

Université catholique de Louvain - Université de Namur

Economics School of Louvain (ESL) - Economics School of Namur (ESN)

L'innovation technologique par les données

D'une économie de la donnée personnelle au big data

Auteur : Cyrille LEQUEUX

Promotrice : Isabelle FERRERAS

Année académique : 2019 - 2020

Master en économie - 60 crédits

Remerciements

Je voudrais remercier toutes les personnes qui ont participé au succès de ce TFE et de ma formation.

Un grand merci à Madame FERRERAS, ma superviseuse de TFE, pour son soutien, ses précieux conseils et encouragements. Elle a su me guider dans la construction et la rédaction de mon travail académique.

Merci à mes proches, notamment ma maman et mes amis pour leur présence tout au long de ces quatre années d'études. Leur soutien pendant la rédaction de ce TFE a été d'une aide inestimable.

Enfin, un grand merci à mes collègues de promotion pour leur bonne humeur.

1. Motivations et périmètre

Le *big data* est à la mode. Il est tant un sujet de débat public et médiatique qu'académique. Qu'il suscite de l'enthousiasme ou de la peur, qu'il soit critiqué ou vénéré, nul ne peut ignorer son impact économique croissant. Certains évoquent la « révolution technologique » - moteur d'innovations qui transformeront de façon structurelle et profonde notre civilisation de demain -, d'autres y dénoncent la « dégradation du savoir-faire humain ». Le débat fait rage, les positions sont tranchées.

Après l'obtention en 2014 de mon Bachelier en sciences économiques, des amis et moi-même avons eu l'occasion de lancer une *startup* dans le domaine de la *fintech* - une industrie associant *innovation* financière et *big data*. Nous y avons travaillé pendant plus de cinq ans. Les motivations sous-jacentes de ce Travail de Fin d'Étude (TFE) sont donc doubles : d'une part, comprendre ce qu'est réellement la donnée numérique et son impact économique dans une société du « tout hyper-connecté », le *big data* s'immisçant dans les rouages d'un capitalisme informationnel où le caractère prédictif de la donnée prend une place prépondérante dans les prises de décision ; d'autre part, s'interroger sur la tension qui s'exerce entre la volonté d'innover, qui requiert toujours plus de données, et celle de garantir le droit à la vie privée.

Nous commencerons donc par définir en termes précis la notion de « data » par une remise en perspective historique et une recherche étymologique. En effet, avant d'être comprise comme une « infobésité » (*big data*), la donnée était avant tout un élément brut, un *fait donné*, connu de lui-même ou par la science. En ce point, l'informatique (néologisme créé par la contraction de *information* et *automatique*) a fortement contribué à rendre la notion de *data* floue, l'assimilant à de *l'information*. Nous distinguerons ces deux concepts.

La donnée (numérique ou non) sera abordée par une approche économique-philosophique, sous l'angle du concept de *finitude*, du rôle qu'elle joue dans notre civilisation contemporaine et de la relation qu'elle entretient avec l'économie. Cette relation - finitude et économie - a été largement étudiée par Christian ARNSPERGER, docteur en sciences économiques de l'UCLouvain qui lui a consacré un chapitre entier dans son ouvrage *Critique de l'existence capitaliste : Pour une éthique existentielle de l'économie*. Éditions du Cerf, Paris (2005).

Nous verrons qu'au cours de la dernière décennie, avec l'avènement du *big data* et de ses acteurs principaux - les *GAF*A (*Google, Amazon, Facebook, Apple, ...*) -, les données numériques ont progressivement donné naissance à une toute nouvelle forme d'économie : celle de la donnée personnelle. Qu'est-ce qu'une donnée personnelle ? Est-elle intimement liée au droit à la vie privée ? Comment et pourquoi valoriser les données personnelles ? Pour répondre à cette dernière question, nous nous appuierons sur les théories pro-libérales d'un grand penseur pluridisciplinaire du XX^e siècle, Friedrich Hayek (1899-1992), économiste récompensé du Prix Nobel de Sciences économiques en 1974, fervent défenseur d'une économie de marché où la propriété privée serait requise pour un calcul économique rationnel.

Après avoir passé en revue l'analyse atomique du *big data* par sa composante de donnée numérique et personnelle, nous traiterons de façon non-exhaustive ses diverses applications concrètes d'un point de vue sectoriel, avec l'appui de statistiques descriptives. L'identification des bénéfices d'une part et des coûts d'autre part, liés au *big data*, fera également l'objet d'une section.

Toujours dans une optique de modélisation, ces coûts et gains seront intégrés dans un modèle d'arbitrage mathématique simplifié : quand une entreprise devrait-elle investir dans une infrastructure de type *big data* ? Enfin, nous ferons la typologie de certains *business models* liés au *big data*.

La deuxième section de ce TFE sera donc consacrée à l'intégration du volet *big data* dans une vision d'innovation technologique. A nouveau, nous prendrons soin de définir ce qu'est une innovation technologique au sens économique du terme, qu'elle soit au niveau de son produit (*product innovation*) ou de son processus (*process innovation*). *Amazon*, par exemple, succès commercial planétaire et pionnier en la matière, fait appel aux dernières technologies, dont une robotisation et une utilisation du *big data* toujours plus poussées au sein de ses processus de production et de management.

Il est à noter que ces innovations technologiques outre-Atlantique ne sont pas toujours bien accueillies par les acteurs économiques européens qui y voient là une concurrence déloyale. Perçues comme anti-concurrentielles, ces innovations seraient sujettes à des réglementations plus souples, notamment celles concernant la gouvernance des données personnelles.

Nous nous efforcerons dès lors d'examiner en profondeur cette *tension* qui règne entre innovation par le *big data*, dynamique mais parfois intrusive, et la protection de la vie privée, requise mais pesante. Alors que l'utilisation du *big data* et les grands principes de précaution européens sont deux logiques s'opposant frontalement selon la professeure Antoinette Rouvroy¹, il sera démontré qu'un encadrement législatif est parfaitement conciliable avec l'innovation si la prise en compte des enjeux d'une protection de la vie privée se fait le plus en amont possible. Les deux notions n'étant donc pas antinomiques, concilier *vie privée* et *innovation* s'inscrit alors dans une démarche non seulement plus responsable et durable mais aussi plus rentable, faisant bénéficier l'entreprise d'avantages compétitifs. Soit une *tension* qui n'aura plus lieu d'être.

Au vu de la conciliation rendue possible par une bonne gouvernance des données, le danger du *big data* ne résiderait alors plus tant dans le fait qu'il pénètre nos vies privées, mais parce qu'il serait plutôt lié au risque que ses algorithmes soient en passe de gouverner nos comportements grâce à leur caractère toujours plus prédictif. La gouvernance par les lois laisse ainsi progressivement la place à une *gouvernance par les nombres*². Un pouvoir impersonnel que certains auteurs qualifient même de « gouvernementalité algorithmique »³, qui s'inscrit alors résolument dans un capitalisme informationnel.

Même si cette *tension* entre *modernité* et *renversement de pouvoir* est plus *implicite* que celle (plus formellement décrite) qui sous-tend la relation entre *innovation* et *vie privée*, en ce sens qu'elle ne suscite pour le moment que peu de réactions de la société, il en va de la survie de notre démocratie de s'interroger sur les solutions possibles pour rendre les algorithmes et leur utilisation légitimes. Hugues Bersini, professeur à l'Université libre de Bruxelles et codirecteur de l'Institut de recherches interdisciplinaires et de développement en intelligence artificielle (*Iridia*), y apportera quelques éléments de réponse.

¹ A. Rouvroy, *Homo juridicus est-il soluble dans les données ? Droit, normes et libertés dans le cybermonde, Liber Amicorum Yves Poulet*, Larcier, 2018, p. 429.

² Alain Supiot, *La gouvernance par les nombres*. Cours au Collège de France (2012-2014), coll. « Poids et Mesures du Monde », Fayard, 2015.

³ Rouvroy Antoinette et Thomas Berns, « Le nouveau pouvoir statistique. Ou quand le contrôle s'exerce sur un réel normé, docile et sans événement car constitué de corps « numériques »... », *Multitudes*, vol. 40, n° 1, 2010, p. 89.

Sommaire

1. Motivations et périmètre	3
2. La donnée numérique et personnelle	7
2.1. Qu'est-ce que la data ?	7
2.2. Le rôle économique-philosophique de la donnée	10
2.3. L'économie de la donnée personnelle	14
2.3.1. Un retour historique sur le droit à la vie privée	14
2.3.2. Qu'est-ce que la donnée personnelle ?	15
2.3.3. Le big data et la valorisation de la donnée personnelle	18
2.3.3.i. Comment valoriser ?	18
2.3.3.ii. Pourquoi valoriser ?	22
2.3.4. Modèle d'arbitrage	26
2.3.5. Le big data et la typologie de ses business models	29
3. Innovation et gouvernamentalité algorithmique	30
3.1. Qu'est-ce qu'une innovation technologique ?	30
3.2. Tension explicite entre vie privée et innovation	32
3.3. Tension implicite entre modernité et renversement de pouvoir	35
4. Conclusion	37
5. Bibliographie	39

2. La donnée numérique et personnelle

2.1. Qu'est-ce que la *data* ?

L'humanité produit toutes les minutes près de 4,5 millions de requêtes sur *Google* et 18 millions de SMS. Toutes les minutes, 188 millions de mails sont envoyés et 232 000 appels téléphoniques via *Skype* sont effectués. Toutes les minutes, 694 000 heures cumulées de vidéo sur *Netflix* sont consommées et 4,5 millions de vidéos sur *YouTube* sont lancées. Ces chiffres donnent le vertige. Pourtant, à cela s'ajoute l'activité de *Facebook*, *LinkedIn*, *Snapchat* et bien d'autres acteurs incontournables du web, aussi bien *Bing* que *Wikipedia*, ou encore *Amazon*. Et tout cela en une minute. C'est en tout cas ce qu'affirme le cabinet de consultance réputé pour la visualisation des tendances internet - *Domo, Inc.* - dans son dernier rapport pour l'année 2019⁴. Toutes ces activités génèrent des données. *Beaucoup* de données. Et comme le révèle l'édition 2019 du *Digital Economy Compass*⁵, c'est l'équivalent de 660 milliards de *Blu-ray* de 50 Go en termes de données numériques qui ont été générées dans le monde en 2018, soit 33 zettaoctets (33×10^{21} octets) de *data*. On estime que ce volume sera même multiplié par 65 d'ici 2035.

Mais au fond, qu'est-ce que la *data* ? *Donnée*, *data* et *information* sont des termes aujourd'hui souvent compris comme synonymes. Or, comme nous le verrons, bien que ces termes soient souvent interchangeables dans une phrase, ils ne sont que des entités d'une structure sémantique plus vaste.

La *donnée*, du latin *dare*, *datum* (donner, donné) est un élément brut, non analysé, non organisé, non lié, non interrompu, qui ne transmet aucune information de lui-même. Elle se définit comme une observation élémentaire d'une réalité. Son contenu peut être qualitatif ou quantitatif. Parfois vue comme une notion abstraite typée, elle est à la fois un concept et une mesure. Protéiforme, elle peut être symbolique, logique, numérique, ou encore textuelle.

⁴ Domo, Inc. *Data Never Sleeps 7.0* (2019) [en ligne] <https://www.domo.com/learn/data-never-sleeps-7> [accès le 19 mars 2020].

⁵ Statista. *Digital Economy Compass* (2019) [en ligne] <https://cdn.statcdn.com/download/pdf/DigitalEconomyCompass2019.pdf> [accès le 19 mars 2020].

La *donnée* ne porte donc pas de sens en elle-même. Pour illustrer ceci, *22112007* pourrait être une *donnée* brute, mais *l'information* qui en serait potentiellement dégagée serait sujette à l'interprétation de chacun-e : certain-es pourraient voir en ce nombre la date du 22 novembre 2007, d'autres, un montant, et d'autres encore, une série de chiffres.

L'information quant à elle est perceptible, interprétée comme un message d'une manière particulière, qui donne un sens aux données. Etymologiquement, le mot *information* vient du latin *informare*, qui littéralement signifie « *mettre en forme* ». *L'information* doit alors prendre une forme par laquelle elle peut soit être transmise, soit être stockée, sinon elle serait perdue. Cette forme doit également présenter un intérêt pour celui ou celle qui la réceptionne, sinon, à nouveau, elle serait perdue car inutilisée. Son contenu est donc sémiotique, dégageant de la signification. *L'information* est donc à la fois matérielle (elle est une somme de ses données) et conceptuelle parce qu'elle transmet une notion qui a un sens particulier, une valeur sémantique, signifiante pour l'émetteur et le récepteur.

La *data*, quant à elle, est un mot emprunté à l'anglais. Tout comme les termes *team*, *buzz*, *vintage*, la *data* est un anglicisme qui se traduit littéralement par *donnée*. Elle est employée surtout lorsque son contexte est associé de près ou de loin à l'informatique. Pourtant, derrière ce simple emprunt lexical se cache une source de confusion, justement en rapport avec la distinction entre *donnée* et *information*.

En informatique, les données sont codées en *bits*, représentées par des 0 et des 1. Lorsque vous écrivez un SMS, vous n'envoyez pas réellement des phrases composées de mots, eux-mêmes composés de lettres. Chaque lettre du message voyage à travers le réseau sous forme de *bit*. Généralement, chaque lettre prend l'espace de 8 *bits*, soit un *octet* de *data*.

En se basant sur la nature même du mécanisme informatique qui vise à codifier les données en binaire, la donnée numérique - la *data* - est alors déjà une représentation arbitraire de la donnée brute, et donc potentiellement une *information* si l'on se réfère à sa forme.

Par analogie, le célèbre tableau de René Magritte, *La Trahison des images* (1929), renforce l'idée qu'une représentation transmet de l'information, et ceci nous pousse davantage dans la réflexion sémiologique. Son oeuvre représente une pipe et s'accompagne de la légende « Ceci n'est pas une pipe ».

Par ce procédé, Magritte démontre que *même* si l'on peint une pipe de la manière la plus fidèle qui soit, la pipe ne restera toujours qu'une idée que l'on se fait de la pipe. La pipe peinte sur un tableau n'est pas une vraie pipe. Elle n'est qu'une image, une représentation (*information*) de pipe (*donnée brute*) qu'on ne peut pas fumer, ce qui aurait été le cas avec une vraie pipe.

La donnée numérique, mise en forme binaire, reste-elle alors une *donnée* ou devient-elle de *l'information* ? Pour répondre à cette question, rappelons que la *donnée* se transforme en *information* lorsqu'elle est communiquée à un être humain capable de l'interpréter, d'absorber son contenu sémiotique. L'être humain donne du sens à la donnée afin de l'interpréter en *information* utilisable. Formellement, *Information* = *Données* + *Signification*.

Selon la chaîne de transformation des données en compétences étudiée par Mack (1995), donner du sens à une *donnée* conduit donc à de *l'information*. En réitérant la logique, donner du sens à de *l'information* conduirait alors à une *connaissance*, qui elle-même interprétée mènerait à une *compétence* (M. Mack, 1995). C'est aussi se rappeler le modèle d'acquisition des compétences de Dreyfus en cinq étapes, qui décrit le chemin parcouru du ou de la novice à l'expert·e (l'expertise).

A la vue de ces éléments de réponse, nous prenons donc progressivement conscience que les systèmes informatiques se contentent de manipuler des représentations de données, la forme que ces données prennent. Bien que ces données ne soient pas brutes à *proprement parler*, le système informatique n'étant non seulement pas humain, il n'en dégage aussi aucun contenu sémiotique de valeur signifiante. C'est pourquoi nous parlons de *donnée* ou de *data* et non *d'information*. Il est donc bien admis par les expert·es du monde entier que nos systèmes d'informations (supportés par des systèmes informatiques) traitent uniquement de la *data*.

A titre de confirmation, Le Petit Robert définit *data* comme une « *représentation conventionnelle d'une information (fait, notion, ordre d'exécution) sous une forme (analogue ou digitale) permettant d'en faire le traitement automatique* » (Le Petit Robert, 2018).

Finalement, l'être humain observe des *données*, interprète des *informations*, associe des *connaissances* entre elles et les utilise pour acquérir des *compétences* qu'il est capable de mettre en oeuvre. Voici donc la structure sémantique dans laquelle nous nous proposons de situer le terme *data*.

La logique voudrait que l'on situe désormais cette structure sémantique dans une perspective historique et, par juxtaposition axiomatique, que l'on replace la donnée dans l'Histoire : du rôle économique qu'elle a et continue de jouer dans les civilisations, au fil du temps, jusqu'à son omniprésence dans la société contemporaine. C'est ainsi que nous amenons le chapitre suivant.

2.2. Le rôle économique-philosophique de la donnée

De tout temps, l'humain a cherché à communiquer : d'abord par la parole et les arts, puis par l'écrit. C'est d'ailleurs ce qui a marqué la fin de la Préhistoire et de début de l'Histoire : la naissance de l'écriture, et de surcroît, la naissance de la transmission de données intra- et inter-générationnelle. *Verba volant, scripta manent* : les paroles s'envolent, les écrits restent.

L'homme préhistorique commença à graver et à peindre il y a 40 000 ans. Ce n'était pas de l'écriture, certes, mais à l'époque déjà, il faut supposer que nos ancêtres souhaitaient transmettre leur pensée à la génération suivante, signe d'un témoignage de leur condition humaine probablement.

Les premiers écrits apparaissent à Sumer (*L'histoire commence à Sumer*⁶) en Mésopotamie, l'Irak actuel, et datent de 3500 ans avant notre ère. Bien sûr, la forme que prenaient ces écrits était quelque peu différente de ce que l'on connaît aujourd'hui. Il s'agissait de petits cailloux en argile, des jetons modelés, des *calculi* sur lesquels étaient gravées des encoches représentant des quantités et permettant le comptage du bétail ou des sacs de grains. Les premiers signes ne transmettant pas d'information sur la nature même des objets, seule l'information du dénombrement de la contenance était transmise. Ainsi, si l'écriture était née, elle ne servait principalement que d'aide-mémoire administratif sur le plan économique intra-générationnel.

⁶ Kramer Samuel Noah, *L'histoire commence à Sumer...*, préface de Jean Bottéro. Coll. Champs Histoire, Paris, Flammarion, 2017.

Plus tard, vers 3400 avant J.C., toujours en Mésopotamie, la découverte du phonétisme avec l'apparition de l'écriture cunéiforme permettra de représenter des sons correspondant à des syllabes écrites, l'écriture syllabique ayant précédé les premières lettres. Des notions complexes telles que la vie ou la mort, qui jusque-là ne pouvaient être communiquées par écrit, sont désormais transmissibles.

Ainsi apparaît la volonté de transmettre des messages au-delà de la génération courante, l'écriture devenant un moyen pour l'humain de raconter des histoires. Son histoire au-delà de sa propre fin. Dès lors que « chacun de nous, riche ou pauvre, puissant ou soumis, reste plus ou moins obscurément conscient de cet horizon d' "impossible possibilité", comme le nomme Heidegger »⁷, pourrions-nous y voir la transmission de données comme une échappatoire à la *finitude existentielle* ?

Aujourd'hui, les investissements à coups de milliards d'euros dans la recherche et le développement (*R&D*) au service de la santé sont légion. Notamment dans des avancées technologiques émergentes appelées couramment *NBIC* pour Nanotechnologie, Biologie, Informatique et sciences Cognitives : exosquelettes, lentilles de contact smart super-connectées, nanorobots guérisseurs, interfaces homme-machine ou encore des cœurs artificiels et des mécanismes de prévention santé via le *big data*, sont quelques exemples d'innovations censées nous offrir une immunité à la maladie, voire à la mort. Selon l'idéologie transhumaniste, l'humain amélioré ou augmenté aurait tout à y gagner. Pourtant, ces innovations sont toutes autant des avancées technologiques qui viseraient à *nier notre vulnérabilité par des moyens limités pour entrevoir l'infinité*.

En ce sens, « Lionel Robbins [...] définissait [déjà] dès les années 1930 la science économique comme la discipline qui étudie l'affectation de moyens limités à la satisfaction de besoins illimités »⁸, soit l'étude de la résolution du « problème économique de base ». S'ensuit une conclusion de C. ARNSPERGER à première vue surprenante mais logique : « [il] apparaît alors assez clairement un principe qui n'est jamais dit dans ces enseignements, à savoir que la science économique fait partie de la philosophie existentielle. »⁹

⁷ Arnsperger, Christian. *Critique de l'existence capitaliste : Pour une éthique existentielle de l'économie*. Éditions du Cerf : Paris (2005) (ISBN:2-204-07694-5) 210 pages, p.49.

⁸ *ibid.*, p.47.

⁹ *ibid.*, p.47.

A dire vrai, il n'est plus tant question de savoir si l'humain tente à tout prix d'atteindre l'infinitude, ce qui nous plongerait dans un schéma argumentaire purement existentiel, mais bien de reconnaître que l'agent économique qui sommeille en lui, au nom de la rationalité économique, tente à tout prix - par les actes économiques qu'il pose - de masquer sa vraie finitude : la consommation étant un moyen de se soustraire à l'angoisse existentielle de la finitude, l'épargne - ou la *consommation différée*¹⁰ - visant à nier la finitude elle-même, et enfin, l'investissement rassurant sur l'existence d'un lendemain. Et c'est précisément dans ce contexte-là que l'ouvrage de C. ARNSPERGER apporte de la substance à notre réflexion : « le jeu économique fausse nos finitudes »¹¹.

Sur base d'une pratique réflexive, l'auteur arrive à définir - avec un certain mérite - de façon scientifique la *fausse finitude*, ou plutôt la finitude *socialement faussée*. Pour ce faire, il renvoie à deux orientations, l'une réductionniste, assimilant la finitude à la mortalité ; l'autre, pluraliste, considérant la finitude comme une finitude socialement vécue parmi d'autres. A la convergence de ces deux orientations, la notion de *limitation* persiste : la mortalité (limitation de l'existence), les limitations intrinsèques à la vie personnelle (limitations physiques), et enfin les limitations intrinsèques à la vie en société (les lois).

La finitude socialisée (« ma finitude [...] [étant] en partie [...] le résultat de vos choix, à leur tour effectués sous des contraintes qui dépendent en partie de mes choix »¹²) serait donc - dans une logique de *réalisme machiavélien* - le fruit de *confrontations* entre agents économiques : certain-es (*la minorité*) mieux loti-es que d'autres (*la majorité*) parviendraient à monopoliser plus de ressources économiques contraignant cette *majorité* à « faire l'expérience d'une finitude sans aucune mesure avec ce qu'aurait impliqué le partage, par tous, de la finitude vraie et inévitable »¹³ pour qu'une *minorité* puisse se dérober à l'angoisse existentielle de la *vraie finitude*. Ceci menant à davantage d'inégalités dans la répartition des finitudes, soit une répartition faussée des finitudes - *une finitude socialement faussée*.

¹⁰ *ibid.*, p.24.

¹¹ *ibid.*, p.47.

¹² *ibid.*, p.48.

¹³ *ibid.*, p.49.

Quant à la *finitude vraie*, l'auteur reconnaît qu'il est plus difficile d'en dégager une définition claire et précise car l'expérience de finitude dépend toujours, entre autres, de l'accès à la richesse et à la technologie, notamment médicale. Dès lors, s'il est relativement facile d'observer la finitude socialement faussée, la vraie finitude, quant à elle, reste toujours à un certain degré socialement construite, de même que nos actes restent économiquement inscrits.

En dehors du jeu économique, l'utilisation même des ressources, qu'elles soient matérielles ou immatérielles, notamment la donnée numérique, renvoie à ce que C. ARNSPERGER appelle une utilisation *authentique* ou *inauthentique* des ressources économiques, définissant respectivement l'attitude de l'agent face à la finitude comme héroïque existentielle authentique ou inauthentique ; la ressource économique étant alors, dans le premier cas, un moyen de repousser la finitude sans jamais la nier, donc l'accepter ; dans le second cas, un moyen d'entretenir le fantasme d'une infinitude sans contraintes. Il nous semble dès lors prudent d'affirmer que l'utilisation massive du *big data* dans notre économie s'apparente à un désir de contrôle existentiel, nous permettant d'entrevoir une illusion de possible fin à notre finitude. Nous laissons néanmoins la liberté à chacun·e d'émettre son hypothèse quant au degré de « fausseté ».

Face à cette illusion d'eldorado du *big data*, certaines voix s'élèveront pourtant, clamant qu'au-delà de notre désir de contrôle existentiel, le risque d'une utilisation toujours plus importante du *big data* résiderait justement dans un renversement de pouvoir (sujet que nous traiterons dans la deuxième partie de ce TFE). En permettant aux algorithmes d'utiliser nos données personnelles (voir prochaine section), nous leur offririons progressivement le contrôle de nos vies. L'humain est donc dual en ce sens : il désire l'infinitude mais refuse catégoriquement son implication.

2.3. L'économie de la donnée personnelle

2.3.1. Un retour historique sur le droit à la vie privée

Avant d'envisager une définition de la ressource principale de l'économie de la donnée personnelle, une prise de conscience sur ce qu'est la vie privée est nécessaire : après tout, *la donnée personnelle est une donnée privée*.

Aujourd'hui, la donnée personnelle est source de nombreux débats animés opposant d'une part l'idéologie libérale, d'origine anglo-saxonne, qui voit en elle l'or noir pour une économie numérique toujours plus innovante ; à celle, d'autre part, protectionniste, d'origine européenne, qui elle craint que son utilisation abusive ne nuise à nos libertés individuelles. Cependant, la notion même de liberté individuelle et du lien qu'elle entretient avec le droit à la vie privée n'a pas toujours existé sous cette forme.

A l'époque de la Grèce antique, ce droit à la vie privée n'était ni nécessaire ni demandé, car cette dernière n'était en rien menacée. La vie privée était la norme, et la sphère publique accessible seulement aux privilégiés. Selon H. Arendt, « être libre ne signifiait originellement rien d'autre que pouvoir aller et venir où bon nous semblait, encore qu'il s'agissait de quelque chose de plus que ce que nous entendons aujourd'hui par la liberté de mouvement. Cela voulait dire non seulement n'être soumis à la contrainte d'aucun homme, mais également pouvoir s'éloigner de la sphère de la contrainte tout entière, du foyer et de la " famille " [...]. Seul le maître de maison disposait de cette liberté, et [...] précisément en vertu de cette domination, il pouvait s'éloigner de sa famille, au sens antique du terme. »¹⁴

De façon surprenante, la liberté était donc comprise comme le privilège de prendre le risque de mettre sa vie en péril en se déplaçant dans la *polis*, la *polis* se traduisant littéralement par « cité », soit la sphère publique dans laquelle seuls les hommes libres (les maîtres de maisons) avaient le droit d'agir et de se mouvoir, excluant notamment les esclaves. A contrario, l'*oikos* (la « maison ») désignait les activités économiques au sein du foyer, qui à l'époque pouvait s'étendre sur un large territoire et qui comprenait les membres de la famille et les esclaves. C'est d'ailleurs en ce sens que l'*économie* y trouve son étymologie.

¹⁴ H. Arendt, *Qu'est-ce que la politique ?*, Seuil, 1958, p. 83.

La sphère privée était donc déjà bien distincte de la sphère publique élitiste, mais elle était surtout principale et hermétique : femmes et esclaves étant obligés d'y passer leur vie, menant au principe suivant : ce qu'il se passe dans la *domus* reste dans la *domus*. Le droit romain formalisera plus tard ce principe en léguant au *pater familias* le pouvoir de vie ou de mort sur les membres qui dépendent de son pouvoir.

Ce n'est que bien plus tard que la sphère publique commencera à empiéter sur la sphère privée, rendant nécessaire la protection de cette dernière. C'est en effet à partir du XVI^e siècle, notamment avec l'affirmation de l'autorité de l'Etat en France, que le besoin de s'abriter de la pression sociale - communautariste - se fait de plus en plus ressentir, donnant naissance à des lieux où il était possible de se retirer, tels que les jardins clos, les études ou encore les bibliothèques¹⁵.

Il faudra attendre le siècle des Lumières pour voir apparaître une formalisation progressive du droit à la vie privée, avec notamment la règle de l'inviolabilité du domicile en France, codifiée à l'article 184 du Code pénal de 1810. Finalement, l'idée que l'on se fait aujourd'hui du droit à la vie privée se rapproche des principes énumérés par Samuel Warren et Louis Brandeis dans un article de référence publié en 1890, sous le titre « The Right to Privacy »¹⁶, où le principe phare est que tout individu a droit à une protection de sa propriété (le droit à la propriété) et de sa personne (le droit à la vie). Ce dernier droit englobant également la protection de l'intégrité de l'individu, de son intimité et de ses émotions, car une atteinte à cette dernière pourrait le rendre de façon indirecte physiquement vulnérable. En bref, le droit à la vie privée pourrait être résumé à celui d'être laissé tranquille.

2.3.2. Qu'est-ce que la donnée personnelle ?

Le règlement n° 2016/679, connu sous l'appellation de *Règlement général sur la protection des données (RGPD)* et entré en vigueur le 24 mai 2016, définit la donnée personnelle comme « toute information se rapportant à une personne physique identifiée ou identifiable [...] ; est réputée être une "personne physique identifiable" une personne physique qui peut être identifiée, directement ou indirectement [...] » (RGPD, art. 4, §1er)

¹⁵ P. Ariès et G. Duby, *Histoire de la vie privée, III De la Renaissance aux Lumières, Seuil, 1985, p.161.*

¹⁶ S.D.Warren & L. D. Brandeis, *The Right to Privacy*, Harvard Law Review, vol. 4, n°5, déc. 15, 1890.

Cette notion est donc vaste dans la mesure où elle s'étend à toute chose qui renvoie indirectement à une personne. Toutes nos traces numériques sont-elles alors des données à caractère personnel ? Tout dépend de leur degré d'anonymisation - et plus généralement de leur forme.

La donnée personnelle peut en effet prendre plusieurs formes : on dit qu'elle est protéiforme. Elle a été soit structurée, donc analysée, soit semi-structurée ou encore laissée brute, telle quelle (non structurée). En parallèle, elle peut être de nature personnelle identifiante ou comportementale. La première catégorie définit la donnée comme rattachée à une personne ou à sa position géographique. C'est notamment le cas d'un nom, d'une adresse email ou d'une adresse IP. Quand elle est comportementale, la donnée n'identifie pas une personne (elle est d'ailleurs pour la plupart du temps anonymisée) ; par contre, elle visera à déceler un comportement (souvent économique) de l'individu qui l'a engendrée. Le marketing d'aujourd'hui n'a en effet plus besoin de connaître les caractéristiques personnelles de l'internaute, comme son nom ou son âge, pour lui vendre ses produits. Seul un profil numérique de ses préférences sera utilisé à des fins d'anticipation sur ses envies ou besoins futurs, extrapolés sur base d'un comportement passé, quitte à l'enfermer dans un avatar comportemental : « on crée des profils, on impute aux gens des désirs et besoins qu'ils n'expriment à aucun moment et on prend le risque de les enfermer dans ces avatars comportementaux. »¹⁷

La donnée personnelle est aussi relative. La valeur qu'une entreprise lui confère est de nature marchande, permettant d'améliorer ses modèles serviciels tant au niveau de son produit que de son processus. Elle enrichit d'un côté l'expérience utilisateur qui utilise le produit, de l'autre, elle améliore le modèle décisionnel en s'infiltrant dans *l'opérationnel*. Du point de vue du client, la donnée peut être vue comme une fuite de son intimité et donc de sa vie privée. Dans ce cas, la confiance qu'accorde le client à l'entreprise est rompue si cette dernière ne délimite pas clairement l'usage qu'elle en fait.

Enfin, le statut juridique de la donnée personnelle sera - en fonction des législations en vigueur dans le pays - tantôt en faveur d'une économie de marché motorisée par l'innovation, tantôt pour la fermeture de toutes les fuites à la vie privée. *Une tension entre innovation par les données et protection à la vie privée qui n'a pas fini de s'exercer.*

¹⁷ CNIL, *Vie privée à l'horizon 2020* (2012) [en ligne] https://linc.cnil.fr/sites/default/files/typo/document/CNIL-CAHIERS_IPn1.pdf [accès le 3 avril 2020], p.19.

Pourtant, parler de fermeture devrait être nuancé. La donnée personnelle d'aujourd'hui n'est jamais fermée. Loin de là, elle est d'ailleurs tout l'opposé. En 2012, la CNIL s'était déjà posé la question au regard de l'ouverture des données dans son ouvrage « Vie privée à l'horizon 2020 »¹⁸ de la collection Cahiers IP. Étant en 2020, voyons ce qu'il ressortait des dires des experts recueillis à l'automne 2011. Selon le sociologue français Dominique Boullier¹⁹, il n'est plus adéquat de voir la donnée personnelle comme limitée à un attribut de la personne dont le domaine serait alors fermé. En outre, il faudrait la considérer comme un « habitèle » : tout comme les habits que nous portons ou l'habitacle de la voiture que nous conduisons, la donnée personnelle n'est pas une chose que nous possédons, mais bien une chose que nous habitons. Nous y sommes enveloppés et nous ne pilotons cet ensemble que dans la relation, par exemple sur les réseaux sociaux. Pour cette raison, Boullier propose de définir la donnée personnelle comme un ensemble de données relationnelles ou transactionnelles, mais dans tous les cas, ouvertes.

Pourquoi conduisons-nous consciemment cet habitacle dans des relations transactionnelles toujours plus ouvertes, quitte à faire fuiter notre vie privée ? Antoinette Rouvroy, docteure en sciences juridiques de l'Institut universitaire européen (Florence, 2006) et chercheuse qualifiée du FNRS au Centre de Recherche en Information, Droit et Société (CRIDS), nous apporte certains éléments de réponse : « premièrement, cela s'explique par une addiction de masse à l'interconnexion, qui passe par un fétichisme des objets connectés, assumant une fonction d'objets transitionnels pour adultes, éloignant l'angoisse de séparation »²⁰.

Deuxièmement, c'est le prix à payer pour pouvoir observer ses proches, d'où un certain goût pour le voyeurisme. Aussi, en se montrant récalcitrant-e à partager ses données, on paraît suspect-e dans une société où la phrase « je n'ai rien à cacher » est communément acceptée. Dans la même logique, parce qu'étant quotidiennement en concurrence dans notre intimité, partager sa vie privée est un moyen de se distinguer à coups de *likes*, de photos, de *comments*, etc. Comme nous le verrons plus tard, c'est aussi finalement en raison de la typologie des *business models* du *big data* : parfois de nature « biface »²¹, où le troc de la donnée personnelle est requise en échange d'un service ; parfois « serviciel »²², à titre de service dans lequel la donnée personnelle est tout simplement la condition *sine qua non* de son existence.

¹⁸ CNIL, *Vie privée à l'horizon 2020* (2012).

¹⁹ *ibid.*, p. 32.

²⁰ A. Rouvroy, *Homo juridicus est-il soluble dans les données ? Droit, normes et libertés dans le cybermonde*, *Liber Amicorum Yves Poulet*, Larcier, 2018, p. 420.

²¹ S. Chignard et L.D. Benyayer, *Datanomics, les nouveaux business models des données*, Fayard, 2014, p. 80.

²² *ibid.*, p. 81.

Par ailleurs, les données personnelles que le *big data* absorbe n'émanent pas toujours directement de la volonté de l'utilisateur de les partager. Par exemple, quand il remplit consciemment un formulaire d'inscription. En d'autres termes, la donnée n'est pas toujours de nature déclarative. Elle peut provenir d'ailleurs parce que l'humain n'est plus le seul producteur de données personnelles. Ses concurrents sont les machines elles-mêmes. Des capteurs en tout genre permettent, via des techniques de data-mining et de mise en corrélation, de créer de nouvelles données à partir de données extraites d'un événement tel qu'une porte qui s'ouvre à une certaine heure, et d'en extrapoler de nouvelles informations : « le propriétaire est rentré du travail » par exemple.

En résumé, la donnée personnelle est protéiforme, relative, indirecte, ouverte - relationnelle ou transactionnelle -, et est donc bien loin d'une simple caractéristique de l'individu ; ce qui amène les politiques à l'avant d'un défi de taille quant à sa réglementation.

2.3.3. Le *big data* et la valorisation de la donnée personnelle

2.3.3.i. Comment valoriser ?

D'un point de vue macroéconomique, il devient clair que les données numériques sont déjà un moteur de croissance dans une économie par ailleurs en déclin²³. Alors que les industries traditionnelles ont décliné de 3,6% entre 2008 et 2011 en Europe, les secteurs à forte densité de données - où l'utilisation de la donnée personnelle est un élément clé des affaires - ont prospéré quant à eux avec des taux de croissance annuels compris entre 15% pour le commerce électronique et jusqu'à 100% pour les communautés « Web 2.0 » (Facebook, Twitter, Instagram, Snapchat, ...).

Dans une étude publiée en 2012²⁴, le cabinet international de conseil en management et leader mondial, le *Boston Consulting Group*, s'est risqué à estimer la valeur économique de ce carburant immatériel dans l'UE pour l'année 2011 et de son potentiel de croissance d'ici 2020. En outre, cette étude distingue le bénéfice de la donnée personnelle dégagé par secteur (8 secteurs) et selon le type d'agent économique auquel il se rapporte (3 acteurs) : d'une part, les consommateurs et les entreprises, d'autre part, les gouvernements. Première constatation : la valeur de la donnée est perçue différemment selon le public.

²³ Boston Consulting Group, *The value of our digital identity*. (2012) [En ligne] <https://2zn23x1nwzzj494slw48aylw-wpengine.netdna-ssl.com/wp-content/uploads/2017/06/The-Value-of-Our-Digital-Identity.pdf>, [accès le 3 avril 2020] p.12.

²⁴ Boston Consulting Group, *The value of our digital identity* (2012).

Fait trivial, il s'agit pour les entreprises du secteur privé d'une valeur monétaire réelle se traduisant par une augmentation du chiffre d'affaire mais aussi par des économies rendues possibles grâce au gain en efficacité qui permet ainsi de réduire leurs coûts de production. S'il est possible pour le secteur privé de faire jouer la donnée dans les deux dimensions (revenus et coûts marginaux), il semblerait que ce soit également le cas pour le secteur public, par l'instauration d'algorithmes de détection d'évasion fiscale ou encore par la mise à disposition en libre service pour le citoyen de démarches administratives, comme le renouvellement du permis de conduire qui nécessitait autrefois un suivi par des fonctionnaires. Autant de techniques visant à augmenter les recettes de l'Etat et à réduire les dépenses publiques.

Toutes organisations confondues - secteur privé et public -, le surplus économique que générerait la donnée personnelle atteindrait alors, pour l'année 2020, respectivement 112 milliards d'euros pour les entreprises et 216 milliards d'euros pour le secteur public et soins de santé : soit approximativement la somme astronomique de 330 milliards d'euros. Qu'en est-il du consommateur ?

Contrairement à l'idée reçue selon laquelle - et par analogie aux pensées marxistes la dénonçant - la donnée (le travail dans la réflexion originelle) serait une commodité, donc un contenu et l'individu son transporteur (le contenant) duquel les multinationales pomperaient la précieuse substance jusqu'à l'en vider, cet individu - étant le consommateur dans notre réflexion - s'avère ici en fait *bénéficiaire net* quand on chiffre le surplus total qu'il retire de l'utilisation de ses données. En effet, le consommateur se positionne *in fine* comme le grand gagnant. Pour lui, la valeur de ses données personnelles est perçue à la fois comme (1) *immatérielle*, émanant alors du fait qu'il bénéficie du développement de nouveaux services gratuits mis à sa disposition ; ou (2) *directement monétaire*, quand on sait qu'il peut réaliser une économie d'argent en faisant le choix d'utiliser les outils numériques du service public, épargnant alors les frais de gestion, voire parce qu'il profite directement des économies d'échelle réalisées par les prestataires, rendues possibles grâce aux données personnelles ; et finalement (3) *indirectement monétaire* du fait qu'il réalise une économie de temps. On le sait, « le temps c'est de l'argent », tout au moins dans une vision capitaliste. Le surplus économique pour le consommateur se chiffrerait ainsi à plus du double de celui cumulé par les entreprises et le secteur public, soit 670 milliards d'euros pour l'année 2020, portant le bénéfice total à près d'un billion d'euros (mille milliards d'euros).

C'est l'équivalent d'environ 6% du PIB-UE28 nominal si l'on se base sur son estimation réalisée en 2019 (source : Eurostat). A noter que les *bénéfices de second tour* n'ont pas été inclus dans la valorisation, sûrement en raison de la complexité à les quantifier de façon non-biaisée. Par *bénéfices de second tour*, nous entendons les conséquences positives indirectes dérivées de l'utilisation des données personnelles. Si, par exemple, l'utilisation d'algorithmes de prévention de maladies - les secteurs public et de soins de santé réalisant à eux seuls 40% du bénéfice organisationnel - venait à augmenter l'espérance de vie (effet positif *direct*), un impact positif sur l'économie se ferait ressentir en second lieu. Cet impact proviendrait du fait qu'il y aurait, à un moment donné, plus d'agents économiques dépensant leur argent (effet positif *indirect* sur l'économie), *ceteris paribus* - en laissant de côté le contre-effet lié à l'augmentation des dépenses publiques en soins de santé en raison d'une population accrue, si ces dépenses mêmes ne sont pas réduites grâce à la numérisation.

Il est donc important de réaliser par ces chiffres - bénéfices de second tour inclus ou non - que la donnée personnelle est pertinente et importante non plus seulement du point de vue de Facebook ou Google, mais pour l'ensemble de l'économie ; la donnée personnelle s'affirmant comme un élément clé d'un capitalisme informationnel toujours plus dépendant de cette dernière, d'où le regroupement sous l'appellation d'*économie de la donnée personnelle*.

Huit secteurs recouvrent à eux seuls l'ensemble de l'économie au sens large : (1) le secteur public et soins de santé, (2) la production traditionnelle, (3) les médias et télécommunications, (4) les services financiers, (5,6,7) le secteur internet [commerce électronique, communautés Web 2.0 et services en ligne de divertissements] et enfin (8) le commerce de détail.

Cependant, tous les secteurs de l'économie ne sont pas au même stade d'avancement dans leur utilisation de la donnée personnelle. Si le commerce de détail et le secteur Internet sont déjà à un niveau relativement avancé - ces organisations exploitant régulièrement les données personnelles en interne pour la R&D afin d'optimiser la livraison de leurs biens et services ou à des fins de revente (*voir section 2.3.5*) -, en revanche la production traditionnelle et le secteur public sont, eux, au tout début de l'exploitation numérique.

Dans l'entre-deux, il y a le secteur des télécommunications et des médias, y compris les câblo-opérateurs qui, en principe, ont accès à une grande quantité de données à caractère personnel mais n'en font jusqu'à présent que peu d'usage. Il est attendu de leur part de se renouveler pour faire face à une concurrence toujours plus ardue, poussée par l'entrée sur le marché de nouveaux acteurs tels que Netflix, Disney Channel ou Apple TV.

Si l'état d'avancement dans l'utilisation des données personnelles est variable selon le secteur, la typologie de ses utilisations diverses reste quant à elle invariable et se retrouve toujours sous une de ces mégatendances : dans (1) *l'automatisation* de processus de production (Amazon étant l'un des pionniers en la matière) qu'il ne faut pas confondre avec (2) *l'autonomisation* de l'individu dans ses démarches administratives affectant positivement le consommateur et les dépenses publiques ; (3) *la personnalisation* qui, elle, sur base de recommandations, peut stimuler à la fois *satisfaction-client* et *rentabilité* de l'entreprise ; (4) la R&D permettant de raccourcir les cycles de développement ; et finalement sous la forme d'une monétisation secondaire (5), quand les données sont suffisamment précieuses pour être revendues à des tiers (modèle *biface*).

Il s'ensuit que cet eldorado de la donnée impute inévitablement des coûts à l'économie. Outre les coûts directs tels que les coûts de collecte, de stockage et traitement des données (standardisation et contrôle de qualité des données non- et semi-structurées), sans oublier les coûts indirects émanant d'externalités négatives, comme la pollution environnementale qu'engendrent les centres de stockage, il existe un coût qui nous intéresse tout particulièrement : celui associé à la fuite de données.

Dans un contexte où seuls 30% des consommateurs ont une compréhension relativement complète des secteurs qui collectent et utilisent leurs informations²⁵, la confiance joue alors un rôle crucial, et cette confiance numérique qu'accorde le consommateur diffère également selon le secteur : les consommateurs sont en moyenne 30% plus enclins à partager des données sensibles avec des entreprises de commerce électronique, des câblo-opérateurs et des constructeurs automobiles qu'avec les communautés du Web 2.0. Dès lors, s'il y a rupture de confiance, le manque à gagner serait tel qu'il s'évaluerait à 440 milliards d'euros de pertes pour l'année 2020, toujours selon la même étude²⁶.

²⁵ Boston Consulting Group, *The value of our digital identity* (2012), p. 4.

²⁶ *ibid.*, p. 4.

Dans une autre étude datant de 2018²⁷, sponsorisée par *IBM Security* et menée par le *Ponemon Institute*, on peut voir qu'il existe - au sein de cet agrégat - des coûts cachés liés à la violation de données, se traduisant pour les entreprises par une augmentation de leurs dépenses à sa lutte. Ainsi sont observés cinq éléments affectant le coût total d'une fuite : d'abord, et en accord avec l'étude précédente, (1) le coût de la perte de confiance se traduisant en une perte imprévue dans le chiffre d'affaire (taux de rotation) ; (2) le coût d'une violation de données étant proportionnel au nombre d'enregistrements perdus, la taille de la brèche a donc son importance ; (3) le temps nécessaire pour identifier et contenir une violation de données [plus il sera long, plus élevé sera son coût] ; (4) en faisant appel à des services externes d'audit et d'évaluation, l'entreprise fait alors face à un coût relatif à la gestion de crise ; enfin, et non des moindres, (5) les coûts postérieurs à la violation des données, dont les coûts imputables à l'informisation des victimes : les activités du centre d'assistance, la résolution de l'incident, les frais juridiques, les remises sur les produits, etc.

Sans trop empiéter sur la deuxième partie de ce TFE, consacrée aux techniques existantes pour une réconciliation entre le respect de la vie privée et l'innovation, nous pouvons déjà dire qu'une gestion transparente des données personnelles aiderait les entreprises à réduire de façon significative ces coûts.

2.3.3.ii. Pourquoi valoriser ?

Nous venons de voir le *comment* de la valorisation de la donnée personnelle, mesurable par les coûts et les revenus qu'elle engendre. Le *pourquoi* d'une telle valorisation est un sujet tout aussi intéressant à débattre, et on peut tirer son origine des pensées pro-libérales d'un grand penseur pluridisciplinaire du XX^e siècle, Friedrich Hayek (1899-1992). Economiste récompensé du Prix Nobel de Sciences économiques en 1974, il a été un fervent défenseur d'une économie de marché où la propriété privée serait requise pour un calcul économique rationnel.

Pour la petite histoire, Hayek est né en 1899 et est d'origine autrichienne. Lui et son professeur Ludwig von Mises sont souvent considérés comme faisant partie des grands penseurs qui ont contribué à l'école autrichienne d'économie : une école de pensée portant la réflexion sur les relations causales qu'entretiennent les individus dans un système économique. Grosso modo, von Mises étudiait d'une part la *praxéologie*, une discipline

²⁷ Ponemon institute LLC, *Étude 2018 sur le coût des violations de données : Impact de la BCM* (2018) [en ligne] <https://www.ibm.com/downloads/cas/PJDB7AB8> [accès le 5 avril 2020].

axiomatico-déductive visant à identifier et analyser les propriétés inhérentes à l'action humaine - « les lois de l'agir humain »²⁸ qu'il développera dans son ouvrage *L'Action Humaine*²⁹. D'autre part, Hayek, en appliquant cette discipline à l'échange marchand, s'insère dans une démarche essentielle à la compréhension de l'agent économique et des rapports d'échange de biens et de services qu'il entretient avec d'autres individus sur le marché - une science des échanges qu'il qualifie de catallaxie³⁰. Influencé par Adam Smith et sa notion d'ordre spontané (la main invisible), Hayek définit la catallaxie comme « l'espèce particulière d'ordre spontané produit par le marché à travers les actes des gens qui se conforment aux règles juridiques concernant la propriété, les dommages et les contrats »³¹.

De la praxéologie à la catallaxie, le *marché* est donc un objet d'étude pour les deux hommes, avec toutefois des nuances quant à la délimitation de son périmètre : « On trouve chez von Mises en germe le marché comme concept politique mais il s'en tient au domaine économique. [...] De l'économique au politique, c'est pourtant Hayek qui franchira le pas. »³²

Politique parce qu'étant anti-constructiviste, Hayek pointe du doigt l'ignorance des partisans du constructivisme à croire en une société où l'économique aurait été politiquement construit, une société initialement imaginée dans son entièreté : politiquement centralisée et économiquement planifiée. Par analogie, comme une machine mise au point d'après un plan où toutes les interconnexions auraient été établies *a priori*, sans tenir compte de son évolution *a posteriori*, et c'est là que le bât blesse. Les défenseurs de l'économie collectiviste pensent en effet - à tort selon Hayek - qu'on peut traiter toute l'information sur les besoins et les capacités de façon centralisée. Cependant, c'est oublier que tout individu possède des connaissances particulières non mécaniquement transmissibles de la base au sommet, le sommet étant l'Etat (le Planificateur central). Croire le contraire serait alors une présomption de *connaissance fatale*.

Aucun planificateur central d'une grande économie ne serait donc capable, selon Hayek, d'accéder aux multiples connaissances dispersées, mais néanmoins requises, pour allouer de façon efficace les moyens de production d'une économie dans laquelle une gamme infinie d'utilisations est possible et potentiellement conflictuelle entre agents économiques poursuivant des intérêts divergents. Seules les petites sociétés - en raison du faible nombre

²⁸ von MISES Ludwig, *L'Action Humaine*, Éditions de l'Institut Coppet, Paris (3e édition 1966) (2011), p.2.

²⁹ von MISES Ludwig, *L'Action Humaine*, Éditions de l'Institut Coppet, Paris (3e édition 1966) (2011).

³⁰ F.A. HAYEK, *Droit, législation et liberté. Le mirage de la justice sociale*, T.2, P.U.F., 1981, 221 p., p. 131.

³¹ *ibid.*

³² Jacob Jean, *Le marché d'Hayek contre la démocratie*. In : *Raison présente*, n°95, 3e trimestre 1990. Bouleversements à l'Est, p. 61.

d'informations à recueillir pour le Planificateur central - parviendraient à survivre au constructivisme, dans la mesure où les relations interindividuelles et les intérêts de chacun seraient délimités. Mais parce que les tissus sociaux évoluent, ces petites sociétés seraient inévitablement amenées à périr, laissant la place aux *sociétés ouvertes* dans lesquelles, selon Hayek, l'individu pourrait poursuivre son intérêt sans se préoccuper de l'impact de celui-ci sur le tissu économique. Finalement, « [seule l'économie de marché] se révèle donc apte à concilier la cohésion du tissu social avec une pluralité d'objectifs individuels »³³.

Si l'accès à l'information faisait à l'époque défaut à l'efficacité du Planificateur central, le *big data* d'aujourd'hui remet-il alors en question les propos d'Hayek ? Si le problème ne se limitait qu'à un manque informationnel, alors en effet, s'il est capable de faire remonter *toutes* les opportunités économiques encore non formulées et *tous* les intérêts implicites des agents économiques jusqu'au sommet (ce qui est très peu probable), le *data-mining* pourrait théoriquement mener à la réconciliation entre Planificateur central et la résolution du *Problème économique de base* (allocation de moyens limités à des besoins illimités) avec - de surcroît - le risque d'un contrôle étatique accru. Mais comme le démontre Hayek, le problème est avant tout d'ordre institutionnel. Les connaissances nécessaires à un calcul économique rationnel ne sont disponibles que dans un cadre institutionnel garantissant la propriété privée, soit le contexte d'une économie de marché. Et von Mises expliquera dès 1920 que c'est parce qu'il y a absence de propriété privée dans le système collectiviste que ce système est voué à l'inefficacité.

Dans une économie de marché, les agents économiques s'échangent des ressources qu'ils possèdent. Le fait qu'une ressource nous appartient - donc privée - nous permet de lui donner une valeur. La valeur de cette même ressource sera toutefois estimée différente selon les agents et leurs intérêts divergents. Cette confrontation d'estimations *subjectives* de valeur débouche alors sur une valeur qui *in fine* sera mécaniquement plus *objective*, à savoir le *prix*. Ainsi, à notre question « pourquoi la donnée personnelle a-t-elle une valeur ? », nous pouvons répondre que c'est en partie parce que la donnée personnelle est privée qu'elle en dégage de la valeur. Dans une étude réalisée par le *Ponemon Institute* en 2015³⁴, nous observons d'ailleurs que plus la donnée personnelle est sensible et privée, plus sa valeur estimée est élevée. Si 56% des participants à l'enquête sont prêts à troquer leurs données

³³ Jacob Jean, *Le marché d'Hayek contre la démocratie*. In : Raison présente, n°95, 3e trimestre 1990. Bouleversements à l'Est, p. 65.

³⁴ Ponemon Institute, *Privacy and Security in a Connected Life* (2015) [en ligne] https://www.trendmicro.de/cloud-content/us/pdfs/security-intelligence/reports/rt_privacy_and_security_in_a_connected_life.pdf [accès le 5 avril 2020].

contre une rémunération financière, le prix varierait toutefois de 2,90\$ pour la vente d'une simple donnée de type genre jusqu'à 59,80\$ pour une information médicale ou encore 75,80\$ pour un mot de passe : soit en moyenne le prix de 19,60\$ pour une donnée personnelle. Et Hayek n'aura de cesse de souligner l'importance de la dimension informationnelle des prix. Les prix relatifs des biens et services transmettent une information sur leur état de rareté. Lorsque le prix d'une ressource augmente, cela signifie généralement que la demande pour cette ressource s'est intensifiée et que l'offre ne peut plus y répondre, peut-être temporairement. Il devient alors plus difficile de s'en approvisionner : la ressource devient donc plus rare. Sans ces prix pour nous informer, nous prendrions des décisions en rupture avec la réalité des raretés, menant alors (généralement) à la catastrophe économique. Les prix permettent donc non seulement de guider l'activité des consommateurs, mais aussi celle des producteurs, en leur permettant de calculer le coût d'opportunité des différents moyens de production qui s'offrent à eux. En l'absence de prix, le moyen de production le plus *technologiquement efficace* l'emporterait toujours sur le plus *économiquement efficient*, repercutant - dans le cas où ces deux moyens de productions seraient différents - son coût plus élevé sur le consommateur final. Ce qui inévitablement réduirait à terme la demande, entraînant alors des pertes pour le producteur.

Et si le producteur réalise des pertes, cela signifie qu'il n'est pas économique : il consomme plus de richesses que ce qu'il en crée pour la communauté. Cette richesse pourrait être allouée plus efficacement ailleurs pour, *in fine*, aboutir à un surplus économique. Et c'est notamment parce que le Planificateur central détient *tous* les moyens de production qu'il lui est impossible de faire ce même calcul économique rationnel.

La logique complète peut maintenant être condensée de la manière suivante : en l'absence de propriété privée, il n'y a pas de marché. Sans marché pour permettre la confrontation d'estimations subjectives de chacun, il n'y aura donc pas de véritable prix. Et sans ces prix objectifs pour informer de l'état de rareté des ressources, l'agent économique ne sera alors plus en mesure d'évaluer de manière rationnelle les moyens de production alternatifs et de choisir la méthode la moins coûteuse pour produire.

En résumé, du fait que la réalité économique est en perpétuelle agitation, que nos connaissances et nous-mêmes sommes dispersés (décentralisés), il est alors impossible pour une unité mère de mimer le marché par une planification centralisée.

2.3.4. Modèle d'arbitrage

Nous avons désormais toute l'intuition requise à la compréhension d'un modèle mathématique simplifié développé par le professeur Badie Farah (Ph.D. Ohio State University)³⁵ pour définir quand une entreprise devrait franchir le pas vers l'utilisation du *big data*. Soient **PI** et **P** dénotant respectivement le profit total généré par l'organisation *avant* et *après* l'instauration de son infrastructure *big data* ; **PBD**, **CBD** et **CBD_o** dénotant respectivement le *gain généré* par l'infrastructure *big data*, son *coût d'utilisation* et son *coût d'investigation*. Alors,

$$\mathbf{P} = \mathbf{PI} + \mathbf{PBD} - \mathbf{CBD} - \mathbf{CBD}_o \quad (1)$$

Mais le *gain généré* par l'infrastructure *big data* (**PBD**) ne pouvant être connu à l'avance, sa distribution de probabilité doit être estimée et servira à l'estimation de son espérance.

Notons **P_i** comme la probabilité que l'organisation réalise le gain **PBD_i** pour **i = 1, 2, ..., n**.

Donc,

$$\mathbf{PBD} = \Sigma[\mathbf{PBD}_i * \mathbf{P}_i] \text{ pour } \mathbf{i} = \mathbf{1, 2, \dots, n} \quad (2)$$

Si l'entreprise croit que ses coûts liés au *big data* suivront la même distribution, alors,

$$\mathbf{CBD} = \Sigma[\mathbf{CBD}_i * \mathbf{P}_i] \text{ pour } \mathbf{i} = \mathbf{1, 2, \dots, n} \quad (3)$$

L'équation (1) peut dès lors être réécrite sous la forme :

$$\mathbf{P} = \mathbf{PI} + \Sigma[\mathbf{PBD}_i * \mathbf{P}_i] - \Sigma[\mathbf{CBD}_i * \mathbf{P}_i] - \mathbf{CBD}_o \text{ pour } \mathbf{i} = \mathbf{1, 2, \dots, n} \quad (4)$$

Finalement, la décision d'implémenter ou non l'infrastructure *big data* dépendra de la relation suivante :

$$\mathbf{P} \geq \mathbf{PI} \quad (5)$$

Si l'équation (5) est vérifiée, cela signifie que le *big data* rapporte plus que ce qu'il ne coûte : son gain excède son coût d'utilisation et d'investigation. Il sera alors avantageux pour l'entreprise d'instaurer un tel système. Sinon, elle ferait mieux d'attendre.

³⁵ Badie Farah, *Profitability and Big Data*. In : Journal of Management Policy and Practice, vol. 18(2) (2017) [en ligne] http://www.na-businesspress.com/JMPP/FarahB_web18_2_.pdf, pp. 49-51 [accès le 10 avril 2020].

Cas spéciaux

Si le coût et gain liés à l'instauration de l'infrastructure *big data* sont connus d'avance avec une probabilité de 100%, alors tous les autres coûts et gains alternatifs ne se réaliseront pas. Formellement, $P_i = 0$ pour tous les i excepté un seul pour lequel $P_i = 1$. L'équation (4) se réécrit alors comme suit :

$$P = PI + PBD_i - CBD_i - CBD_0 \text{ avec } i \text{ pour lequel } P_i = 1 \quad (6)$$

Remarquons que si $P_i = 0$ pour tous les i ($i = 1, 2, \dots, n$), alors l'infrastructure *big data* ne contribuera pas à la profitabilité de l'entreprise, le profit d'avant et d'après son utilisation étant égal, et l'équation (4) devient dans ce cas :

$$P = PI \quad (7)$$

Pour plus de concret, supposons que les deux variables aléatoires (gain et coût liés à l'utilisation du *big data*) suivent une loi uniforme, alors toutes leurs estimations auront la même probabilité de se réaliser. Formellement, si $P_i = 1/n$ où n est le nombre de valeurs que peut prendre i , alors l'équation (4) s'écrit sous cette forme :

$$P = PI + 1/n * \Sigma[PBD_i - CBD_i] - CBD_0 \text{ pour } i = 1, 2, \dots, n \quad (8)$$

L'auteur propose finalement de calculer la rentabilité d'un projet *big data* en faisant appel à trois indicateurs, au choix ou de façon combinée. Le premier est l'indicateur ROI (Return On Investment), mettant en relation directe tous ses gains nets et coûts estimés :

$$ROI = 100 * [(\Sigma PBD_i - \Sigma CBD_i - CBD_0) / (\Sigma CBD_i + CBD_0)] \text{ pour } i = 1, 2, \dots, n \quad (9)$$

L'entreprise peut aussi utiliser l'indicateur RR (Rate of Return). C'est un deuxième indicateur pour mesurer la rentabilité d'un investissement sur une période donnée (un an par exemple) vis-à-vis de son profit initial. Une formule dans laquelle PI est alors le profit réalisé en début d'année et P le profit total en fin d'année incorporant la valeur ajoutée du *big data*. Sa rentabilité peut alors être comparée à d'autres projets d'investissement que l'entreprise aurait choisi de poursuivre durant l'année.

$$RR = 100 * (P - PI) / PI \quad (10)$$

En remplaçant les termes, on obtient alors :

$$RR = 100 * (\Sigma[PBD_i * P_i] - \Sigma[CBD_i * P_i] - CBD_0) / PI \text{ pour } i = 1, 2, \dots, n \quad (11)$$

Enfin, il existe un troisième indicateur appelé taux de rentabilité interne ou IRR en anglais (Internal Rate of Return), soit le taux d'actualisation **R** pour lequel la valeur actuelle nette de l'investissement (NPV en anglais pour Net Present Value) est égale à 0. Quand la NPV est égale à 0, cela signifie que le coût initial de l'investissement est égal à la valeur actualisée de ses cash flow futurs. Pour un projet d'investissement générant des flux monétaires sur **T** périodes données, plus son IRR sera élevé (et positif), plus le projet sera rentable. Formellement, la NPV appliquée au projet du *big data* s'écrit sous la forme suivante :

$$\text{NPV} = \sum_{i=1}^n \sum_{t=1}^T \frac{(\text{PBD}_i - \text{CBD}_i)_t}{(1+R)^t} - \text{CBD}_0 \text{ pour } i = 1, 2, \dots, n ; t = 1, 2, \dots, T \quad (12)$$

Pour calculer le IRR, il faut donc fixer la NPV à 0, donc :

$$0 = \sum_{i=1}^n \sum_{t=1}^T \frac{(\text{PBD}_i - \text{CBD}_i)_t}{(1+\text{IRR})^t} - \text{CBD}_0 \text{ pour } i = 1, 2, \dots, n ; t = 1, 2, \dots, T \quad (13)$$

Dès lors, s'il n'y a qu'une période (1 an par exemple), soit $T = 1$, alors le IRR est calculé comme suit :

$$0 = \sum [\text{PBD}_i - \text{CBD}_i] / (1+\text{IRR}) - \text{CBD}_0 \text{ pour } i = 1, 2, \dots, n \quad (14)$$

$$\text{CBD}_0 = \sum [\text{PBD}_i - \text{CBD}_i] / (1+\text{IRR}) \text{ pour } i = 1, 2, \dots, n \quad (15)$$

$$\text{CBD}_0 * (1+\text{IRR}) = \sum [\text{PBD}_i - \text{CBD}_i] \text{ pour } i = 1, 2, \dots, n \quad (16)$$

$$\text{CBD}_0 + \text{IRR} * \text{CBD}_0 = \sum [\text{PBD}_i - \text{CBD}_i] \text{ pour } i = 1, 2, \dots, n \quad (16a)$$

$$\text{IRR} * \text{CBD}_0 = \sum [\text{PBD}_i - \text{CBD}_i] - \text{CBD}_0 \text{ pour } i = 1, 2, \dots, n \quad (16b)$$

$$\text{IRR} = [\sum [\text{PBD}_i - \text{CBD}_i] - \text{CBD}_0] / \text{CBD}_0 \text{ pour } i = 1, 2, \dots, n \quad (17)$$

$$\text{IRR} = \sum [\text{PBD}_i - \text{CBD}_i] / \text{CBD}_0 - 1 \text{ pour } i = 1, 2, \dots, n \quad (18)$$

2.3.5. Le *big data* et la typologie de ses business models

Nous pensons qu'il est intéressant de clôturer cette première partie consacrée à l'économie de la donnée personnelle en présentant brièvement deux modèles économiques liés au *big data* qui ont été finement observés par Simon Chignard et Louis-David Benyayer. Ils permettront, nous l'espérons, d'être utilisés par le lecteur ou la lectrice de ce TFE comme référentiels dans leur utilisation quotidienne de services impliquant leurs données personnelles. Après tout, on connaît tous cet adage : « Si quelque chose est gratuit sur internet, c'est que vous en êtes le produit ».

Il y a donc d'une part les modèles dits *serviciels* pour lesquels nos données sont le carburant. Ces services existent et permettent d'être vendus comme produits parce que justement ils utilisent nos données. Le cas des assurances de voitures est marquant. Au lieu de parier avec la spéculation d'accidentologie, le tarif de l'assurance pourra être diminué ou augmenté selon l'attitude du conducteur ou de la conductrice sur la route : moins le comportement de conduite sera agressif, moins la facture sera élevée. Les modèles *serviciels* servent aussi à enrichir l'expérience de l'utilisateur final. En partageant ses données, le client bénéficiera d'analyse de pointe quant à son utilisation dudit service, augmentant ainsi son confort. Il y a donc là un « *troc implicite, données contre service [...], les données ne servent pas uniquement à l'entreprise qui les collecte, [mais aussi] aux utilisateurs par un meilleur service* »³⁶.

Il existe aussi les modèles *bifaces* comme ceux qu'utilisent Google ou Facebook en plus de leurs modèles *serviciels*. Ces entreprises retirent du profit de vos données en les partageant avec des tiers contre une rémunération monétaire. Il s'agit là bien souvent d'une *monétisation secondaire*. Les données détenues par l'organisation ou les informations qui en découlent (donnée *versus* information) sont précieuses pour les autres parties - et suffisamment précieuses pour les revendre. Ces entreprises servent d'intermédiaire entre deux acteurs, à savoir les utilisateurs d'un côté et les annonceurs de l'autre. Ils sont donc en ce sens *bifaces* car « *c'est l'existence d'une face qui représente une valeur pour l'autre face* »³⁷. Fait surprenant, cela ne se limite pas au secteur privé. Même le secteur public a profité de la mise à disposition de ses données. Aux États-Unis, l'État de Floride a récolté plus de 60

³⁶ S.Chignard et L.D. Benyayer, *Datanomics, les nouveaux business models des données*, Fayard, 2014, p. 83.

³⁷ *ibid.*, p. 80.

millions de dollars en 2010 en vendant des informations qu'il recueillait auprès des conducteurs³⁸.

3. Innovation et gouvernamentalité algorithmique

3.1. Qu'est-ce qu'une innovation technologique ?

Les deux types d'innovation les plus populaires sont celles liées au produit et au processus. L'innovation de produit (*product innovation*) se réfère à une amélioration d'un produit existant ou à la création d'un nouveau produit. Quand elle vise à augmenter la capacité d'un produit existant, on dit qu'elle est incrémentale. En effet, *l'incrément se rapporte à l'existant*. Par exemple, une augmentation de la résolution d'un téléviseur est une innovation de produit incrémentale. La délimitation de l'incrément n'est pourtant pas aussi précise qu'on le souhaiterait. Le concept même de l'innovation incrémentale est tout aussi relatif. Savoir si le fait de rajouter des fonctionnalités à un produit relève *toujours* de l'incrémental est une question qui pose encore aujourd'hui débat. Certains diront que *l'iPhone 7* et ses deux caméras qui n'existaient pas dans les iPhones précédents était une vraie révolution pour l'époque, marquant une rupture technologique. D'autres y verront là une simple continuité technologique. Par contre, tous s'accordent à dire que le *premier iPhone* était de toute évidence une innovation de produit de *rupture*, un genre de produit totalement nouveau sur son marché. *En ce sens, les innovations de rupture se distinguent alors aisément des innovations incrémentales.*

Pourtant, ni Amazon ni Google ni Facebook ni Netflix n'ont atteint des milliards de dollars de capitalisation boursière grâce à la seule innovation de produit³⁹. Le deuxième type d'innovation est donc appelé *process innovation* ou *innovation de procédé*. Il comprend des changements dans toutes les activités principales ou de soutien de la chaîne de valeur, parmi lesquelles une logistique améliorée dans la gestion des stocks, une meilleure planification des campagnes *marketing (media planification)* ou encore un processus de fabrication amélioré. Par exemple, l'utilisation de données en temps réel sur la demande pour planifier le cycle de production est une amélioration de processus. Ces deux types d'innovation divergent donc tant dans le but recherché que dans leur visibilité. Quand les innovations de produits sont souvent visibles pour les clients finaux, les innovations de processus ne le sont pas. Les

³⁸ Boston Consulting Group, *The value of our digital identity* (2012), p. 12.

³⁹ Sorescu, Alina. (2017), *Data-Driven Business Model Innovation : BUSINESS MODEL INNOVATION*. In : *Journal of Product Innovation Management.*, p. 691.

innovations de produits visent l'amélioration des produits, tandis que l'innovation de processus vise l'optimisation des coûts.

Cependant, toutes se rejoignent sur un point, car elles partagent généralement le concept de *valeur* : de sa création à sa transmission en passant par son appropriation. La *création de valeur* se réfère aux ressources et processus qui sous-tendent le développement du bien ou du service. La *transmission de valeur* se concentre justement sur la valeur ajoutée qu'apportent ces produits et services au consommateur ainsi que sur l'environnement dans lequel ils sont vendus (transmis). Enfin, *l'appropriation de valeur* fait référence aux fonctions de coûts et de revenus de l'entreprise. On dit alors que l'entreprise s'approprie la valeur.

On parlera plus généralement d'innovation liée au modèle d'entreprise (*business model innovation*) quand l'un de ces trois maillons est amélioré. Ce méga-type d'innovation englobant dès lors les innovations de type produit et procédé. « A business model innovation is defined as a change in the value creation, value appropriation, or value delivery function of a firm that results in a significant change to the firm's value proposition. »⁴⁰ Il est à noter que les innovations de produit ou procédé peuvent impacter un ou plusieurs maillons sans nécessairement avoir un effet sur la proposition de valeur globale d'une entreprise.

Alors que l'innovation de produit trouve généralement son origine dans le maillon de *création de valeur* et est nécessaire pour de nombreuses entreprises dans le renouvellement de leur assortiment de produits, lui-même associé au maillon de *transmission de valeur*, l'innovation de procédé consiste quant à elle non seulement dans un changement dans la *création de valeur* mais aussi dans son appropriation, ne faisant alors pas toujours bénéficier directement le consommateur final (transfert de valeur). Supposons une amélioration dans la chaîne d'approvisionnement qui rende le canal simplement plus efficace. Ce gain d'efficacité sera dit fermé car *approprié* par l'entreprise, et aucun transfert de valeur au consommateur final ne sera alors effectué. Par contre, la mise au point par l'entreprise Zara d'une chaîne d'approvisionnement rapide et dynamique produisant une nouvelle gamme de vêtements en 2 à 4 semaines, aura un impact significatif sur la valeur transmise aux clients (Girotra & Netessine, 2011), permettant alors de combler leurs besoins de variétés. Ce modèle est rendu possible grâce au *big data* qui, d'une part, analyse les dernières tendances de mode en temps réel pour plus de réactivité, et d'autre part contrôle les inventaires des

⁴⁰ Sorescu, Alina. (2017), *Data-Driven Business Model Innovation : BUSINESS MODEL INNOVATION*. In : *Journal of Product Innovation Management.*, p. 692.

magasins pour plus de flexibilité. Ceci se traduit à la fois par une réduction des coûts de stockage pour l'entreprise et comme un outil précieux à la prévention de rupture de stock.

3.2. Tension explicite entre vie privée et innovation

Nous observons donc qu'il existe une intime relation entre innovation et *big data* qui, rappelons-le, nécessite une quantité significative de données personnelles, allant - à première vue - à l'encontre du droit à la vie privée. Dans une Europe voulant temporiser son utilisation - une Europe même craintive, diront certains -, la donnée personnelle est sujette à des réglementations toujours plus strictes. Récemment, le *Règlement général sur la protection des données (RGPD)*, entré en vigueur le 24 mai 2016, est parfois vu comme un frein à l'innovation. Cet encadrement législatif est-il conciliable avec l'innovation ? Pour répondre à cette question, il est nécessaire d'examiner la tension qui règne entre les grands principes européens au regard de la protection des données personnelles et la logique du *big data*, deux logiques s'opposant frontalement selon la professeure Antoinette Rouvroy⁴¹.

Là où la nature même du *big data* est de collecter des quantités massives de données, soit la *maximisation* de la collecte sans discriminant, s'oppose la logique de protection prônant le principe de *minimisation* de cette dernière, statuant que seules les données strictement nécessaires au projet peuvent légitimement être collectées. Ainsi, quand la réglementation préconise que la finalité de la collecte doit être *a priori* déclarée, précisée et justifiée en raison d'un objectif bien défini et limité dans le temps - les données devant être effacées une fois le but atteint -, le *big data*, au contraire, tire sa force d'une collecte maximale de données brutes : peu importe leur qualité, de façon illimitée dans le temps pour un cumul croissant ; peu importe leur finalité établie *a priori*, l'utilité de la donnée pouvant se manifester *a posteriori* en recourant à des pratiques statistiques de *data-mining* qui font transpirer la donnée *a priori* inutile en information de valeur à mesure que croît le *set* de données.

L'opposition entre ces deux mécaniques est donc flagrante. Leur objet d'étude est cependant différent. Si le *big data* se concentre avant tout sur l'abondance des données, soit la *quantité*, c'est le caractère *qualitatif* de la donnée personnelle qui polarise les activités des régimes juridiques. Sa qualité justifiant sa protection. Ainsi, ce ne serait plus tant l'utilisation inappropriée de la données personnelle, fût-elle impersonnelle, qui poserait problème

⁴¹ A. Rouvroy, *Homo juridicus est-il soluble dans les données ? Droit, normes et libertés dans le cybermonde*, Liber Amicorum Yves Poulet, Larcier, 2018, p. 429.

prioritairement ; ce serait plutôt sa prolifération outre mesure et sa disponibilité en grande quantité. Serait-ce là le signe timide d'une compatibilité naissante ?

« Le défi qui serait le nôtre [...] pourrait donc s'énoncer ainsi : comment tenir compte de la nature relationnelle et, donc, aussi collective à la fois de la donnée [...] et de ce qui mérite d'être protégé, y compris à travers la protection des données ? »⁴² Plus généralement, le débat d'aujourd'hui porte sur l'équilibre à trouver entre innovation, dynamique mais parfois *intrusive*, et protection de la vie privée, requise mais *pesante*.

S'ensuit que contester le bien-fondé de la protection de la vie privée en affirmant que celle-ci entrave l'innovation et brime le développement économique, manquerait cruellement de finesse d'analyse. Si les *GAFAs* américains sont parvenus à innover dans leur marché, ce n'est pas tant grâce à un cadre législatif plus laxiste, mais parce qu'ils se sont insérés dans une longue lignée entrepreneuriale expérimentée qui s'est développée dans la Silicon Valley⁴³.

De même, si la réglementation est perçue comme fort contraignante pour le développement économique, une bonne gouvernance des données semble pourtant bien plus rejoindre les intérêts des entreprises qu'il n'y paraît. En effet, comme déjà stipulé dans une section précédente⁴⁴, le risque encouru par les entreprises en cas de violation de données serait estimé à 440 milliards d'euros de pertes pour l'année 2020⁴⁵. Pour cause, une rupture de confiance du consommateur qui se traduirait par une baisse du chiffre d'affaire, entre autres.

Si le traitement des données personnelles est donc vecteur de création de valeur, exploiter ces données peut aussi faire subir à l'entreprise de gros dommages et être source de destruction de valeur. Les conséquences pour l'entreprise sont alors catastrophiques, non seulement économiquement parlant mais aussi en termes d'image. Gagner la confiance des clients peut prendre plusieurs années, la perdre ne prend qu'un instant. Conscientes que leur stabilité économique dépend donc d'une meilleure gouvernance des données, les entreprises ont alors intérêt à adapter leurs modèles économiques et, de surcroît, à transformer les contraintes émanant des réglementations en avantages compétitifs, *sous forme de label*

⁴² *ibid.*

⁴³ C. de Salle, *La vie privée à l'ère des big data : dangers & opportunités de la révolution numérique (2018) [en ligne] cjjg.be/wp-content/uploads/2018/12/CJG-ETUDE-Big-Data-Page-par-page.pdf [accès le 20 avril 2020] p.13.*

⁴⁴ 2.3.3.i *Comment valoriser la donnée personnelle.*

⁴⁵ Boston Consulting Group, *The value of our digital identity* (2012), p. 4.

qualité⁴⁶ ou label de confiance. Pour accéder à ce statut ultime d'acteur de confiance, il convient pour l'entreprise de garantir sécurité et vie privée à tous les échelons composant ses systèmes d'informations (SI) - dès sa création, tout comme il est nécessaire pour l'architecte conventionnel de s'efforcer de définir les normes de stabilité d'une construction pour éviter sa ruine. C'est ce qu'affirme l'ancienne commissaire à l'information et à la protection de la vie privée de la province canadienne de l'Ontario, Ann Cavoukian⁴⁷.

Ann Cavoukian part du principe qu'il ne faut pas voir la relation qui existe entre technologie et vie privée comme un jeu à somme nulle, mais au contraire comme des interactions à somme positive mutuellement bénéfiques. Il faut donc chercher à atteindre un optimum au sens de Pareto. Comme la sécurité, la confidentialité ne doit pas diminuer la fonctionnalité de la technologie, et de surcroît l'innovation.

En ce sens, elle et Mark Dixon (architecte d'entreprise chez Oracle) ont coécrit le livre blanc *Privacy and Security by Design : An Enterprise Architecture Approach*⁴⁸ dans lequel les approches *privacy by design* (initialement imaginée par Ann Cavoukian) et *security by design* sont mises côte à côte, chacune d'elle visant à l'intégration harmonieuse du respect de la vie privée dans un cadre technologique innovant. Deux approches inséparables (la *privacy* est en effet rattachée aux défaillances de la *security* par son risque de piratage) comprenant des principes tels que la proactivité (il faut anticiper à l'avance, dès la conception du système, *by design*, les problèmes de sécurité ou de confidentialité), la règle par défaut (la protection au sens large n'est pas optionnelle), ou encore le choix de la transparence totale pour permettre une vérification indépendante, augmentant ainsi le climat de confiance.

Faire le choix pour une technicité conforme *by design* aux législations ne semble donc pas insurmontable. Si la prise en compte des enjeux d'une protection de la vie privée se fait le plus en amont possible, il devient alors parfaitement possible d'innover de façon durable via le *big data* tout en garantissant le respect à la vie privée. C'est même encouragé. N'étant pas antinomique, concilier droit à la vie privée et innovation s'inscrit dès lors dans une démarche non seulement plus responsable et durable mais aussi plus rentable, faisant bénéficier l'entreprise d'avantages compétitifs. Finalement, une tension qui n'a plus lieu d'être.

⁴⁶ CIGREF. *L'Économie des données personnelles : les enjeux d'un business éthique. (2015)* [en ligne] cigref.fr/wp/wp-content/uploads/2015/11/CIGREF-Economie-donnees-perso-Enjeux-business-ethique-2015.pdf [accès le 9 avril 2020], p. 19.

⁴⁷ Ann Cavoukian et Mark Chanliau, *Privacy and Security by Design : A convergence of Paradigms*, janvier 2013, p.3.

⁴⁸ Ann Cavoukian et Mark Dixon, *Privacy and Security by Design : An Enterprise Architecture Approach*, septembre 2013.

3.3. Tension implicite entre modernité et renversement de pouvoir

La tension qui s'exerce entre innovation et droit à la vie privée semble donc être plus ou moins explicite du fait qu'elle peut être décrite formellement par ses mécanismes et les logiques contraires de ses deux extrémités. Pourtant se cache une tension beaucoup plus implicite de par sa progression lente et silencieuse. Une tension naissante entre la volonté des individus de vivre dans une société toujours plus moderne, une société de confort impliquant *de facto* la gouvernance des algorithmes, et le désir de mainmise sur leur existence (garder le contrôle sur leur vie auquel ils aspirent tant). Un malaise existentiel menant à une tension s'exerçant entre modernité et renversement de pouvoir.

Alain Supiot, juriste nantais spécialiste du droit du travail, parle même de gouvernance par les nombres. Dans son livre *La gouvernance par les nombres*, il écrit que « depuis le début des Temps modernes, le vieil idéal grec d'une cité régie par les lois et non par les hommes a pris une forme nouvelle : celui d'un gouvernement conçu sur le modèle de la machine [...]. L'idéal d'un pouvoir impersonnel a pris un nouveau visage depuis la Seconde Guerre mondiale, le nombre remplaçant progressivement la loi comme fondement des obligations entre les hommes. »⁴⁹

Nombres et données étant le carburant algorithmique, nous serions donc en passe d'être gouvernés par les algorithmes (si pas déjà). C'est aussi la thèse d'Antoinette Rouvroy. Avant le *big data*, les statisticiens se contentaient d'agréger des données à un niveau macro et d'en déduire des informations. Aujourd'hui, le *big data* perpétue les mêmes calculs, mais grâce à des corrélations entre données à première vue anodines, il lui est désormais possible de continuellement « zoomer » sur l'individu et, de façon dynamique, jusqu'à lui induire des comportements futurs. Un comportementalisme de l'utilisateur qui est analysé à un niveau indirectement micro.

Dès lors, si le *data-mining* et systèmes de profilage découlant de ces algorithmes se disent *user centric*, plaçant l'utilisateur *au centre* de toutes les attentions, il s'avère bien regrettable que tout ce qui n'est pas mesurable chez l'individu ne puisse être pris en compte. Ainsi donc,

⁴⁹ Alain Supiot, *La gouvernance par les nombres*. Cours au Collège de France (2012-2014), coll. « Poids et Mesures du Monde », Fayard, 2015, p. 305.

ses intentions ou désirs sont - s'ils n'ont pas été inférés automatiquement - tout bonnement ignorés. Là est tout le paradoxe : l'individu est le *sujet* mais en même temps *l'objet*⁵⁰.

Parce qu'il délimite *a priori* le champ d'action des individus, le *big data* incarne alors une nouvelle « gouvernamentalité algorithmique »⁵¹ et il s'inscrit résolument dans un capitalisme informationnel qui privilégie le caractère inductif et prédictif de la donnée sur l'intégrité émotionnelle de l'individu. Notre société le tolère. Tout du moins, s'il y a rébellion, celle-ci est, semble-t-il, marginale. C'est aussi pour cette raison que nous défendons l'idée selon laquelle cette tension entre modernité et renversement de pouvoir est implicite car tacite.

Pourquoi aussi peu de récalcitrance ? La réflexion d'Antoinette Rouvroy nous apporte là quelques éléments de réponse : « La force du gouvernement algorithmique, les raisons pour lesquelles il ne rencontre que peu ou pas d'obstacle et très peu de récalcitrance, tient aux rapports inédits qu'il noue avec la temporalité (s'attachant à gouverner le potentiel, le virtuel plutôt que l'actuel), avec les sujets (auxquels il ne s'adresse que très indirectement, et auxquels il paraît donc inoffensif), et avec le « réel » à partir duquel il gouverne, et dont il semble émaner spontanément. »⁵²

Récemment, Hugues Bersini, professeur à l'Université libre de Bruxelles et codirecteur de l'Institut de recherches interdisciplinaires et de développement en intelligence artificielle (*Iridia*), s'est penché sur la question de la *gouvernamentalité algorithmique* dans une tribune du « Monde »⁵³. Selon lui, la solution à notre malaise résiderait dans la mise en place de deux conditions démocratiques auxquelles tout algorithme devrait se soumettre pour être légitime. Primo, l'algorithme doit être transparent et connu de tous, sa finalité doit avoir été déclarée *a priori*. Secundo, il aura été mis en place, discuté et délibéré par toutes celles et ceux qui le subiront ou en bénéficieront.

Finalement, tout comme il est nécessaire de recourir à une bonne gouvernance des données pour une moindre friction entre innovation et droit à la vie privée, il en va de même pour une bonne gouvernamentalité algorithmique qui, si elle n'efface en rien la tension, peut tout du moins adoucir les moeurs.

⁵⁰ CNIL, *Vie privée à l'horizon 2020* (2012), p.20.

⁵¹ Rouvroy Antoinette et Thomas Berns, « Le nouveau pouvoir statistique. Ou quand le contrôle s'exerce sur un réel normé, docile et sans événement car constitué de corps « numériques »... », *Multitudes*, vol. 40, n° 1, 2010, p. 89.

⁵² *ibid.*, p. 90.

⁵³ Hugues Bersini. (2020), *Libertés numériques* : « La gouvernance algorithmique est d'une efficacité redoutable ». Tribune dans *Le Monde*.

4. Conclusion

Notre introduction a débuté avec une sentence d'actualité : *le big data est à la mode*. Pour la conclusion, nous dirons que le *big data*, c'est avant tout un point de rupture dans le temps, marquant *un avant* et *un après*. Avant, la donnée était *grossièrement* statistique, elle-même étant une statistique d'antan. Jusqu'à récemment, avant l'arrivée des ordinateurs dans les années 60, nos décideurs politiques se contentaient de prendre des décisions sur base de moyennes. Elles mêmes étaient calculées à partir d'échantillons de taille modérée, allant de quelques milliers à quelques centaines de milliers d'observations, elles mêmes imparfaites et dépréciées. L'Etat faisait alors avec les moyens du bord. L'appartenance de l'individu à tel quantile était approximative. Par contre, *l'après*, c'est la possibilité de mettre des machines au service du pouvoir. Des bases de données incalculables, allant du million aux milliards d'observations, rapidement triées en fonction d'un discriminant choisi. C'est donc la possibilité de regrouper et de diviser selon un nombre quasiment infini de critères, fragilisant ainsi les représentations classiques des populations. La moyenne laisse alors la place au profil numérique. Finalement, *l'après*, c'est surtout d'avoir permis au *big data* de diviser pour mieux régner. Et le règne peut être politique ou économique.

Les enjeux que soulève la donnée au sens large du terme sont effet autant politiques qu'économiques ou sociétaux. Dans cette imbrication de sphères se dérobe continuellement le précieux pouvoir que la donnée leur confère. Tantôt il sera un outil de pouvoir politique impersonnel, les lois se faisant remplacer progressivement par les nombres, tantôt il sera économique, tirant alors de l'individu la seule caractéristique d'agent économique. Dépossédé de son intégrité émotionnelle, le sujet objetisé est poussé au service d'un capitalisme informationnel. Emerge alors une lueur d'espoir, malheureusement elle aussi faussée quand la donnée vient nourrir un fantasme illusoire d'infinitude. Tel un héros existentiel inauthentique, l'individu se contentera alors de la grosse part (70%) du bénéfice économique qui lui revient de droit. Car il faut rappeler que sans lui et sa précieuse substance, le bénéfice potentiel aurait été tout simplement inexistant, *caduc*.

Que son rôle économique-philosophique soit perçu de façon optimiste ou non, c'est un état de fait : la donnée d'aujourd'hui est numérique, analysée et calculée. Si bien qu'elle rentre parfaitement dans des modèles d'arbitrage. La donnée étant relative, son modèle le sera tout autant, permettant alors à tout un chacun d'y voir son propre intérêt, d'où le maintien du *statu quo*.

En effet, alors que plusieurs scandales liés aux fuites de données ont inquiété l'opinion publique, notamment l'affaire *Cambridge Analytica* qui, en 2015, a éclaté au grand jour, rares sont encore celles et ceux qui ont pris l'initiative de quitter définitivement les réseaux sociaux. Pour cause, une addiction au partage des données personnelles qui s'expliquerait selon le psychologue Michael Stora⁵⁴ par le fait que les réseaux sociaux nous procurent un sentiment de réconfort. Nous avons besoin d'être complimentés et rassurés à coups de *likes*. Aussi, ayant beau savoir que nos données sont traitées, nous n'en sommes relativement que peu conscient·es, n'acceptant tout simplement pas l'idée que nous pourrions être utilisé·es. Se défaire d'addictions qui altèrent notre liberté reste donc quelque chose de difficile, et plus encore si elles n'ont pas été conscientisées - quitte à simplement avoir été observées.

De même, s'il est admis que la lutte contre le réchauffement climatique est aujourd'hui encore inefficace, c'est parce que l'individu se contente d'observer *l'agir* de la collectivité, de façon détachée, qui elle-même est la somme des individualités. Il en va aussi de même pour le désir de voir un changement radical d'attitude collective face aux traitements du *big data* des données personnelles, mais la causalité étant bidirectionnelle, l'individu est alors coincé dans son malaise de circularité propre à tout changement sociétal. Il aspire à un changement systémique mais n'est pas prêt à sacrifier son propre confort et réconfort que lui procurent bien souvent les remèdes à la solitude existentielle, tels Google, Facebook ou Instagram. En d'autres mots, « aucun individu ne changera seul si le système dans lequel il vit ne se modifie pas. A l'inverse, aucun système ne changera si les individus qui le composent ne modifient pas leurs modes de vie »⁵⁵.

⁵⁴ Le Figaro, *Pourquoi nous sommes incapables de quitter Facebook : le regard d'un psy. Entretien avec Michael Stora* (2018) [en ligne]

<https://www.lefigaro.fr/vox/societe/2018/04/13/31003-20180413ARTFIG00187-pourquoi-nous-sommes-incapables-de-quitter-facebook-le-regard-d-un-psy.php> [accès le 2 mai 2020].

⁵⁵ Arnsperger, C. *Critique de l'existence capitaliste : Pour une éthique existentielle de l'économie*. Éditions du Cerf : Paris (2005) (ISBN : 2-204-07694-5), 210 pages, p.54.

5. Bibliographie

- Arnsperger, Christian. *Critique de l'existence capitaliste : Pour une éthique existentielle de l'économie*. Éditions du Cerf : Paris (2005, ISBN : 2-204-07694-5), 210 pages, pp. 18-55.
- Ariès, P. et Duby, G. *Histoire de la vie privée, III De la Renaissance aux Lumières*, Seuil, 1985, p.161.
- Bersini, Hugues. (2020), *Libertés numériques : « La gouvernance algorithmique est d'une efficacité redoutable »*. Tribune dans *Le Monde*.
- Boston Consulting Group. *The value of our digital identity*. (2012) [En ligne] <https://2zn23x1nwzzj494slw48aylw-wpengine.netdna-ssl.com/wp-content/uploads/2017/06/The-Value-of-Our-Digital-Identity.pdf>, [accès le 3 avril 2020].
- Boubaker, Leila, Leila Mellal et Mébarek Djebabra. « *Modèle DIC (Données – Informations – Connaissances), Outil support pour le développement des mémoires projets* », *La Revue des Sciences de Gestion*, vol. 243-244, no. 3, 2010, pp. 153-159.
- Boullier, D. (2019) « Rendre le numérique habitable : l'habitable », in Calbérac Y, Lazzarotti O., Levy J., Lussault M., *Carte d'identités. L'espace au singulier*, Paris, Hermann (Les colloques de Cerisy), pp. 151-174.
- Cavoukian, Ann et Chanliau, Mark. *Privacy and Security by Design : A convergence of Paradigms*, janvier 2013, p.3.
- Cavoukian, Ann et Dixon, Mark. *Privacy and Security by Design : An Enterprise Architecture Approach*, septembre 2013.
- CNIL, *Vie privée à l'horizon 2020*. (2012) [en ligne] https://linc.cnil.fr/sites/default/files/typo/document/CNIL-CAHIERS_IPn1.pdf [accès le 3 avril 2020], pp. 18-20.
- CIGREF. *L'Économie des données personnelles : les enjeux d'un business éthique (2015)* [en ligne] cigref.fr/wp/wp-content/uploads/2015/11/CIGREF-Economie-donnees-perso-Enjeux-business-ethique-2015.pdf [accès le 9 avril 2020], pp. 1-30.
- Chignard, S. et Benyayer, L.D. *Datanomics, les nouveaux business models des données*, Fayard, 2014, pp. 80-81.
- Domo, Inc. *Data Never Sleeps 7.0* (2019) [en ligne] <https://www.domo.com/learn/data-never-sleeps-7> [accès le 19 mars 2020].
- Fillol, Charlotte. « *Apprentissage et systémique. Une perspective intégrée* », *Revue française de gestion*, vol. no 149, no. 2, 2004, p. 35.
- Farah, Badie. *Profitability and Big Data*. In : *Journal of Management Policy and Practice* Vol. 18(2) (2017) [en ligne] http://www.na-businesspress.com/JMPP/FarahB_web18_2_.pdf, pp. 49-51 [accès le 10 avril 2020].

- Hayek, F.A. *Droit, législation et liberté, Le mirage de la justice sociale*, T.2, P.U.F., 1981, 221 pages, p. 131.
- Jacob, Jean. *Le marché d'Hayek contre la démocratie*. In : Raison présente, n°95, 3e trimestre 1990. *Bouleversements à l'Est*, p. 61.
- Kramer, Samuel Noah. *L'histoire commence à Sumer...* préface de Jean Bottéro. Coll. Champs Histoire, Paris. Flammarion, 2017.
- Le Figaro. *Pourquoi nous sommes incapables de quitter Facebook : le regard d'un psy*. Entretien avec Michael Stora (2018) [en ligne] <https://www.lefigaro.fr/vox/societe/2018/04/13/31003-20180413ARTFIG00187-pourquoi-nous-sommes-incapables-de-quitter-facebook-le-regard-d-un-psy.php> [accès le 2 mai 2020].
- von Mises, Ludwig. *L'Action Humaine*, Éditions de l'Institut Coppel, Paris (3e édition 1966) (2011), p.2.
- Ponemon Institute. *Privacy and Security in a Connected Life* (2015) [en ligne] https://www.trendmicro.de/cloud-content/us/pdfs/security-intelligence/reports/rt_privacy_and_security_in_a_connected_life.pdf [accès le 5 avril 2020].
- Ponemon institute LLC. *Étude 2018 sur le coût des violations de données : Impact de la BCM* (2018) [en ligne] <https://www.ibm.com/downloads/cas/PJDB7AB8> [accès le 5 avril 2020].
- Rouvroy, Antoinette. *Homo juridicus est-il soluble dans les données ? Droit, normes et libertés dans le cybermonde, Liber Amicorum Yves Pouillet*, Larcier, 2018, pp. 420 - 429.
- Rouvroy, Antoinette et Berns, Thomas. « Le nouveau pouvoir statistique. Ou quand le contrôle s'exerce sur un réel normé, docile et sans événement car constitué de corps « numériques »... », *Multitudes*, vol. 40, n° 1, 2010, pp. 89-91.
- de Salle, C. *La vie privée à l'ère des big data : dangers & opportunités de la révolution numérique* (2018) [en ligne] cjjg.be/wp-content/uploads/2018/12/CJG-ETUDE-Big-Data-Page-par-page.pdf [accès le 20 avril 2020] p.13.
- Statista. *Digital Economy Compass* (2019) [en ligne] <https://cdn.statcdn.com/download/pdf/DigitalEconomyCompass2019.pdf> [accès le 19 mars 2020].
- Supiot, Alain. *La gouvernance par les nombres*. Cours au Collège de France (2012-2014). Coll. « Poids et Mesures du Monde », Fayard, 2015, p. 305.
- Sorescu, Alina (2017). *Data-Driven Business Model Innovation : BUSINESS MODEL INNOVATION*. In : *Journal of Product Innovation Management*, pp. 691 – 692.