

Université de Technologie de Compiègne

---

# La marchandisation des données personnelles

---

Penser la technique aujourd'hui - PH03  
20 juin 2017

**Auteurs:**

Arnaud Vanhuele  
Pierre-Louis Lacorte

**Suiveur:**

François-Xavier Guchet

# Table des matières

<b>Table des matières</b>	<b>2</b>
<b>Introduction</b>	<b>4</b>
<b>1. Une nouvelle forme de capitalisme autour des data</b>	<b>5</b>
1.1 L'entrée dans une nouvelle ère: Une nouvelle définition du capitalisme	5
1.2 Le capitalisme cognitif comme base de création de valeur économique des données personnelles ?	6
1.2.1 Qu'est ce que le capitalisme cognitif?	6
1.2.1.1 En bref	6
1.2.1.2 Les facteurs d'évolution à l'origine du concept	6
1.2.2 Un nouveau rapport capital/travail	7
<b>2. Création de valeur économique aux données personnelles.</b>	<b>8</b>
2.1 Qu'est ce qu'une donnée personnelle ?	8
2.1.1 Une définition complexe ?	8
2.1.2 D'où vient cette complexification ?	9
2.1.2.1 Les données sont de plusieurs natures	9
2.1.2.2 Les données sont relatives	9
2.1.2.3 Les données sont ouvertes	10
2.2 Valorisation des données personnelles	10
2.2.1 Les potentiels de la valorisation des données personnelles	10
2.2.1.2 Un potentiel stratégique	11
2.2.1.3 Un potentiel opérationnel	11
2.2.1.4 Un potentiel de recherche et développement	11
2.2.1.5 Un potentiel marketing	12
2.3 Les modèles d'affaires autour des données personnelles	13
2.3.1 Deux modèles économiques en expansion	13
2.3.2 Peut-on se protéger ou concurrencer de tels modèles?	14
2.4 L'anonymisation et la valorisation des données est-elle compatible?	14
2.4.1 Comment réaliser une anonymisation ?	14
2.5 Un Big Data sans données personnelles non anonymes?	15
<b>3. Constater la création de valeur, pour qui a telle une plus-value? Captation d'une plus-value par le client?</b>	<b>16</b>
3.1 Principe de la création d'une plus-value	16
3.2 Fonctionnement de l'analyse de données	17
3.2.1 Format des données	17
3.2.2 Traitements des données	17
3.2.2.1 Quelques méthodes d'analyse de données	17
3.2.2.2 L'évaluation des performances des classifieurs	20

3.2.2.3 Origine des données	20
3.2.3 Expliquer les tendances observées	20
3.3 Plus-value pour l'entreprise	21
3.3.1 Caractérisation de la plus-value	21
3.3.1.1 Mesure de la satisfaction des clients	21
3.3.1.2 Le côté abstrait de la création de valeur	22
3.3.2 Pouvoir de prédiction	22
3.3.2.1 Conséquences de ces plus-values sur les sciences sociales	23
3.4 Plus-value pour le client	23
3.4.1 Les bénéfices pour les clients	23
3.4.2 Conséquences de la personnalisation des espaces numériques	24
<b>Conclusion</b>	<b>26</b>
<b>Bibliographie</b>	<b>27</b>

## Introduction

Nous sommes aujourd'hui plongés dans l'ère du Big data, de l'interconnexion des données et de leur valorisation. Dans ce contexte Big Data, les données personnelles occupent une place de plus en plus importante. En effet, leurs modes de production, leurs moyens de collecte et d'analyse ont explosé ces dernières années et offrent des potentiels de valorisation gigantesques, mettant notamment l'expérience client au cœur des stratégies. La relation client devient de plus en plus proactive, ce qui permet d'anticiper les besoins, d'adapter les stratégies et les opérations.

Il semble donc exister une vraie marchandisation des données personnelle. Comment expliquer ce phénomène? Sur quels concepts se base t-elle? Comment créer de la valeur à une donnée? Qui perçoit la plus-value, l'entreprise ou le client? Tout autant de questions soulevées que nous essayerons de traiter de ce mémoire.

# 1. Une nouvelle forme de capitalisme autour des data

## 1.1 L'entrée dans une nouvelle ère: Une nouvelle définition du capitalisme

En effet depuis ses premières apparitions dans le langage au XVIIIe siècle, la définition du capitalisme s'est vu bouleversé afin de mieux correspondre à notre ère et aux nouveaux usages que l'on en fait.

Lors de son acceptation en 1753, il désignait "l'état de la personne qui possède des biens", et jusqu'au début du XIXe siècle la définition est extrêmement lié au concept de richesse et à la possession de biens.

Encore en 2005, on définissait le Capitalisme par "l'état de celui qui est riche".

Aujourd'hui les concepts important qui définisse le Capitalisme sont l'entreprise privée des moyens de production et la liberté du marché:

- *"statut juridique d'une société humaine caractérisée par la propriété privée des moyens de production et leur mise en œuvre par des travailleurs qui n'en sont pas propriétaires"*
- *"système de production dont les fondements sont l'entreprise privée et la liberté du marché"*

Ces définitions montrent bien l'importance qu'est la privatisation des moyens de production, c'est ensuite aux individus détenteurs de ces propriétés d'en choisir librement la finalité que ce soit servir le consommateur, la recherche du profit ou l'accumulation du capital.

Ces définitions dénotent fortement face aux anciennes définitions du capitalisme qui se réduisait à la l'accumulation de richesse auxquelles la définition de Marx se rapproche.

En effet il définit le Capitalisme par la recherche de profit, l'exploitation des travailleurs par les détenteurs des moyens de production.

Il semblerait donc qu'on vive en ce moment une révolution en terme de définition du capitalisme. Comme le dit Yann Moulier Boutang, *"Notre époque n'est assurément pas celle d'une transition vers le socialisme. L'ironie de l'histoire est que, si transition il y a, comme nous le pensons, il s'agit d'une transition vers un nouveau type de capitalisme. De ce point de vue, le socialisme et la gauche semblent en retard d'une révolution. La « mondialisation » actuelle correspond en effet à l'émergence, depuis*

1975, d'un troisième type de capitalisme. Celui-ci n'a plus grand chose à voir avec le capitalisme industriel qui, à sa naissance (1750-1820), rompit avec le capitalisme mercantiliste et esclavagiste." Cette troisième révolution du capitalisme, le capitalisme cognitif, ne poserait-il pas les bases propices à la création de valeur économique des données personnelles?

## 1.2 Le capitalisme cognitif comme base de création de valeur économique des données personnelles ?

### 1.2.1 Qu'est ce que le capitalisme cognitif?

#### 1.2.1.1 En bref

Le capitalisme cognitif est une forme nouvelle de capitalisme, dans laquelle la production de connaissances et, plus généralement, " la production de l'homme par l'homme " jouent le rôle principal, alors que, dans le fordisme, prédominaient la spécialisation des tâches, la recherche d'économies d'échelle et l'investissement matériel. L'efficacité ne réside plus dans les gains de temps de travail, mais dans les processus d'apprentissage et d'innovation. La capacité créatrice est dans la tête des détenteurs des connaissances, non dans celles du propriétaire des machines. Lequel, par le biais des droits de propriété et la privatisation du savoir et du vivant, vise à capter la part la plus importante possible de cette richesse produite par l'intelligence collective.

#### 1.2.1.2 Les facteurs d'évolution à l'origine du concept

Il est communément admis que les facteurs permettant au capitalisme cognitif, c'est à dire à la nouvelle forme d'économie du savoir, d'être en plein essor sont le développement du secteur tertiaire et des activités immatérielles, du progrès des NTIC (Nouvelles Technologies de l'Information et de la Communication), de la mondialisation qui réduit la rémunération du travail physique, et qui accroît au contraire le profit des « idées », en leur fournissant des débouchés élargis : il est notable que vendre une idée à 10 personnes coûte autant au « fabricant » que de la vendre à des milliers, mais que les profits dégagés sont eux infiniment supérieurs dans le second cas.

Tous ces facteurs permettent donc un essor de cette nouvelle économie du savoir qui a pour matière première non plus une matière matérielle (les denrées, matières premières et sources d'énergie, même si les enjeux sont grands sur ces sujets) mais dans l'immatérielle (l'information, le savoir-faire et la connaissance).

Un point qu'il serait bon de souligner est aussi le cadre de l'éducation dans le développement de cette nouvelle économie. Notre génération entend depuis notre

enfance qu'il faut travailler à l'école et qu'il faut réussir à obtenir un travail "intellectuelle". Bien entendu cette éducation n'est pas transmise à tout le monde mais sa part ne cesse de croître dans le marasme des éducations inculquées. On pousse donc les jeunes, à développer des compétences et des capacités à l'innovation qui nourrissent cette économie du savoir. Est-ce cette économie croissante qui pousse cette éducation ou l'éducation qui nourrit cette économie?

### 1.2.2 Un nouveau rapport capital/travail

Pour citer Carlo Vercellone, un des défenseurs du capitalisme cognitif : *"La montée en puissance de la dimension cognitive du travail correspond à l'affirmation d'une nouvelle hégémonie des connaissances mobilisées par le travail, par rapport aux savoirs incorporés dans le capital fixe et l'organisation managériale des firmes. Plus encore, c'est le travail vivant qui joue désormais un grand nombre des fonctions principales jouées jadis par le capital fixe."* La stratégie n'est plus aujourd'hui d'avoir "les plus grands locaux avec le plus de machines" mais d'avoir des collaborateurs avec la tête bien faites qui seront à même de concevoir les désirs des clients de demain. Ce nouveau rapport capital/travail induit deux conséquences fondamentales. D'une part, à l'échelle de l'entreprise, l'activité créatrice de valeur (la valeur ajoutée) coïncide de moins en moins avec l'unité de lieu et de temps propre aux réglages des temps collectifs de la période fordiste. D'autre part, à l'échelle sociale, la production de richesses et de connaissances s'opère de plus en plus en amont du systèmes des entreprises et des sphères marchandes.

La question de création de valeur économique aux données personnelles semble être basée donc sur cette nouvelle forme de capitalisme. La question n'étant plus forcément d'avoir le plus de données possibles mais plutôt de pouvoir leur faire prendre un sens. Pouvoir par la suite les étudier et en tirer un profit marchand. La valorisation des données personnelles ne se limite plus au stockage et à la gestion de fichiers clients, elle n'a aujourd'hui de sens que dans la circulation et l'interaction des données avec tout un écosystème qui peut être interne et/ou externe à l'entreprise

## 2. Création de valeur économique aux données personnelles.

La donnée se retrouve être aujourd'hui au cœur des priorités stratégiques des entreprises. La donnée est partout, ses sources sont multiples, provenant à la fois des entreprises, des individus ou même des puissances publiques. La donnée a la particularité de pouvoir circuler librement, se reproduire, se stocker, s'agréger ou même être corrélé. Elle devient la matière première de nombreux métiers et la raison d'être de nouveaux marchés, témoignant d'une tendance généralisée qui est la *data driven economy* (l'économie guidée par les données). Les mutations économiques, stratégiques, politiques et sociales n'en sont qu'à leur début, le déluge des données s'accroissant de manière exponentielle avec le développement récent de nouvelles technologies (objets connectés, *machine learning*, intelligence artificielle...).

### 2.1 Qu'est ce qu'une donnée personnelle ?

#### 2.1.1 Une définition complexe ?

Avant toute étude de création de valeur économique des données personnelles il est justifié de définir ce qu'est une donnée personnelle.

La définition de la loi Informatique et Libertés couvre un périmètre très large car comme le précise son article 2: « *Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne* ».

Cette notion de donnée personnelle est donc très vaste puisqu'elle englobe tout ce qui peut renvoyer même indirectement à une personne. Mais dans ce cas, les données personnelles seraient en fait partout? Est-ce que toutes les informations qui peuvent renvoyer (même indirectement) à une personne le sont au même niveau? C'est-à-dire, devons nous traiter au même niveau des informations provenant d'une adresse IP ou des messages postés sur notre mur Facebook? On constate donc une vraie complexification de la nature des données personnelles.



## 2.1.2 D'où vient cette complexification ?

### 2.1.2.1 Les données sont de plusieurs natures

Avant tout raccourcis sur la nature d'une donnée personnelle il convient de distinguer deux catégories distinctes de données :

les données personnelles identifiantes : Ce sont des données qui sont directement rattachées à l'identité d'une personne. Ces données sont toutes les données rattachées à un nom (compte Facebook, Twitter), une adresse email ou toutes les données qui permettent d'identifier indirectement une personne comme une adresse IP ou encore un identifiant.

les données comportementales : ces données ne sont pas rattachées directement à la personne mais plutôt au comportement de ce dernier. Ces données sont collectées via le suivi de ses navigations (où l'utilisateur passe le plus de temps, où il clique, ce qu'il regarde), mettant en exergue ses comportements d'achats ce qui permet d'affiner le profil d'une personne, surtout si ces données comportementales sont croisées avec des données personnelles identifiantes. En cernant mieux les centres d'intérêt d'un individu et ses préférences, il est possible d'optimiser son parcours d'achat et de lui proposer des offres personnalisées.

### 2.1.2.2 Les données sont relatives

Toute donnée qui soit est nécessairement liée au référentiel culturel des individus et à la réglementation en vigueur dans chaque pays. Le regard à porter sur telle ou telle donnée est différent en fonction d'une de sa provenance (pas seulement la source d'origine de la donnée, mais aussi dans quel contexte celle-ci a été générée). Nous pouvons questionner trois perceptions relatives de la donnée personnelle qui impliquent trois approches différentes de la problématique :

**Point de vue entreprise :** La donnée a une valeur stratégique et marchande. Elle permet de développer ou d'améliorer les modèles serviciels, d'enrichir l'expérience client, d'être analysée à des fins décisionnelles, d'opérationnalisation, de revente etc.

**Point de vue client :** ses données personnelles sont une émanation de sa personnalité et de sa vie privée, une « trace » de ses comportements. C'est évidemment un point sensible. La confiance qu'il accorde dans les services numériques est fragile si la transparence, les finalités de traitement et les conditions de consentement ne sont pas claires.

Du point de vue juridique : le statut juridique de la donnée personnelle permet-il de faire face aux tensions qui existent entre la protection des personnes et les besoins

du marché d'utiliser leurs données pour développer de nouveaux services ? Un Code de la Donnée est-il souhaitable pour prendre en compte les enjeux de la société du XXI<sup>ème</sup> siècle, tout comme l'on a créé un Code de la Propriété Intellectuelle au XX<sup>ème</sup> siècle ?

Le point de vue que nous étudierons en profondeur est celui de l'entreprise puisqu'il s'agit d'étudier les mécanismes permettant de créer de la valeur à des données personnelles.

### 2.1.2.3 Les données sont ouvertes

Aujourd'hui, les données sont ouvertes, elles circulent et se partagent en permanence. La notion même de donnée personnelle ne correspond plus à la réalité. Parler de données personnelles n'a plus de sens, il faudrait plutôt parler de données relationnelles, transactionnelles puisque les données circulent en permanence et qu'il est possible de déduire des informations personnelles à partir de données non personnelles.

Il faut comprendre ici nous ne sommes plus forcément définis par juste des attributs mais plutôt par les interactions que nous avons avec le monde. Ces interactions sont de plus en plus nombreuses et sont facilitées par le contexte de Big Data.

Une des grands challenge auquel nous devons faire face est de ne plus uniquement sécuriser les données personnelles en temps qu'entité propre mais d'identifier les traitements qui sont réalisés ces dernières.

## 2.2 Valorisation des données personnelles

Dans le contexte de Big Data, les données personnelles et leur valorisation prennent une importance croissante. D'après les prévisions d'IBM, d'ici 5 à 10 ans la plus grosse partie des données viendra non pas des usages d'internet mais directement de l'usage des objets (Internet des objets), qui encadrent notre quotidien. Ces tendances économiques et technologiques témoignent d'un avenir prometteur pour l'économie des données personnelles, mais celle-ci ne se fera pas sans le concours des consommateurs/clients qui prennent une part de plus en plus importante dans les processus de valorisation.

### 2.2.1 Les potentiels de la valorisation des données personnelles

Les grandes entreprises ont pris conscience du potentiel de création de valeur que leur apportait l'usage des données ciblées appliquées à la masse croissante de données qu'elles peuvent mobiliser sur leurs clients. Mieux connaître ses clients, comprendre leurs comportements et leurs attentes, anticiper leurs réactions permet de les fidéliser et de leur proposer les offres personnalisées les mieux adaptées.

C'est un enjeu majeur pour toutes les entreprises, enjeu souvent inscrit au premier rang dans leur stratégie.

Mais quels sont les potentiels de création de valeur de données personnelles?

#### 2.2.1.2 Un potentiel stratégique

Le modèle de valorisation des données personnelles ouvrent de plus en plus la voix à des alliances stratégiques et de co-crédation entre entreprise. Ce modèle est un modèle qui sert de passerelles, les secteurs de l'automobile et de l'assurance sont par exemple amenés à travailler ensemble sur la voiture connectée. Dans un contexte de mondialisation on constate que les données personnelles possèdent un potentiel d'ouverture au monde qui va au delà de l'information intrinsèque qu'elles portent.

#### 2.2.1.3 Un potentiel opérationnel

Le potentiel stratégique n'est pas l'unique potentiel d'une valorisation des données personnelles. On peut aussi constater un potentiel opérationnel. Afin de répondre à certaines contraintes de vie privée il existe un mécanisme très utilisé lors de la valorisation des données qui est celui de l'anonymisation des données. La donnée personnelle est traitée de telle manière qu'elle ne permet plus d'identifier directement ou indirectement une personne. La donnée devient une donnée comportementale. Une donnée comportementale pose moins de problème de sécurité ou d'éthique puisqu'elle représente des individus mais sans distinctions entre chacun. L'exploitation de ces données permet d'optimiser des processus opérationnels. C'est comme ça que fonctionne les services de transport qui mesurent les flux de passagers pour adapter leurs services en fonction. La connaissance des axes les plus fréquentés ou les plus désertés à certaines tranches horaires, permet d'adapter, parfois en temps réel, les services de surveillance, ouverture de guichets, etc..

#### 2.2.1.4 Un potentiel de recherche et développement

Il est courant que l'entreprise crée de la valeur sur les données via l'expérimentation, et donc selon un usage non défini à l'avance. Or, un des principes fondamentaux de la protection des données personnelles est d'informer la personne concernée de la finalité du traitement au moment de la collecte de ses données afin d'en obtenir un consentement éclairé. Cette finalité n'est pas toujours simple à définir à l'avance dans le cadre de l'innovation ou de la R&D. Le recours à une définition de cette finalité de manière relativement générique permet de se laisser une marge de manœuvre suffisante pour valoriser au mieux les expérimentations. De plus, tant que le contexte n'est pas lié au business, les données personnelles peuvent faire l'objet d'analyses réalisées en silos, et permettre par exemple d'améliorer la qualité d'autres

données, d'affiner un modèle algorithmique, et ainsi de perfectionner les savoir-faire internes des entreprises.

### 2.2.1.5 Un potentiel marketing

La personnalisation des services et des produits à des fins marketing est un pilier de l'économie des données personnelles. De nombreuses plateformes Web sont fondées exclusivement sur la vente de *retail*. L'analyse comportementale peut porter sur des données non-structurées et permettre de développer des modèles prédictifs pour être proactif dans la relation et la connaissance client. De nombreux modèles existent, mais nous pouvons citer en particulier les résultats rendus publics des chercheurs iraniens des universités Azad et Alzahra à Téhéran : ils ont mis au point un algorithme de modélisation (par apprentissage semi-automatisé) des comportements clients par l'intégration de données non balisées à un classificateur de données balisées. Ce modèle a pour vocation à affiner progressivement les prédictions sur le comportement des clients en vue d'un usage marketing.

Comme le montre le schéma ci-après, l'algorithme repose sur un réseau neuronal utilisant des données déjà annotées. Les données non balisées sont alors traitées et classifiées par rapport au classificateur initial de données balisées. Un nouveau classificateur est ainsi créé, regroupant les données balisées et celles nouvellement balisées. Ce système permet d'acquérir un meilleur degré de confiance et de précision pour l'évaluation des modèles, notamment des modèles prédictifs.

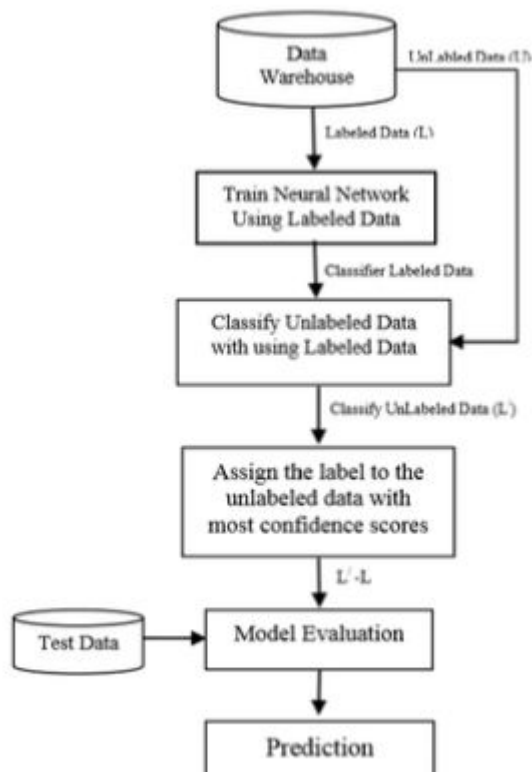


Figure 1. An overview of the Algorithm

Ce modèle nécessite un modèle de base mais sera affiné au fur et à mesure que des données utilisateurs seront ajoutées. Un usage marketing des données personnelles ne se fera cependant sans le concours de consentements des individus. Selon une enquête menée en février 2014 par serdaLAB auprès de 533 individus de 15 à 65 ans, 90% des personnes acceptent les pratiques commerciales à partir de leurs données personnelles, à condition d'exprimer au préalable leur consentement et de pouvoir maîtriser les conditions de cette réutilisation des données.

Finalement toutes les données ne sont pas soumises au même jugement de confidentialité de la part des utilisateurs. Les

données de consommation sont jugées par les trois quarts des répondants comme étant les moins sensibles pour une réutilisation commerciale, tandis que les données de géolocalisation, de navigation internet, des réseaux sociaux, de santé et les données bancaires sont considérées comme étant les plus confidentielles.

Ces potentiels sont donc nombreux et on comprend aujourd'hui que l'économie des données personnelle représente un enjeu majeur. Mais ces données n'ont pas uniquement des potentiels. Elles ont carrément un impact sur les modèles d'affaires.

## 2.3 Les modèles d'affaires autour des données personnelles

Les données ont un impact sur les modèles d'affaires, qui peuvent ou doivent dans certains cas se redéfinir ou se réinterroger par rapport à ce nouveau contexte et/ou environnement. En effet l'exploitation des données personnelles a permis à certains *business models* de se réinventer, à d'autre de voir le jour, impulsant ainsi une nouvelle dynamique dans l'économie numérique.

Deux modèles économiques, en totale expansion, témoigne d'ailleurs de ces intérêts grandissants :

### 2.3.1 Deux modèles économiques en expansion

Le modèle biface : C'est le cas de Google et Facebook qui sont des entreprises bifaces. Ces entreprises font l'intermédiation entre deux types d'acteurs, les utilisateurs et les annonceurs. Ces deux acteurs forment les deux faces, l'une des faces représente une valeur pour l'autre face. Ce modèle économique n'est bien entendu pas propre au numérique, puisqu'il existait déjà dans l'économie des médias (télévision ou radio). Finalement les géants courtiers collectes des milliards de données, analysent et revendent leurs résultats à des sociétés visant essentiellement des fins marketing.

Les modèles serviciels : les données participent grandement au mouvement de *servicialisation*, c'est-à-dire à l'avènement de la vente non plus de bien mais de service. L'exemple des assurances est instructif : plutôt que de spéculer sur l'accidentologie pour établir leurs tarifs, les sociétés peuvent se baser désormais sur l'analyse des usages réels des dispositifs, grâce aux capteurs et objets connectés qui transmettent les données de conduite, de kilométrage etc. Le temps réel est une vraie révolution dans la conduite des *business models* autour des données. Les modèles serviciels permettent également d'enrichir l'expérience utilisateur : il y a un troc implicite entre les données et le service. Les données ne servent pas uniquement à l'entreprise qui les collecte, elles bénéficient également indirectement aux utilisateurs par un meilleur service, c'est-à-dire plus personnalisé, plus efficace et

pertinent. Nous reviendrons sur des exemples précis de retours de valeurs pour les clients par la suite.

### 2.3.2 Peut-on se protéger ou concurrencer de tels modèles?

Mais comment concurrencer ces modèles serviciels des géants américains? Les services proposés deviennent plus performant et plus personnels, proposent aujourd'hui des services qui sont plus traditionnels. Comment éviter une domination totale? Simplement, en misant sur cette confiance des services qui fait défaut. Il faut arriver à développer des services à forte valeur ajoutée en intégrant cette confiance qui fait aujourd'hui défaut, réussir à créer la même valeur de service, mais de façon différente.

Ces données qui sont donc de plus en plus personnelles, les usages de plus en plus précis et des rendements toujours plus importants. Comment protéger ces utilisations? Comment au moins les limiter? Et un bien un modèle existant et permettant une utilisation plus admise par les utilisateurs est l'anonymisation des données. Mais à ce moment là, quand on sait que la valeur est souvent créée à partir d'une personnalisation des services, comment faire?

## 2.4 L'anonymisation et la valorisation des données est-elle compatible?

Comme exprimé précédemment par le potentiel opérationnels de la valorisation des données, comment valoriser une donnée anonyme?

Il existe pourtant des méthodes qui permettent de rendre les données cohérentes, statistiquement significatives, sans être nécessairement nominatives : c'est ce qui est attendu par les entreprises et cadré par la législation.

### 2.4.1 Comment réaliser une anonymisation ?

Mais comment réaliser une bonne anonymisation?

Pour aider à évaluer une bonne solution d'anonymisation, le G29 définit trois critères qui déterminent les risques de réidentification :

- L'individualisation : est-il toujours possible d'isoler un individu ?
- La corrélation : est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
- L'inférence : peut-on déduire de l'information sur un individu ?

En France, la CNIL a travaillé par exemple avec de nombreux acteurs autour du Pass Navigo, dans des objectifs de conformité et d'innovation. L'exemple du Pass Navigo

permet d'illustrer la démarche de la CNIL : dès 2009, les Pass Navigo étaient tous dotés d'un numéro unique, personnel et infalsifiable, permettant d'effectuer des statistiques fines sur la fréquentation du réseau. La puce du Navigo contenait des informations sur le « qui », le « quand » et le « où ». La CNIL et la STIF (Société de Transport d'Ile de France) ont donc décidé ensemble d'anonymiser ces données, sans que cela n'entrave la finalité poursuivie qui était l'optimisation du réseau. La STIF pouvait donc travailler sur des données qui n'étaient plus soumises à la loi « Informatique et Libertés » car anonymes.

Il doit être cependant souligné que l'anonymisation sera chaque jour plus complexe, notamment avec l'émergence du temps réel. Les progrès techniques et la multiplicité des sources des données fragilisent en effet la durabilité de l'anonymisation des données et mettent en cause les investissements qui seraient faits dans ce sens. Il faut notamment prendre garde à la notion de « jeu de données anonymes » car il est toujours possible de le « désanonymiser » dans un contexte *Big Data*. Un jeu de données anonymes peut donc être de nouveau soumis à la loi Informatique et Libertés. L'anonymisation est quelque chose qui doit être sans cesse révisé dans le temps.

## 2.5 Un Big Data sans données personnelles non anonymes?

Cependant une telle anonymisation, qui rend impossible la personnalisation des services aux clients et limite leurs valorisations à de l'opérationnalité des services, présentes donc des inconvénients concrets pour les entreprises, comme pour les utilisateurs.

Un moyen utilisé par les entreprises pour passer outre ces limites et donc de naviguer avec l'accord des utilisateurs. Ces entreprises doivent obtenir l'accord des utilisateurs pour utiliser leurs données de manière personnelle. De nombreux exemples sont notifiables dans la vie courante où une application nous demande la permission d'envoyer les informations d'utilisations, ou de nous géolocaliser. Grâce à cet accord les entreprises peuvent déployer et analyser nos données afin de les valoriser au maximum. C'est du Big Data "privé".

Pour que ce système de Big Data avec privacy soit viable il faut que l'utilisateur est un dernier levier dispositif qui est celui permettant de revenir sur les accords précédents définis de privatisation. Garder une maîtrise sur sa personne, sur ce que l'on représente pour les autres.

En outre, comme le souligne Ann Cavoukian dans le Livre Blanc « *Using Privacy By Design to Achieve Big Data Innovation* » : « Savez-vous pourquoi on ne peut pas avoir des données de qualité avec le Big Data sans privacy ? Parce que le contexte est la clé du Big Data ». Les données personnelles font partie de ce contexte qui va permettre de valoriser tout un écosystème de données. Les données personnelles sont bien un facteur de réussite du Big Data. Voilà pourquoi, le consentement des personnes et la mise en oeuvre de dispositifs leur permettant d'exercer leurs droits seront certainement toujours plus rentables.

### 3. Constater la création de valeur, pour qui a telle une plus-value? Captation d'une plus-value par le client?

#### 3.1 Principe de la création d'une plus-value

Le but de la collecte de données est de créer de la valeur qui aura une plus-value pour l'entreprise et pour le client. On utilise les données afin de trouver des résultats qui vont apporter des connaissances à partir desquelles on va créer de la valeur. La création d'une plus-value se fait par l'identification de tendances. On a en entrée des paramètres caractérisant des personnes et en sortie des corrélations entre ces paramètres. On analyse des critères concernant des personnes et leur environnement, on les manipule à l'aide d'algorithmes et de méthodes statistiques afin de connaître les influences des paramètres les uns sur les autres. Une fois la connaissance acquise on va pouvoir s'appuyer dessus afin d'orienter des choix, en outre des stratégies marketing, et ainsi créer de la valeur.

Voici un exemple très caricatural de ce principe :

Données personnelles :

- Couleur de t-shirt souvent portée
- Classe de voiture possédée
- Sexe
- Année de naissance

Après analyse on pourrait trouver comme résultat que les hommes portant des t-shirts bleus et nés dans les années 70 ont tendance à acheter des berlines, tandis que les femmes nées dans les années 80 et portant des t-shirts rouges se tournent plutôt vers les citadines.

Une marque de voitures pourrait alors s'appuyer sur ces connaissances afin de proposer aux personnes dont elle connaît ces caractéristiques des offres sur des modèles de voiture adaptés.

Cet exemple est évidemment caricatural mais présente bien le principe de la création de valeur. Dans la pratique on s'appuie souvent sur un nombre plus important de critères dont les valeurs peuvent se situer dans de grands intervalles. On travaille les données de façon à tester une multitude de combinaisons et on joue sur le nombre de critères utilisés pour perfectionner la vraisemblance des tendances. Connaissant



les tendances l'objectif sera à partir d'un nouvel individu de pouvoir déterminer en quelque sorte à quelle catégorie il appartient pour ensuite en connaître plus sur lui.

## 3.2 Fonctionnement de l'analyse de données

### 3.2.1 Format des données

Toute analyse de données, ou *Data mining*, s'appuie sur les mêmes structures de données : des tableaux individus-variables. On a donc un certain nombre d'individus représentés par un certain nombre de caractéristiques, les variables. Les individus ne représentent pas nécessairement des êtres humains, ils peuvent être de toute sorte. Les caractéristiques des individus peuvent être des variables aussi bien quantitatives que qualitatives. De plus les individus sont généralement associés à une classe dans laquelle les individus ont des caractéristiques proches. Un exemple de données que l'on pourrait avoir serait des informations caractérisant des êtres vivants, par exemple des crabes. A chaque crabe on pourrait associer des caractéristiques quantitatives telles que la taille de différents membres, qualitatives telle que la couleur et une classe qui serait l'espèce biologique.

Si les méthodes d'analyse de données s'appliquent quelle que soit le nombre d'individus, on obtient de meilleurs résultats d'analyse en s'appuyant sur de grands jeux de données, d'où le fameux terme « big data ».

### 3.2.2 Traitements des données

Une fois que l'on a un tableau individus variables on peut appliquer de nombreuses méthodes statistiques. Le principal objectif de ces méthodes est de construire un modèle qui permettra à partir d'un nouvel individu de deviner à quelle classe il appartient. En reprenant l'exemple précédent, l'idée serait qu'en étudiant un crabe dont on ne connaît pas l'espèce, on puisse déduire à partir de ses caractéristiques l'espèce à laquelle il a le plus de probabilité d'appartenir.

#### 3.2.2.1 Quelques méthodes d'analyse de données

Une première méthode est l'analyse en composantes principales, elle permet de bien comprendre son jeu de données en étudiant les caractéristiques quantitatives. En appliquant des méthodes mathématiques sur le tableau individus-variables (matrices, leur produits, leurs valeurs propres, leurs vecteurs propres...) on va pouvoir construire un nouveau tableau individus variables, où les variables seront artificielles et représenteront en quelque sorte une synthèse des données que l'on avait précédemment. La plupart de l'information se trouve dans la première variable artificielle, une part moins importante dans la seconde et ainsi de

suite, de façon à ce qu'avec seulement les 2 ou 3 premières variables on puisse caractériser un individu presque aussi bien qu'avec les variables réelles. On va ainsi pouvoir représenter les individus dans un plan où l'axe des abscisses sera la première variable artificielle et l'axe des ordonnées la seconde ou la troisième. On pourra ainsi visuellement voir quels individus sont proches les uns des autres et repérer les classes de façon plus compréhensible puisqu'on n'a besoin que de deux variables pour comprendre.

Avec les méthodes mathématiques on va également pouvoir étudier l'influence des variables réelles sur les valeurs des variables artificielles, et représenter ces influences dans un plan où chaque variable réelle est représentée par un vecteur. Voici un exemple de cette dernière représentation :

Les variables corrélées du jeu de données seront alors représentées par des vecteurs ayant une orientation proche. Cette représentation permet ainsi d'avoir une idée de la corrélation entre les facteurs.

L'analyse en composantes principales et d'autres méthodes permettent d'étudier un jeu de données et de bien le comprendre. On a ensuite des méthodes qui vont permettre de créer des modèles de décision, permettant ainsi de deviner la classe d'un nouvel individu à partir de ses caractéristiques. Ces méthodes sont appelées des classifieurs et se basent sur l'apprentissage de données d'individus dont on connaît la classe.

Un classifieur simple est le classifieur euclidien. Il s'appuie sur la distance euclidienne entre les individus, qui permet de dire numériquement si deux individus sont proches ou éloignés en fonction de toutes les caractéristiques. Dans un cas à 2 variables où l'on représenterait les individus dans un plan cela revient à donner la distance entre 2 points.

Le principe du classifieur euclidien est de calculer les centres de gravité de chaque classe d'individus. Considérant ensuite un nouvel individu on va calculer les distances euclidiennes à chacun des centres de gravité, et décider ainsi que la classe de l'individu est celle correspondant au centre de gravité qui lui est le plus proche. Cette méthode se révèle plus ou moins efficace en fonction de la différence qu'il y a entre les classes et donc la distance qu'il y a entre les centres de gravité. Si deux centres sont proches on a toute une zone entre les deux où il est difficile de dire à laquelle des deux espèces appartiennent les individus.

Un autre classifieur très célèbre est celui des « K plus proches voisins » ou « K-means », qui s'appuie sur la distance entre les individus en considérant toutes les variables. Considérons un nouvel individu dont on ne connaît pas la classe et introduisons-le parmi d'autres individus dont on connaît l'espèce. On va regarder ses

$k$  (nombre entier fixé) voisins les plus proches en considérant les caractéristiques de manière générale, et regarder à quelle espèce appartiennent ces voisins. La classe devinée pour l'individu introduit sera celle qui est la plus représentée parmi les voisins les plus proches. On pourra faire varier le nombre de voisins  $k$  afin d'obtenir un classifieur le plus performant possible. L'évaluation de la performance d'un classifieur sera détaillée plus loin.

Enfin une troisième méthode d'apprentissage des données est celle des arbres de décision. Pour simplifier les explications on va considérer un ensemble de données où il n'existe que deux classes. Pour construire l'arbre on va à chaque fois séparer un ensemble de données en 2 suivant des critères. Considérant par exemple une variable quantitative on va pouvoir dire que les individus ayant pour cette variable une valeur supérieure à un certain nombre appartiennent à une classe et les autres individus à la seconde classe. Visuellement, sur un plan où seraient représentés des individus à l'aide de 2 variables on aurait une droite qui séparerait le jeu de données en 2, avec d'un côté les individus d'une espèce et de l'autre ceux de la seconde espèce. Le découpage ne pouvant généralement pas être parfait, on a des individus qui se trouvent du mauvais côté de la frontière d'après leur classe. On va donc trouver un nouveau critère de décision afin de pouvoir, dans le sous-espace où se trouvent des individus « intrus », effectuer une nouvelle séparation entre les espèces. A l'aide d'un second critère on pourra dire si un individu du sous-espace appartient bien à la classe dominante ou à l'autre.

On va ainsi trouver des séparations jusqu'à ce qu'on puisse de manière certaine déterminer la classe d'un individu du jeu de données. On va ensuite répertorier les différentes séparations dans un arbre de décision. A la racine on aura 2 branches correspondant aux deux sous espaces issus de la première séparation. A l'extrémité des branches on aura soit des feuilles, signifiant qu'il n'y a plus de séparations et que l'on est donc sûr de l'espèce, soit un nœud à partir duquel vont partir deux autres branches correspondant à une nouvelle séparation. Une fois l'arbre de décision terminé on va peut vérifier la classe d'un individu. On le place au niveau de la racine et on applique le premier critère de séparation. En fonction de la valeur de la variable considérée pour la séparation on va suivre l'une ou l'autre branche. Si le nœud à l'autre extrémité de la branche est une feuille alors on connaît la classe de l'individu. Si deux nouvelles branches partent du nœud alors on applique le critère de décision correspondant. On continue ainsi de suite jusqu'à arriver à une feuille, c'est-à-dire jusqu'à avoir déterminé de manière certaine la classe de l'individu.

Un critère de décision peut s'appliquer sur des variables quantitatives aussi bien que qualitatives. Pour trouver le critère de décision on va tester toutes les possibilités et choisir celle qui sépare le mieux les individus. Une fois l'arbre de décision construit on peut l'utiliser pour deviner la classe de nouveaux individus.

### 3.2.2.2 L'évaluation des performances des classifieurs

Une partie essentielle de la construction d'un modèle de prédiction à partir de l'analyse de données est l'évaluation des performances : on veut savoir quelle est la probabilité qu'il a de se tromper lorsqu'il devine une classe pour un individu. Pour tester l'efficacité d'un modèle on l'utilise sur des individus dont on connaît la classe et on regarde le taux de succès de prédiction. Il est essentiel de ne pas utiliser les mêmes données pour les tests que celles qui ont servi à élaborer le modèle, sinon il est évident que le classifieur sera performant. Avant la construction on sépare donc les données dont on dispose en un ensemble d'apprentissage et un ensemble de test. Pour élaborer un bon modèle on va tester différentes méthodes de différentes manières et à chaque fois évaluer ses performances. A force d'essayer et de jouer sur les méthodes on finit par trouver un classifieur performant.

La liste des méthodes énoncée n'est pas exhaustive mais la plupart se basent sur les mêmes principes. Pour élaborer un modèle de prédiction performant on combine généralement différentes méthodes.

### 3.2.2.3 Origine des données

Pour construire un modèle on peut se baser sur des individus de différentes sortes qui ne sont pas nécessairement caractérisés par les mêmes variables. On peut alors utiliser différents modèles de prédiction afin d'en apprendre plus sur l'influence qu'ont certains paramètres les uns sur les autres. Dans le cadre d'un projet d'analyse de données en particulier, on peut très bien se servir en partie de données ou de résultats sans lien direct avec le sujet.

Si les données caractérisent souvent des individus et des objets elles peuvent également être des actions d'un utilisateur sur internet, comme un clic. Grâce à cela on a de plus en plus de temps réel dans l'exploitation des données, qui permet aux sites web d'afficher certains contenus en fonction des actions faites par un utilisateur, croisées avec ses caractéristiques.

## 3.2.3 Expliquer les tendances observées

L'analyse de données permet donc de déterminer des tendances qui permettent de deviner des faits, d'anticiper des événements... Cependant elle ne permet pas de déterminer la causalité des tendances observées. Les méthodes de *data mining* traitent les informations de manière mathématique sans s'intéresser à ce qu'elles représentent. Si on veut expliquer les résultats on ne peut se baser que

sur nos réflexions et il est souvent très difficile voire impossible de vérifier nos hypothèses. Parfois on ne peut pas donner d'explication rationnelle à une tendance observée, dans certains cas les résultats peuvent même être contre-intuitifs. Cela s'explique par le fait qu'il n'y a pas forcément de cause, ce sont simplement des statistiques qui nous permettent de dire par exemple qu'un individu va agir de telle manière avec une certaine probabilité, suivant ses caractéristiques. C'est d'ailleurs ce qui fait la force des méthodes d'analyse de données, c'est qu'elles ne cherchent pas à expliquer des phénomènes et fonctionnent de manière arbitraire là où des êtres humains se bornent à chercher des causes et à trouver des résultats qui font sens pour eux.

Si l'explication d'une tendance peut s'avérer intéressante dans certains domaines comme la biologie, parfois cela n'a pas vraiment d'intérêt. En marketing par exemple, les résultats obtenus avec l'analyse de données suffisent à établir des stratégies efficaces et connaître la cause de ces résultats n'a pas d'intérêt pour une entreprise.

### 3.3 Plus-value pour l'entreprise

#### 3.3.1 Caractérisation de la plus-value

Grâce aux tendances qu'une entreprise aura pu identifier, celle-ci va pouvoir créer de la valeur en effectuant des choix et des actions adaptés. Les stratégies qui suivent auront toutes pour but final de gagner de l'argent, que ce soit en évitant des coûts ou en faisant plus de profits.

Un exemple bien concret est le cross-selling : cela consiste sur un site web marchand à associer des produits à d'autres afin de favoriser leur achat commun. Le choix est personnalisé en fonction des caractéristiques connues de l'utilisateur, du produit et d'autres critères tels que la météo, le jour de l'achat. On peut noter par exemple que sur Amazon, 35% des produits vendus font suite à une recommandation faite à partir de l'analyse de données.

Une autre application très utilisée est la publicité. Grâce à l'analyse de données les entreprises arrivent à mieux identifier les profils des individus et peuvent ainsi leur proposer des publicités ciblées et personnalisées. Facebook entre autres utilise l'analyse de données pour savoir quelles publicités afficher en fonction du profil de l'utilisateur. Le choix de la publicité montrée est fait de manière à maximiser les chances que l'utilisateur clique dessus.

##### 3.3.1.1 Mesure de la satisfaction des clients

Une autre plus-value qui peut être apportée à partir de données est la mesure de la satisfaction des clients. Beaucoup d'entreprises, notamment les opérateurs

téléphoniques analysent des données de leurs clients afin de savoir lesquels ont le plus de chance de changer de fournisseur. Elles associent également une valeur pour chaque client indiquant combien cela a coûté et combien il a rapporté. Un opérateur téléphonique par exemple peut ainsi savoir quels clients ont le plus de chances de se désabonner et ainsi proposer des offres ou des avantages s'ils valent la peine d'être gardés.

On a à faire à quelque chose de nouveau, on arrive à repérer des sentiments d'individus sans même interagir avec eux. Les entreprises ne sont plus obligées de dialoguer avec leurs clients pour détecter leur satisfaction. Certains sentiments sont détectés par ordinateur ce qui est fondamentalement nouveau.

### 3.3.1.2 Le côté abstrait de la création de valeur

Le processus de création de valeur à partir de l'analyse de données s'appuie sur des concepts abstraits. Auparavant une entreprise créait de la valeur uniquement à partir de produits, de processus, de services, sur des choses concrètes. Maintenant elle peut s'appuyer sur des données qui prises sans analyse n'apportent rien. Ce sont de simples informations qui n'ont pas forcément de liens entre elles. C'est seulement en utilisant des algorithmes et en les traitant que l'on va réussir à créer de la valeur. Sans la puissance de calcul des ordinateurs l'analyse des données n'aurait apporté que très peu d'information.

### 3.3.2 Pouvoir de prédiction

On peut donc constater que l'analyse des données a un pouvoir de prédiction. On peut en manipulant de simples données qui n'ont pas d'intérêt toutes seules deviner les futures actions ou états d'individus. C'est là une nouvelle dimension de ce que l'on peut obtenir à l'aide de la technique. Jamais auparavant on n'avait réussi à prédire de façon aussi précise et certaine les futurs comportements d'individus. On arrivait à prédire plus ou moins la météo ou des probabilités d'occurrence de certains événements, alors qu'aujourd'hui à l'aide de cette nouvelle forme de technique on prédit des événements ou on devine des faits de façon relativement précise dans de nombreux domaines.

Les prédictions faites à partir de données utilisateur peuvent s'appliquer dans énormément de domaines, pas nécessairement sur des individus en particulier. Voici quelques exemples de prédictions effectuées à l'aide de l'analyse de données :

- Airbnb calcule pour un logement et une période donnés le prix optimal à fixer pour une location. Les statistiques montrent que les utilisateurs qui

appliquent ce prix à 5% près ont 4 fois plus de chance de trouver des locataires.

- UK supermarket arrive à prédire pour 19% de ses clients quand ils vont revenir et combien ils vont dépenser à 10\$ près.
- Une autre enseigne de supermarché arrive à partir de comportements d'achat à dire si une personne est enceinte
- Les universités de Brigham et de l'Utah prédisent avec un taux de réussite de 80% les naissances prématurées en se basant sur des biomarqueurs et ceci dès la 24<sup>ème</sup> semaine de grossesse.

Les entreprises utilisent les données de leurs utilisateurs, de leurs comportements de leur environnement, éventuellement des données de sources extérieures qui peuvent les aider à identifier des tendances et les manipulent de manière à trouver des résultats qui vont leur permettre de prédire certaines informations ou comportements.

On est donc avec ces pouvoirs de prédiction dans une nouvelle ère. On peut prédire des faits sur des individus à partir de caractéristiques qui si on les avait considérées seules n'auraient rien pu nous apprendre. C'est en croisant ces données avec celles de nouveaux individus que l'on y arrive.

### 3.3.2.1 Conséquences de ces plus-values sur les sciences sociales

Il est important de constater que le traitement des données a un pouvoir de prédiction. On arrive à prédire des comportements et des actions d'individus, parfois alors que ceux-ci n'étaient même pas encore conscients qu'ils allaient agir ensuite. Il convient d'étudier l'impact que cela peut avoir sur les sciences sociales.

Si les sciences sociales avaient presque toujours su s'approprier le numérique depuis sa création, le big data pourrait être en voie d'écraser les sciences sociales par leur pouvoir de prédiction. On ne s'intéresse plus à la cause du comportement de l'internaute mais à connaître des tendances.

## 3.4 Plus-value pour le client

### 3.4.1 Les bénéfices pour les clients

Si la valeur apportée par l'analyse de données apporte toujours une plus-value à l'entreprise qui la pratique, elle peut aussi en apporter une à ses clients.

Dans le cas des opérateurs téléphoniques que l'on a vu ci-dessus on voit qu'un client insatisfait peut ainsi recevoir des offres avantageuses.

Dans le domaine des assurances, les entreprises peuvent personnaliser les tarifs. Les prix des assurances étant fixés à partir de taux de risques et de coûts moyens, un assuré ne paye pas par rapport à son propre risque. Avec la valeur que l'on dégage de données les assureurs peuvent personnaliser le taux de risque pour leurs clients et leur faire payer le prix correspondant à leur risque.

D'une manière générale la principale plus-value dégagée des données pour l'utilisateur est la personnalisation des services, des offres et des produits. En travaillant les données recueillies sur internet les entreprises peuvent présenter aux internautes ce qui les correspond le mieux. C'est là encore un aspect nouveau : on arrive à mieux satisfaire des individus grâce à une personnalisation effectuée sans aucune interaction directe ou indirecte avec eux et des connaissances limitées sur leur profil. Si on voulait en connaître plus sur les besoins et les désirs d'un individu il fallait rentrer en contact avec lui, même de si c'était de manière indirecte.

Une autre plus-value est l'amélioration de la relation client entre l'utilisateur et l'entreprise qui lui propose des produits et services. L'entreprise connaît mieux ses clients ou ses usagers et peut leur accorder des privilèges, des recommandations personnalisées etc.

L'analyse des données peut aussi apporter à un internaute des informations qui peuvent lui être utiles comme des réductions sur sa marque préférée, des offres avantageuses sur ce qu'il recherche, ou des indications sur le moment opportun d'acheter un produit, comme un billet d'avion.

### 3.4.2 Conséquences de la personnalisation des espaces numériques

Les entreprises semblent apporter aux internautes une personnalisation à l'aide du traitement de données, cependant il convient d'étudier plus en profondeur cet aspect.

Ce qui intéresse les entreprises c'est de détecter automatiquement nos potentialités, ce que nous pouvons désirer et comment nous pourrions agir, parfois alors que l'on en est même pas conscients. Or ces éléments ne forment pas une personne.

On pourrait penser que la personnalisation des environnements numériques et des offres commerciales menacent une disparition de la vie privée, hors le problème n'est pas là. Le problème est que la sphère privée se développe considérablement au détriment de l'espace public. Les individus sont de plus en plus confrontés à des choses qui ont été prévues pour eux, et ont donc moins d'interactions avec l'espace commun. C'est l'espace public qui est menacé plus que la sphère privée. Les



internauts n'ont en quelque sort plus rien à dire ou à faire, tout est déjà « pré-dit » ou pré-fait.

Sur un site marchand on court-circuite nos étapes de recherche pour nous proposer directement ce dont nous avons besoin. Quelles que soient nos volontés, celles-ci ne sont pas présumées ou requises pour nous proposer ce dont nous avons besoin.

On pourrait penser que le processus de personnalisation permet de faire des choix moins arbitraires et que l'on est donc mieux éclairé. Cependant même si on ne maîtrise pas toutes les circonstances qui nous ont amené à faire un choix il est important de rendre compte de ses décisions. Dans le cas d'un achat suite à une recommandation faite par un site marchand par exemple, on subit en quelque sorte l'action menée.

Un autre aspect problématique avec cette personnalisation est qu'elle n'émancipe pas nécessairement les individus. Par un processus de recommandations comme celui fait par des sites web marchands certains pourraient être victimes de leurs pulsions tabagiques ou alcooliques.

Les algorithmes d'analyse de données arrivent à deviner des choses sur nous alors même que beaucoup d'éléments qui font la complexité d'une personne échappent à la numérisation. La valeur créée à partir des données s'appuie sur des résultats statistiques qui n'ont aucun sens pour nous.

## Conclusion

Nous avons dans ce document l'ambition de répondre à des questions soulevées par le nouveau contexte de Big Data dans lequel l'économie des données personnelles prends de plus en plus de place.

Alors qu'un marché est en train de se construire autour de la valorisation des données personnelles, sur les bases d'un nouveau capitalisme (le capitalisme cognitif), les entreprises ont aussi leur rôle à jouer dans cette économie *via* la redéfinition de leurs modèles d'affaire. L'économie des données personnelles est bien une économie du partage et de la cohabitation puisque qu'elle ne vaut pas grand chose seule. Mais c'est aussi et surtout une économie de la confiance.

Les enjeux techniques, juridiques, éthiques et de sécurité vont de pair avec la construction de cette confiance. Bénéficier des nombreux potentiels de valorisation liés aux données personnelles est tout à fait possible au regard de la réglementation et représente le plus grand challenge des entreprises.

L'analyse des données permet la création de valeurs très précieuses pour les entreprises qui peuvent désormais être plus proches de leurs clients et leur proposer des services de plus en plus personnalisés. Les plus-values générées permettent aux entreprises de s'enrichir en se basant sur des prédictions, mais peuvent aussi être bénéfiques aux clients qui reçoivent des services personnalisés.

Si la personnalisation des services aux clients se fait au prix de la divulgation de certaines données personnelles, il faut surtout se méfier des situations que cela pourrait générer terme. On pourrait se retrouver dans un monde où tout serait préfabriqué pour correspondre aux individus, qui seraient donc moins confrontés à l'espace public. On deviendrait ainsi passif dans le sens où notre curiosité, nos facultés de recherche et notre intérêt pour l'inconnu seraient étouffés par la bulle personnalisée créée autour de nous.

La valeur créée à partir des données marchandes est donc exceptionnelle et profite aux individus même si l'essentiel va à l'entreprise. Il faut cependant comprendre les risques liés aux stratégies appliquées par ces dernières pour que les changements sur notre monde soient le plus bénéfiques possible.

## Bibliographie

Capitalisme : <https://fr.wikipedia.org/wiki/Capitalisme> - Wikipedia

Éthique des données :

<https://www.lesechos.fr/idees-debats/cercle/cercle-156360-une-ethique-des-donnees-personnelles-doit-etre-le-fondement-du-marketing-data-driven-1215449.php> - Les Echos

Capitalisme cognitif :

<https://www.cairn.info/revue-multitudes-2008-1-page-39.htm> - Yann Moulier Boutang

[https://fr.wikipedia.org/wiki/Économie\\_du\\_savoir](https://fr.wikipedia.org/wiki/Économie_du_savoir) - Wikipedia

<https://halshs.archives-ouvertes.fr/halshs-00401880/document> - Carlo Vercellone

Économie des données :

<http://www.cigref.fr/wp/wp-content/uploads/2015/11/CIGREF-Economie-donnees-perso-Enjeux-business-ethique-2015.pdf> - CIGREF

Création de valeur:

<https://www.inserm.fr/index.php/thematiques/technologies-pour-la-sante/dossiers-d-information/big-data-en-sante>

<https://lelab-social.humanis.com/paroles-dexperts/big-data-indispensable-lamelioration-de-l-analyse-risques/>

Vie et mort des sciences sociales avec le big data - Dominique Boullier

<http://www.economiematin.fr/news-090514-big-data-banque-analyse-donnee-relation-client-ele>

<https://dataintrusion.wordpress.com/2015/04/11/big-data-quels-benefices-pour-les-consommateurs/>

<http://binaire.blog.lemonde.fr/2016/01/22/le-sujet-de-droit-au-peril-de-la-gouvernementalite-algorithmique/>

<http://binaire.blog.lemonde.fr/2016/01/22/le-sujet-de-droit-au-peril-de-la-gouvernementalite-algorithmique/>

Predictive Analytics - Eric Siegel

Poly de cours UTC SY09 2017