rapportent des conceptions différentes. Sans entrer dans les détails, je peux ainsi donner trois exemples de travaux traitant ce sujet très différemment.

Dans son article « Modèles économiques des données : une relation complexe entre demande et offre », Paul Belleflamme (2018) affirme qu'à ce jour, « *il n'existe pas de "marché des données" à proprement parler* ». Il existe des plateformes d'échanges marchands spécifiques à certains secteurs ou industries, et qui sont donc restreintes. Philippe Steiner et Marie Trespeuch (2014) quant à eux, consacrent un chapitre entier au « marché des données », dans leur ouvrage *Marchées contestées : quand le marché rencontre la morale*. Ils présentent les différentes formes de « mise en marché » de la donnée, qui sont en fait différentes manières de valoriser économiquement les données (personnelles uniquement), sans forcément passer par des échanges marchands. Finalement, Grazia Cecere et Matthieu Manant (2018), dans « Vie privée, valeur des données personnelles et régulation » reprennent le travail de l'économiste Alessandro Acquisti pour montrer qu'il existe trois marchés des données : « *un marché où les entreprises proposent des services aux utilisateurs en échange de leurs données personnelles, un marché où les individus paient pour se protéger, un marché des Data Brokers¹* ».

¹ Les Data Brokers sont des courtiers en données, j'explique dans le 3) de cette partie qui ils sont précisément.

Le « marché des données » est une notion problématique, sur laquelle il est difficile de se mettre d'accord. De quelles données s'agit-il ? Qui sont les vendeurs et les acheteurs ? Comment les échanges sont-ils organisés ?

A travers mes recherches, j'ai pu remarquer que, aussi bien dans les médias, que dans les travaux scientifiques ou les rapports, on parle généralement du « marché des données personnelles ». En effet, la recherche de données aujourd'hui est principalement animée par des motivations relevant du marketing, de la publicité, ou de la sécurité. Plus problématique, plus médiatisé, c'est le « marché des données personnelles » que je vais donc étudier.

Selon moi, ce qu'on appelle « marché des données personnelles » correspond en réalité à trois types d'échanges qui coexistent. Si les deux premiers échanges ne font pas de la donnée une marchandise, ils sont indispensables à la mise en place du troisième type d'échange, présenté comme une organisation d'échanges marchands. Je vais donc présenter ces trois activités qui, pour moi, correspondent en réalité à ce qu'on entend par « marché des données personnelles ».

Dans un premier temps, je vais m'intéresser à l'activité de production des données par le consommateur. Cette activité est caractérisée par un type d'échange non-marchand, où les données personnelles entrent en jeu. Je vais en montrer les caractéristiques.

Ensuite, j'étudierai le second type d'échange, qui est un échange marchand, orchestré par les plateformes numériques (dont je donnerai une définition). Les données personnelles alimentent ce marché, mais n'en représentent néanmoins pas l'enjeu principal.

Enfin, je présenterai les Data Brokers, ces acteurs de l'ombre qui récupèrent des quantités massives de données, et qui les traitent afin de les revendre. C'est dans ce troisième niveau que l'on pourrait, dans une certaine mesure, parler d'une marchandisation des données personnelles.

1. L'ACTIVITE DE PRODUCTION DES DONNEES PAR LES CONSOMMATEURS.

Avec la montée en puissance du numérique vient la datafication, c'est-à-dire la prolifération des données. Une grande partie correspond aux données dites personnelles. On l'a vu, ce sont les traces laissées par les individus lors de la navigation sur internet, ainsi que les contenus qu'ils partagent, qu'ils visionnent, les formulaires qu'ils remplissent etc. Sans le savoir, les individus produisent un grand nombre de données, qui, à l'état brut ne présentent pas forcément d'intérêt, mais qui peuvent obtenir une grande valeur si celles-ci sont récoltées et réutilisées. Ainsi, les individus consommateurs deviennent également producteur, un peu malgré eux. C'est, à mon sens, le premier niveau d'échange organisé autour des données personnelles.

Pour l'étudier, je vais m'appuyer sur plusieurs travaux. Cet échange a plusieurs caractéristiques. Il repose sur ce qu'on pourrait appeler le *digital labor*, qui sous-entend une forme d'exploitation inconsciente de l'internaute par les acteurs du numérique, afin de récupérer des données personnelles. Pour étudier cette question j'utiliserai l'ouvrage *Qu'est-ce que le Digital Labor ?* de Dominique Cardon et Antonio Casilli (2015).

Une autre caractéristique de cet échange est, qu'à ce stade, la donnée personnelle à l'état brut¹ ne peut pas être considérée comme une marchandise. Pour justifier cela, j'utiliserai l'ouvrage *Datanomics, les nouveaux business models des données* (Benyayer, Chignard, 2015), et les articles « La donnée, une marchandise comme les autres ? » (Isaac, 2018) et « L'instauration de droits de propriété sur les données personnelles : une légitimité économique contestable » (Anciaux, Farchy, Méadel, 2017).

Enfin, j'étudierai, grâce à l'exemple de la plateforme Datacoup, que cet échange pourrait tendre à devenir marchand, à travers la monétisation des droits de collecte et d'utilisation des données personnelles, rétribuant ainsi le travail des internautes producteurs de données.

¹ Ici « brut » est à comprendre dans le sens « en vrac », « pas traité », et pas comme les données « brutes » dans le cadre de l'Open Data, où il y a un traitement visant à rendre neutres et utilisables les données. Ici « brut » sous-entend « inutilisable ».

a. Le *digital labor* : quand le consommateur devient producteur.

Dans *Qu'est-ce que le digital labor ?* (Cardon, Casilli, 2015), Casilli essaye de montrer et définir ce nouveau phénomène qui fait du consommateur un producteur malgré lui. D'après lui, notre utilisation quotidienne des plateformes et outils numériques s'apparente à une forme de travail car elle est « *productrice de valeur, faisant l'objet d'un quelconque encadrement contractuel et soumise à des métriques de performance* ». En effet, ces traces que l'on laisse sont activement collectées par les acteurs du numérique du fait de leur valeur potentielle (« *productrice de valeur* »). Cet encadrement contractuel dont il parle se retrouve dans les conditions générales d'utilisation que chaque utilisateur doit accepter avant de pouvoir accéder à du contenu et des services (« *faisant l'objet d'un quelconque encadrement contractuel* »). De plus, nos performances sont évaluées via des mesures de popularité (nombre de *likes* ou de partages, nombre d'amis ou de *followers* etc.) (« *soumise à des métriques de performance* »). La consommation du service d'une plateforme ou d'un outil produit des données qui seront valorisées ensuite par la plateforme. Le *digital labor* pourrait alors être défini comme « *une contribution à faible intensité et à faible expertise mise à profit via des algorithmes et des fouilles de données* » (*Ibid.*, 2015).

Le consommateur ne se rend pas compte du travail qu'il produit. Les fondements libéraux du web, qui le montrent comme un espace neutre et participatif (bien qu'ils soient remis en cause aujourd'hui), participent, à mon avis, au fait que cette forme de travail passe si inaperçue.

La notion de *digital labor* permet de remettre en cause cette vision idéalisée du web, et de mettre en évidence les logiques capitalistes et marchandes des plateformes numériques¹. Elle souligne l'asymétrie d'information qui joue en la faveur des acteurs du numériques. Le consommateur-producteur ne se rend pas compte du travail qu'il fournit, mais la répartition des gains est très inégalitaire. Casilli va même jusqu'à dire que les utilisateurs sont « *exploités et heureux* ».

¹<https://www.inaglobal.fr/numerique/article/quest-ce-que-le-digital-labor-8475>

b. A ce niveau d'échange, les données personnelles peuvent-elles être considérées comme des marchandises ?

La donnée présente en des caractéristiques économiques particulières. Dans leur ouvrage *Datanomics : les nouveaux business models des données*, Simon Chignard et Louis-David Benyayer (2015) présentent ses caractéristiques. La donnée est tout d'abord un bien non rival, c'est-à-dire qu'elle n'est pas détruite après avoir été consommée, et qu'elle peut être consommée simultanément par plusieurs acteurs. Sa quantité est toujours croissante, du fait de la production de données en continue et de la baisse des coûts de stockage. Elle est très facile à reproduire, et prend de la valeur dans sa circulation, plutôt que dans sa thésaurisation.

Le travail, la rareté et l'utilité sont les fondements de économiques de la valeur. Un bien ou un service a de la valeur parce qu'il a demandé des efforts nécessaires à sa production, parce qu'il est rare, ou parce qu'il répond à un besoin.

La donnée n'est pas forcément produite pour elle-même. Comme on l'a vu dans le cadre du *digital labor*, Elle est le résultat de traces laissées par des individus, d'informations publiées etc. Dans tous les cas, le fait de créer une donnée, nécessite un travail très réduit, dont le coût marginal s'approche de zéro.

Pour ce qui est de la rareté, il est vrai que certaines données sont plus rares que d'autres, parce que plus confidentielles (des données financières par exemple). Mais le fait est que la donnée en elle-même existe en quantité infinie (puisqu'il se crée des données personnelles en très grand nombre chaque jour) et qu'il est très simple de la reproduire.

Finalement, quand on prend la donnée brute, qui vient d'être produite, elle ne répond à aucun besoin, n'a donc pas de valeur d'usage.

Ainsi, d'après les fondements économiques de la valeur, la donnée personnelle, au moment de cet échange ne peut être considérée comme une marchandise.

Pour aller plus loin, on peut voir qu'au niveau juridique, la nature des données personnelles ne leur permet pas non plus d'être considérées comme des marchandises.

D'après Henri Isaac, « *le fait que certaines données soient soumises à des réglementations juridiques particulières, notamment en Europe, les rend totalement étrangères à la notion de marchandise* » (Isaac, 2018). Si un encadrement juridique existe pour les

bases de données¹, les données personnelles en elles-mêmes n'appartiennent pas à l'individu auquel elles sont rattachées (Anciaux, Farchy, Méadel, 2017).

Je vais tenter d'aller plus loin en prenant l'exemple d'une plateforme numérique. Elle collecte les données de ses utilisateurs avec leur consentement explicite². Ces données sont recueillies dans des bases de données créées par la plateforme. Donc, selon le droit français, ces bases de données relèveraient de la propriété intellectuelle de la plateforme.

On voit donc bien que ce niveau d'échange n'est pas marchand. Il ne consiste qu'en un échange de données contre un service. Les données sont produites par les consommateurs sans même qu'ils ne s'en rendent forcément compte (malgré les alertes pour les cookies par exemple), et qui n'ont l'impression que de profiter d'un service gratuit.

c. Une réappropriation de ces données en passant par leur monétisation.

Il est vrai qu'aujourd'hui, on peut lire des articles traitant de la volonté qu'auraient les individus de reprendre le contrôle de leurs données personnelles, afin de les vendre eux-mêmes. On pourrait donc imaginer la création d'un marché des données, avec des producteurs conscients du travail qu'ils fournissent, et donc de toutes les traces qu'ils laissent à chaque utilisation de services ou d'outils du numérique.

On voit des initiatives se mettre en place dans ce sens. Elles tentent de rééquilibrer cette asymétrie entre les « consommateurs-producteurs » et les « récoltants ». Par exemple, l'entreprise Datacoup propose, pour huit à dix dollars par mois, de collecter les données des utilisateurs provenant des réseaux sociaux, ainsi que leurs données bancaires afin de les vendre aux entreprises. Elle laisse aux individus, le contrôle sur ce qui peut être collecté. On peut choisir de donner accès à ses données Facebook pendant un temps, puis de couper cet accès³.

¹ Article L 341-1 du Code de la propriété intellectuelle : le producteur d'une base de données bénéficie du droit d'auteur et du droit *sui generis*.

² Le « consentement explicite d'un individu », correspond à son accord sur les conditions générales d'utilisation et à ses réglages de confidentialité.

³ Voir le site de Datacoup : : <http://datacoup.com/docs#how-it-works> et l'article qui en parle : <https://atelier.bnpparibas/smart-city/article/datacoup-remunere-donnees-personnelles-transactions-cartographie>

Ici, on a l'exemple d'une entreprise qui possède les technologies nécessaires permettant la récupération massive de données, ainsi que leur traitement afin d'en tirer des informations pertinentes pour d'autres entreprises. Ce genre d'initiatives pourrait créer une forme de « marché des données personnelles » à partir d'un échange à la base non-marchand. Cela semble donner un peu plus de pouvoir aux individus. Néanmoins, je ne pense pas qu'elles soient disposées à grandir au point de rivaliser avec les géants du web qui réutilisent nos données. De plus, même si les individus sont rémunérés pour l'utilisation de leurs données par Datacoup, ça n'empêche pas les autres acteurs du numériques de récupérer ces données gratuitement, et de s'en servir ensuite¹.

2. L'ACTIVITE DE VALORISATION DES DONNEES PAR LES PLATEFORMES NUMERIQUES.

Nos données personnelles sont récoltées par des entreprises gérant des outils numériques (smartphone, objets connectés), ou des plateformes numériques.

Dans le rapport « Ambition numérique. Pour une politique française et européenne de la transition numérique » de 2015², présenté par Benoît Thieulin, président du Conseil national du numérique (CNNum), on peut lire : « *Une plateforme est un service occupant une fonction d'intermédiaire dans l'accès aux informations, contenus, services ou biens édités ou fournis par des tiers. Au-delà de sa seule interface technique, elle organise et hiérarchise les contenus en vue de leur présentation et leur mise en relation aux utilisateurs finaux. A cette caractéristique commune s'ajoute parfois une dimension écosystémique caractérisée par des relations entre services convergents* »¹.

D'après l'article « L'instauration de droits de propriété sur les données personnelles : une légitimité économique contestable », de Arnaud Anciaux, Joëlle Farchy et Cécile Méadel (2017), on peut lire une autre définition : « *Une première acception couvrirait notamment les "écosystèmes d'applications", les sites de partage de contenus et la place de marché, bref tous les sites qui permettent à des tiers de proposer des contenus, des services ou des biens ; une seconde acception, plus large, [...] couvrirait également tous les sites*

¹ Voir le reste de la partie II.

² Benoît Thieulin, « Ambition numérique. Pour une politique française et européenne de la transition numérique », Rapport remis au Premier Ministre, juin 2015. <https://cnumerique.fr/files/2018-04/CNNum--rapport-ambition-numerique.pdf>

qui servent de point de passage pour accéder à d'autres contenus, notamment les moteurs de recherche, les agrégateurs ou les comparateurs de prix. Tous ces sites ont en commun d'être des portes d'entrée, soit pour l'expression des internautes, soit pour l'accès des internautes à d'autres biens et services, soit les deux »².

La plateforme numérique est donc un intermédiaire. Elle met en contact des acteurs, et organise les relations entre eux. Ces relations peuvent être de différentes natures. Elles peuvent être de l'ordre de la sociabilité, en permettant aux individus de communiquer entre eux ou partager du contenu. Elles peuvent être commerciales, en mettant en lien des acheteurs et des vendeurs, ou des producteurs de contenu (culturel, journalistique) et un public, etc. Ces services que proposent les plateformes numériques sont souvent gratuits, ou alors requièrent un paiement de la part du consommateur. Ce dernier est souvent très peu élevé par rapport au service rendu³.

Elles possèdent un fonctionnement économique particulier, dans lequel les données personnelles jouent un rôle central.

Les plateformes numériques régulent le deuxième type d'échange qui met en jeu les données personnelles. On arrive ici à un échange de type marchand, qui est certes alimenté par les données, mais qui pour moi, ne peut être appelé « marché des données ». Je vais ainsi montrer les caractéristiques de ce marché.

Selon moi, la marchandise en jeu ici est l'attention. Grâce à l'article « Les économies de l'attention : fidélisation, alerte ou immersion » (Boulier, 2009), je justifierai le fait que cet échange repose sur une économie de l'attention.

Je présenterai à la suite le type de marché qu'articule la plateforme numérique : le marché à deux versants, caractéristiques de l'économie de l'attention.

Je présenterai ensuite le modèle d'affaire de Facebook afin d'illustrer mes propos.

¹ Rapport p 395.

² Conseil d'État, *Le numérique et les droits fondamentaux*, Étude annuelle 2014, pp 215-216.

³ Les plateformes de vidéos à la demande, par exemple, donnent accès à une grande quantité de films et de séries pour un prix très réduits par rapport au cinéma ou aux supports physiques.

a. Un marché de l'attention, plus qu'un marché des données.

Comme le rappelle Dominique Boullier dans « Les économies de l'attention : fidélisation, alerte ou immersion » (2009), « *tout mécanisme économique se fonde sur la rareté* ». Aujourd'hui, avec la montée du numérique, l'offre de contenu est devenue abondante. Les plateformes de VOD (Vidéo à la demande) mettent à disposition une quantité très importante de films à très bas coûts, idem pour les plateformes de streaming de musique. Les médias sont de plus en plus accessibles sur le net, les consommateurs n'ont plus le même rapport à la télévision ou aux journaux papiers. Il y a « *moins de fidélité et de "concentration" dans la lecture ou la réception des médias* » (Ibid., 2009). Dans ce contexte d'offre supérieure à la demande, ce qui devient rare c'est l'attention. On parle alors d'économie de l'attention. L'objectif va être de cibler le mieux possible les individus, le plus personnellement possible, afin de les attirer vers un contenu particulier.

C'est dans ce contexte que les données personnelles deviennent une « mine d'or », pour reprendre une expression courante. Il ne s'agit plus de créer un bien pour attirer un consommateur, mais de lancer les bons signaux pour attirer le consommateur vers ce bien. Les données personnelles récupérées et valorisées par les plateformes et autres acteurs du numériques permettent de mieux connaître les pratiques des utilisateurs. Cette valorisation passe par l'anonymisation de ces données, leur traitement, leur regroupement dans des catégories en fonction de leur sexe, âge, pratiques, goûts, etc. Elle permet de cibler les bons utilisateurs, grâce aux bons signaux pour les attirer vers certains contenus plus que d'autres.

Pour garder l'attention de ses utilisateurs, une plateforme se sert des informations qu'elle récupère à leur propos. Par rapport aux pages qu'ils ont consultés, aux contenus qu'ils ont visionnés et *likés*, la plateforme sait quoi leur montrer en priorité afin de conserver leur l'attention.

b. Le marché à deux versants, caractéristique des économies de plateforme.

Le modèle d'affaire de ces plateformes numérique est caractérisé par un « marché à deux versants » (Gabszewicz, Sonnac, 2013). Deux types d'acteurs se retrouvent sur la plateforme. Ils ont des intérêts différents mais qui se rejoignent. Par exemple, dans le cas

d'un réseau social, les individus utilisent les services mis à leur disposition. Les utilisateurs souhaitent continuer à utiliser le réseau social et à accéder à du contenu gratuitement.

En face, sur l'autre versant du marché, on trouve les diffuseurs de contenus (les médias, les marques), ou des annonceurs. Le réseau social va leur permettre de mettre en ligne du contenu, de la publicité, afin d'assurer leur visibilité.

La plateforme articule les attentes des acteurs de ce marché à deux versants, caractérisé par des effets de réseaux croisés.

L'effet de réseau direct est le fait que l'utilité d'un bien dépend du nombre d'utilisateurs. Pour ce qui est de l'effet de réseau croisé (ou indirect), c'est lorsque l'utilité des acteurs d'un versant dépend du nombre d'acteurs sur l'autre versant. Ainsi, plus il y a d'utilisateurs sur la plateforme, plus les diffuseurs ont intérêt à proposer du contenu, et les annonceurs à investir pour mettre en ligne de la publicité ; et plus il y a de contenu, et de publicité, plus les utilisateurs seront sûrs d'avoir accès à un large choix d'articles, de vidéos etc., et ce, gratuitement (grâce aux revenus publicitaires de la plateforme).

Les données sont mises en valeurs grâce aux technologies, aux algorithmes des plateformes. Elles sont anonymisées, traitées, classées dans des bases de données, qui sont la propriété des plateformes. Ces données obtiennent, certes, une valeur à la fois de travail, d'usage et de rareté. Cependant, pour moi, parler d'un marché des données personnelles à ce niveau camoufle la réalité qu'il y a derrière. Les données ne sont pas vendues pour elles-mêmes. Ce sont des espaces publicitaires, associés à des services de ciblage (possibles grâce au travail d'analytics et de data mining des plateformes) qui sont vendus. Ce qui est recherché, ce qui possède réellement de la valeur, c'est l'attention des utilisateurs. Pour moi, il s'agit donc plus d'un marché de l'attention, de la visibilité plutôt qu'un « marché des données personnelles ».

c. L'exemple du modèle d'affaire de Facebook.

Une plateforme comme Facebook regroupe plus de deux milliards d'utilisateurs, et récolte des informations sur les goûts, les habitudes, la navigation de tous ces individus. Comme on peut le lire dans sa politique de confidentialité des données : « *Nous ne vendons*

aucune de vos informations à quiconque et nous ne le ferons jamais »¹. Traitées, anonymisées et classifiées, ces elles sont utilisées par Facebook dans le cadre des services qu'il vend aux annonceurs.

Facebook vend des espaces publicitaires associés à des services de ciblage. Grâce aux données récoltées, et à leur traitement la plateforme permet aux annonceurs de créer des annonces ciblant des catégories d'individus spécifiques. Les publicités deviennent alors plus pertinentes et efficaces². Des outils permettent ensuite de connaître les performances des différentes campagnes et postes publicitaires : *« Nous fournissons aux annonceurs des rapports sur les types de personnes qui voient leurs publicités et sur les performances de leurs publicités, mais nous ne partageons pas d'informations permettant de vous identifier personnellement (des informations telles que votre nom ou votre adresse e-mail qui peuvent être en elles-mêmes utilisées pour vous contacter ou vous identifier). [...] nous communiquerons aux annonceurs des informations générales concernant la démographie et les centres d'intérêt [...] pour les aider à mieux comprendre leur audience. Nous vérifions également les publicités Facebook qui vous ont incité(e) à effectuer un achat ou à entreprendre une action auprès d'un annonceur. »*³.

Facebook offre un service d'analyse qui permet de connaître les profils et les habitudes des internautes qui naviguent sur les pages des marques qui désirent faire de la publicité. *« Nous fournissons des statistiques et des informations (insights) agrégées qui permettent aux individus et aux entreprises de comprendre la façon dont les gens interagissent avec leurs publications, leurs annonces, leurs Pages, leurs vidéos »*⁴.

La grande quantité d'informations personnelles récoltée par la plateforme permet ainsi un ciblage très qualitatif et précis pour les annonceurs. Facebook ne vend donc pas les données personnelles de ses utilisateurs. Son modèle d'affaire s'appuie sur les revenus publicitaires, qui représentent plus de 97% du chiffre d'affaire de Facebook en 2017.⁵

¹ <https://fr-fr.facebook.com/privacy/explanation>

² Cet article de blog présente les techniques publicitaires offertes par Facebook et leur bonne utilisation : <https://daniloduchesnes.com/blog/les-6-erreurs-de-debutant-en-publicite-facebook-et-comment-les-eviter/>

³ <https://www.facebook.com/about/privacy/update#sharing-partner-information>

⁴ <https://www.facebook.com/about/privacy/update#sharing-partner-information>

⁵ https://www.lesechos.fr/02/02/2017/lesechos.fr/0211760299478_les-revenus-publicitaires-de-facebook-en-forte-hausse.htm

3. L'ACTIVITE DE MARCHANDISATION DES DONNEES PAR LES DATA BROKERS.

Les deux premiers types d'échanges que j'ai présentés ne permettent pas de parler d'un « marché des données personnelles ». Pourtant, ils sont importants à avoir en tête, puisque ce fameux marché fait, en partie, référence à ces activités de récolte, valorisation des données puis vente de services reposant sur ces données. L'expression demeure néanmoins problématique dans ce contexte, puisque, comme j'ai pu le montrer, il s'agit plus d'un marché de l'attention que d'un marché des données.

Un autre type d'activité coexiste cependant aux deux premières, correspondant plus à une mise en marché des données. C'est l'activité des Data Brokers, les courtiers en données.

Lors de notre entretien, Michael Chevalier m'a présenté ces acteurs comme « *[jouant] un rôle de courtiers de l'information, un peu comme un courtier immobilier aide à vendre une propriété qui n'est pas la sienne, le courtier d'information vend l'information qui ne lui appartient pas* ». Ils existent depuis le milieu du XXème siècle, créant déjà à ce moment-là, de grandes bases de données. Avec l'arrivée du numérique, et l'essor du Big Data, les Data Brokers ont adapté leurs modèles d'affaires et leur fonctionnement, afin de devenir des acteurs centraux dans une activité commerciale basée sur les données personnelles.

Que ce soit pour des stratégies de marketing ou de sécurité, de nombreux acteurs font appel à eux pour se doter de l'information ou des outils qui leur manquent.

Je vais présenter les Data Brokers, leur fonctionnement et leur activité en détail dans cette partie. Pour ce faire je vais m'appuyer sur l'entretien que j'ai mené avec Michael Chevalier, qui avait réalisé le mémoire « Les enjeux juridiques concernant les nouveaux modèles d'affaires basés sur la commercialisation des données » (2015), ainsi que sur le rapport de 2014 de la Federal Trade Commission : *Data Brokers, a call for transparency*¹. Ce document est une référence, il est la source d'information la plus complète à propos de ces acteurs peu connus. Il présente les activités, caractéristiques et

¹ Federal Trade Commission, *Data Brokers, a call for transparency and accountability*, CreateSpace Independent Publishing Platform 19 mai 2014, 110p.

enjeux de neuf Data Brokers : Acxion, Corelogic, Datalogix, eBureau, ID Analytics, Intelius, PeekYou, Rapleaf, Recorded Future.

A la fin de cette partie, je pourrai alors donner un regard critique sur l'expression de « marché des données personnelles ».

a. Qui sont les Data Brokers ?

Les Data Brokers sont des entreprises plus ou moins récentes, touchant des secteurs variés (crédit, banque, audiovisuel, sécurité etc.). Elles sont spécialisées dans la collecte, l'utilisation et la revente de données personnelles et de services associés. Elles enrichissent leurs bases de données en tirant l'information depuis un grand nombre de sources variées. Certaines entreprises préexistent à l'avènement du numérique.

Nielsen est une société de management de la performance dans le secteur de l'audiovisuel qui étudie les pratiques de visionnage et de consommation des individus au niveau international. Fondée en 1923, elle récolte depuis presque un siècle des données sur un grand nombre d'individus, ce qui lui donne aujourd'hui un avantage très important : une base de données très fournie. Les entreprises peuvent faire appel à elle afin d'acheter des données dont elles ont besoin¹.

Acxion est une entreprise fondée en 1969 pour collecter des données destinées à la politique locale. Elle est la première entreprise qui allie collecte, segmentation et revente de données. Dès les années 1980, elle met en place des partenariats avec de nombreux acteurs comme des banques et des commerçants, afin de récupérer des données. Aujourd'hui, l'entreprise est un leader mondial dans le domaine du marketing et de la revente de données². En au moment de l'étude du FTC (Federal Trade Commission), elle possédait déjà des renseignements concernant plus de sept cents millions d'individus³.

Si ces deux entreprises sont américaines, il existe également des Data Brokers français. Mediaprism en est un par exemple. L'entreprise propose des services et des données destinées aux annonceurs, en s'appuyant sur une base de données touchant plus de trente-six millions d'individus⁴.

¹ <https://www.lebigdata.fr/top-5-data-brokers-vendeurs-donnees> et [https://fr.wikipedia.org/wiki/Nielsen_\(entreprise\)](https://fr.wikipedia.org/wiki/Nielsen_(entreprise))

² <https://www.lebigdata.fr/top-5-data-brokers-vendeurs-donnees> et <https://fr.wikipedia.org/wiki/Acxion>

³ Federal Trade Commission, *Data Brokers, a call for transparency and accountability*, CreateSpace Independent Publishing Platform 19 mai 2014, p3.

⁴ <http://www.mediadatalive.com/acteurs/mediaprism.html>

L'écosystème des Data Brokers est complexe, et s'articule autour de plusieurs strates. En effet il existe une variété de courtiers en données. Certains ne font de la collecte de données massives que pour les revendre à d'autres Data Brokers, qui ciblent des secteurs et des clients plus précis, et développent des outils et des services en s'appuyant sur le data mining et l'analytics. Il existe également plusieurs catégories dans lesquelles les Data Brokers offrent leurs services : principalement le marketing et la sécurité.

Ces acteurs sont très peu connus du grand public. Ils commencent à l'être à cause de scandales qui éclatent dans les médias, concernant nos données personnelles et leur réutilisation. Pour Michael Chevalier, il ne s'agit pas là d'une volonté des Data Brokers : *« En fait je crois que l'obscurité vis-à-vis du grand public c'est plus par rapport au fait que le grand public n'est pas un client du Data Broker. Les Data Brokers n'ont pas besoin de beaucoup d'exposition pour pouvoir vendre ou acheter. C'est surtout du BtoB. [...] une fois qu'on tombe dans le domaine du marketing, ou de la collecte d'information, dans le cadre privé, ces types de services sont bien connus, et ne sont pas nécessairement cachés ».*

b. Comment se déroule la collecte des données ?

L'objectif des Data Brokers est de récolter un maximum de données. Ils privilégient la quantité à la qualité. Ils collectent des données personnelles via une multitude de sources. Le rapport du FTC précise qu'elles peuvent venir de registres publics (gouvernementaux, locaux etc.). Les données ainsi récoltées concernent alors les différents permis délivrés aux individus, les possessions immobilières, les casiers judiciaires etc.¹

Elles peuvent être des informations commerciales, à propos de clients, récupérées directement chez des entreprises (financières, bancaires, commerciales, de grande distribution etc.) ou via d'autres Data Brokers. Beaucoup de courtiers en données partagent les mêmes sources, ou se rachètent des informations en eux².

¹ Federal Trade Commission, *Data Brokers, a call for transparency and accountability*, CreateSpace Independent Publishing Platform 19 mai 2014, pp11-12.

² *Ibid.*, p 13.

Finalement, ils ont également accès aux données disponibles librement sur internet, comme des numéros de téléphones, des articles de presse, des données issues des réseaux sociaux, lorsque les paramètres de confidentialité choisis par les utilisateurs le permettent¹.

Dans ces trois cas, soit les Data Brokers obtiennent les informations directement via les sources primaires, soit (dans la majeure partie des cas, il semblerait) ils les récupèrent chez d'autres Data Brokers.

Les données récoltées peuvent apporter assez peu d'information en elles-mêmes. Mais en collectant massivement, depuis un grand nombre de sources, les Data Brokers parviennent à les mettre en commun, à les classer, les raffiner : *« il y a énormément de données qui ne sont pas utiles au moment de la collecte pour les Data Brokers. Donc le raffinement c'est plus un filtre, ou une classification qui va permettre la revente des données »*².

Les méthodes de collecte varient en fonction des données et des sources. Les Data Brokers peuvent passer par des contrats pour collecter des données, afin d'acquérir la propriété sur la donnée, le droit de l'utiliser et de la vendre. Ces contrats contiennent aussi généralement les restrictions d'utilisation des données. Certaines sources demandent à ajouter aux contrats les cas spécifiques dans lesquels les données peuvent être utilisées, ou bien des interdictions d'utiliser les données sans en demander l'autorisation au préalable³.

Ils peuvent avoir recours à des *« web crawlers »*, des programmes qui permettent aux Data Brokers de récupérer les informations disponibles sur internet, et de les enregistrer sur leurs serveurs. Ils récupèrent également des documents imprimés (dans les administrations par exemple), qu'ils numérisent afin de pouvoir les conserver dans leurs bases de données. Ils s'arrangent également avec leurs sources afin de récupérer des API qui permettent de collecter les données⁴.

Les Data Brokers peuvent également avoir recours à une autre technique peu connue dont Michael Chevalier m'a parlé : *« il y a aussi des activités de brokerage dans le cadre de faillites »*. Certains courtiers en données vont racheter les actifs d'une entreprise qui a fait faillite, afin de récupérer des fiches clients par exemple. Cela devient intéressant

¹ *Ibid.*

² Entretien avec Michael Chevalier.

³ Federal Trade Commission, *op. Cit.*, pp16-17.

⁴ *Ibid.*, p 17.

lorsque ces données sont confidentielles, ce qui confère une forme d'exclusivité aux acteurs qui les achètent.

Les acteurs servant de sources aux Data Brokers imposent, en général, la fréquence avec laquelle ils peuvent collecter des données et les mettre à jour¹.

c. Comment se déroule la revente des données ?

Les Data Brokers sont des intermédiaires entre des vendeurs et des acheteurs : ils collectent, traitent, puis revendent des données ou services associés. La demande et l'offre de données personnelles s'organisent donc autour d'eux. Les entreprises et acteurs qui ont besoin de nouvelles données ou de services vont s'adresser aux Data Brokers pour voir ce qu'ils ont et récupérer ce qui les intéresse : « [...] comme je l'ai vu c'est que le Data Broker collecte massivement de l'information et les gens viennent le voir un peu comme on vient consulter une bibliothèque : "est-ce que vous avez tel type d'information ?". "Oui, non, voici ce que j'ai pour tel prix". Voilà. La collecte se fait avant même que la demande ne soit formulée »².

Les acteurs (entreprises, administrations etc.) qui font appel aux Data Brokers peuvent tout à la fois être acheteurs et demandeurs. Une entreprise peut vendre ses données et en acheter auprès du même Data Broker. N'importe quelle entreprise peut acheter ou vendre des données aux Data Brokers. Il en existe de toutes tailles, proposant tous types de services et de données, permettant de répondre à la demande de n'importe quelle entreprise. L'offre sera adaptée à ses moyens. Dans l'article « The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information » (Otto *et al.*, 2006), un très bon schéma résume cette idée d'un marché composé d'acteurs très nombreux et variés gravitant autour d'un Data Broker (ici, l'entreprise ChoicePoint)³.

Les Data Brokers sont des courtiers en données, ils n'utilisent pas les données pour eux-mêmes, ils les « centralisent », d'une certaine manière, les font simplement circuler entre plusieurs acteurs.

¹ *Ibid.*, p 18.

² Entretien avec Michael Chevalier.

³ Voir le schéma en Annexe 2.

Cette idée de centralisation des données chez les Data Brokers, pour ensuite les revendre, est à nuancer. En effet, on n'a pas un nombre restreint de courtiers autour desquels gravitent les acheteurs et les vendeurs, avec chaque courtier spécialisé et ciblant un secteur spécifique. La réalité est, d'après mon entretien avec Michael Chevalier et le rapport du FTC, plus complexe et floue que cela.

Les Data Brokers se revendent et se rachètent des données entre eux. A travers mes lectures j'ai pu déduire l'organisation suivante. On aurait des courtiers spécialisés dans la collecte massive de données, via des sources très nombreuses et variées. On pourrait les appeler les Data Brokers « primaires ». Après avoir traité et classé un minimum leurs données, ils peuvent les revendre à d'autres Data Brokers, plus spécialisés dans un secteur par exemple, ayant plus de contrats commerciaux, ou utilisant les Data Sciences afin de créer des services et des produits basés sur les données. Ainsi, on peut avoir plusieurs Data Brokers « secondaires » qui s'adressent au même Data Broker « primaire » afin de récupérer de l'information, mais avec une fin totalement différente. Un courtier spécialisé dans le marketing va avoir intérêt à récupérer des données de transaction bancaire, ou de navigation afin de pouvoir les revendre à des annonceurs par exemple. Un courtier spécialisé dans la détection de fraude, aura peut-être intérêt à récupérer ces mêmes données, mais pour les utiliser d'une autre manière, dans un contexte relevant de la sécurité.

Les Data Brokers proposent des services qui, d'après le rapport du FTC, ont trois finalités principales : marketing, sécurité et gestion des risques, recherche d'individus¹.

Les services et outils concernant la recherche d'individus sont les moins lucratifs pour les Data Brokers. L'addition des revenus annuels des neuf Data Brokers du rapport du FTC pour cette catégorie de produits s'élève à environ cinquante-trois millions de dollars en 2012².

Ces produits vendus dans le cadre de la recherche d'individus sont composés d'informations concernant les individus, les consommateurs. Elles sont récupérées principalement via des sources dont l'accès est libre ou public. Ces produits s'adressent autant à des entreprises et organisations, qu'à des individus seuls (ce sont les seuls à avoir cette particularité). Ils sont développés afin de permettre de tracer certaines activités de

¹ Voir en Annexe 3, le graphique issu du rapport du FTC (p23).

² Voir en Annexe 3.

concurrents, afin de retrouver des amis, de tisser des liens, de chercher les antécédents judiciaires d'une personne etc.¹

Les produits vendus pour du marketing représentent la majorité des revenus annuels des neuf Data Brokers étudiés dans le rapport du FTC. En les additionnant, on arrive à environ cent-quatre-vingt-seize millions de dollars, en 2012².

Le rapport du FTC met en avant trois catégories de services et outils marketing : ceux destinés au marketing direct³, ceux pour le marketing en ligne⁴, et ceux pour le marketing analytique⁵. Ces produits vont permettre aux entreprises qui les achètent de mieux connaître et cibler leur clientèle, d'être plus efficaces dans leurs stratégies marketing⁶.

Les produits destinés au marketing direct s'apparentent à des jeux de données classées, destinées aux entreprises qui souhaitent mieux connaître leurs clients, ainsi que des listes regroupant les caractéristiques particulières de la clientèle, pour savoir quels clients les partagent⁷.

Dans le contexte de stratégie marketing en ligne, il s'agit plutôt de services vendus par les Data Brokers. On trouve des services de ciblage à destination de sites internet sur lesquels les utilisateurs doivent s'inscrire, comme les sites de vente en ligne (« *Registration Targeting* »)⁸. Ils permettent de proposer des articles aux consommateurs qui correspondent à leurs goûts, par exemple. Il y a également des services permettant aux annonceurs de cibler des clients potentiels, le tout, en collaboration avec un site internet (« *Collaborative Targeting* »)⁹. Le site donne une liste de ses utilisateurs, l'annonceur donne une liste de ses clients et de ses objectifs. Seul le Data Broker a accès à ces deux jeux d'information. Il peut les mettre en commun afin de proposer un service personnalisé

¹ Federal Trade Commission, *op. Cit.*, p34.

² Voir en Annexe 3.

³ La définition de Wikipédia du marketing direct est : « une technique de communication et de vente qui consiste à diffuser un message personnalisé et incitatif vers une cible [...], dans le but d'obtenir une réaction immédiate et mesurable. [...] le marketing direct est celui qui touche directement la cible », https://fr.wikipedia.org/wiki/Marketing_direct.

⁴ La définition du *Journal Du Net* du marketing en ligne, ou marketing digital est : « Le marketing digital utilise les canaux numériques pour vendre un produit ou promouvoir une marque auprès de consommateurs », <https://www.journaldunet.fr/business/dictionnaire-du-marketing/1198319-marketing-digital-definition-traduction-et-synonymes/>.

⁵ D'après [definition-marketing.com](https://www.definition-marketing.com) : « Le marketing analytique est une notion [...] qui regroupe l'ensemble des techniques destinées à traiter l'information client pour optimiser la mise en place des actions marketing » (ça passe l'analyse de la valeur client, la mise en place de campagnes de ciblage etc.) : <https://www.definitions-marketing.com/definition/marketing-analytique/>.

⁶ Federal Trade Commission, *op. Cit.*, p 23.

⁷ *Ibid.*, pp 23-24-25.

⁸ *Ibid.*, p 26.

⁹ *Ibid.*, pp 26-27.

à l'annonceur. Finalement, le Data Broker peut également proposer un service de ciblage de clients potentiels grâce à des données *online* et *offline* (« *Onboarding* »)¹.

Les produits de marketing analytique sont des services basés sur le traitement des données par des algorithmes, permettant de prévoir les comportements des individus².

Les revenus des produits vendus dans le cadre de la gestion des risques s'élèvent, pour les neuf Data Brokers du rapport, à environ cent-soixante-dix-sept millions de dollars en 2012³.

Il existe deux catégories de services et d'outils, s'appuyant sur les données personnelles récoltées par les Data Brokers, pour la détection des risques, des fraudes et des failles de sécurité. La première concerne la vérification d'identité⁴. Il s'agit d'un produit permettant de confirmer ou d'infirmer l'identité d'un individu, se basant sur les données personnelles collectées par les Data Brokers. Ce genre de produit peut être utilisé par une banque, par exemple. Il peut alors prendre plusieurs formes de services et d'outils. Il peut s'apparenter à la notation de profils permettant de connaître le niveau de risque associé à un individu ; ou à un questionnaire, dont les réponses sont censées être évidentes pour l'utilisateur.

Pour la seconde, il s'agit de la détection de la fraude⁵. Les outils proposés par les Data Brokers vont permettre, grâce à l'analyse des données personnelles, de détecter des anomalies liées à un profil, à une adresse mail, etc. Alors que dans le cadre de produits destinés au marketing, il est important d'avoir des informations fiables, dans le contexte de la détection de fraude, la moindre erreur ou anomalie peut être importante à identifier. Elle peut en effet correspondre à un indice, ou une preuve à la preuve d'une faille.

d. Pour conclure sur les Data Brokers ...

Ainsi, on a pu le voir dans cette partie, une forme de « marché des données personnelles » s'organise autour des Data Brokers. Ce sont des intermédiaires entre les acteurs qui souhaitent vendre et acheter des données. Le marché est très hétérogène, et tous

¹ *Ibid.*, pp 27-28-29.

² *Ibid.*, p 31.

³ Voir le graphique en Annexe 3.

⁴ *Ibid.*, pp 32-33.

⁵ *Ibid.*, p 33.

les acteurs qui y participent peuvent être à la fois vendeurs et acheteurs, les Data Brokers eux-mêmes faisant commerce de données avec d'autres Data Brokers.

Ils utilisent un grand nombre de sources différentes afin de recueillir un maximum de données sur les individus. Ces données sont ensuite traitées, classées, afin d'être soit revendues à d'autres Data Brokers, soit directement à des clients, comme des entreprises, des administrations ...

Les Data Brokers ne revendent pas forcément des données, mais surtout des services, des produits, des outils qui reposent sur le traitement, voire l'analyse de ces données. Les acteurs qui font appel aux Data Brokers sont motivés par plusieurs finalités qui relèvent du marketing, de la sécurité, ainsi que de la recherche d'individus.

4. CRITIQUE DE LA NOTION DE « MARCHÉ DES DONNEES PERSONNELLES ».

Dans cette partie, j'ai voulu interroger la notion de « marché des données ». Je me suis rendu compte, qu'en général, lorsque l'on parle de « marché des données », on pense tout de suite à Facebook, aux GAFA, aux scandales qui touchent à la réutilisation de nos données personnelles...

Au fur et à mesure de mes recherches, j'ai réalisé que ce fameux « marché des données », qui concerne principalement les données à caractère personnel, est en fait une expression qui mélange plusieurs réalités, plusieurs types d'échanges, qui coexistent les uns avec les autres. On pourrait dire que les trois types d'échanges que j'ai présentés plus haut « composent le marché des données ». Dans ce cas, on a affaire à un marché qui ne correspond en rien à un marché traditionnel. Il est divisé en trois niveaux d'échanges, au sein desquels, les acheteurs peuvent aussi bien être des vendeurs, les producteurs, des consommateurs. Ces échanges s'organisent autour de plateformes et d'intermédiaires. Les biens et services échangés sont très hétérogènes, ce sont à la fois des jeux de données (Data Brokers), mais aussi, et surtout des services et des produits développés autour de la donnée (Data Brokers, Plateformes numériques etc.).

Pour reprendre les mots de Jean-Marc Lazard, co-fondateur d'OpenDataSoft, lors de notre entretien : *« Est-ce qu'il y a une économie de la donnée ou irriguée par la donnée ? Ce n'est pas la même chose en fait. La donnée est extrêmement importante mais*

aujourd'hui c'est quoi les challenges ? C'est une économie basée, construite avec une meilleure exploitation, une meilleure circulation de la donnée ».

A l'heure actuelle, d'après mes recherches, je ne pense pas que nous ayons des modèles d'affaires reposant uniquement sur la vente de données, pour elles-mêmes. L'intérêt n'est pas là. Les acteurs ne cherchent pas à vendre ou acheter à tout prix de la donnée brute. Ce qui compte le plus c'est d'être intégré dans cette économie portée par la donnée.

Ma conclusion, pour cette partie sur le marché des données pourrait donc être la suivante. Le « marché des données » est un terme utilisé par les médias et le grand public, qui leur permet de se saisir d'une réalité complexe et floue, mais qui touche néanmoins chacun d'entre nous. Parler de « marché de l'attention », d'« économie irriguée par la donnée » etc., ça n'est pas très évocateur. Au contraire, le « marché des données personnelles » est beaucoup plus parlant. On (le grand public) sait que nos données sont utilisées, on ne sait pas exactement comment, mais on sait que c'est dans une logique économique, capitaliste. A mon sens, le terme « marché » permet donc, au plus grand nombre, de se saisir de cette réalité complexe et hétérogène, tout en la simplifiant et en la floutant dangereusement.

Il camoufle et fausse la réalité. Parmi les discussions que j'ai pu avoir avec mon entourage, le fait de dire que Facebook ne vend pas nos données éveillait souvent le même type de réactions : de l'incompréhension, pensant que Facebook vendait nos données sur le « marché des données ».

D'après mes recherches et mes réflexions, je pense que l'expression de « marché des données » est une expression trompeuse. Il serait peut-être plus intéressant de parler ici d'une forme d'économie capitaliste irriguée par la donnée.

Dans cette partie j'ai déconstruit l'idée de « marché des données », en montrant que ce terme est utilisé pour résumer une réalité bien plus complexe que l'on peut croire.

Aujourd'hui l'idée d'une « marchandisation » des données éveille toutes sortes de fantasmes. Il est vrai que grâce au traitement et à l'utilisation des données personnelles, des acteurs comme les GAFA ou les Data Brokers parviennent à construire des modèles d'affaires alléchants.

Il me semble cependant très important, pour contrebalancer cette vision très capitaliste de l'utilisation des données, de présenter d'autres formes de valorisation, ne passant pas par l'échange marchand, et ne reposant pas exclusivement sur les données à caractère personnel.

Dans la partie suivante, je vais présenter, de manière non exhaustive, des modèles que j'ai découverts à travers mes recherches et mes entretiens. Il serait intéressant de les faire connaître au grand public, qui pourrait devenir un acteur conscient, et y prendre part de manière active.

Partie 3 : Des modèles tournés vers l'ouverture pour d'autres formes de valorisation des données.

En m'intéressant à la question du « marché des données », j'ai découvert des autres manières de valoriser les données autrement que via des échanges marchands. L'Open Data et les modèles d'affaires qui peuvent en découler m'ont semblé pertinents à étudier. Ils permettent de mettre en lumière des alternatives à l'exploitation économique de nos données personnelles, en faisant apparaître de nouveaux modèles économiques, allant presque à l'encontre de ceux présentés précédemment.

Au cours de mes recherches, j'ai eu la chance de mener des entretiens avec des personnes venant d'entreprises ou de collectivités concernées par l'Open Data, et les problématiques qui en résultent. Contrairement à la revente de données, l'ouverture de celles-ci est un sujet sur lequel beaucoup plus de monde souhaite communiquer, sans doute est-il beaucoup moins sujet à des scandales et des *a priori*.

J'ai pu mener trois entretiens dans lesquels il était question d'Open Data (considéré dans des contextes à chaque fois différents, du fait de la diversité des profils). Le premier était avec Norbert Friant, responsable du service numérique au sein de Rennes Métropole, le second avec Jean-Marc Lazard, co-fondateur et président d'OpenDataSoft, et le troisième avec Jean-François Montagne, responsable de la gouvernance des données chez Enedis. A l'issue de ces entretiens, l'image que j'ai de l'Open Data est vraiment celle d'un mouvement assez idéaliste, grâce auquel il serait possible d'améliorer un grand nombre de processus (tant au niveau des villes, que des entreprises ou des administrations), mais dont la culture n'est pas encore suffisamment implantée en France pour se développer correctement.

Je reviendrai plus en détail sur cette idéologie caractérisée par l'ouverture et la collaboration, contrebalançant l'exploitation économique des données personnelles.

Au cours de ces entretiens, j'ai eu peu d'informations sur la valorisation monétaire concrète des données ouvertes. Comme l'Open Data est un phénomène assez récent, les

effets économiques réels semblent encore assez flous, il existe peu de littérature à ce propos, et l'on y traite principalement de l'Open Data d'un point de vue juridique ou éthique (transparence démocratique). J'ai cependant découvert un recueil d'articles très intéressant et assez unique dans sa façon de présenter les modèles d'affaires reposant sur l'Open Data : *Open models : les business models de l'économie ouverte*, coordonné par Louis-David Benyayer (2015). Le chapitre « Les Open Models de la Data » présenté par Chloé Bonnet (Benyayer, 2015, p 66-93) regroupe plusieurs articles qui m'ont permis d'avoir une meilleure compréhension de la valorisation des données gratuites, que je présenterai plus tard.

Je souhaite néanmoins aller plus loin que l'Open Data comme seule alternative à l'exploitation des données personnelles. Au fil de mes recherches des modèles très intéressants, qui s'appuient sur les données et les valorisent dans des contextes différents se sont démarqués.

Je me suis intéressée à la Smart City, comme organisation urbaine basée sur l'utilisation des données produites par les individus et les acteurs de la ville. Je vais en présenter un aperçu dans cette partie, en m'appuyant sur les entretiens que j'ai pu réaliser, ainsi que sur des travaux. « Villes intelligentes sources d'inspiration » de Vincent Rialle (2017) ou « La ville intelligente. Origine, définitions, forces et limites d'une expression polysémique » de Sandra Breux et Jérémy Diaz (2015) m'ont notamment aidé à mieux cerner ce qu'est une Smart City et de comprendre quels en sont les enjeux.

En m'intéressant à la ville intelligente, j'ai également découvert la notion de « troc de données »¹. J'ai découvert cette expression dans le contexte du projet de Smart City Rennaise. Un article du *Journal Du Net* présentait une initiative lancée par la ville. Elle permettrait le troc de données entre différents acteurs. L'entreprise Enedis, chargée de la gestion et de l'aménagement du réseau de distribution d'électricité, était citée comme un de ces acteurs. Mon entretien à ce sujet avec Jean-François Montagne, pour essayer de mieux saisir les enjeux de ce troc de données.

¹ <https://www.journaldunet.com/economie/services/1206939-rennes-metropole-echange-donnes-privées/>.

1. COMMENT CREER DE LA VALEUR AVEC L'OPEN DATA ?

Comme je l'ai montré dans la première partie du mémoire, l'ouverture des données nécessite une organisation importante avec la mobilisation d'une équipe Open Data, et l'investissement de nombreux outils. Et ce pour u processus est assez long. Depuis 2016, les administrations sont tenues de diffuser des données en libre accès sur des plateformes prévues. Cette démarche représente pour elles des mises en œuvre chronophages, associées à des coûts importants, avec, des retombées économiques assez anecdotiques pour l'instant. Parmi les « contraintes » inhérentes à l'Open Data, on peut par exemple citer l'indispensable maintenance informatique qui doit permettre aux plateformes de s'adapter et de perdurer. On peut également penser aux problématiques liées au partage des données personnelles. Pour respecter les préceptes de la CNIL, les administrations qui ouvrent leur données doivent mener une analyse rigoureuse des jeux de données, ce qui implique du temps et des coûts en plus¹.

Les données ouvertes ne peuvent pas faire l'objet d'un échange marchand : « *La licence associée aux données publiques ne permet pas aux administrations de les revendre. Elles sont en revanche susceptibles d'être à la base d'un service marchand de la part du réutilisateur. Il n'est donc pas possible de considérer ces données comme des marchandises* » (Isaac, 2018). Ainsi, les retombées économiques de l'Open Data ne sont pas évidentes à mesurer ni même à envisager au premier abord.

Dans un article issu de l'ouvrage sur les Open Models, intitulé « Open Data. Show me the money ! », Louis-David Benyayer et Simon Chignard (2015) donnent trois hypothèses pour expliquer pourquoi l'impact économique de l'Open Data est encore difficile à mesurer. La première est qu'il est encore trop tôt pour voir des retombées économiques de l'Open Data. Il y aurait trois phases dans l'ouverture des données : la phase d'ensemencement (l'ouverture des données), la phase de culture (la réutilisation des données ouvertes) et la phase de récolte (les retombées économiques). Cette phase de récolte ne serait pas encore d'actualité, on ne pourrait donc pas concrètement voir les effets produits par l'Open Data sur l'économie. Une seconde hypothèse serait que les impacts sont trop diffus et complexes à saisir, en prenant en compte le fait que les données d'Open Data peuvent être utilisées et mises en valeur en étant associées à des données privées lors de leur traitement, par exemple. Pour « faire parler les données » ouvertes, une entreprise

¹<http://www.lagazettedescommunes.com/438803/open-data-quels-couts-pour-les-collectivites-locales/>

peut les recouper avec ses données privées. L'Open Data ne crée par toute la valeur. Une troisième hypothèse serait la mauvaise connaissance des données ouvertes par les utilisateurs.

Ainsi dans ce premier axe je vais présenter deux façons de produire de la valeur (monétaire ou non) grâce à l'Open Data. Elles ne sont certainement pas les deux seules qui existent, ce sont néanmoins celles qui m'ont semblées les plus intéressantes et que je souhaite approfondir ici.

Dans ce premier axe, je vais m'appuyer sur trois articles : « La trajectoire des innovations ascendantes : inventivité coproduction et collectifs sur Internet » (Cardon, 2006), « Intéresser à la réutilisation des données : la participation des développeurs aux concours d'applications open data » (Goëta, 2013), « Open data, usagers du numérique et entreprises publiques » (Asdourian, Zimmerli, 2015). Je vais grâce à eux montrer en quoi l'Open Data est une source d'innovation pour les entreprises et pour la société.

Je vais ensuite utiliser plusieurs articles du chapitre « Les Open Models de la Data », de l'ouvrage *Open models : les business models de l'économie ouverte* (Benyayer, 2015), afin de montrer que de nouveaux acteurs intermédiaires émergent. Ils créent de la valeur en permettant une circulation et une réutilisation plus simple des données brutes en libre accès.

Je montrerai ensuite quelles sont les limites, à l'heure actuelle, de la mise en place et de la rentabilité de l'Open Data, en m'appuyant principalement sur les entretiens que j'ai mené avec Jean-Marc Lazard d'OpenDataSoft et Norbert Friant de la Renne Métropole qui côtoient des acteurs qui ouvrent leurs données au quotidien dans le cadre de leur travail.

a. L'Open Data comme source d'innovation.

L'Open Data peut tout d'abord être une source d'innovation qui peut partir aussi bien des usagers du numérique que des professionnels. Les uns et les autres s'emparent des données et créent de nouveaux produits se basant sur ces dernières.

Bruno Asdourian et Virginie Zimmerli présentent la notion de « *pro-am* » (Asdourian, Zimmerli, 2015), des professionnels-amateurs, qui maîtrisent la technologie, et qui vont créer des services, des applications en se basant sur les données disponibles

gratuitement. Ils peuvent également être appelés « *lead-users* », selon un concept du professeur Eric von Hippel. Ce sont des utilisateurs avant-gardistes de technologies. Ils adoptent avant les autres un service, un produit ou une tendance, et deviennent source d'innovation de par leur intérêt à développer des nouveaux usages ou nouvelles solutions eux-mêmes¹. Les « *lead-users* » ne se satisfont pas d'un usage qui leur est imposé, ils cherchent plutôt à adapter le produit à leur usage.

Le concept d'innovation ascendante (Cardon, 2006) reprend cette idée. Les usagers s'approprient les produits en réinventant l'utilisation qui doit en être faite. On peut citer l'exemple classique du téléphone portable, qui n'était au départ fait que pour téléphoner. Peu à peu les utilisateurs se sont approprié la fonctionnalité d'envoi de SMS. Cet usage s'est massifié, de sorte que les terminaux ont évolué pour s'adapter à cette nouvelle pratique. Ainsi, contrairement à un processus d'innovation descendante, avec cette idée d'usages imposés aux utilisateurs par ceux qui fabriquent les produits, l'innovation ascendante part des usagers et remonte pour faire évoluer les produits. L'appropriation et les usages qui en découlent ne sont généralement pas prévisibles (*Ibid.*, 2009), obligeant les fabricants, chercheurs, ingénieurs à adapter leur produit *a posteriori*.

Le concept d'Open Innovation du professeur Henry Chesbrough (2003) fait référence à un mode d'innovation qui se base sur la coopération et le partage entre les acteurs économiques. Dans un contexte d'innovation dite « fermée », l'entreprise mène ses projets en interne, en utilisant ses connaissances et ses technologies propres. Dans le processus d'innovation ouverte, les projets se basent tant sur des technologies internes, qu'externes. L'entreprise s'ouvre aux connaissances extérieures¹.

L'Open Innovation associée à une innovation ascendante permet la mise en place d'un phénomène de *crowdsourcing*. « *Le crowdsourcing consiste littéralement à externaliser (to outsource) une activité vers la foule (crowd) c'est-à-dire vers un grand nombre d'acteurs anonymes* » (Burger-Helmchen, 2011). C'est une activité qui n'existe pas que dans le cadre de l'Open Data, mais dans ce contexte, l'entreprise va fournir des données aux individus qui proposeront des innovations reposant sur ces données (Asdourian, Zimmerli, 2015). C'est un échange entre les entreprises et des acteurs extérieurs.

Procéder via une démarche de *crowdsourcing* permet aux entreprises de se décharger de certaines contraintes et certains risques. Elles externalisent la Recherche et

¹<https://www.lescahiersdelinnovation.com/2016/03/lead-users-role-linnovation/>

Développement, ce qui leur permet un gain de temps et d'argent. Cela permet également de tester l'innovation à l'extérieur, externalisant ainsi la prise de risque. Si les résultats sont satisfaisants, l'entreprise peut décider d'intégrer l'innovation dans ses offres. En procédant ainsi, elle se prémunit de risques liés à l'innovation d'une part, mais travaille également sur son image. En mettant en avant cette ouverture, cette communication avec le public, l'entreprise peut se montrer comme porteuse de valeurs de transparence, de liberté, jouant en sa faveur d'un point de vue communicationnel (*Ibid.*, 2015).

D'autres acteurs vont également s'emparer des données disponibles gratuitement afin de créer des nouveaux services permettant d'améliorer le quotidien des individus. « *Le potentiel d'innovation de la réutilisation des données publiques repose sur les développeurs, entrepreneurs, fondateurs de startups. Inventer une seconde vie aux données publiques permettra de créer des services innovants, de faire émerger de nouveaux usages, de nouveaux services en ligne, de nouvelles applications* », d'après Séverin Naudet, ancien directeur de la mission Etalab (Goëta, 2013). Ces données peuvent ainsi être récupérées dans un cadre professionnel dans le but de créer un produit ou un service se basant sur ces informations.

Les administrations cherchent à inciter ces acteurs (professionnels, startups etc.) à se saisir des données pour créer des services, des applications apportant de la valeur : « *Les développeurs sont ainsi conçus comme des innovateurs qui créeront de nouveaux usages et services dès lors que les données seront mises à disposition.* » (*Ibid.*, 2013).

Pour ce faire, Samuel Goëta (2013) montre que des concours sont lancés afin d'attirer les développeurs, les entreprises à innover grâce aux données ouvertes. Il donne notamment l'exemple de l'évènement Moov'In The City lancé par la ville de Paris le 21 mai 2013, un grand concours ayant pour thème le développement de la mobilité dans une ville en mouvement. L'objectif était la réutilisation des données présentes sur le portail d'Open Data de Paris, le tout en partenariat avec les entreprises symboles de mobilité comme la SNCF, la RATP ou JCDecaux (publicité sur supports urbains), pour la création d'une application répondant à un besoin. L'incitation était notamment financière, et presque cent équipes se sont inscrites pour participer. Sept prix ont été décernés à la fin, pour les applications les plus innovantes.

¹ <https://openinnovation.wordpress.com/2015/01/01/definition-de-lopen-innovation-selon-henry-chesbrough/>

Au-delà d'une motivation liée à la volonté d'innover, les concours autour des données ouvertes ont pour but de justifier la politique d'Open Data, de prouver que de la valeur peut être créée à partir de données gratuites. La création d'applications destinées au grand public doit permettre à ce dernier de prendre conscience de l'importance de l'ouverture des données, des bienfaits que cela peut lui procurer.

b. L'émergence d'intermédiaires pour une meilleure circulation des données.

La création d'applications est la première forme de valorisation des données ouvertes à laquelle on pourrait penser. Cette idée est à nuancer. Aujourd'hui, la durée de vie économique des applications est très faible. Une minorité seulement sont ce qu'on appelle des « *killer apps* » (Bonnet, Lalanne, 2015), c'est-à-dire, les applications très utilisées, qui font le buzz.

Chloé Bonnet et Romain Lalanne (2015) reprennent l'idée qui dit que la valeur de la donnée ne réside pas dans sa thésaurisation, mais bien dans sa circulation. Au-delà du développement d'applications plus ou moins populaires, le véritable intérêt pour les acteurs du numérique aujourd'hui est de gérer le flux des données. Ils mettent donc en avant un maillon intermédiaire dans l'utilisation des données ouvertes, qui se situe entre la plateforme d'Open Data et les développeurs d'applications.

Des plateformes de services apparaissent : « *Ces plateformes de services qui se placent entre le producteur et le réutilisateur consolident et accélèrent le potentiel d'innovation des données ouvertes en simplifiant leur accès, leur transformation et leur consommation* » (Ibid., p 70).

Des plateformes émergent donc. Elles récupèrent des jeux de données brutes gratuits, et proposent des services variés, simplifiant ainsi la réutilisation des données pour d'autres acteurs. Elles s'appuient sur des modèles freemium, qui combinent une offre payante et une offre gratuite, et qui peuvent prendre des formes assez variées.

Dans l'article « Open Data : de la *killer app* à l'économie des plates-formes » (Ibid., 2015), les auteurs présentent trois modèles très parlants, que je vais reprendre dans cette partie afin de montrer quelle valorisation des données ouvertes et gratuites est possible.

La première catégorie d'intermédiaires est celle qui propose une offre qualifiée de "*freemium à fonctionnalités variées*" (*Ibid.*, p 71). Une couche d'outils et de design est ajouté sur les données brutes. L'utilisateur doit payer pour y accéder, et accéder aux services associés.

La seconde catégorie est celle qui offre un "*freemium fondé sur les finalités d'usage*" (*Ibid.*, p 71). L'article donne l'exemple de OpenCorporates, qui est une entreprise qui propose de fournir des données aux acteurs qui en ont besoin, qui peuvent également les enrichir, mais dont la finalité est de les laisser ouvertes. Si les acteurs souhaitent ensuite privatiser ces données (licence commerciale), leur accès devient payant.

La dernière catégorie est le "*freemium comme cheval de troie*" (*Ibid.*, p 72). Ici, l'objectif est d'offrir un certain nombre de services basés sur des données ouvertes et gratuites, afin de mettre en avant les capacités techniques et scientifiques, le savoir-faire de l'entreprise. L'entreprise vend ensuite des services d'expertise en Data Sciences. La gratuité est ici une « vitrine » (*Ibid.*, p72).

Dans ces trois modèles, ce n'est pas la donnée qui prend une valeur économique, mais bien le service qui y est rattaché, et qui facilite le flux de données (*Ibid.*, p 71). En s'appuyant sur l'ouvrage *Free : the future of a radical price* de Chris Anderson (2009), les auteurs précisent que ce n'est pas la donnée qui a de la valeur, mais bien les compétences en Data Sciences qui permettent de l'exploiter. Pour eux, ce sont des modèles économiques plus viables que le développement d'applications qui ne parviennent pas, au final, à se détacher du lot, et qui finissent par disparaître faute d'utilisateurs.

c. Les limites de ces modèles de valorisation des données en libre accès.

L'Open Data promet donc de créer de la valeur grâce à des initiatives individuelles, à la participation de professionnels au développement de services, ou encore par l'arrivée d'intermédiaires s'appropriant la gestion du flux de données à travers des modèles freemium.

Néanmoins, ces initiatives sont assez timides en France. Samuel Goëta (2015) montrait que sans les concours lancés par les institutions, les entreprises ne se saisissent pas vraiment des données en libre accès, et ne cherchent pas à innover d'elles-mêmes :

« Encouragée par la dotation financière et la mobilisation de l'institution vers les résultats du concours, la participation ne parvient pas dans les cas étudiés à s'inscrire sur le long terme. L'intéressement des réutilisateurs ne semble s'inscrire que sur la période du concours à la suite duquel il est difficile de mobiliser les acteurs internes et externes pour la poursuite de la dynamique engagée » (Goëta, 2015). Norbert Friant confirme cette idée, dans notre entretien : « En fait [le site d'Open Data de Rennes Métropole] n'est pas énormément utilisé pour vous dire les choses très franchement. C'est resté une affaire de spécialistes en fait l'Open Data. Il est utilisé par quelques entreprises qui ont créé des applications dans le domaine des transports notamment », mais d'après lui, c'est très restreint, ce n'est pas une problématique qui semble s'adresser à tout le monde.

Les entretiens que j'ai réalisés m'ont en effet confirmé que la culture de l'Open Data n'est pas très bien implantée en France. Pour Norbert Friant « *accepter la décentralisation de l'innovation, [...] culturellement, ce n'est pas évident* » pour les institutions ou même les entreprises. D'après Jean-Marc Lazard, d'OpenDataSoft, la culture de l'Open Data en Europe est assez faible, et en France elle est même « *bloquée* », contrairement à la « *culture du secret* » qui est « *très forte* ».

En essayant de montrer, à travers des concours, la valorisation potentielle des données en libre accès, les institutions et administrations voudraient développer cette culture de l'ouverture. Nous sommes néanmoins très en retard à ce niveau-là. La plupart des personnes de mon entourage avec lesquelles j'ai pu parler d'Open Data ne se sentaient absolument pas concernées par le sujet. Moi-même en découvrant les plateformes d'Open Data existantes (notamment celle de Rennes Métropole, que j'ai découvert en premier), j'ai été surprise de voir quelles données étaient partagées, quels outils étaient mis à disposition sur ces plateformes (quelques outils de Data Visualisation, mais le tout est assez limité). La volonté de démocratiser l'Open Data semble assez présente chez certains acteurs, mais n'est pas assez forte pour permettre de développer une culture de l'ouverture forte.

2. LA SMART CITY : CREER DE LA VALEUR ENSEMBLE, EN COLLABORATION.

Mes recherches sur le marché des données m'ont poussée à m'intéresser à la question de l'Open Data, qui m'a elle-même fait découvrir le sujet de la Smart City et les enjeux qui en découlent. J'ai abordé ce sujet un peu par hasard. Lors du salon Big Data Paris 2018¹, j'ai rencontré une personne travaillant chez Orange qui présentait une plateforme d'échange de données, applicable dans le cadre de la Smart City. C'était un sujet qui m'était étranger, mais après avoir discuté, j'ai décidé d'approfondir le sujet. J'ai eu peu de réponses lors de mes demandes pour mener des entretiens, mais ces derniers ont néanmoins été très intéressants. Je vais en grande partie m'appuyer sur les entretiens avec Norbert Friant, et Jean-Marc Lazard pour cette partie.

La Smart City est pour moi un exemple concret de la mise en valeur non marchande des données.

Dans cette sous-partie, je vais tout d'abord présenter ce qu'est une Smart City en m'appuyant sur les articles « Villes intelligentes sources d'inspiration » (Rialle, 2017) et ou « La ville intelligente. Origine, définitions, forces et limites d'une expression polysémique » de (Breux, Diaz, 2015).

En m'appuyant sur les entretiens que j'ai réalisés, je montrerai en quoi la Smart City est un espace d'utilisation et de valorisation des données.

Je ferai enfin un point sur les valeurs rattachées à la Smart City, sur lesquelles les différentes personnes que j'ai pu rencontrer avaient insisté, et qu'il me semble important de présenter. Pour ce faire, je vais notamment m'appuyer sur l'article « L'économie collaborative, entre utopie et *big business* » de Simon Borel, David Massé et Damien Demailly (2015).

a. Qu'est-ce que la Smart City ?

Le terme de Smart City, ou « ville intelligente » en français, est apparu dans les années 1990, popularisée par l'entreprise d'informatique IBM. Dans une époque marquée

¹ Les 12 et 13 mars 2018.

par la récession économique, l'entreprise envisage la ville, en l'associant aux TIC, comme un marché potentiel important (Breux, Diaz, 2015), lui permettant de faire de nouveaux profits. IBM diffuse et popularise l'expression petit à petit, jusqu'à ce que les villes et les acteurs publics se l'approprient.

Aujourd'hui, la définition de la Smart City la plus simple que l'on puisse donner est celle de Wikipédia : « *une ville utilisant les technologies de l'information et de la communication pour améliorer la qualité des services urbains ou encore réduire ses coûts* »¹. Les enjeux autour de la ville intelligente sont de plus en plus importants du fait de la croissance de la population urbaine. Selon le sociologue Edgar Morin, qui a justement beaucoup écrit sur la ville, le processus d'urbanisation, « *d'après les prévisions actuelles, risque d'englober plus de 80% de la population dans quelques décennies* » (Morin, 2014, p1).

Deux conceptions de la ville intelligente existent. Dans un premier temps, la Smart City peut être vue comme une ville où tout est automatisé, contrôlé par des entreprises à la pointe des nouvelles technologies, qui récupèrent les données afin d'optimiser les processus quotidiens, du ramassage des ordures, à la consommation énergétique, en passant par la durée des feux tricolores (Rialle, 2017).

La ville de Toronto prévoit, par exemple, de créer un quartier géré par Google. Les habitants auront un compte utilisateur qui leur permettra d'accéder aux services proposés par les acteurs chargés de la gestion du quartier. Des capteurs seront placés partout afin de mesurer le taux de fréquentation des parcs, des magasins, afin de mieux organiser l'espace urbain etc.²

Ce type d'organisation de l'espace urbain peut inquiéter. Vincent Rialle (2017) cite Edgar Morin : « *Il est certain que nous sommes dans un processus de déshumanisation de villes, c'est-à-dire de désolidarisation. Parce qu'on est dans une civilisation, une société qui privilégie l'individualisme, l'intérêt, le calcul, la chronométrie, toutes ces choses-là, il y a une tendance à la déshumanisation de la ville* ».

¹ https://fr.wikipedia.org/wiki/Ville_intelligente

² <https://usbeketrica.com/article/sidewalk-labs-quartier-google-toronto>

La Smart City peut donc rapidement sembler assez dystopique, dénuée de toute humanité, centrée sur l'efficacité, l'optimisation des processus, censés procurer plus de bien être aux individus.

D'autre part, la Smart City peut aussi être envisagée comme un espace d'ouverture et de collaboration entre des acteurs aux compétences et connaissances variées. C'est la vision que m'en a donnée Norbert Friant, responsable du service numérique au sein de Rennes Métropole, lors de notre entretien. Souhaitant rompre avec une vision trop technocentrée, qui donnerait à l'ingénieur le rôle clé, la Smart City Rennaise semble mettre l'individu au cœur de son fonctionnement. Pour Norbert Friant, aujourd'hui, le numérique « *ce n'est plus uniquement le monde de l'ingénieur informatique* ». La métropole cherche à mettre en avant des individus avec des compétences variées. Il insiste justement sur le fait qu'elle repose sur « *plus de 300 métiers* » : « *On va avoir un chauffeur, un journaliste, une assistance, un chanteur d'Opéra, un maître-nageur, une documentaliste.... A chaque fois avec des compétences variées. On voit quelle est leur appropriation du numérique, de la question de la data.* » La Smart City peut ainsi être conçue comme un espace qui permet d'innover et de collaborer.

Il n'existe pas une seule conception de la Smart City, elle peut varier en fonction de la taille de la ville, des attentes des responsables, du niveau d'investissement des habitants etc. L'image que m'en ont donnée les personnes avec qui j'ai pu m'entretenir présentait la ville intelligente comme un espace de collaboration et d'innovation, mettant les individus au cœur du processus, je vais donc plus me concentrer sur cette conception de la Smart City, à l'image de Rennes Métropole.

b. La Smart City est un espace de circulation et de valorisation des données.

La Smart City est un concept qui date des années 1990 mais qui commence à véritablement prendre sens aujourd'hui. La ville intelligente s'appuie sur les données récoltées par les entreprises et les administrations. Le développement des objets connectés

et de l'IoT permet donc de récupérer un maximum d'informations sur les activités quotidiennes des individus au sein de la ville.

Les données permettent aux différents acteurs de mieux comprendre les habitudes des habitants, de mieux comprendre comment fonctionne la ville, pour ainsi adapter leurs services aux usages effectifs. Dans le cadre de la Smart City, l'intérêt pour un acteur se situe même au-delà de sa simple compréhension et adaptation personnelle. Il s'agit, dans une certaine mesure, de mettre en commun ses données avec les autres acteurs de la ville, afin d'affiner ses analyses et d'optimiser les processus du quotidien. Dans ce cadre, les entreprises et les administrations peuvent utiliser les données que les autres choisissent de mettre à leur disposition afin de mieux comprendre les enjeux spécifiques à la ville.

D'après les entretiens et les recherches que j'ai menés, j'ai déduit que l'intérêt est vraiment de considérer la ville comme un écosystème, dans lequel les acteurs sont tous en liens les uns avec les autres : *« C'est là qu'on arrive dans une logique de plateforme de données, soit sectorielle, soit trans-sectorielle, où on va mixer tous les éléments de la Smart City, c'est ce qu'essaye de faire Rennes métropole »*¹.

La mise en valeur des données passe donc par une ouverture plus ou moins importante, ou du moins une mise en commun de ses informations. Il faut pouvoir coordonner au mieux les activités des différents acteurs de la ville, en prenant en compte les us et coutumes.

Les acteurs de la Smart City ne se concentrent pas que sur l'optimisation des activités au sein de la ville. Le fait de mettre certaines de leurs données en commun permet également d'obtenir des retours, des critiques constructives de la part des autres acteurs sur leur fonctionnement. Jean-Marc Lazard, dans notre entretien revient d'ailleurs sur le manque de transparence de la SNCF envers les régions : *« la SNCF aurait dû se servir dès le début de ses données comme d'un élément de dialogue. Mais bon ça, ça vient souvent d'un déficit de compréhension, de connaissances de ces sujets-là. Rennes et Kéolis², vous avez deux parties qui ont compris les enjeux, les sujets et pour qui c'est normal, que des fois ils ne soient pas d'accord etc. »*. Ouvrir ses données permet d'engager un dialogue afin d'améliorer son propre fonctionnement.

¹ Entretien avec Jean-François Montagne.

² Opérateur des transports en communs à Rennes.

Dans notre entretien, Jean-Marc Lazard insiste sur cette effet réseau indispensable à la Smart City. Dans le cadre « *d'un projet de rénovation d'une place à Paris, comme la place de la Nation où il y a plein de capteurs, mais tous opérés par des acteurs différents, les uns par General Electric, l'autre par la ville de Paris etc. Ca veut dire que ces acteurs-là qui ont chacun leur système, ces acteurs ont un besoin de collaborer, pour collaborer et développer des services qu'ils vont pouvoir proposer à leurs propres clients, qui sont les mêmes finalement. Ce qui est intéressant c'est que tout ça, ça amène à formuler, pour ces acteurs, le fait que l'utilisateur de leurs services, c'est la même personne. Derrière, la donnée sous-jacente est un élément essentiel qui doit circuler et s'échanger pour pouvoir aider ces entreprises à comprendre comment on utilise leur service, qu'est-ce qu'on fait après etc.* »

Les données permettent donc d'en apprendre plus sur les habitants de la ville, sur leurs habitudes, leurs activités. Il faut faire attention cependant. Norbert Friant souligne la dimension stigmatisante que peuvent avoir les pratiques d'ouverture, de partage et d'analyse des données. Il donne l'exemple des données des bibliothèques d'une ville : « *on pourrait [...] analyser, par quartier, quels types de livres sont lus ? En fait ça peut être super stigmatisant* ». Les données qui mettent en lumière des inégalités sociales devraient justement servir à lutter contre ces inégalités au niveau de la ville.

Pour ouvrir leurs données, les acteurs de la Smart City peuvent, par exemple, compter sur des plateformes mises en place par OpenDataSoft¹. L'offre de plateforme Smart City permet de partager ses données de manière ciblée, sans tomber dans une ouverture totale. Elle offre « *des fonctionnalités liées la sécurité, des fonctionnalités liées à une diffusion plus ciblée de la data, donc pas forcément open à tous, une partie open une partie en moyen open et une partie en privée* »². Elle permet de collecter les données des capteurs de la ville en temps réel et de les réunir sur le même "hub". Elle invite les habitants à participer à la construction de la ville intelligente³.

Néanmoins, on l'a vu, la culture de l'ouverture et du partage des données n'est pas très développée en France. D'après les entretiens que j'ai pu faire, je peux déduire que les

¹ OpenDataSoft est une entreprise d'édition de logiciel fondée en 2011. Elle propose différentes solutions orientées Open Data, fonctionnant comme des SaaS (Software as a Service). Elles permettent aux entreprises et acteurs publics d'ouvrir leurs données, d'utiliser des outils de visualisation etc.

² Entretien avec Jean-Marc Lazard.

³ Pour en savoir plus sur OpenDataSoft : <https://www.opendatasoft.fr/donnees-smart-city/>.

outils mis en place ne sont pas forcément utilisés de manière optimale. C'est une des limites actuelles de la Smart City, comme de l'Open Data.

c. Les valeurs sur lesquelles repose la Smart City.

Les valeurs rattachées à la Smart City, prônées par Jean-Marc Lazard comme Norbert Friant sont similaires à celles inhérentes à l'Open Data : l'ouverture, la transparence, la collaboration, l'inclusion de tous les individus, et pas seulement des entreprises et administrations.

Pour OpenDataSoft, la Smart City doit inclure le citoyen au processus, favoriser « *l'innovation civique* » :

« Donnez la possibilité à votre communauté de développer et de co-construire les solutions dont elle a besoin à partir de vos données, tout en améliorant votre stratégie de communication.

OpenDataSoft vous permet de créer un écosystème inclusif et actif de données en publiant les données de toutes les parties prenantes : service financiers et services SIG (Système d'Information Géographique), opérateurs de transport et producteurs d'énergie peuvent tous partager leurs données via votre hub.

Offrez la possibilité à votre collectivité d'être plus responsable et transparente en publiant des tableaux de bords financiers dynamiques et contextuels. Donnez l'opportunité aux développeurs et aux citoyens technophiles qui vivent sur votre territoire de construire des applications basées sur vos données ouvertes pour améliorer l'accès aux services publics. Tout est question de collaboration ! »¹.

Pour Norbert Friant, la ville intelligente doit s'appuyer sur un modèle d'économie circulaire, et d'ouverture : « *on voit aujourd'hui une ville qui s'oriente vers des modèles collaboratifs, de production locale, d'économie sociale et solidaire, de rupture avec certains modèles* ».

Au cours des entretiens, il ressort que l'échange et l'utilisation des données au sein de la Smart City sont caractérisés par l'économie collaborative. On passe d'une organisation verticale (c'est-à-dire hiérarchisée, où les entreprises et autres acteurs

¹ C'est ce qu'on peut lire sur le site d'OpenDataSoft, dans la partie qui décrit l'offre « Smart City » : <https://www.opendatasoft.fr/donnees-smart-city/>

économiques produisent et utilisent les données pour créer des produits à destination des individus), à une organisation horizontale où les individus sont intégrés dans les processus de valorisation des données (Borel *et al.*, 2015). L'objectif est de démocratiser la réutilisation des données. Norbert Friant le dit dans notre réflexion : on n'est plus « *dans une logique impérialiste, l'impérialisme de l'informaticien dans les années 1980-90, non, il faut être vraiment à l'écoute* ».

L'Open Data et la Smart City favoriseraient ainsi le « *Do it yourself* », le fait que les individus s'approprient eux-mêmes les techniques et les données mises à disposition, qu'ils s'impliquent dans la transformation de la ville afin qu'elle devienne un véritable espace de collaboration. On retrouve cette idée dans l'article « *L'économie collaborative, entre utopie et big business* » de Simon Borel, David Massé et Damien Demailly : « *les pratiques collaboratives s'inscrivent souvent dans le passage de la propriété à l'accès⁷. L'enjeu n'est plus de posséder un bien en propre (logique propriétaire) mais de pouvoir s'en servir et en disposer quand on en a besoin (logique d'usage)* » (*ibid.*, 2015, p 12). C'est bien le cas dans la ville intelligente. Cela n'a pas tellement d'intérêt de posséder des données concernant une partie précise du fonctionnement de la ville. Il faut pouvoir les traiter, les analyser pour en tirer quelque chose, ainsi que les partager pour les rendre encore plus intéressantes et utiles.

Je pense que cette vision de la Smart City dépend beaucoup de la culture au sein même de la ville concernée, de sa taille, des liens entre les habitants. Norbert Friant insiste sur cette dimension afin d'expliquer pourquoi Rennes est une pionnière en termes d'ouverture et de ville intelligente : « *[Rennes]c'est un gros village et il y a une intensité d'échanges, car les gens se connaissent, sont tous à deux trois poignées de main les uns des autres. Et donc l'intelligence de la ville, parce que qui dit "smart" dit notion d'intelligence, je pense qu'elle est dans l'intensité d'échange qui produit un apprentissage collectif. Cet apprentissage collectif, c'est à dire la naissance d'informations, c'est un peu comme un cerveau hyper connecté en fait et c'est cette rapidité, cette agilité qui a fait qu'on a pu être pionnier, premier sur ce sujet* ». Mais toutes les Smart Cities ne s'appuient pas forcément sur ce même état d'esprit. La dimension collaborative doit varier en fonction de la ville.

Face à une économie motivée par des logiques capitalistes de profits, où l'utilisation principale des données est faite dans un cadre de marketing ou de publicité, on

retrouve, en parallèle, une économie collaborative, qui semble mettre l'individu au cœur du processus de traitement des données. La valorisation des données peut donc répondre à des logiques opposées. Je pense qu'il est important d'avoir en tête les différents modèles, marchands ou non, afin de mieux comprendre comment la donnée obtient de la valeur, mais aussi comment, en tant qu'individu, il serait possible de prendre part à ces transformations, de ne plus être passif face au matraquage publicitaire, mais de s'inscrire dans un mouvement d'ouverture et de changement. Je pense que la Smart City, du moins la vision que l'on en a dans les métropoles françaises, pourrait permettre aux citoyens de se réappropriier les données.

3. DU « TROC » A LA MUTUALISATION DES DONNEES.

Mes recherches sur la Smart City m'ont fait découvrir la pratique du « troc de données ». C'est dans un article du *Journal Du Net* que j'ai pu découvrir cette stratégie, qui doit être déployée par Rennes Métropole et ses quelques entreprises partenaires¹. La ville se pose comme intermédiaire entre des entreprises qui ont intérêt à « troquer » certaines de leurs données privées contre d'autres afin d'améliorer leurs services au sein de la ville de Rennes.

Ce projet m'a semblé très intéressant, notamment du fait qu'il se détache totalement de la marchandisation des données. J'ai donc tenté d'en apprendre plus à ce propos. Enedis fait partie, d'après l'article du *Journal Du Net*, des partenaires de la ville de Rennes dans cette mise en place de troc de données. Enedis est une entreprise chargée de mission de service public, qui a donc l'obligation, depuis 2016, de partager des jeux de données en libre accès sur internet². J'ai donc pris contact avec Jean-François Montagne, responsable de la gouvernance des données chez Enedis.

Il serait intéressant ici de voir à quoi correspond le concept de troc. D'après Jacques Sapir (2002), il se développe « *comme une réaction à une perte de confiance dans une monnaie, dans le cadre d'une hyper-inflation, poussant les agents à refuser cette monnaie comme moyen de paiement ou instrument d'épargne en raison du risque de perte de valeur de cette monnaie* ». Cette définition m'évoque la situation actuelle de croissance

¹ <https://www.journaldunet.com/economie/services/1206939-rennes-metropole-echange-donnees-privées/>

² <https://data.enedis.fr/explore/?sort=modified>

exponentielle et infinie des données brutes¹, entraînant une baisse de confiance face à un volume d'informations non traitées, dont la valeur n'est pas assurée.

L'idée de troc pourrait donc avoir germé avec l'apparition du Big Data et de ses problématiques. En réponse à cette perte de confiance dans la donnée de masse, on pourrait imaginer des échanges d'informations ciblées et précises entre des entreprises dont les besoins se recoupent. Jacques Sapir (2002) dit également que « *le troc permet d'imposer des prix relatifs différents de ceux des échanges monétaires* ». Ainsi, la valeur ne se fonde plus sur les concepts de rareté, travail, ou utilité, ni même sur le jeu de l'offre et de la demande, mais plutôt sur une forme d'utilité subjective de chaque acteur. Ainsi deux entreprises peuvent s'échanger des données qui auront une valeur subjective réciproque, dans l'utilisation que veulent en faire les acteurs à un instant précis (ces jeux de données n'auront pas la même valeur pour différents acteurs, à différents moments).

En théorie, cette idée semble très intéressante pour la valorisation des données, puisqu'elle permet de donner une valeur non marchande, mais elle dépend des usages des différents acteurs. Je vais présenter, dans les faits, comment ce troc de données pourrait se mettre en place concrètement.

Tout d'abord je vais montrer, à partir de mes réflexions et de l'entretien avec Jean-François Montagne, que le « troc de données » est une idée assez vague et problématique, qui penche plutôt vers une mutualisation des données.

Je donnerai ensuite un exemple concret de mutualisation des données par un acteur du secteur privé : Airbus.

a. Dans les faits, le troc est difficile à envisager.

Ma source principale pour me renseigner sur les pratiques concrètes de troc de données était mon entretien avec Jean-François Montagne. Je l'ai donc questionné sur le projet en partenariat avec Rennes Métropole, mais il n'a pas su me répondre. Il a d'ailleurs été assez surpris en découvrant l'article du *Journal Du Net*, qu'il n'avait jamais lu. Il semblait que c'était la première fois qu'il entendait parler de ce projet. Au final, plus que des informations précises, nous avons surtout questionné ensemble cette idée de troc de données.

¹ Ici, « brutes » dans le sens « en vrac ».

En tant qu'entreprise chargée de service public, Enedis se doit d'ouvrir une partie de ses données depuis 2016. Je ne reviendrai pas sur cette question de l'ouverture, déjà étudiée plus haut, mais je peux juste re-dire que les données accessibles en ligne gratuitement sont des données non sensibles ou confidentielles : *« on ne peut pas transmettre les fiches de paye, ou les données personnelles, sensibles. En revanche c'est notre devoir, on l'a compris, avec cette démarche Open Data, que maintenant on doit être en partenariat, en symbiose avec l'écosystème des parties prenantes »*¹.

Pour Jean-François Montagne, plutôt que du troc, on aurait des exemples de mutualisation des données. On a déjà parlé de la Smart City, où des plateformes trans-sectorielles peuvent permettre à des acteurs différents de partager leurs données (comme la plateforme Smart City d'OpenDataSoft). Il y a également d'autres exemples de mutualisation qui ne découlent pas forcément d'initiatives rattachées à la ville intelligente, mais bien des acteurs eux-mêmes. Jean-François Montagne donne l'exemple du partage de données entre Enedis et les territoires sur lesquels on retrouve des postes d'Enedis : *« On a des capteurs pour les inondations dans nos postes, et par ailleurs, le territoire, sur lequel ces postes sont, a des capteurs aussi. Si on met en commun les données de ces capteurs, on peut voir plus précisément l'inondation qui est en train de monter. C'est à peu près la même donnée, mais chacun enrichit l'autre en la partageant, donc c'est plus de la mise en commun que du troc »*.

Très clairement, si l'entretien mené avec Jean-François Montagne a été très intéressant, il nous a été impossible de définir clairement la notion de troc de données, ni de trouver d'exemple concret. Il m'a néanmoins parlé de Dawex (dont je parle rapidement dans la partie I du mémoire), et de leur *« data marketplace »* : *« je crois qu'il y avait la possibilité de passer par une forme de troc. Il proposait une forme de contractualisation pour mettre à disposition des données, en les échangeant ou en les vendant. C'est la seule fois où j'ai entendu parler de troc de données »*. Ainsi la notion de *« troc de données »* que j'avais découverte dans l'article du *Journal Du Net* et que j'avais trouvée très intéressante et prometteuse ne semble pas se rattacher à une réalité tangible.

Comment le troc de données pourrait-il fonctionner ? Comme le soulignait Jean-François Montagne : *« ce n'est pas évident de se dire "je t'échange une donnée météo*

¹ Entretien avec Jean-François Montagne.

contre une donnée innovation" ». Je pense néanmoins qu'il serait intéressant d'approfondir cette idée de troc, ou voir si des initiatives de ce type se développent, et comment elles gèrent la question de la valeur subjective de certains jeux de données par rapport à d'autres.

b. Skywise : un exemple de mutualisation des données privées.

La Smart City présentée plus haut peut-être un exemple concret de mutualisation des données, dans une certaine mesure. Il existe également des initiatives privées, avec des objectifs plus concrets motivant le partage des données entre différents acteurs.

Les entreprises privées ne mettent pas leurs données à disposition. Elles peuvent néanmoins être à l'origine d'initiative de mise en commun « privée » des données dans le cadre d'un service particulier.

Airbus a lancé en 2017 sa plateforme Skywise¹. C'est une base de données qui regroupe les informations de plusieurs compagnies aériennes. L'objectif de cette plateforme est de réunir le maximum d'acteurs du secteur de l'aéronautique, afin de mettre en place des services de maintenance, de gestion des risques, de prévention grâce aux données mises en commun, permettant ainsi des économies importantes pour les compagnies aériennes.

Au moins douze acteurs de l'aéronautique ont rejoint la plateforme depuis son lancement, dont Easyjet et Emirates. Elle récupère donc les données concernant plus d'un millier d'avions, sachant que pour chaque vol, elle peut collecter deux-cents mégaoctets d'informations².

Les données récoltées sont stockées sur un cloud fourni par Amazon et traitées par l'entreprise américaine d'analytics Palantir qui, d'après le président exécutif d'Airbus Group, Tom Enders « [...] est la meilleure entreprise de data analyse du monde »¹. Sa technologie permet de traiter les données hétérogènes collectées, de les visualiser, de les rendre intelligibles et utilisables.

D'après Palantir, la plateforme Skywise permet concrètement d'améliorer la maintenance des systèmes, de mieux coordonner les horaires des vols, ainsi que les

¹ <https://www.airbus.com/aircraft/support-services/skywise.html>

² <https://www.usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963>

réservations des passagers. Les informations collectées et traitées vont aider à identifier des défauts ignorés avant (pour les régler plus rapidement), à éviter des dépenses inutiles, à mieux gérer les changements d'horaires des avions etc.

La mutualisation des informations des compagnies aériennes est donc bénéfique pour celles-ci car elle permet d'assurer une plus grande sécurité, et une plus grande efficacité dans tous les processus, de la maintenance à l'expérience client.

A partir de cet exemple, on peut voir en quoi la mise en commun des données, sans passer par une logique marchande, permet de créer de la valeur.

4. POUR CONCLURE SUR LES AUTRES FORMES DE VALORISATION DES DONNEES ...

Dans cette partie j'ai voulu montrer d'autres exemples de valorisation des données qui ne passent pas par une marchandisation, ou du moins qui s'éloignent d'une logique purement capitaliste reposant sur la publicité et le marketing.

Ces exemples reposent principalement sur les entretiens que j'ai menés, sur la lecture de travaux scientifiques, et d'articles de presse. Il serait intéressant de plus les développer, pour en montrer les limites principales, présenter les enjeux etc. Ce n'était pas mon objectif ici.

Pour moi, il était indispensable de présenter ces alternatives à la vision purement capitaliste du « marché des données », afin de se rendre compte qu'il est possible d'intégrer les individus, de créer de la collaboration entre les acteurs grâce aux données. Qu'une économie alimentée par les données n'est pas forcément une économie qui met les plus petits à la merci des GAFAs, et autres acteurs importants du numérique.

¹ <https://www.usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963>

Il serait peut-être intéressant de déplacer l'attention des individus du « marché des données », vers l'utilisation collaborative des informations dans la construction d'une société véritablement¹ centrée sur les citoyens et les utilisateurs.

¹ Je dis « véritablement » dans le sens où les GAFA (entre autres) justifient leurs activités en disant que l'individu est au centre alors qu'en réalité, ils cherchent principalement à récupérer l'attention de ces derniers afin de récolter des bénéfices liés aux revenus publicitaires, comme je l'ai montré dans la deuxième partie

Conclusion

L'expression « marché des données », aussi répandue soit-elle, est problématique car elle fait référence à un écosystème bien plus complexe qu'il n'y paraît. Des données de différentes natures sont collectées puis utilisées par des acteurs nombreux et variés, dans des modèles divers. Ainsi, le fait de questionner le bien fondé d'un « marché des données », m'a permis de m'intéresser à différents modèles de valorisation des données.

Dans ce mémoire, j'ai donc tout d'abord déconstruit la vision homogène que l'on a des données, en les présentant sous trois catégories : les données personnelles, les données ouvertes dans le cadre de l'Open Data et les données privées d'entreprise. Elles ne sont pas soumises aux mêmes lois, et ne sont pas toutes récoltées puis traitées de la même manière.

En présentant cet écosystème hétérogène, j'ai, dans un sens, posé le décor, la base afin d'étudier ensuite les différents types d'échanges et de mise en valeur en fonction des spécificités de ces données.

J'ai ensuite voulu étudier la « marchandisation des données ». En réalité, j'ai surtout déconstruit le « marché des données », afin de montrer qu'il s'agit d'un ensemble d'échanges organisés autour d'acteurs différents (internauts, GAFAs, Data Brokers, annonceurs, etc.), qui ont des utilisations différentes des données (production, collecte, traitement, production de services à partir de la donnée et revente de ceux-ci etc.). Il ne s'agit donc pas d'un marché à proprement parler, mais plutôt d'une économie capitaliste irriguée par les données personnelles. Celles-ci sont exploitées par les différents acteurs du numériques, et mises en valeur avec pour seul objectif principal : le profit économique.

Face à cela, d'autres modèles émergent, qui se fondent sur l'ouverture des données. L'Open Data peut permettre de créer de la valeur, lorsque des acteurs (professionnels, ou amateurs) s'emparent des données pour les transformer en services.

Au sein de la Smart City, la collaboration entre les habitants, la ville et les entreprises se base sur la réutilisation des données. Elle s'oriente vers un plus grand bien-être et une plus grande efficacité des services.

Certains acteurs privés, choisissent également d'ouvrir, dans une certaine limite, leurs données afin de partager des connaissances, et d'améliorer leurs processus.

Ces modèles de valorisation sont motivés en partie par des volontés économiques, mais aussi par des valeurs d'ouverture, de transparence et de collaboration.

Ces modèles de valorisation des données ne s'organisent pas de la même manière, ne visent pas les mêmes objectifs, mais fonctionnent selon une idée similaire : la donnée doit circuler. Que ça soit en incitant les individus à créer à partir de données en libre accès, où en collectant, transformant en services les données personnelles, on voit bien que ce qui a de la valeur, ce n'est jamais la donnée « en vrac », qui reste dans sa base. Elle doit être traitée, raffinée, partagée, utilisée. La donnée ne vaut pas grand-chose en elle-même. Ce sont les services que l'on peut créer en l'utilisant qui possèdent de la valeur.

La donnée est la matière première de cette économie, tant du côté de l'exploitation des données personnelles, que du côté de l'incitation à l'innovation. On a affaire à une économie irriguée par les données sur laquelle se créent différents modèles de valorisation, certains plus tournés vers une exploitation capitaliste de la matière première, d'autres ayant plus une logique d'ouverture et de collaboration.

Avec les différents scandales qui éclatent, le sujet de l'utilisation des données est assez sensible. La façon dont il est traité par les médias et présenté au grand public n'arrange pas les choses. Je pense qu'il est important de déconstruire l'idée d'un « marché des données ». Il faut montrer comment elles sont réellement utilisées et mises en valeur, pour favoriser la prise d'initiatives citoyennes dans ce domaine.

Les modèles économiques développés par les GAFAs et les Data Brokers sont hégémoniques dans l'exploitation des données personnelles. En parallèle, l'ouverture des données est une opportunité pour les individus qui souhaitent innover et s'inscrire dans cette nouvelle économie irriguée par la donnée.

L'utilisation des données, de manière générale, ne doit donc pas forcément susciter de l'inquiétude. Il faut bien prendre en compte le type de données dont on parle, ainsi que le modèle qui les met en valeur.

Ce mémoire m'a permis d'y voir plus clair sur l'organisation de l'économie irriguée par les données. Déconstruire l'idée de « marché des données » m'a incité à m'intéresser aux autres formes d'utilisation des données. La Smart City est un modèle que j'ai trouvé très riche à étudier, et qui mériterait une analyse plus profonde.

Les initiatives portées par les villes et les communes françaises sont variées, et dépendent, entre-autre, de la culture qui existe à la base. Il serait intéressant de voir, par exemple, le niveau d'intégration des citoyens à l'organisation de la ville, et si grâce à leur expérience du quotidien, ils peuvent aider à améliorer les services.

Les entreprises qui participent au fonctionnement des services de la ville seront sûrement de plus en plus incitées à ouvrir une partie de leurs données. Il faudrait s'intéresser à la façon dont cela se met en place. A cela des questions de sécurité se poseront, auxquelles répondront des technologies nouvelles comme le Cloud ou la Blockchain. Les professionnels, les startups, les citoyens seront encore incités à s'emparer des données pour les transformer en services.

Pour que tout cela se mette en place de manière efficace, il serait bon de diffuser cette culture de l'ouverture des données, aussi bien parmi les citoyens, que dans les entreprises et les administrations. L'adhésion de tous les acteurs de la Smart City doit accompagner les avancées technologiques, pour permettre à ces nouveaux modèles d'organisation de s'implanter durablement.

Bibliographie

Ouvrages :

BENYAYER Louis-David (2015), *Open Models : les business modèles de l'économie ouverte*, 4ePub, 227 pages.

Chapitre : « Les Open Models de la Data », coordonné par BONNET Chloé pp 66 - 102.

BENYAYER Louis-David, CHIGNARD Simon (2015), *Datanomics, Les nouveaux business models des données*, FYP EDITION, 192 pages.

CARDON Dominique, CASILLI Antonio (2016), *Qu'est-ce que le digital labor ? Les enjeux de la production de valeur sur Internet et la qualification des usages numériques ordinaires comme travail*, INA, 104 pages.

CARMES Maryse, NOYER Jean-Max (2013), *Les débats du numérique*, Transvalor - Presse des Mines, 280 pages.

Chapitre : « Le mouvement "Open Data" dans la grande transformation des intelligences collectives », pp 137-168.

ROCHELANDET Fabrice (2010), *Economie des données personnelles et de la vie privée*, La Découverte, 128 pages

Sonnac Nathalie, Gabszewicz Jean (2013), *L'industrie des médias à l'ère numérique*, La Découverte (3^{ème} édition), 128 pages.

STEINER Philippe, TRESPEUCH Marie (2015), *Marchés contestés : quand le marché rencontre la morale*, Presse Universitaire du Mirail, 450 pages.

Chapitre 7 : « La mise en marché des "données personnelles" ou la difficile extension du marché à la personne » pp 219-250.

Articles :

ANCI AUX Arnaud, FARCHY Joëlle, MEADEL Cécile (2017), « L'instauration de droits de propriété sur les données personnelles : une légitimité économique contestable », *Revue d'économie industrielle*.

ASDOURIAN Bruno, ZIMMERLI Virginie (2015), « *Open data*, usagers du numérique et entreprises publiques », *Revue française des sciences de l'information et de la communication*.

BELLEFLAMME Paul (2018), « Modèles économiques des données : une relation complexe entre demande et offre », *Enjeux Numériques* (n°2), pp 9-13.

BENYAYER Louis-David, CHIGNARD Simon, « Open Data : Show me the money ! » (2015), *Open Models : les business modèles de l'économie ouverte*, 4ePub, pp 74-78.

BONNET Chloé, LALANNE Romain (2015), « De la *Killer app* à l'économie des plateformes », *Open Models : les business modèles de l'économie ouverte*, 4ePub, pp 68-73.

BOULLIER Dominique (2009), « Les industries de l'attention : fidélisation, alerte ou immersion », *Réseaux* (vol 154, n°2), pp 231-246

BOURDONCLE François (2014), « Peut-on créer un écosystème français du Big Data ? », *Le journal de l'école de Paris du management* (vol. 108, no. 4), pp 8-15

BURGER-HELMCHEN, Thierry (2011), « Crowdsourcing : définition, enjeux, typologie », *Management & Avenir* (vol. 41, no. 1), pp. 254-269.

CARDON Dominique (2006), « La trajectoire des innovations ascendantes : inventivité, coproduction et collectifs sur Internet », *Charnet, C. Innovations, usages, réseaux*.

CECERE Grazia et al. (2015), « Les modèles d'affaires numériques sont-ils trop indiscrets ? Une analyse empirique », *Réseaux* (n° 189), pp 77-101.

CECERE Grazia, MANANT Matthieu (2018), « Vie privée, valeur des données personnelles et régulation », *Enjeux Numériques* (n°2), pp 14-19.

CYTERMANN Laurent (2015), « Promesses et risques de l'open et du Big Data : les réponses du droit », *Informations sociales* (n° 191), pp 80-90.

DENIS Jérôme, GOËTA Samuel (2013), « La fabrique des données brutes. Le travail en coulisses de l'open data », *Penser l'écosystème des données. Les enjeux scientifiques et politiques des données numériques*.

D'ELIA Danilo (2014), « La guerre économique à l'ère du cyberspace », *Hérodote* (n° 152-153), pp 240-260

DUMOULIN Régis Dumoulin, LANCELOT MILTGEN Caroline (2012), « Entreprise et respect de la vie privée du consommateur. De l'usage autorisé à l'utilisation souhaitable des données personnelles », *Revue française de gestion* (N° 224), pp 95-109.

GOËTA Samuel, MABI Clément (2014), « L'open data peut-il (encore) servir les citoyens ? », *Mouvements* (n° 79), pp 81-91.

ISAAC Henri (2018), « La donnée, une marchandise comme les autres ? », *Enjeux Numériques* (n°2), pp 20-24.

OTTO Paul N. (2006), « The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information », 5-5 *IEEE Security and privacy*.

PEUGEOT Valérie (2010), « Les enjeux publics, économiques et citoyens de l'ouverture des données : l'expérience britannique », *DocSoc*.

RIALLE Vincent, « Villes intelligentes sources d'inspiration » (2017), *L'Officiel des Aînés* (3e édition).

SAPIR Jacques (2002), "Le troc et le paradoxe de la monnaie", *Journal des anthropologues*, pp 283-303.

Mémoires :

CHEVALIER Michael (2016), « Les enjeux juridiques concernant les nouveaux modèles d'affaires basés sur la commercialisation des données », Université de Montréal, 118 pages.

GALOUL Alessia (2015), « "Les villes intelligentes" : l'Open Data contribue-t-il à leur développement ? », Louvain School of Management (Université Catholique de Louvain), 91 pages.

Évènements :

GOËTA Samuel, « Intéresser à la réutilisation des données : la participation des développeurs aux concours d'applications open data », in GIS Démocratie et Participation, *Actes des 3^{èmes} journées doctorales sur la participation et la démocratie participative*, Bordeaux, 22-23 novembre 2013

VAN CUYCK Alain, « L'open data comme nouvelle forme de gouvernance numérique : enjeux, marchés, modèles, idéologies ». XIX^o colloque international franco-roumain « Culture et Responsabilité sociale dans la communication des organisations », Mars 2013, Bucarest, Roumanie.

Rapports :

BREUX Sandra, DIAZ Jérémy, « La ville intelligente : origines, définitions, forces et limites d'une expression polysémique », Institut National de la recherche scientifique Centre – Urbanisation, culture, société, Rapport remis à la ville de Repentigny, Janvier 2017, 37 pages.

THIEULIN Benoît, « Ambition numérique. Pour une politique française et européenne de la transition numérique », Rapport remis au Premier Ministre, juin 2015, 399 pages.

Conseil d'État, *Le numérique et les droits fondamentaux*, Étude annuelle 2014, 446 pages.

Federal Trade Commission, *Data Brokers, a call for transparency and accountability*, CreateSpace Independent Publishing Platform 19 mai 2014, 110 pages.

Webographie :

<https://www.blogdumoderateur.com/>

<https://www.lebigdata.fr/>

<https://www.cnil.fr/>

<https://www.journaldunet.com/>

<https://wikipedia.org/>

<https://www.inaglobal.fr/>

<https://www.definitions-marketing.com/>

<https://www.lescahiersdelinnovation.com/>

<https://openinnovation.wordpress.com/>

<https://www.gouvernement.fr/>

<https://www.etalab.gouv.fr/>

<https://donneesouvertes.info/> (Blog de Simon Chignard).

Sources

Entretiens :

- Entretien avec Laurent Rouillot le 20 novembre 2017, durée : 01:01:06
- Entretien avec Théo Roux le 20 février 2018, durée : 01:08:59
- Entretien avec Grégoire Guilmin le 23 avril 2018, durée : 00:27:49
- Entretien avec Jean-Marc Lazard le 4 mai 2018, durée : 01:21:00
- Entretien avec Jean-François Montagne le 5 juin 2018, durée : 00:44:19
- Entretien avec Michael Chevrier le 9 juin 2018, durée : 00:31:24

« Marché des données », « nouvel or noir » :

<https://www.journaldunet.com/solutions/expert/67484/l-or-noir-du-numerique--le-big-data-nous-livre-ses-secrets.shtml>

https://www.francetvinfo.fr/replay-radio/nouveau-monde/le-big-data-nouvel-or-noir-de-l-economie-numerique_1773873.html

<https://www.lesechos.fr/partenaire/le-comptoir-mm-de-la-nouvelle-entreprise/partenaire-1185-data-le-nouvel-or-noir-2113238.php>

<https://www.lesechos.fr/idees-debats/cercle/cercle-183763-le-siecle-de-la-donnee-2182542.php>

https://www.lesechos.fr/07/03/2013/lesechos.fr/0202599460114_la-ruée-vers-l-or-des-donnees-personnelles.htm#

<https://www.latribune.fr/supplement/internet-et-revente-de-donnees-personnelles-un-marche-encore-timide-mais-qui-gagne-chaque-jour-du-terrain-751177.html>

<https://www.usine-digitale.fr/article/le-marche-de-la-data-en-france-estime-a-1-5-milliard-d-euros.N365258>

https://www.francetvinfo.fr/societe/cash-investigation-le-marche-colossal-des-donnees-personnelles_1115585.html (reportage)

https://www.lemonde.fr/economie/article/2017/05/28/des-donnees-personnelles-tres-convoitees_5135092_3234.html

<https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>

SAFARI :

<https://donneesouvertes.info/2018/01/26/safari-la-chasse-aux-francais-40-ans-apres/>
<https://sites.ina.fr/cnil-40-ans/focus/chapitre/4>

Modèle d'affaire de Facebook :

<https://fr-fr.facebook.com/about/privacy/>
<https://daniloduchesnes.com/blog/les-6-erreurs-de-debutant-en-publicite-facebook-et-comment-les-eviter/>
<https://www.facebook.com/about/privacy/update#sharing-partner-information>
https://www.lesechos.fr/02/02/2017/lesechos.fr/0211760299478_les-revenus-publicitaires-de-facebook-en-forte-hausse.htm

Les données privées d'entreprises :

https://www.entreprises.gouv.fr/files/files/directions_services/information-strategique-sisse/Guide_LA_PROTECTION_DES_INFORMATIONS_SENSIBLES_DES_ENTREPRISES.pdf
<https://www.sekurigi.com/2015/09/la-protection-des-donnees-sensibles-dentreprise/>
<https://www.solutions-numeriques.com/deploiement-g-suite-airbus-repond-aux-critiques-quant-a-la-securite-de-ses-donnees/>

La SNCF :

<https://www.nouvelobs.com/rue89/rue89-nos-vies-connectees/20141211.RUE6955/mais-pourquoi-la-sncf-planque-t-elle-les-horaires-des-tgv.html>
<https://www.usinenouvelle.com/article/sncf-ratp-transdev-et-blablacar-souhaitent-mettre-en-commun-leurs-donnees.N581278>

Dawex :

<http://www.bpifrance.fr/A-la-une/Actualites/Dawex-futur-airbnb-de-la-data-28607>
<https://www.usine-digitale.fr/article/le-francais-dawex-veut-creer-une-bourse-mondiale-de-la-donnee.N577983>

<https://www.maddyness.com/2016/06/02/datas-dawex-leve-2-millions-euros/>
<http://www.larevuedudigital.com/partager-et-monetiser-ses-donnees-en-toute-legalite-avec-dawex/>

Datacoup :

<http://datacoup.com/docs#how-it-works>
<https://atelier.bnpparibas/smart-city/article/datacoup-remunere-donnees-personnelles-transactions-cartographie>

Les Data Brokers :

<https://www.lebigdata.fr/top-5-data-brokers-vendeurs-donnees>
[https://fr.wikipedia.org/wiki/Nielsen_\(entreprise\)](https://fr.wikipedia.org/wiki/Nielsen_(entreprise))
<https://fr.wikipedia.org/wiki/Axiom>
<http://www.mediadatalive.com/acteurs/mediaprism.html>

Open Data et Smart City :

<http://www.lagazettedescommunes.com/438803/open-data-quels-couts-pour-les-collectivites-locales/>
<https://data.rennesmetropole.fr/>
<https://www.journaldunet.com/economie/services/1206939-rennes-metropole-echange-donnees-privees/>
<https://usbeketrica.com/article/sidewalk-labs-quartier-google-toronto>
<http://data.ratp.fr/>
<http://data.enedis.fr/>
<https://www.opendatasoft.fr/donnees-smart-city/>

Troc de données :

<https://www.airbus.com/>
<https://www.airbus.com/aircraft/support-services/skywise.html>
<https://www.usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963>
<https://www.usinenouvelle.com/article/avec-skywise-airbus-lance-l-internet-des-avions.N618963>

Table des matières

<i>Introduction</i>	5
<i>Partie I : Déconstruction de l'image homogène des données.</i>	15
1. Présentation des données personnelles et de leurs modes de collecte.	17
a. Quelle définition des données à caractère personnel ?.....	18
b. Quelles techniques de collecte des données à caractère personnel ?	20
c. Les données personnelles sont-elles privées ?.....	25
d. Pour conclure sur les données personnelles	27
2. Présentation du mouvement d'Open Data et des données sujettes à l'ouverture.	28
a. Pourquoi ouvrir ses données ?.....	29
b. Qui ouvre ses données ?	32
c. Comment ouvrir ses données ?.....	33
d. Pour conclure sur les données d'Open Data	37
3. Les données privées d'entreprises : quels enjeux et quelles utilisations ?	38
a. Les données d'entreprises sensibles et confidentielles.	38
b. Les données privées moins sensibles et leur utilisation.	40
c. Pour conclure sur les données privées d'entreprises	41
<i>Partie II : « Le marché des données » : une expression problématique pour désigner l'exploitation économique des données.</i>	43
1. L'activité de production des données par les consommateurs.	45
a. Le <i>digital labor</i> : quand le consommateur devient producteur.....	46
b. A ce niveau d'échange, les données personnelles peuvent-elles être considérées comme des marchandises ?	47
c. Une réappropriation de ces données en passant par leur monétisation.	48
2. L'activité de valorisation des données par les plateformes numériques.	49
a. Un marché de l'attention, plus qu'un marché des données.	51
b. Le marché à deux versants, caractéristique des économies de plateforme.....	51
c. L'exemple du modèle d'affaire de Facebook.....	52
3. L'activité de marchandisation des données par les Data Brokers.	54
a. Qui sont les Data Brokers ?.....	55
b. Comment se déroule la collecte des données ?.....	56
c. Comment se déroule la revente des données ?	58
d. Pour conclure sur les Data Brokers	61

4. Critique de la notion de « marché des données personnelles ».....	62
<i>Partie 3 : Des modèles tournés vers l'ouverture pour d'autres formes de valorisation des données.....</i>	65
1. Comment créer de la valeur avec l'Open Data ?	67
a. L'Open Data comme source d'innovation.	68
b. L'émergence d'intermédiaires pour une meilleure circulation des données.....	71
c. Les limites de ces modèles de valorisation des données en libre accès.....	72
2. La Smart City : créer de la valeur ensemble, en collaboration.	74
a. Qu'est-ce que la Smart City ?	74
b. La Smart City est un espace de circulation et de valorisation des données.....	76
c. Les valeurs sur lesquelles repose la Smart City.....	79
3. Du « troc » à la mutualisation des données.	81
a. Dans les faits, le troc est difficile à envisager.	82
b. Skywise : un exemple de mutualisation des données privées.....	84
4. Pour conclure sur les autres formes de valorisation des données	85
<i>Conclusion</i>	87
<i>Bibliographie.....</i>	90
<i>Sources.....</i>	95
<i>Annexe 1</i>	101
<i>Annexe 2</i>	102
<i>Annexe 3</i>	103

Annexe 1

Traceurs	Définition
Cookies	Les « cookies » sont des fichiers texte stockés sur l'ordinateur de l'internaute lorsqu'il consulte des sites. À l'origine, ces fichiers permettaient d'enregistrer des informations sur les préférences de l'internaute afin de faciliter son « expérience de navigation ». Ces fichiers sont désormais controversés, car ils peuvent contenir des données personnelles parfois accessibles aux parties tierces.
IP dynamique	Une adresse IP permet d'identifier de façon unique un ordinateur sur le réseau. Cela rend possible la constitution de bases de données sur les comportements de navigation d'une connexion à une autre. Lorsque cette adresse devient dynamique, elle change au fur et à mesure des connexions, ce qui rend le <i>tracking</i> plus difficile. Néanmoins, il est désormais possible d'utiliser les caractéristiques de l'ordinateur (résolution de l'écran, système d'exploitation, navigateur, etc.) pour identifier un ordinateur de façon unique avec 99 % de chances de succès. Plus largement, l'internet des objets, par définition mobiles, nécessite un identifiant unique et permanent pour chaque appareil, ce qui va rendre caduc l'usage de l'adresse IP.
DOM storage	Le « DOM storage » est une technologie similaire à celle des cookies, mais elle permet de stocker une plus grande quantité d'information de manière pérenne.
Web bugs	Les « Web bugs » sont des objets invisibles (par exemple des images d'un pixel – 1x1 gif) incorporés sur les sites internet afin de s'assurer qu'un individu a bien accédé à un contenu spécifique.
evercookies	Les cookies classiques sont des fichiers texte qui peuvent être effacés de l'ordinateur de façon automatique ou au bout d'un certain temps. Les « evercookies » correspondent à des technologies qui rendent la destruction de ces fichiers pratiquement impossible.

Cecere *et al.*, 2015, p 97

Annexe 2

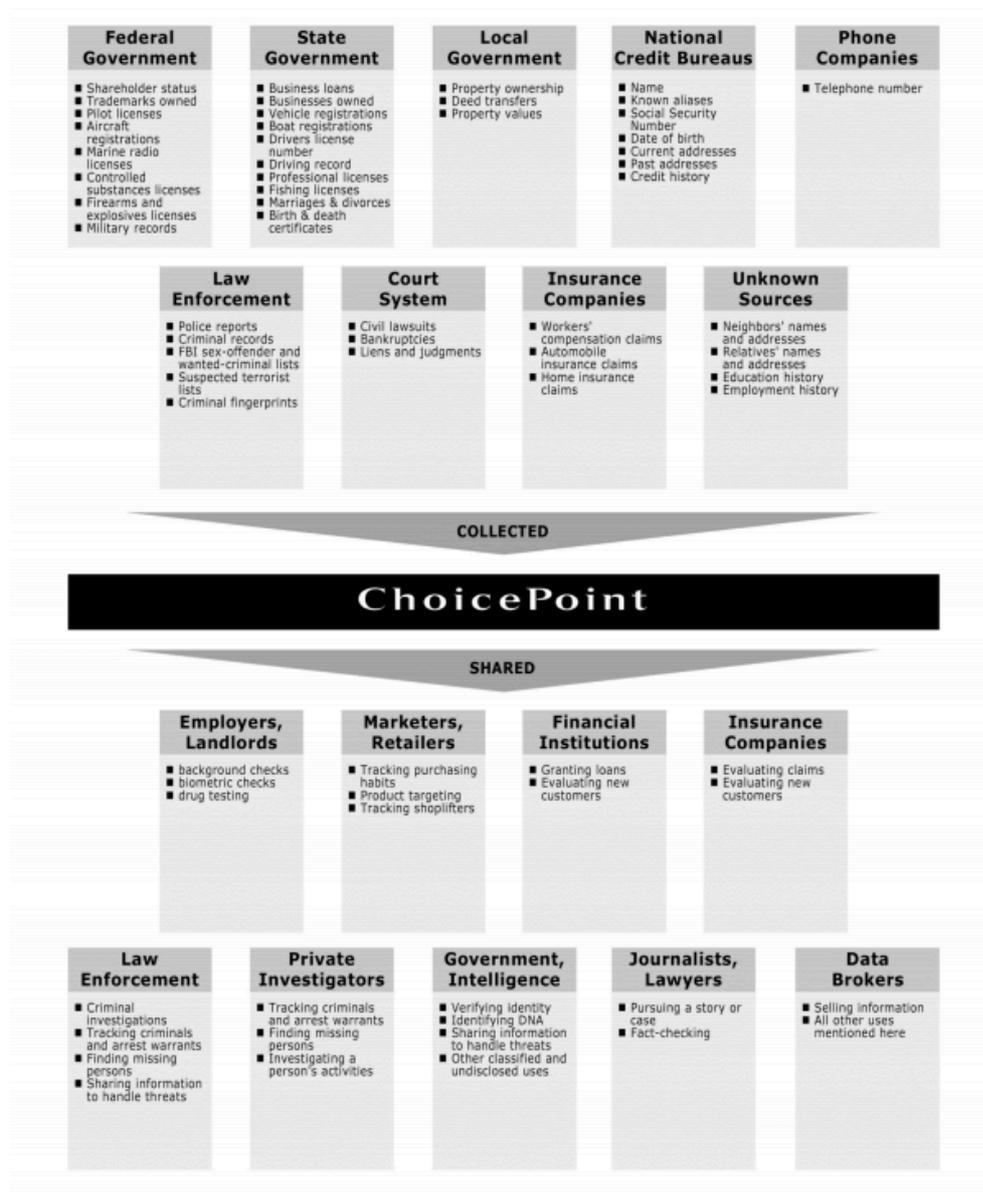
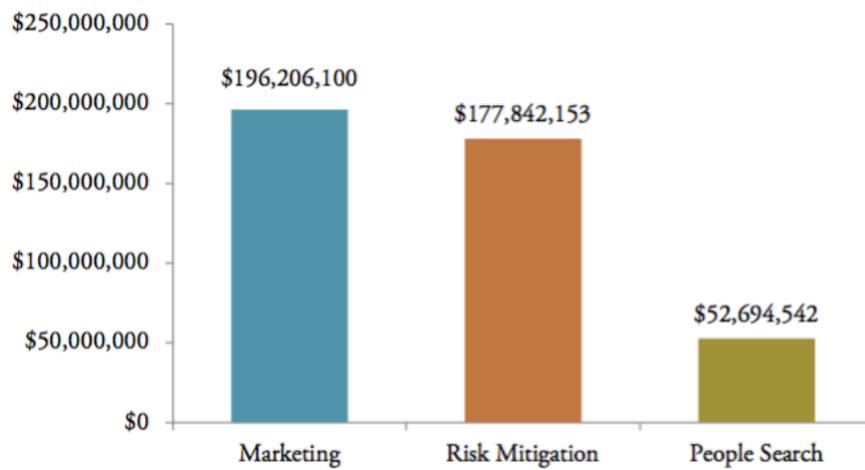


Figure 1: The Information Flow Model of ChoicePoint's Operations. The rounded boxes below the center ChoicePoint rectangle represent data leaving ChoicePoint, while the rectangles above ChoicePoint's name reflect sources of data entering the company.

Otto Paul N., *et al.*, 2006, p 3.

Annexe 3

Exhibit 3: Revenue of Nine Data Brokers by Product Category



Federal Trade Commission, 2014, p 23.



Résumé :

L'expression de « marché des données », très répandue dans les médias et dans l'imaginaire collectif, suscite aujourd'hui un grand nombre de fantasmes. Entre les énormes profits réalisés par les GAFAs, et les scandales à propos du commerce des données personnelles, c'est toujours du même « marché » dont il est question. Seulement, dès que l'on tente de comprendre ce qu'il y a vraiment derrière ce « marché des données » (les acteurs, l'organisation des échanges), on découvre une réalité complexe et difficile à saisir. En déconstruisant cette expression, on découvre un écosystème riche, composé d'une grande diversité d'acteurs qui s'approprient et valorisent les données selon des modèles différents.

Les données ne forment pas un tout homogène. Données personnelles, données ouvertes (Open Data), ou encore données privées des entreprises (confidentielles ou non) ne sont pas soumises aux mêmes régulations, et répondent à des usages différents.

Derrière le terme de « marché », des échanges variés, pas forcément marchands, entre différents acteurs (internauts, GAFAs, Data Brokers, annonceurs etc.) se basent sur l'exploitation intensive des données personnelles (presque exclusivement). Permettant d'offrir des services publicitaires et marketing, la donnée n'est pas échangée pour elle-même, mais bien pour ce qu'elle permet de faire. Elle obtient de la valeur à travers les services qu'elle permet de produire (ciblage publicitaire, marketing analytique etc.) Plus qu'un « marché des données », il faudrait parler d'une économie capitaliste irriguée par la donnée.

Enfin, au-delà de cette économie fondée sur le profit, d'autres types de valorisation des données émergent, reposant sur l'ouverture de celles-ci (Open Data, Smart City). En effet, ouvrir ne veut pas dire détruire, et donner accès à ses données librement peut générer de la valeur, parallèlement à l'exploitation capitaliste des données.

Mots clés : Marché des données, Data, Big Data, Open Data, Smart City, économie circulaire, données personnelles, données privées

Nota : cette page, dernière de couverture, sera retournée avant reliure.