

## Big Data et Data marketing : opportunités et limites

Si la problématique de gestion des données n'est pas récente, le déluge de données produites par le monde digital (e-commerce, requêtes Internet, réseaux sociaux, capteurs, smartphones et de plus en plus d'objets connectés) conduit les entreprises -ou du moins certaines d'entre elles- à une utilisation radicalement différente des données. Poussée par les besoins de stockage et de traitement, l'évolution technologique ces dernières années est telle qu'elle ouvre des possibilités inenvisageables jusque-là : la gestion du « Big Data ». Si pour certaines entreprises il n'y a pas de réelle rupture dans la façon de gérer les données mais simplement une meilleure performance (plus de données gérées à moindre coût), d'autres ont réellement défini une stratégie centrée sur les données, qu'elles proviennent de l'interne ou soient obtenues en externe auprès de fournisseurs de données. Le marketing et la relation client font partie des fonctions de l'entreprise fortement impactées qu'il s'agisse d'affiner la connaissance client, développer du chiffre d'affaires par un meilleur ciblage, diminuer l'attrition ou créer de nouvelles offres. La mise en place de tels projets doit cependant surmonter un certain nombre de freins et comporte des contraintes et des limites tant juridiques, qu'en termes d'acceptation client, d'organisation ou de ressources humaines.

Mots clefs :

Big Data, Data Marketing, Data driven, Données massives, Données personnelles, Data scientist, Digital, GAFAs, Données non structurées, Objets connectés, Text-mining, Recyclage, Personnalisation, VRM

### I. Big Data et concepts associés

#### 1. La notion de Big Data

Le terme de **Big Data** est apparu pour la première fois en 2000 lors d'un congrès d'économétrie puis repris en 2008 et 2010 en couverture des revues Nature et Sciences, et s'est imposé dans les entreprises suite à un rapport de MC Kinsey en 2011. La traduction française recommandée par la DGLFLF<sup>1</sup> est « **mégadonnées** » mais on trouve parfois l'expression **données massives**. D'après Pierre Delort<sup>2</sup>, il n'existe pas de définition sérieuse de la notion. Celle-ci étant utilisée parfois de manière abusive. Trois termes lui sont cependant toujours associés : Volume, Vitesse, Variété symbolisés par les 3 V. Ces termes ont été employés par le groupe Gartner<sup>3</sup> en 2001 qui constatait une production exponentielle de données, de formats de plus en plus divers et nécessitant une amélioration des systèmes de gestion.

Le **volume** de données créées quotidiennement ces dernières années est impressionnant. On parle aujourd'hui couramment de pétaoctets<sup>4</sup> (milliard de méga d'octets). Entre 2010 et 2012 le volume de données récoltées aurait été équivalent aux données générées depuis le début de l'humanité. Selon une étude du cabinet IDC, le volume de données va être multiplié par 50 entre 2005 et 2020<sup>5</sup>. Cette explosion de données est liée au développement du digital facilitant la création, le stockage et la transmission des données. Celles-ci proviennent du développement de l'Internet, des pratiques de communication permanente des internautes et des mobinautes sous forme de production de contenu mais également de réactions, commentaires à ces contenus. 150 milliards de mails et 500

---

<sup>1</sup> La délégation générale à la langue française et aux langues de France élabore la politique linguistique du Gouvernement.

<sup>2</sup> Pierre Delort. Le Big Data PUF Avril 2015.

<sup>3</sup> Leader américain du conseil et de la recherche en technologies avancées.

<sup>4</sup> Les unités sont les suivantes : octet, kilo, méga, giga, tera, péta, exa, zetta.

<sup>5</sup> « The digital universe in 2020 » IDC 2012 cité in La révolution Big Data J C Cointot et Y Eychenne.

millions de tweets sont émis chaque jour, 2 millions de requêtes formulées chaque minute<sup>6</sup>. Mais plus encore ce sont les données de mobilité (géolocalisation) et dans le futur des données de l'Internet des objets (IoT) c'est -à-dire des objets du quotidien reliés à Internet grâce aux technologies des puces RFID<sup>7</sup> ou autres technologies sans fil qui constitueront des flux massifs de données. La possibilité de les maîtriser ouvre de nombreuses perspectives aux entreprises en termes de nouveaux services à offrir aux consommateurs ou clients professionnels.

La **variété** : les données créées sont de formats de plus en plus variés. Il s'agit de plus en plus de données non structurées, textes bruts issus d'un échange d'e-mail ou de conversations sur les réseaux sociaux, images, vidéos, enregistrements vocaux, traces de connexion à des sites web (logs), signaux transmis par les objets connectés grâce aux multiples capteurs dont ils disposent. Ces formats variés rendent les traitements plus complexes.

La **vélocité** : parallèlement à cette complexification, les besoins des entreprises évoluent vers une prise de décision de plus en plus rapide, voire en temps réel, de l'ordre de la fraction de seconde. Les données doivent donc être collectées et traitées de plus en plus rapidement.

## 2. Technologies Big Data

Ce sont **Google** et **Yahoo** qui sont à l'origine des progrès spectaculaires réalisés en traitement de données. Le fonctionnement du moteur de recherche de Google basé sur des calculs de « **Pagerank** » (algorithme<sup>8</sup> de classement de la popularité des pages web) et le stockage d'informations diverses sur ces pages a amené très rapidement Google à innover en infrastructures logicielles du fait du volume de données accumulées. Google a développé les composants essentiels du stockage et traitement des données massives (Map Reduce, Google Big Table, Google Big Files). Par ailleurs, dès 2001, Google était disponible en 26 langues s'appuyant pour se faire sur des technologies de **traduction automatique**, puis ont été ajoutés des modèles de **reconnaissance vocale** créés à partir de millions d'échantillons de voix et d'intonation. Doug Cutting employé par Yahoo ! a créé la première version de **Hadoop**, technologie Open Source, permettant d'effectuer des requêtes dans des puits de données distribués c'est-à-dire dont les informations sont situées sur des serveurs distants les uns des autres. Les traitements des données sont fragmentés sur différents serveurs afin d'optimiser le temps de traitement. Les requêtes peuvent être de nature très diverses, les algorithmes très complexes, les données très nombreuses, non structurées et les ressources de calcul mobilisées en divers lieux, là où se trouvent les données. La technologie permet de coordonner les traitements et de gérer les incohérences et les redondances. Dès lors, selon Gilles Babinet<sup>9</sup>, « on obtient un système dont l'efficacité est sans commune mesure avec ce qu'il était possible de faire auparavant dans des environnements de données traditionnels ».

La principale différence entre les données traditionnelles et les données massives ne porte donc pas sur le volume même si celui-ci a explosé mais sur le type de données et la façon dont elles sont stockées.

Traditionnellement les données d'entreprise étaient stockées dans des entrepôts de données internes au sein de **bases de données relationnelles**. Dans une base de données relationnelle, les données sont rangées de manière structurée : 1 ligne = 1 enregistrement ; 1 colonne = 1 attribut ; chaque cellule au croisement d'une ligne et d'une colonne a un format défini par avance. Le langage

---

<sup>6</sup> Enjeux les Echos Big Data grande chance 1<sup>er</sup> mai 2015.

<sup>7</sup> Composants électroniques qui envoient des informations par onde radio.

<sup>8</sup> Un algorithme est une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre un problème. Wikipedia.org

<sup>9</sup> Gilles Babinet. Big Data penser l'homme et le monde autrement. Le passeur 2015.

**SQL**<sup>10</sup> permet de formuler des requêtes. Cela suppose d'avoir défini par avance les types d'informations qui doivent être stockées et établi un modèle permettant de relier ses informations entre elles.

A contrario, lorsque l'on parle de Big Data il s'agit de données de types divers dont une partie au moins est constituée de **données non structurées**. Il peut s'agir par exemple du contenu des mails adressés par les clients au service réclamation, de conversations téléphoniques enregistrées, de conversations sur les forums, de traces laissées par les connections aux sites Internet etc. Ces données sont stockées telles quelles, sans construction d'un modèle préalable de rangement, dans un « **lac** » de données. On parle de données **NoSQL**<sup>11</sup> car le langage SQL ne permet pas de les traiter. Des outils permettent d'indexer et de catégoriser les informations non ou peu structurées en temps réel ; des algorithmes de traitement spécifiques permettent de trouver des liens entre les données et de découvrir des modèles. Le stockage des données Big Data se fait généralement non plus dans un entrepôt de données au sein de l'entreprise mais dans le **Cloud** au sein de Data Centers. Différents types d'acteurs interviennent : ceux qui développent et intègrent les bases de données, ceux qui les hébergent et les maintiennent, ceux qui apportent la puissance de calcul et ceux qui les utilisent. Des architectures plus agiles et plus puissantes permettent d'optimiser les ressources, de limiter les investissements et la maintenance et faire évoluer les infrastructures progressivement. L'entreprise utilisatrice des données peut choisir d'externaliser tout ou partie des opérations.

Les technologies Hadoop et NoSQL sont précurseurs dans le domaine du Big Data mais leur conception dans un environnement Open Source a favorisé le développement de multiples technologies similaires ou complémentaires en matière de stockage ou d'analyse des données non structurées. Aux trois V, Volume, Variété, Vitesse, les acteurs du marché ont parfois ajouté d'autres V : Valeur des données pour ce qu'elles sont susceptibles de contenir comme signaux ou en référence au fait qu'elles sont commercialisables, Vérité pour insister sur la nécessaire qualité des données.

### 3. Traitements en Big Data

Selon Charles Huot<sup>12</sup>, cinq familles de technologies sont clés pour le secteur des Big Data : text-mining, graph-mining, machine learning, data-vizualisation, ontologies. Toutes convergent vers un même objectif : simplifier l'analyse de vastes ensembles de données (donner du sens) et permettre la découverte de nouvelles connaissances.

**Le text-mining** consiste à dénombrer des occurrences (apparition d'un terme) mais également analyser le sens. On parle de plus en plus de sens-mining. Les techniques utilisées s'appuient sur celles développées pour le **traitement automatique des langues (TAL)**. Elles reposent soit sur des méthodes à base de règles explicites (grammaire) et données (dictionnaires), soit sur des méthodes par apprentissage ou machine learning.

**Le machine-learning** consiste à concevoir des systèmes apprenants qui ont pour caractéristique d'être de plus en plus performants au fil du temps car les algorithmes de calcul s'améliorent automatiquement au fur et à mesure. Par exemple, lors de l'analyse de textes, le principe consiste à partir d'un corpus de textes donnés en exemple dans lequel le système d'apprentissage va rechercher des régularités pour créer un modèle. On distingue les méthodes supervisées, semi-supervisées et non supervisées. Dans le système supervisé, les documents de base sont annotés

---

<sup>10</sup> Structured Query Language.

<sup>11</sup> Not only SQL.

<sup>12</sup> Charles Huot. Le Big Data si nous en parlions ? Actes du séminaire IST Inria (Institut public français de recherches en sciences du numérique) octobre 2014.

manuellement pour souligner et expliciter les éléments devant servir de modèle. Selon les objectifs de l'analyse et le niveau de détail recherché, les différentes méthodes seront plus ou moins pertinentes et plus ou moins coûteuses à mettre en œuvre.

Le graph mining permet de créer des graphes relationnels en isolant des groupes d'information. Cette méthode est utilisée par les moteurs de recherche sur le web pour établir le « ranking » (positionnement) mais aussi par des applications métiers spécialisées comme la détection de communautés et d'influenceurs sur les réseaux sociaux.

**La data visualisation** facilite la lecture des données massives. En effet la mise en forme graphique appropriée permet de communiquer beaucoup d'informations en peu d'espace et met en évidence des phénomènes non visibles à la lecture de tableaux tels que les corrélations<sup>13</sup> par exemple. La visualisation en temps réel est une technologie clé.

**L'ontologie** c'est-à-dire la création de référentiels métiers<sup>14</sup> est nécessaire au développement d'applications métiers pour des contextes professionnels ciblés.

Pour Pierre Delort<sup>15</sup> **l'exploitation du Big Data « consiste à créer en exploratoire et par induction sur des masses de données à faible densité en information des modèles à capacité prédictive »**. La grande masse de données analysées permet en effet de **détecter des signaux faibles**, de suivre leur évolution et ainsi d'analyser des **tendances**. A partir des corrélations observées entre différentes données des **modèles prédictifs** peuvent être élaborés. Le Big Data permet en effet de mettre au jour des corrélations « massivement multifactorielles » et ainsi confirmer des corrélations de type impact de la météo sur les ventes ou détecter des corrélations inattendues. Celles-ci ne doivent pas pour autant être interprétées comme des causalités (la variation en parallèle de deux phénomènes n'indiquant en rien qu'il existe une causalité ni si elle existe le sens de la causalité). Par ailleurs, de plus en plus, les systèmes mis en œuvre permettent de **réagir en temps réel** qu'il s'agisse d'automatisation ou d'aide à la décision.

## II. Big Data et Data marketing

Si l'exploitation du **Big Data** tel que défini précédemment ne concerne encore en marketing qu'une fraction des entreprises, l'explosion des données disponibles et le contexte concurrentiel accroît la tendance à faire reposer les décisions marketing sur des données plutôt que sur de simples intuitions. La mesure du retour sur investissement et les pratiques d'optimisation se multiplient. Les pratiques d'automatisation se développent pour faciliter la personnalisation. Le baromètre<sup>16</sup> visant à mesurer la perception du Big Data et la maturité des entreprises en France montre une forte évolution entre 2012 et 2014. L'index de maturité a augmenté de 66 % et plus particulièrement 80 % des équipes marketing interrogées se positionnent comme la première partie prenante d'une initiative Big Data. Les études font état de nombreux projets.

### 1. Les secteurs utilisateurs du Big Data

Aujourd'hui, tous les secteurs d'activité peuvent être concernés par l'exploitation du Big Data à des fins marketing et commerciales même si certains le sont plus particulièrement du fait de leur accès privilégié aux données clients.

---

<sup>13</sup> Voir les outils « Google trend » et « Google correlation ».

<sup>14</sup> Représentation des différents concepts propres à un métier et de leurs relations.

<sup>15</sup> Pierre Delors, Le Big Data PUF Avril 2015.

<sup>16</sup> « Big Data index » baromètre EMC et IDC cité dans IT for business février 2015.

Les premiers utilisateurs du Big Data ont été les **GAF**A (Google, Amazon, Facebook, Apple) du fait de la quantité de données accumulées par leurs activités respectives et de leur fonctionnement en ligne. D'après Pierre Delort<sup>17</sup>, Google, ayant l'intuition de la valeur de ces données a probablement stocké depuis 2003 les traces des recherches sur le web ainsi que des informations complémentaires (retour ou non de l'internaute sur la page de résultats après consultation d'un site et délai pour le faire, adresse IP correspondant à la recherche). Les données seraient anonymisées au bout de neuf mois puis agrégées à la semaine. Par suite, on peut considérer que l'ensemble des **plateformes** de mise en relation sont au cœur de l'exploitation du Big Data et l'on trouve aujourd'hui l'acronyme GAF**A** complété du M de Microsoft, du Y de Yahoo, auxquels on pourrait rajouter le L de LinkedIn et le T de Twitter.

La **publicité digitale** fait partie des secteurs précurseurs, les « intentions » clients détectées à partir des mots clefs de recherche (Search Engine Marketing) ou à partir de sites et pages visitées permettent d'affiner le choix des affichages et améliorent très sensiblement les gains. Au-delà de Google qui détenait en 2013 la moitié du chiffre d'affaires publicitaire Internet mondial, grâce à son programme d'Adwords d'autres acteurs tel **Critéo**, entreprise française, leader du reciblage se sont développés. Le « programmatique » c'est à dire le processus d'achat automatisé d'espaces publicitaires sur Internet devient peu à peu majoritaire.

La **distribution** est également un secteur où l'exploitation du Big Data est relativement développée. En particulier, les entreprises de vente à distance Pure Player, les places de marchés ou les entreprises multicanales ont, grâce à l'analyse des données digitales, mis en place des mécaniques très élaborées de ciblage et de recommandation. Le système de recommandation d'Amazon a été précurseur en la matière. La grande distribution dispose à travers ses systèmes d'encassements et ses cartes fidélité de grandes quantités de données, retraçant l'ensemble des achats clients. Ces données peuvent être utilisées pour son compte propre et sont susceptibles d'intéresser également l'ensemble des industriels. Par suite, les sociétés de service gérant ces données à des fins d'études telles les panélistes ou les sociétés gérant les opérations promotionnelles des distributeurs telles **Catalina**<sup>18</sup> détiennent de très gros volumes de données. Cette dernière affirme avoir accès aux données d'achat en grande distribution de 88 % des foyers français.

D'autres secteurs tels que le transport aérien, les opérateurs de téléphonie mais aussi les assurances et les banques, disposent de nombreuses informations sur leurs clients et ont des habitudes de traitement de celles-ci dans le cadre de leurs outils CRM. Certaines entreprises de ces secteurs ont une réelle avance, d'autres tardent cependant dans la mise en œuvre de modèles Big Data c'est-à-dire allant au-delà de l'exploitation de données internes structurées.

Enfin, le développement des objets connectés ouvre des opportunités en ce domaine à des acteurs de secteurs très divers.

## 2. Sources de données et Data Management Platform

La gestion du Big Data consiste donc à exploiter différentes sources de données, internes et externes en les faisant converger dans l'idéal vers une plateforme de gestion de données (DMP) afin d'enrichir les analyses et d'affiner les actions. Les premières DMP étaient centrées sur les données de navigation Internet mais aujourd'hui elles intègrent les différents points de contacts internes et externes. Ainsi les données internes transactionnelles ou les caractéristiques clients pourront être enrichies de données de marché plus riches. L'ensemble de ces flux de données pourront être traités

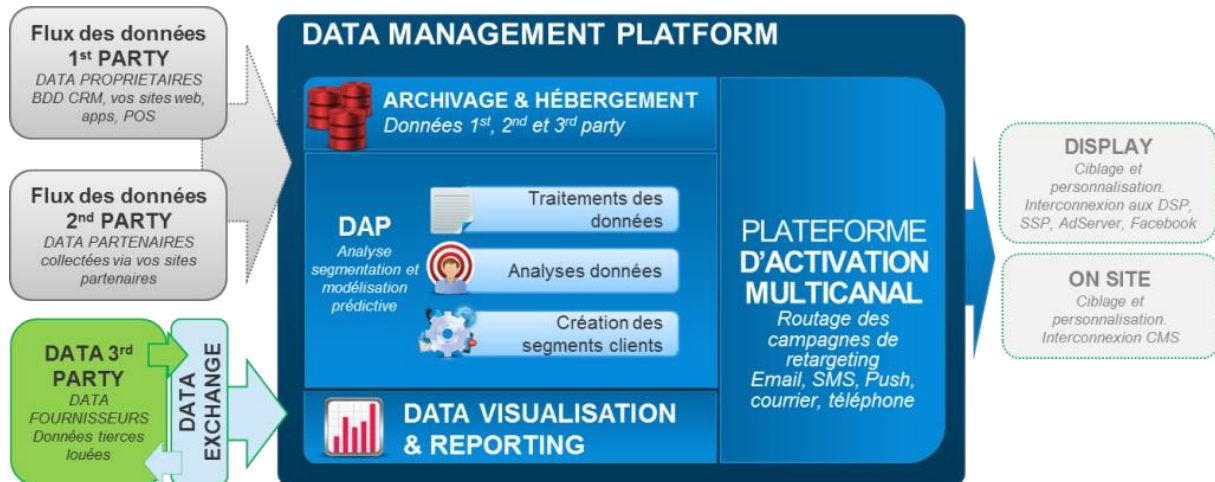
---

<sup>17</sup> Pierre Delort Le Big Data PUF Avril 2015.

<sup>18</sup> Catalinamarketing.fr.

**en différé** pour affiner des analyses de marché **ou en temps réel** pour optimiser des transactions. Les données peuvent dans certains cas être rattachées à un client en particulier et constituer des données personnelles ou avoir été anonymées et agrégées. Les différentes sources permettent de disposer de **données transactionnelles, comportementales et attitudinales**.

On peut distinguer des flux de données first party, c'est-à-dire propriétaires, second party, soit collectées auprès de site partenaires, et third party louées à des prestataires externes.



Source Cabestan – Filiale du groupe La poste

Actuellement, les stratégies DMP sont souvent réservées à l'achat et l'optimisation médias en temps réel mais leur intérêt va bien au-delà puisque cela permet d'agréger des données consommateurs pour personnaliser l'expérience digitale mais également sur d'autres canaux et apporte un réel avantage concurrentiel en matière de connaissance client.

- **En interne**

- L'outil CRM lui-même alimenté par les transactions clients au travers des différents canaux de distribution (physique, courrier, téléphonique, web) permet de connaître de façon détaillée les achats clients et leur historique. Les données personnelles confiées dans le cadre de formulaires d'inscription permettent de connaître les caractéristiques clients.
- L'enregistrement des contacts avec le SAV et des conversations liées (appel sur plateforme téléphonique, e-mail, chat, contact en magasin) permet un historique des événements mais aussi des réactions et opinions émises et de l'évolution de l'état d'esprit du client. Dans le cadre d'un contact téléphonique par exemple, la conversation enregistrée permet non seulement d'analyser le contenu mais aussi le ton employé et en déduire des émotions.
- Les outils de web analytics permettent d'analyser de manière détaillée les comportements clients sur le site en ligne (pages vues, entonnoir de transaction...) et ainsi d'optimiser ses fonctionnalités et son design, mais aussi de mieux comprendre les comportements clients et d'utiliser cette information dans le cadre des campagnes marketing. Les données de connexion (logs) aux pages Internet enregistrées grâce aux cookies déposés permettent de connaître les centres d'intérêt du ou des clients.
- De même, l'analyse des campagnes promotionnelles et notamment la pratique de l'A/B testing, c'est-à-dire la mise en place simultanée de deux campagnes présentant une différence permet

d'optimiser celles-ci et d'apporter des informations complémentaires en matière de connaissance du client.

- La mise en place de divers systèmes permettant de localiser les clients au sein des unités commerciales telles que les puces RFID sur les chariots ou les boîtiers ibeacon permettent d'analyser la fréquentation de manière précise. Les premières permettent en effet de retracer les parcours clients au sein d'un magasin, les seconds, grâce aux applications téléchargées par les clients sur leurs téléphones, permettent de les identifier, de garder un historique de leur venue et de communiquer au bon moment.

- Les signaux émis par les objets connectés, grâce aux capteurs dont ils disposent, permettent d'analyser les usages des produits et peuvent contribuer à l'évolution de ceux-ci, amener la création de nouveaux services ou l'amélioration de l'expérience client. Cette source d'information est encore aujourd'hui à ses balbutiements mais son exploitation ouvre des possibilités immenses.

- L'analyse des conversations au sein des communautés créées par l'enseigne et notamment sur les pages Facebook ou Twitter permet d'enrichir assez aisément les bases de données transactionnelles et comportementales de données attitudinales, on parle alors de social CRM.

- **En externe**

Des données de marché peuvent être achetées pour enrichir la connaissance client auprès des détenteurs de celles-ci (plateformes, opérateurs de téléphonies, agences spécialisées dans le traitement de données Big Data). Les données collectées sont généralement traitées par les entreprises qui en disposent puis anonymées afin de respecter la réglementation. Les données sont fournies aux entreprises clientes sous forme de comptage ou de statistiques et en fonction de profils clients.

- Les données de recherche sur les moteurs tels Google et l'analyse des retours sur la page de recherche permettent de mesurer les centres d'intérêts de la population. De même, les données de connexion et de navigation sur des sites partenaires permettent de détecter des intentions.

- Les données issues des réseaux sociaux (Facebook, Twitter, LinkedIn,) permettent de détecter des tendances en analysant les conversations. L'utilisation des lieux privés doit respecter la législation concernant les données personnelles. Twitter est une source particulièrement performante pour la veille du fait de son utilisation très répandue<sup>19</sup> et de son format. En effet, du fait du format limité à 140 caractères, les messages publics sont plus simples et l'indexation par hashtags est facilitante. Les forums sont généralement difficilement exploitables car il est nécessaire de comprendre les contextes ; l'orthographe est souvent trop fantaisiste pour permettre une utilisation des outils de traitement automatique des langues.

- Les données de géolocalisation produites par les téléphones mobiles, les tablettes, les GPS ou autres objets connectés permettent de situer le client potentiel mais aussi de mesurer des fréquentations.

La combinaison des différents types de données ou la mise en parallèle de celles-ci permet d'enrichir considérablement les analyses et d'affiner les actions.

---

<sup>19</sup> Selon Enquête IPSOS mars 2013, 89 % de la population française connaît Twitter, 5% ont un compte Twitter actif dont 60 % lisent des tweets au moins tous les deux jours.

### 3. Les modèles d'utilisation des données

Selon Christophe Bénavent<sup>20</sup>, il existe quatre modèles d'utilisation de ces données massives.

Pour lui, le premier modèle qui consiste à mettre à disposition des **services marketing études** des informations plus fouillées, plus nombreuses, plus fines n'est pas révolutionnaire. Les analyses menées sont de même type que celles menées précédemment (descriptives ou explicatives). La différence réside dans le niveau de granularité. Plus de données permet plus de finesse dans l'analyse.

Le deuxième modèle consiste à **automatiser des décisions** basées sur l'analyse des données en temps réel comme l'affichage d'une publicité sur une page en fonction d'un algorithme ou l'affichage d'un prix sur une étiquette électronique. Ces analyses sont menées en un temps très court de l'ordre de fractions de secondes, il s'agit de millions de micro décisions.

Le troisième modèle, qu'il appelle modèle de l'empowerment, consiste à **fournir de l'information personnalisée** aux individus afin de les aider dans leur prise de décision. Ceci peut concerner par exemple les forces de vente ou le service relation client d'une entreprise ou bien directement des clients par exemple au travers des sites d'avis consommateurs.

Le quatrième modèle consiste à fournir de **nouveaux services** grâce aux données collectées tels les services de coaching en lien avec les bracelets fitness connectés ou encore les services de mise en relation dans le cadre des activités de consommation collaborative comme le covoiturage ou la location d'appartements.

### 4. Les applications marketing

La gestion du Big Data, qu'elle entraîne une réelle rupture dans les méthodes ou simplement permette d'optimiser le travail des équipes et améliorer le ROI (Return On Investment ou rentabilité du capital investi) peut impacter chacune des étapes de la démarche marketing, qu'il s'agisse des études, de la segmentation, de la définition de l'offre, de la communication ou de la gestion de la relation client.

- **Une nouvelle posture pour les études**

L'impact principal du Big Data en matière d'études consiste en un changement de posture, l'entreprise devenant **observatrice plutôt qu'enquêtrice**.

Il s'agit dorénavant d'analyser les données existantes plutôt que d'aller en chercher de nouvelles. L'analyse des bases de données gigantesques sur les achats clients, leurs réactions aux promotions, l'impact de publicités, couplée aux informations météo et géographiques, la possibilité de mener et d'analyser de multiples tests et celle par ailleurs d'analyser des attitudes à partir de conversations sur les réseaux sociaux rend dans de nombreux cas inutiles les études traditionnelles. La position d'observateur peut par ailleurs laisser supposer que l'on a accès à des données présentant **moins de biais**. Elles posent cependant dans certains cas la question de la représentativité, ainsi que de la fiabilité. Les nouvelles méthodes de l'analyse d'opinion permettent de combiner les avantages de méthodes traditionnelles quantitatives et qualitatives puisqu'elles permettent d'obtenir des informations spontanées, exprimées librement comme cela pourrait être le cas dans le cadre d'un focus group et auprès d'une grande quantité de répondants. Dans un contexte de Big Data, la distinction études quantitative et qualitative devient de fait dépassée. Dans le cadre d'études quantitatives, la notion d'échantillonnage et les calculs de probabilités associées perdent également

---

<sup>20</sup> « Les quatre chemins du Big Data : no best way » Xerfi canal.



de leur importance du fait des volumes analysés, voire de l'exhaustivité de l'étude. Par ailleurs, les outils de géolocalisation permettent d'obtenir des données de fréquentation en temps réel très précises.

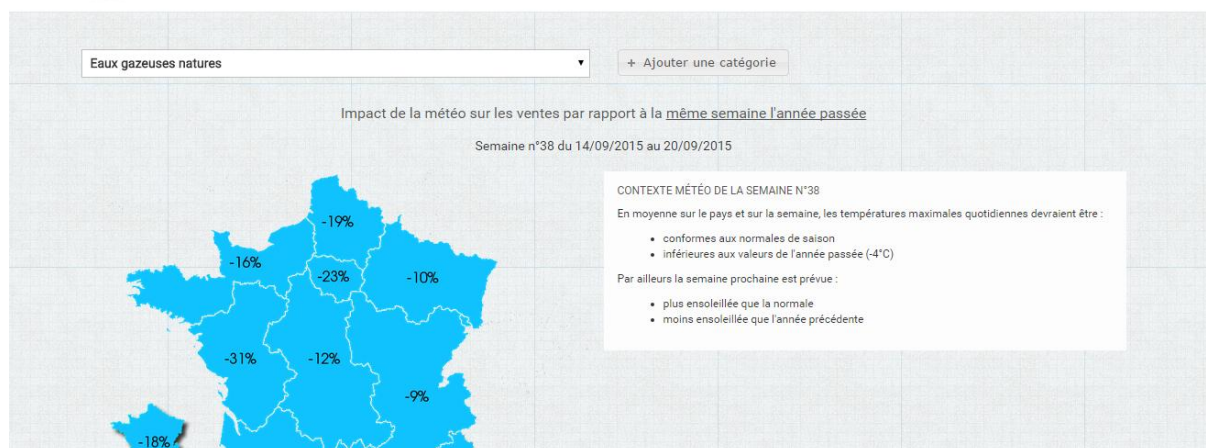
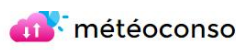
- **La connaissance des comportements, des tendances et l'amélioration des prévisions**

Différents prestataires proposent des outils permettant de faire des prévisions de volume ou d'envisager les tendances d'un marché ou d'une catégorie de produit généralement par zone géographique. Nous présenterons ici quelques exemples sans pour autant prétendre à l'exhaustivité.

L'outil **Google trend** permet de mesurer l'occurrence des recherches menées sur un terme donné et Google corrélation permet de représenter sur un même graphe l'occurrence de plusieurs termes pour visualiser des corrélations. Ces informations sont cependant plus utiles pour affiner une stratégie de communication (choix de mots clefs pertinents) que pour réaliser des prévisions.

Plus précisément, l'outil **Google « Universal analytics »**, a permis aux Galeries Lafayette de connaître le parcours numérique des clients possesseurs de la carte de l'enseigne et constater par exemple que 20 % des actes d'achats en magasin sont précédés d'une visite sur un des sites et que pour un cinquième d'entre eux, cette visite sur le web se fait par l'intermédiaire d'un smartphone dont pour moitié le jour même de la transaction en magasin. Cette connaissance fine des parcours ROPO (Research on line, Purchase off line) et des indicateurs liés permet à l'enseigne de remettre à plat son offre digitale.

La société **Climpactmetnext**<sup>21</sup> permet de mesurer la sensibilité météo des produits et par la suite de réaliser des prévisions de ventes en fonction de la météo afin d'éviter les ruptures et de pousser le bon produit au bon moment. En partenariat avec IRI (<http://www.iri.centrepompidou.fr/>), elle a mis à disposition des professionnels un site de prévisions. Celui compare les prévisions de températures maximales et d'ensoleillement aux moyennes saisonnières et à l'année précédente et calcule l'impact prévisionnel sur les ventes des différentes catégories de produits.



**Orange business Service** propose depuis fin 2013 une offre d'analyse des déplacements de population à destination des acteurs du tourisme, **Flux Vision**<sup>22</sup>. Ces données sont anonymées mais permettent de compter trois types de population : les résidents, les touristes (qui passent au moins

<sup>21</sup> [meteoconso.climpact-metnext.com](http://meteoconso.climpact-metnext.com)

<sup>22</sup> Olivier Ondet, « Le Big Data au service du tourisme » Annales des Mines Réalités industrielles. 2015.

une nuit) et les excursionnistes (qui passent seulement une journée). Les statistiques sont quotidiennes et dynamiques (par plages de 2 heures). Les flux sont analysés par origine et destination pour comprendre les interactions, et sont complétés par des caractéristiques sociodémographiques (hommes, femmes, tranches d'âges...). Ce type d'étude permet de mieux comprendre les fréquentations et ainsi d'adapter les infrastructures ou la mise à disposition de personnels.

Représentation dynamique de la fréquentation touristique du département des bouches du Rhône.



Par ailleurs, la taille des bases de données accessibles permet de détecter des tendances émergentes. Des événements rares peuvent être identifiés. L'analyse prédictive permettant de déceler de nouvelles tendances de marché s'appuie de plus en plus sur les outils d'analyse d'opinion. En effet, en analysant une très grande quantité de messages sur les blogs, les médias sociaux et les sites de presse, il est possible d'anticiper les évolutions d'une mode. Par exemple, dans le domaine de l'industrie de la mode, où les tendances évoluent très vite et des choix pertinents de collection impactent fortement le chiffre d'affaires annuel, la société **StyleSight**<sup>23</sup> analyse de façon automatisée pour le compte de marques de prêt à porter à partir des photos prises lors des défilés de modes. Elle met ainsi en évidence ce qui a inspiré les collections (formes, couleurs, accessoires). Par ailleurs, la perception des clientes est analysée au travers des photos partagées sur **Pinterest** et des commentaires associés.

- **La mesure de la notoriété, de la réputation, des opinions et des attentes**

L'importance prise par les échanges entre internautes concernant les marques ou les produits et l'influence que cela peut avoir sur les comportements des acheteurs amène les entreprises à s'intéresser à ce qui s'écrit sur elles, leurs produits, leurs marchés, leurs concurrents. L'étude de la **notoriété** plus encore de la **réputation** ou des **opinions** peut être menée grâce à l'analyse des avis sur Internet et des conversations sur les réseaux sociaux<sup>24</sup>. **Le text mining** permettait déjà depuis quelques années de mesurer des informations objectives et ainsi savoir de qui et de quoi les consommateurs parlaient sur la toile en repérant des mots et mesurant la fréquence d'apparition. L'analyse automatique sémantique permet d'aller plus loin et constitue un aspect essentiel de l'intérêt du Big Data en marketing.

**L'analyse des opinions** permet de classer les textes en fonction de leur avis positif ou négatif (la polarité) de mesurer l'intensité d'une opinion et d'identifier des émotions de base comme la colère, la peur, la joie, la tristesse, le dégoût voire d'autres émotions plus fines. Ces techniques nécessitent cependant une réelle expertise et un travail préalable poussé d'identification des concepts clefs à

---

<sup>23</sup> Decideo.fr «Au-delà du glamour, l'industrie de la mode embrasse le Big Data » février 2014.

<sup>24</sup> Bernard Normier « Analyser les avis sur internet et les réseaux sociaux pour valoriser votre notoriété : sentiment analysis et Opinion mining » Vitrac éditeur 2014.

analyser. L'opinion <sup>25</sup>porte sur un **objet** et une ou des **caractéristiques** de cet objet. Un **sentiment** est exprimé sur cette caractéristique par un **auteur** à un **moment** donné. Il est nécessaire de définir les objets utiles à l'analyse et de préciser si l'opinion nous intéresse globalement ou pour chaque caractéristique en particulier. On peut s'interroger sur la fiabilité des sources d'informations et sur la valeur de ces opinions selon l'influence que peuvent avoir les auteurs, voire les sites ou canaux sur lesquels les opinions sont exprimées. Certains logiciels permettent de calculer des scores d'influence des internautes en fonction de leur activité, de leur pertinence sur tel ou tel thème, de leur audience (nombre de followers, d'amis, d'abonnés) de leur affiliation (journal, parti, entreprise) de leur résonance (nombre de retweets, commentaires).

Pour mener une analyse de l'opinion, les entreprises peuvent s'appuyer d'abord sur leurs données internes (courrier, courriels, centre d'appels, enquêtes de satisfaction, commentaires sur le site de l'entreprise). Par exemple, **Auchan**<sup>26</sup> a mis en place une solution depuis 2011 **d'analyse sémantique de « la voix du client »** (sollicitations magasins, téléphone ou e-mail dont les remontées sont centralisées) qui est capable d'interpréter les phrases et apprend le langage des clients, les noms des produits et les marques. Les « catégorie managers » ont été formés à l'outil. L'analyse sémantique permet de mettre en évidence les **sources d'irritation** des clients et les **demandes de produits**. Cet outil a permis à Auchan d'augmenter son chiffre d'affaires en proposant en rayon les produits attendus par les clients comme par exemple en proposant des agendas jusque début février ou en offrant des pains individuels au rayon boulangerie ou des produits Halal avant les autres distributeurs. Les messages clients deviennent ainsi une source permanente d'étude clients.

L'analyse d'opinion s'appuie par ailleurs sur des données externes : sites d'avis et sites de e-commerce publiant des avis comme ceux incontournables dans le secteur du tourisme mais aussi sites de grandes enseignes comme Darty ou la Fnac, sites de presse et blogs, Twitter. Les outils d'analyse de l'influence mettent aussi en évidence les communautés, généralement en utilisant les techniques de visualisation de données statiques ou dynamiques comme par exemple l'outil inMaps de LinkedIn ou Linkfluence très utilisé en marketing.

De nombreux acteurs proposent des solutions d'analyse d'opinion. Certains sont spécialisés sur cette activité comme les plateformes d'e-réputation telles Talkwalker voire spécialisées sur un réseau comme les outils d'analyse de Twitter, d'autres sont des plateformes généralistes de la veille (ex : Radian6), des plateformes de CRM comme Salesforce ou Eptica qui ont ajouté ces fonctionnalités à leur offre pour analyser les opinions des clients. Enfin de grands acteurs du progiciel de gestion intégré (PGI / ERP) comme SAP, IBM ou Oracle intègrent dans leurs systèmes des modules d'analyse d'opinion.

- **Une segmentation à l'infini**

La notion de segmentation est fortement impactée par le Big Data. Les technologies Big Data permettent de segmenter à l'extrême la clientèle dans une optique opérationnelle, d'affiner toujours plus les caractéristiques clients afin d'automatiser la proposition d'offres adaptées. La question de la taille minimum du segment n'a plus de raison d'être. Il n'y a pas réellement de limite à la segmentation. Promovacances<sup>27</sup> par exemple annonce avoir identifié 30 000 profils clients. Le Big

---

<sup>25</sup> D'après le professeur Bing Liu dans Bernard Normier « Analyser les avis sur internet et les réseaux sociaux pour valoriser votre notoriété ».

<sup>26</sup> Bernard Normier.

<sup>27</sup> Mcfactory.fr Table ronde « User expérience centrix ».

Data amène également une évolution en termes de critères de segmentation. Au-delà de l'utilisation de données socio- démographiques, et des bases de données transactionnelles permettant de classer les clients en fonction de critères de récence, fréquence et montant des achats (RFM), il permet en effet une véritable montée en puissance des techniques de segmentation comportementales. La fréquentation du site web (pages consultées) et les réactions aux newsletters et e-mailing précédemment adressés (ouverture ou non, actions ultérieures) sont analysées afin de réaliser un véritable ciblage comportemental. Enfin des critères d'attitudes vis-à-vis de la marque ou de certaines valeurs issues des données des réseaux sociaux peuvent être ajoutés.

Dans certains cas, des données d'usage sont intégrées à l'analyse des comportements afin d'affiner les segments. Par exemple Tellmemore<sup>28</sup> analyse le temps passé à étudier par les utilisateurs de la plateforme en ligne et utilise ce critère pour constituer ces segments. De même l'analyse des consommations menée par les opérateurs de téléphonie ou les fournisseurs d'énergie permet d'obtenir des segments beaucoup plus fins que précédemment.

- **L'amélioration des produits, de nouveaux services basés sur les données et une personnalisation de l'offre**

La gestion du Big Data et particulièrement des données issues des objets connectés permet de mieux connaître les usages des produits par les consommateurs et ainsi adapter les produits aux besoins réels ou proposer des versions adaptées à différents profils de consommateurs. Ainsi le constructeur automobile PSA s'est allié à IBM<sup>29</sup> pour exploiter les données des véhicules connectés (1,5 millions de voitures). Grâce à l'analyse des usages, PSA peut améliorer la conception et la qualité des véhicules ainsi qu'offrir des services personnalisés aux clients. La fréquence d'utilisation des différentes fonctionnalités (comme par exemple l'utilisation du toit ouvrant) permet d'affiner les modèles ainsi que les prix. Par ailleurs PSA commercialise déjà cinq services, dont l'appel d'urgence, la localisation du véhicule en cas de vol, ou encore la gestion de flotte pour les entreprises.

De même, la compagnie d'assurance Allianz propose à ses clients automobiles d'embarquer un boîtier analysant les conditions et comportements de conduite. Celui-ci leur permet d'offrir des services d'assistance et conseils mais également de bénéficier éventuellement d'une réduction de tarif pour bonne conduite.

Dans le même esprit le bouton de SAV Darty ou la gestion des cartouches d'encre d'une imprimante connectée sont de nouveaux services proposés aux clients particuliers. Pour aller plus loin encore, SEB et Coheris se sont alliés pour imaginer une transformation numérique de la cuisine et créer un moteur de recommandation de recettes tenant compte de diverses données en provenance des objets connectés de la cuisine mais aussi du profil, de la météo, de « like » de recettes issus de Facebook. Ce moteur s'appuie sur un algorithme de machine-learning et permet de faire du prédictif.

Dans un certain nombre de secteurs d'activité, le **produit** proposé peut être adapté à chaque individu. La personnalisation de l'offre repose sur une connaissance client approfondie, une gestion des profils clients et une personnalisation dynamique<sup>30</sup> basée sur une offre modulaire.

---

<sup>28</sup> Mcfactory.fr Table ronde « User expérience centrix ».

<sup>29</sup> L'usine digitale, « pourquoi PSA s'associe à IBM dans les voitures connectées » avril 2015.

<sup>30</sup> Issac et Volle 2011 cité dans l' e-marketing Grégory Bressoles Dunod 2012.

Ainsi pour Air France<sup>31</sup>, le niveau de personnalisation possible est infini et les traitements individuels réservés aux services de luxe sont accessibles aux services de masse. La connaissance client en temps réel combinée à l'automatisation permet de pousser à son extrême la personnalisation. Le « One to one » devient effectif. Par exemple la connaissance des centres d'intérêt du client grâce aux cookies déposés lors d'une visite du site permet de proposer une page d'accueil personnalisée sur un site web à commencer par la langue d'expression. Selon le directeur du développement digital d'Air France, l'ensemble des éléments de personnalisation aurait amélioré le taux de conversion de 8 %. La connaissance client basée sur le web analytics, le dossier d'achat et l'historique client permet d'améliorer l'expérience client. Les e-mails adressés avant le vol sont personnalisés. Un client voyageant avec un chien trouvera dans l'e-mail un encart lui rappelant comment voyager avec un animal. Ces encarts spécifiques sont ouverts dans 70 % des cas par les clients. La personnalisation pourrait aller plus loin dans les mois à venir : notification de porte d'embarquement par SMS adaptée au profil du client (pressé de rentrer dans l'avion ou souhaitant y entrer au dernier moment), en cours de vol, playlist de musique ou de films en fonction de ce qu'il a écouté précédemment.

La personnalisation de l'offre se traduit pour les enseignes proposant une offre large telle La Fnac ou Netflix par exemple, par un processus de recommandation qui peut être basé sur un filtrage collaboratif grâce aux informations données par l'internaute lui-même mais le plus souvent sur une gestion des profils. La navigation et le comportement de l'internaute sont enregistrés par le site afin d'effectuer des recoupements et des comparaisons avec des profils similaires stockés dans la base de données et ainsi personnaliser l'offre de manière dynamique. Cette technique a pour origine les pratiques d'Amazon qui recommande des produits en fonction de celui affiché à l'écran et des produits achetés par les internautes ayant commandé ce dernier. Ces pratiques favorisent les ventes croisées.

- **Une communication ciblée en fonction des probabilités d'achat**

L'automatisation des opérations de **communication** (marketing programmatique) repose également sur un processus de création de scénarios correspondant aux différents segments. Ceux-ci sont utilisables pour l'envoi de newsletters ou e-mailing de relance. Il existe par exemple un scénario d'abandon de panier (trigger marketing), d'abandon de site, d'anniversaires, de fête des mères. Lorsqu'il s'agit d'un nouveau client, un e-mailing de bienvenue peut être envoyé. Le travail des équipes marketing consiste à créer et optimiser les scénarios adaptés aux différents segments identifiés. Dans un certain nombre de cas, les scénarios peuvent être mis en œuvre en temps réel, une probabilité forte d'abandon de panier décelée pouvant ainsi conduire à l'affichage d'une promotion particulière afin de favoriser la conclusion de la transaction.

Le marketing programmatique fait disparaître le décalage entre les phases amont de préparation d'une campagne et son exécution, les actions sont maintenant consécutives, il est possible de tester différents scénarios (A/B testing) et de suivre leurs performances en temps réel afin de décider de les amplifier ou de les limiter en fonction de leurs performances.

La Fnac par exemple a testé l'approche « predictive Big Data » de la société Tinyclues<sup>32</sup> pour optimiser la cible de ses newsletters. Grâce aux données d'achat en magasin, sur Internet, sur les applications mobiles, les historiques de navigation, les click sur les e-mail précédents, les taux d'ouverture, les informations sociodémographiques, les données de géolocalisation, les avis produits concernant 3,5 millions d'adhérents, 20 millions de clients, 12 millions de visites uniques sur le site

---

<sup>31</sup> Table ronde User expérience centrix et Table ronde les priorités pour 2016 vidéos MC Factory.

<sup>32</sup> Points de vente N° 1174 . Avril 2015.

Internet chaque mois, elle a identifié une cinquantaine de segments, des signaux précurseurs d'achat et scoré la probabilité pour un client de répondre à une offre. Ainsi elle cerne les 5 à 10 % de clients qui représentent la plus grande part des achats pour limiter l'envoi de communication et éviter de « spammer » les autres utilisateurs. Cette technique lui permet également de sélectionner la meilleure offre à faire à un client lorsque plusieurs campagnes entrent en concurrence. Le ciblage est optimisé en permanence grâce aux algorithmes de machine learning.

- **Un ciblage en temps réel**

Le développement des données de géolocalisation permet également d'adresser aux consommateurs une publicité ciblée en fonction de leur proximité avec certaines enseignes. Ainsi SFR<sup>33</sup> a développé une solution orientant les messages en temps réel en fonction des données de géolocalisation et selon des segmentations effectuées au préalable par un adserver<sup>34</sup>. Ces opérations sont menées avec l'accord de l'utilisateur qui signale par un Opt In<sup>35</sup> initial son consentement à recevoir ce type de message.

Les pratiques de geofencing<sup>36</sup> permettent également de cibler un client potentiel lorsqu'il entre dans une zone définie comme par exemple une zone commerciale et se trouve donc à proximité d'une enseigne. Les pratiques de géofencing sont encore assez rudimentaires et méritent d'être enrichies de données comportementales et affinées afin d'être plus pertinentes (par exemple la publicité devrait être adaptée aux tranches horaires et à la fréquentation des enseignes par les clients).

En magasin, la mise en place de balises ibeacon permet également d'adresser des messages ciblés aux consommateurs qui l'ont préalablement accepté en téléchargeant une application de l'enseigne.

Le Retargeting ou reciblage consiste à cibler un consommateur intentionniste, c'est-à-dire qui a manifesté un intérêt pour un produit en cliquant sur celui-ci sans aller jusqu'à l'achat. Un cookie déposé sur son ordinateur permet de conserver des données transactionnelles ou comportementales (nombre de pages vues, temps passé, langue, horaire, produits cliqués). Une autre technique est parfois utilisée, le finger printing qui tente d'identifier précisément un internaute et de suivre ses connexions sur l'ordinateur (à travers l'adresse IP mais aussi d'autres identifiants enregistrés sur son ordinateur ou sa tablette) même si l'on ne connaît pas son identité réelle. Ces données de navigation peuvent être croisées avec la base CRM de l'annonceur pour définir des scores d'appétence et l'intérêt de recibler le client lors de sa navigation future sur un autre site. Le reciblage est généralement confié à un prestataire spécialisé en placement de bannières tel Critéo. Ce dernier achète par enchère en temps réel (Real Time Bidding) soit quelques millisecondes, un espace sur une page visitée et présente ainsi aux internautes des bannières adaptées à leur intérêt à l'occasion de visites sur différents sites Internet. Grâce à ce modèle le taux de clic serait multiplié par 5.

- **Une relation client assistée pour limiter l'attrition et un support aux équipes de ventes**

L'utilisation des systèmes apprenants permet par exemple de prédire l'attrition. Augustin Huret et Jean-Michel Huet<sup>37</sup> expliquent comment l'analyse « hypercube »<sup>38</sup> menée pour un opérateur de

---

<sup>33</sup> Synthèse du marché Big Data Paris.com.

<sup>34</sup> L'adserver est un logiciel de gestion de campagnes de publicité en ligne.

<sup>35</sup> Option d'adhésion en français : terme marketing ou légal pouvant concerner les adresses courriel, ou plus globalement les services informatiques.

<sup>36</sup> Cela consiste à définir des frontières virtuelles permettant la détection de consommateurs dès qu'ils se rendent dans la zone.

<sup>37</sup> « L'intelligence artificielle au service du marketing » Augustin Huret et Jean Miche-Huet.

téléphonie sur une base contenant à la fois des caractéristiques clients (adresse, profession, revenus, sexe, âge), des usages (type de contrat, ancienneté, évolution, usage voix, SMS, data, subvention, type de terminaux) et l'historique des relations de l'opérateur avec chacun (campagne de marketing direct précédente, appel entrant au service client, appel sortant, emailing etc.) a permis à cet opérateur d'isoler les clients risquant de le quitter pour un concurrent et de focaliser les efforts de marketing direct. L'analyse exhaustive des données a permis de délimiter les profils des plus forts « churners ». L'analyse a été couplée avec celle des canaux de contacts pertinents et permis de déterminer par exemple qu'une femme mariée habitant dans telle région et ayant opté pour telle offre est deux fois plus fidèle quand elle reçoit d'abord un email de type 1 puis un appel téléphonique de type C.

Les modèles permettent également d'assister les opérateurs en fournissant des suggestions de réponse ou de compensation adaptée à la réclamation client mais aussi à son potentiel et son comportement probable. Par exemple Canal Plus a mis en place une DMP permettant de faire converger les données des différentes sources de contact client et des données externes afin que les opérateurs du centre d'appel répondent au mieux aux réclamations clients.

La connaissance de données techniques de réseaux par exemple pour les fournisseurs d'accès Internet et l'intégration de ces données dans la base de données utilisée par le centre d'appel relation clients permet aux opérateurs de vérifier les causes de problème voire la véracité des propos clients en temps réel et ainsi d'adapter leur proposition.

L'assistance à la vente est aujourd'hui également disponible pour les vendeurs présents en magasin comme par exemple chez Séphora qui met à disposition des équipes des outils de recommandation au travers de l'application « My Séphora » disponible sur i-pad.

### III. Freins et limites à la mise en place d'une approche Big Data en marketing

Une étude menée en 2014 par le cabinet Ernst & Young<sup>39</sup> auprès de 150 entreprises françaises a mis en évidence que le « Big Data bang » n'avait pas encore eu lieu bien que la perception soit majoritairement positive. Pour les consultants, ce retard s'explique par dix principaux freins d'ordre psychologique, stratégique, organisationnel ou technologique aux différentes phases de l'exploitation des données.

#### 1. Les principaux freins

- **Collecte des données clients** : les canaux digitaux sont sous-exploités et les données proviennent essentiellement des systèmes de facturation (84 %) et des outils CRM (66 %) ;
- **Données non structurées peu exploitées** : par 45 % seulement des entreprises essentiellement du secteur de la grande distribution ou des média et technologie de l'information ;
- **Manque de compétences analytiques** au sein de l'entreprise ;
- **Carence des outils de traitement de données** : les technologies et outils de traitement spécifiques nécessaires sont encore trop peu utilisés ;
- **L'analyse de la Data est encore trop peu orientée vers le prédictif** (10 % seulement des entreprises) et le **temps réel** (5 %) ;

---

<sup>38</sup> Les fondements de l'approche hypercube ont été développés par l'école Polytechnique. Il s'agit d'un système apprenant capable à partir de données de différents types de déduire des règles.

<sup>39</sup> (Big) data : où en sont les entreprises françaises ? 2014.

- **Manque de transversalité** dans la gestion des projets Big Data. Les silos en interne sont un frein à l'exploitation de la data ;
- **Absence de mesure du ROI des projets Big Data.** Or mesurer la valeur des projets sur la base de KPI faciliterait leur mise en place ;
- **Manque de soutien de la direction générale ;**
- **Réticence** des clients à **partager des données personnelles** pour des raisons de sécurité et de protection de la vie privée (70 % des consommateurs sont réticents) ;
- **Faible sensibilisation aux enjeux de sécurité et de protection de la data.**

## 2. Les contraintes juridiques en matière de données personnelles<sup>40</sup>

Dans le cadre du marketing, une utilisation respectueuse de la législation en matière de **données personnelles** paraît fondamentale pour la pérennité de tels outils.

Du fait des très nombreuses informations collectées sur Internet, du développement permanent des technologies qui facilitent le quotidien mais créent une traçabilité des comportements de plus en plus importante (géolocalisation, marketing ciblé...) de nouveaux enjeux en termes de protection de la vie privée apparaissent et amènent certains à penser qu'il faudrait renforcer les règles actuelles.

Les pratiques en matière de données personnelles sont très encadrées. C'est la Loi « Informatique et libertés » du 6 janvier 1978 modifiée le 6 août 2004 qui pose pour la France les règles et les principes de la protection des données personnelles. La directive européenne du 24 octobre 1995 puis celle du 12 juillet 2002, modifiée en 2009, permettent d'harmoniser les règles en Europe. Cette dernière est transposée en France le 24 août 2011.

La « donnée à caractère personnel » est une information qui va permettre d'identifier directement une personne comme son nom, son prénom, une photo mais aussi celle qui permet de l'identifier indirectement comme un numéro de carte bancaire ou une adresse IP.

Les personnes doivent être informées par écrit et de façon très accessible que leurs données font l'objet d'une collecte, d'un traitement, de la finalité de celui-ci et du fait que les données sont susceptibles d'être cédées. Tout formulaire de collecte doit comporter une mention d'information. Pour le moment, sauf cas spécifique (prospection commerciale, données de santé) les sites peuvent se contenter d'informer les internautes et n'ont pas l'obligation d'obtenir leur consentement. En ce qui concerne les cookies, déposés sur les sites afin de mémoriser les comportements des internautes, ceux-ci doivent être informés de leur utilisation ou installation ainsi que de la façon dont ils peuvent empêcher que ces cookies soient installés. Les sites n'ont pas l'obligation d'obtenir le consentement de l'internaute à travers une case à cocher (selon une étude 98 % ne la cocheraient pas) mais doivent obtenir l'accord de l'internaute au travers par exemple (selon une recommandation de la CNIL) de la mise en place d'un bandeau sur la page d'accueil explicitant que des cookies sont installés et leur finalités et qu'ils demeurent jusqu'à ce que l'internaute ait changé de page et donc tacitement donné son accord. La CNIL recommande par ailleurs que les historiques de navigation ne soient pas conservés au-delà de 13 mois. Les internautes doivent donner leur consentement pour que les données fournies soient cédées à des tiers.

Lorsqu'elles pratiquent une segmentation de la clientèle grâce aux données collectées, les entreprises doivent faire savoir aux clients qu'ils sont susceptibles d'être classés par exemple comme bon ou moins bon client. Leur classement doit par ailleurs leur être communiqué s'ils le souhaitent.

<sup>40</sup> Source France- Université- Numérique-Mooc.fr « Informatique et libertés sur internet » juin/juillet 2015.



Dans le cadre de la collecte de données, les entreprises doivent respecter une obligation de proportionnalité (ne collecter que des données pertinentes par rapport à leur finalité) ce qui pose des limites. En 2010, suite à des commentaires déplacés sur son fichier client un avertissement a été donné par la CNIL à Acadomia avec pour conséquence la chute immédiate de son cours de bourse.

La collecte doit être loyale (les personnes doivent savoir que l'on collecte de l'information les concernant et cette collecte doit être faite directement auprès d'elles). De fait, la collecte de données personnelles sur les espaces publics de l'Internet (pages Facebook par exemple) est strictement interdite. Ainsi les analyses réalisées à partir des données des réseaux sociaux sont généralement anonymées en étant agrégées ou rattachées à des pseudonymes. L'anonymisation réelle pose cependant question.

La collecte doit être licite (les informations collectées doivent être autorisées par la loi). En France, il est interdit de collecter des données sensibles, la segmentation par exemple ne peut être bâtie sur un critère de lieu de naissance qui pourrait traduire une caractéristique ethnique. Il est également spécifiquement interdit de collecter le numéro de sécurité sociale sauf dans les cas d'usages définis par la loi. Le numéro de sécurité sociale étant un identifiant unique, il permettrait aisément de faire des recoupements de fichiers. Certains acteurs regrettent cette interdiction.

La durée de conservation des données doit être limitée (cohérente par rapport à l'usage). Les données de connexion (jour et heure de connexion à tel ou tel site Internet, appels téléphoniques etc.) doivent en théorie être anonymées sauf par les opérateurs fournisseurs d'accès. Les données de contenu ne doivent théoriquement pas être conservées. Pour les mesures d'audience, la durée maximum recommandée par la CNIL est de 6 mois.

La pratique de l'IP tracking qui aboutirait à augmenter un prix lorsque l'on constate qu'un internaute revient à plusieurs reprises voir un produit afin de l'inciter à acheter immédiatement est strictement interdite.

Par ailleurs, tout site Internet qui va collecter des données doit assurer leur sécurité (veiller à ce qu'elles ne soient ni détruites ni modifiées) et leur confidentialité (qu'elles ne soient pas accessibles par des tiers non autorisés). Cette obligation s'applique y compris lorsque les données sont stockées chez des sous-traitants. Un opérateur de téléphonie a été condamné à ce sujet.

Un projet de règlement européen de protection des données personnelles est en cours de discussion depuis 2012 et pourrait être adopté prochainement. L'objectif est de l'adapter à l'évolution technologique et aux nouveaux usages en matière de données numériques. Plusieurs points pourraient avoir un impact sur la gestion du Big Data. Un des éléments du projet consisterait à mettre en place un principe de « minimisation de la collecte de données » en lieu et place du principe actuel de proportionnalité ce qui de fait limiterait le recueil de données. Les questions d'anonymisation sont débattues. Bien souvent, en effet, les données de connexion collectées permettent indirectement d'identifier la personne. Si la réglementation est renforcée cela pourrait donc modifier considérablement les méthodes de gestion des données.

Il faut noter cependant que seuls dix pays ont mis en place des règles de protection équivalentes aux règles européennes et que les entreprises implantées hors Europe y expérimentent parfois d'autres pratiques. Un point clef par exemple est l'identifiant unique du consommateur qui permet dans certains pays (d'Amérique du sud par exemple) d'agréger systématiquement l'ensemble des données le concernant quels que soient les canaux d'obtention de ces informations.

### 3. Les limites liées à la perception client

Au-delà de l'aspect légal, la question se pose de la perception par le consommateur de l'ensemble de ces pratiques. Par exemple, dans les pays anglo-saxons la notion de « privacy » consiste à mettre en place des règles de protection des données clients afin de gagner leur confiance.

Maria Mercanti-Guerin a mené une recherche afin de mesurer l'impact des pratiques de reciblage<sup>41</sup> sur l'image des marques. Les consommateurs sont prêts à fournir des informations les concernant en échange de bénéfices perçus. La contextualisation joue un rôle primordial, il est particulièrement important que le message soit délivré au moment le plus approprié. Le ciblage comportemental améliore très sensiblement le taux de clic, cependant le retargeting peut déclencher une perception d'intrusion qui peut nuire à l'image de la marque. Une étude qualitative préalable menée auprès d'étudiants de formation continue en marketing a montré que la technologie permettant le reciblage est incomprise et l'extrême personnalisation assimilée à de la magie. Il y a une inquiétude quant à l'utilisation des données personnelles en particulier celles issues des réseaux sociaux. L'image de marque souffre du retargeting car la marque est perçue comme harcelant le consommateur ; l'envie d'acheter est cependant accrue lorsque la publicité est personnalisée et en particulier le cross selling est bien perçu. L'étude quantitative ultérieure menée, en soumettant des hommes et des femmes à des scénarios de retargeting, a confirmé que la répétition de l'annonce accroît l'intrusion perçue qui elle-même dégrade l'image de la marque. L'acceptation des nouvelles techniques de ciblage par les consommateurs est facilitée par la pertinence des annonces reçues qui correspondent à un vrai service et nécessite la mise en place d'un système permettant de limiter la répétition à laquelle est soumis l'internaute (capping).

De même Stéphanie Hérault et Bertrand Belvaux<sup>42</sup> cherchent à évaluer le rôle de l'intrusion perçue dans la vie privée de l'utilisation de services mobiles géo localisés. Ils soulignent que ce frein, peut être compensé par certains bénéfices (utilité et facilité à l'utilisation). Les entreprises peuvent donc proposer davantage de bénéfices et mieux gérer les problèmes de vie privée par une éducation numérique des utilisateurs.

Les pratiques montrent que certains consommateurs essaient aujourd'hui de se soustraire aux possibilités de pistage en ligne en utilisant par exemple un navigateur respectueux de leur vie privée ne stockant ni historique, ni adresse IP, ni cookies comme DuckDuckGO. La barre de navigation de la société Privowny permet de gérer les cookies et de générer des adresses e-mails à usage unique, de remplir automatiquement les formulaires en ligne à partir des informations personnelles stockées sur un compte Privowny. Au-delà, certains acteurs visent la réappropriation par les consommateurs de leurs données personnelles et le développement du principe de **déclaration** d'intention d'achat. Ces derniers phénomènes peuvent être regroupés sous l'acronyme VRM (Vendeur Relationship Management)<sup>43</sup>. Le principe est d'offrir aux consommateurs un ensemble d'outils pour gérer la relation entre vendeur et client en leur permettant de contrôler leurs données personnelles afin de rééquilibrer la relation client-vendeur, lutter contre les utilisations potentiellement abusives des données et renforcer la confiance. Pour Sylvain Willart, la façon dont les entreprises peuvent intégrer le VRM dans le cadre de leur relation avec leurs clients est cependant encore à construire.

La qualité et la fiabilité des données collectées peuvent également être questionnées. En effet pour diverses raisons dont la méfiance à l'égard de l'utilisation ultérieure des données, un certain nombre

---

<sup>41</sup> L'amélioration du reciblage par les Big Data : une aide à la décision qui menace l'image des marques ? Maria Mercanti Guerin Cairn.

<sup>42</sup> Privacy paradox et adoption de technologies intrusives Le cas de la géolocalisation mobile. Stéphanie HERAULT, Bertrand BELVAUX. DM n° 74 (2014).

<sup>43</sup> Le VRM : un nouveau paradigme pour la relation client ? Sylvain WILLARD DM oct-dec 2013.

de consommateurs travestissent leur identité numérique. Renforcer ou rétablir la confiance permettrait d'obtenir des données de meilleure qualité. D'autre part, l'utilisation par les consommateurs de multiples terminaux de connexion (devices) ordinateurs, téléphone mobile, tablettes, parfois partagés par exemple entre plusieurs membres d'une même famille rend particulièrement difficile le fait de rattacher un cookie à un identifiant unique. Le défi est d'éviter la déperdition d'individus liée au changement de canal.

#### 4. Les problématiques d'organisation

Le Big Data n'est pas qu'une question d'outils mais de stratégie et nécessite réflexions et choix en termes d'organisation. La question doit être traitée comme un véritable projet et l'organisation devra être choisie en lien avec les besoins réels de l'entreprise. La finalité de cette gestion Big Data et les usages envisagés doivent être définis.

Tout d'abord les entreprises ont le choix entre mettre en place des structures en interne ou faire appel à des prestataires externes qui gèreront pour elle le Big Data. Il peut s'agir de généralistes ou de spécialistes. Par exemple, l'agence Captain Dash<sup>44</sup> propose un outil à destination des équipes marketing permettant d'intégrer l'ensemble des données clients et de produire des tableaux de bord graphiques et dynamiques intégrant éventuellement des données externes. D'autres sociétés peuvent simplement traiter l'aspect réseaux sociaux par exemple ou fournir des analyses et du conseil issus des transactions clients comme le propose Catalina. Dans les faits, seules les entreprises de taille significative ont intérêt et les moyens de mettre en place leur propre structure de gestion du Big Data.

Pour les acteurs choisissant de mettre en œuvre un projet Big Data en interne, des choix se posent en termes de technologie : conserver les outils existants (CRM...) et ajouter une brique Big Data, mettre en place des systèmes hybrides, ou mettre en place des systèmes centrés sur la gestion du Big Data<sup>45</sup>.

La mise en place d'une gestion Big Data, comme tout projet d'entreprise, nécessite un engagement des directions d'entreprises mais aussi une capacité à faire travailler ensemble différents services bien souvent cloisonnés. Il s'agit d'une part d'impliquer tant les services informatiques que les services marketing ventes, mais également parfois de faire partager des référentiels communs aux différents services marketing, commerciaux, relations clients. Des objectifs parfois contradictoires entre la direction financière, le marketing et la supply chain doivent être mis à plat et réconciliés. L'organisation en silos est un véritable frein. La mise en place d'une DMP pour les données internes suppose une véritable coordination entre tous les intervenants.

Selon l'étude menée par Markess<sup>46</sup> en 2014, la **gouvernance** des données est essentielle pour tirer parti du Big Data. Il s'agit de recenser les données disponibles en validant leur qualité, de déterminer les données externes nécessaires, de définir comment, par quel prestataire accéder à ces données en choisissant un prestataire pérenne.

Par ailleurs, la question des ressources humaines est aujourd'hui cruciale. Elle se pose d'abord en termes de pilotage. Dans une entreprise « data driven », outils et culture permettent de récupérer les données, les structurer et produire de la valeur. La coordination de l'activité Big Data peut être

---

<sup>44</sup> Créée par Gilles Babinet Captaindash.com.

<sup>45</sup> Jean-Charles Cointot et Yves Eychenne, « La révolution Big Data » Dunod 2014.

<sup>46</sup> **Markess** « meilleures approches pour tirer parti du Big Data, référentiel de pratiques 2014-2016 ».

confiée à un « Chief Data Officer » qui travaillera en lien avec les directions métiers. D'autre part, les traitements et analyses de données Big Data nécessitent des profils particuliers qui sont rares aujourd'hui. Il existe très peu de personnes formées à l'analyse de ces données non structurées. Le profil de « data scientist » qui suppose des compétences à la fois en mathématique, statistique et informatique est aujourd'hui très recherché. Enfin les équipes marketing et les managers marketing sont bien souvent trop peu formés à ces outils pour être de véritables interlocuteurs des analystes. Pour combler ces lacunes des formations se mettent en place, de niveaux très divers. En 2014, en partenariat avec IBM, HEC a mis en place une formation sur le Big Data destinée aux étudiants du MBA, visant à combiner une formation généraliste de dirigeants d'entreprise avec une compétence plus pointue en matière d'analyse et d'interprétation des données. France – Université-numérique propose un MOOC « Fondamentaux pour le Big Data » destiné à un public ayant des bases en mathématique et algorithmique nécessitant d'être rafraichies pour suivre des formations en Data sciences et Big Data. Ce Mooc prépare à un Mastère spécialisé « Big Data : gestion et analyse des données massives », un certificat d'études spécialisées « Data scientist » ou aux formations sur le machine-learning. Enfin, une redéfinition des emplois et un accompagnement en interne de certains salariés pourrait sans doute permettre d'adapter les compétences aux nouveaux besoins.

## Bibliographie / sitographie

- **Pierre Delort**, « Le Big Data » PUF Que sais-je ? Avril 2015
- **Jean-Charles Cointot et Yves Eychenne**, « La révolution Big Data » Dunod 2014
- **Gilles Babinet**, « Big Data, penser l'homme et le monde autrement » Le Passeur 2015
- **Charles Huot**, « Le Big Data si nous en parlions ? » Actes du séminaire IST Inria (Institut public français de recherches en sciences du numérique) de boeck octobre 2014.
- **Bernard Normier** « Analyser les avis sur Internet et les réseaux sociaux pour valoriser votre notoriété sentiment analysis et Opinion mining » Vitrac éditeur 2014
- **Grégory Bressoles**, « L' e-marketing » Dunod 2012
- **Augustin Huret et Jean Michel Huet** « L'intelligence artificielle au service du marketing » Expansion Management Review 2012
- **Olivier Ondet**, « Le Big Data au service du tourisme » Annales des Mines Réalités industrielles. 2015
- **Maria Mercanti Guerin** « L'amélioration du reciblage par les Big Data : une aide à la décision qui menace l'image des marques ? » Revue internationale d'intelligence économique -2013
- **Stéphanie Herault, Bertrand Belvaux**, « Privacy paradox et adoption de technologies intrusives : Le cas de la géolocalisation mobile » DM n° 74 (2014)
- **Sylvain Willard**, « Le VRM : un nouveau paradigme pour la relation client ? » DM oct-dec 2013
- **Christophe Bénavent**, « Les quatre chemins du Big Data : no best way » **Xerfi Canal**
- **France- Université- Numérique-Mooc.fr** « Informatique et libertés sur Internet » juin/juillet 2015
- **Mcfactory.fr** Table ronde « User expérience centrée » et Table ronde « les priorités pour 2016 »
- **Points de vente**, Dossier Big Data N° 1174 . Avril 2015
- **Bigdataparis.com** Synthèse du marché 2014
- **Ernst & Young**, (Big) data : où en sont les entreprises françaises ? 2014
- **Markess** « meilleures approches pour tirer parti du Big Data, référentiel de pratiques 2014-2016 »
- **Cabestan.com**
- **Catalinamarketing.fr**
- **meteoconso.climpact-metnext.com**
- **IT for Business**, Le tandem DSI-Marketing 1<sup>er</sup> février 2015
- **Stratégies** « Google montre le chemin pour les Galeries Lafayette » 30 avril 2015
- **Decideo.fr** « Au-delà du glamour, l'industrie de la mode embrasse le Big Data » février 2014
- **L'usine digitale**, « pourquoi PSA s'associe à IBM dans les voitures connectées » avril 2015
- **Enjeux les Echos**, « Big Data grande chance » 1<sup>er</sup> mai 2015
- **L'expansion** « mégadonnées, sortir de la pensée magique » mai 2015

