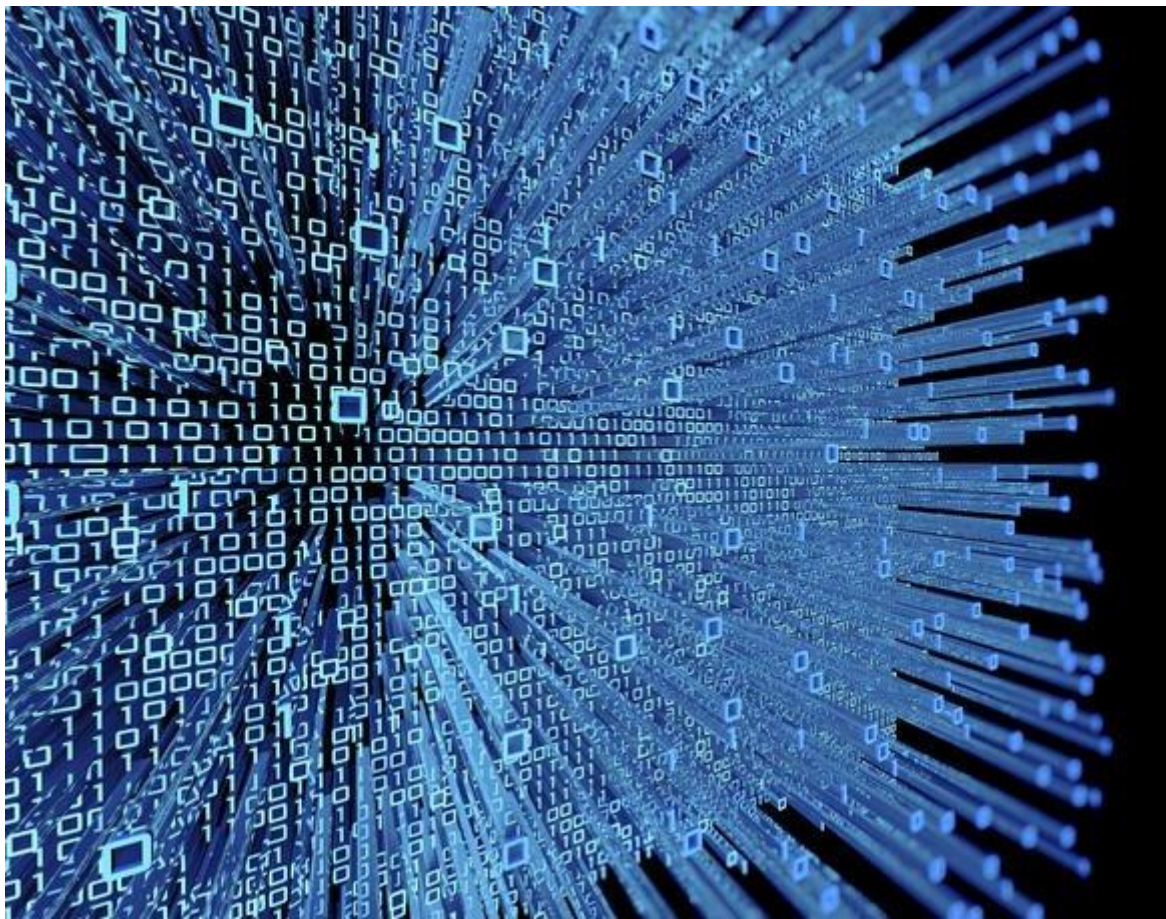


# Propos sur les big data

Les apports à l'analyse, à l'IA, à la société, les compétences, les difficultés, les fournisseurs



Michel Bruley

novembre 2021

## Préface

La société d'information a vu son développement s'accélérer avec la numérisation des textes, des sons, des films et les capacités de les stocker, de les communiquer, de les traiter ... Avec ces big data, une nouvelle ère pour l'approfondissement des connaissances et de leurs diffusions s'est ouverte.

D'aucuns ont dit que le développement de la société de l'information ou de la connaissance, selon les auteurs, était pour l'humanité un moment aussi important que le développement de l'agriculture au néolithique. Comme nous en sommes au début, il est difficile de préjuger toutes les conséquences pour les hommes de cette nouvelle ère.

Cependant, il est clair que nous allons vivre de nombreux changements plus ou moins positifs. Il y a quelques années, j'ai déjà décrit de nombreuses possibilités apportées aux entreprises par les big data dans un document intitulé « Premiers pas dans les Big Data », ci-dessous j'ai regroupé 35 articles parus chez DECIDEO en six sous thèmes :

### **1 - Big data et analyse de données : révolution ou évolution ?**

- 11 - Big Data, la fin des analyses qualitatives ?
- 12 - Pour aller des Big Data aux solutions, il faut savoir identifier les problèmes
- 13 - De l'analyse à l'action : 50 nuances à prendre en compte
- 14 - Seul l'apprentissage automatique permettra à de nombreuses entreprises de valoriser leurs mégadonnées
- 15 - De la difficulté, des échecs et de l'importance des projets Big Data / Intelligence Artificielle
- 16 - Nous sommes tous impactés par nos propres biais cognitifs
- 17 - Big Data : à propos des études bidon qui rapportent gros notamment en pharmacie

### **2 - Quelques aspects des big data et de l'intelligence artificielle**

- 21 - Big Data : bientôt les algorithmes vous connaîtront mieux que vous-même
- 22 - Le Big data et l'Intelligence économique
- 23 - Big data et Prospective : « Il n'y a pas de vent favorable pour celui qui ne sait où aller » Sénèque
- 24 - Big data et Lobbying : « Le pouvoir n'est rien ... seule compte l'influence » Hérodote
- 25 - L'intelligence artificielle nous conduit-elle vers une prolétarisation de la société ?

### **3 - De la maîtrise des données dans la société d'aujourd'hui**

- 31 - De l'analyse de données aux choix politiques
- 32 - Les données sont le carburant indispensable de la dynamique démocratique
- 33 - L'open data au service des citoyens pour contrôler les politiques et l'administration
- 34 - Les données quantitatives sont des arguments de poids
- 35 - Big Data, statistiques et infographie : les chiffres disent-ils toujours la vérité ?

### **4 - De la nécessité de disposer d'hommes compétents pour tirer parti des mégadonnées**

- 41 - Propos sur les conseils en big data
- 42 - Mégadonnées : la construction d'algorithme nécessite toujours du jus de cervelle
- 43 - L'analyste de mégadonnées ou big data doit se méfier de son enthousiasme
- 44 - Big data : au-delà des incontournables solutions open source, il faut savoir mobiliser les bonnes compétences
- 45 - Big data : Il faut un bon équipage pour ne pas se noyer dans le data lake
- 46 - L'avenir est radieux pour les spécialistes de l'analyse des données

### **5 - La multiplication des technologies du big data amène son lot d'inconvénients**

- 51 - L'internet des objets ou la consécration du Big Data
- 52 - Le monde merveilleux de l'internet et des big data : GAFAM, HACKERS, NSA ... et usurpation d'identité
- 53 - Big Data et protection de la vie privée : mégadonnées => méga-problèmes
- 54 - Lutte antiterroriste et mégadonnées vont-elles conduire à des atteintes à la vie privée ?
- 55 - Vers une révolution de la cryptographie avec le quantique

### **6 - Propos sur les entreprises qui développent les technologies big data**

- 61 - Big data et open source
- 62 - Big data : Hadoop doit être au cœur de votre système d'information décisionnel
- 63 - Installez votre système décisionnel dans les nuages à vos risques et périls
- 64 - Big data, analyse de données et modèles métiers
- 65 - Fournisseurs de solutions Big Data : la vie est un « long » torrent déchaîné
- 66 - Big Data : les Systèmes d'Information décisionnels migrent dans les nuages
- 67 - La diversification, une voie pour la pérennité

## 1 - Big data et analyse de données : révolution ou évolution ?

Certes, la mise à disposition rapide d'un grand volume de données variées de qualité ... est de nature à profondément améliorer les analyses de données en général et les études quantitatives en particulier. Cependant, nous avons affaire, plus à une évolution qu'à une révolution, car l'approche Big Data ne transforme pas radicalement la façon de faire des analyses de données et la plupart des savoir-faire antérieurs restent nécessaires.

### 11 - Big Data, la fin des analyses qualitatives ?

Face à un cas à résoudre, la bonne démarche pour y arriver consiste à identifier le ou les problèmes, voir la littérature, spécifier un but, collecter des données, les analyser et les interpréter, dégager des informations. On a l'habitude en matière de données de distinguer deux grandes approches : l'approche qualitative et l'approche quantitative.

L'approche qualitative met l'accent sur une vue holistique de ce qui est étudié (via des documents, des histoires de cas, des observations et des entretiens). Les plans de recherche quantitatifs s'efforcent eux d'identifier et d'isoler des variables spécifiques dans un contexte d'étude (recherche de corrélation, relations, causalité).

La recherche qualitative, qui met l'accent sur l'exploration, la compréhension, la contextualisation, l'introspection et la construction de théories, fournit une base solide pour des mesures quantitatives, une mise à l'échelle et une généralisation plus larges. La recherche quantitative, qui met l'accent sur de grands échantillons, peut fournir une vue d'ensemble d'un domaine pouvant révéler des tendances, des incohérences, etc., pouvant être analysée plus avant à l'aide de méthodes qualitatives.

Approche qualitative ou quantitative ? Le choix dépend du type de données nécessaire pour un problème de recherche particulier. La principale différence entre recherche qualitative et quantitative n'est pas la « qualité » mais la procédure. Il est tout à fait possible de quantifier des données qualitatives ou d'exercer une analyse qualitative des données quantifiées. Les méthodes qualitatives et quantitatives ne s'excluent pas mutuellement.

La recherche qualitative est une approche d'enquête utile pour explorer et comprendre un phénomène central. Pour en savoir plus sur ce phénomène, l'enquêteur pose aux participants des questions ouvertes et générales, recueille les vues détaillées des participants sous forme de mots ou d'images et analyse les informations à des fins de description et pour définir des thématiques. À partir de ces données, le chercheur interprète la signification des informations en s'inspirant de réflexions personnelles et de recherches antérieures. La structure du rapport final est flexible et affiche les préjugés et les réflexions du chercheur.

Il faut avoir bien conscience que le but d'une recherche qualitative est le plus souvent de découvrir des idées, de développer une compréhension détaillée et approfondie d'un phénomène, d'acquérir une compréhension qualitative des raisons et motivations sous-jacentes, de poser un large éventail de questions, de développer une compréhension initiale du paradigme qualitatif du champ étudié. L'approche qualitative met l'accent sur la compréhension, se concentrant sur la compréhension du point de vue du répondant, et ouvre la voie à des généralisations par comparaison de propriétés et de contextes individuels.

La recherche quantitative est une approche d'enquête utile pour décrire les tendances et expliquer la relation entre les variables trouvées dans la littérature ou des études qualitatives antérieures. Pour mener cette recherche, l'enquêteur spécifie des questions fermées, localise ou développe des mesures à l'aide de statistiques. À partir des résultats de ces analyses, le chercheur interprète les données par comparaison avec des prévisions et des études antérieures. Le rapport final, présenté dans un format standard, doit entre autres montrer l'objectivité et l'impartialité du chercheur.

Le but des recherches quantitatives est de tester des hypothèses ou des questions de recherche spécifiques. Il s'agit de mesurer, de quantifier des données et généraliser les résultats d'un échantillon à la population. Pour cela il faut un grand nombre de cas représentatifs. Il faut se concentrer sur des faits ou des raisons pour les événements sociaux et tester des hypothèses. Concrètement il convient de découper la réalité en morceaux distincts, qui sont ensuite combinés en grappes

statistiques. Il s'agit de résumer et classer les observations, les prédictions de comportement, de façon à expliciter clairement et ce qui est recherché et construire une structure statistique.

Le Big Data se situe clairement dans le domaine des approches quantitatives, mais pour certains il ne se compose pas que de données, mais aussi de traces, c'est-à-dire d'éléments quantitatifs dont la pertinence n'est pas totalement assurée et dont il faut se méfier. Par exemple, une transaction effectuée sur internet est clairement une donnée, mais une page vue, les traces d'un parcours sur le web ne seraient que des éléments, dont les significations ne seraient pas certaines. Dans un contexte de numérisation de la société, une avalanche de traces est produite, qui autorise la multiplication « d'approches corrélacionnistes radicales » qui oublient fréquemment que corrélation n'est pas causalité.

Dans le monde du Big Data d'aujourd'hui, les approches qualitatives restent indispensables pour traiter des problèmes humains qui ont pratiquement toujours des dimensions relatives difficiles à appréhender quantitativement. Par exemple, on a, depuis de nombreuses années, identifié statistiquement en France des travailleurs pauvres, mais par un défaut d'analyse qualitative et d'attention à cette notion de pauvreté qui est relative dans notre société d'hyperconsommation, on n'a pas vu venir le mouvement des gilets jaunes aux fins de mois difficiles.

## **12 - Pour aller des Big Data aux solutions, il faut savoir identifier les problèmes**

Les systèmes d'aide à la décision supportent le processus qui va de l'analyse à l'action, à sa conduite et au contrôle de son résultat. Dans ce processus, la première grande étape est celle de l'identification d'un problème qui appelle une solution. L'expérience nous dit qu'un problème bien posé est un problème à moitié résolu et surtout qu'il n'y a pas de bonne réponse à une mauvaise question. Dans la vie de tous les jours, les difficultés, les situations instables ou dangereuses, les questions en désaccord sont nombreuses. Mais il faut prendre en considération qu'un problème est généralement relatif, qu'il n'existe pas en soi, mais pour soi, que c'est une question de perception qui sera différente par exemple selon les personnes, les entreprises, et selon les moments. Ce qui n'était pas hier un souci peut l'être aujourd'hui, ce qui aujourd'hui n'est pas un souci peut l'être demain, du fait d'un changement : de contexte, de volume, d'objectif ...

Pour identifier un éventuel problème en entreprise, il faut explorer la situation, la comprendre, poser un diagnostic et formuler le problème. L'exploration de la situation nécessite des informations sur les faits, les perceptions, les opinions des individus, des groupes, de caractériser les besoins de chacun et de l'organisation. Il faut analyser les causes et conséquences de la situation, préciser les faits, mettre en relief les écarts significatifs entre les situations (hier, aujourd'hui, désirée, satisfaisante ...). Pour expliquer, comprendre la situation il faut tenir compte de beaucoup d'éléments (conjoncture générale, contexte particulier, stratégie des acteurs ...), faire une analyse systémique avec objectivité et rigueur, bien contrôler, confronter les informations récoltées.

Pour bien comprendre la situation, avant de caractériser le problème, il y a des pièges à éviter comme de considérer que la situation est simple, de ne pas regarder les détails, de confondre opinions & faits, symptôme & mal, d'avoir des préjugés, de rapidement privilégier un point de vue ... La formulation du problème doit décrire la situation définie comme insatisfaisante, ses manifestations, ses effets, les personnes, les résultats impactés, décrire son historique et esquisser une situation plus désirable envisageable, ses objectifs, ses avantages, les voies possibles pour l'atteindre en évitant d'être idéaliste ou irréaliste.

Un problème est donc un décalage entre une situation actuelle jugée insatisfaisante et une situation définie comme satisfaisante. Résoudre le problème, va donc consister à passer d'une situation à l'autre, c'est-à-dire supprimer (ou réduire) l'écart entre les deux en trouvant, en mettant en œuvre des solutions. Mais pour cela il faut d'abord bien décrire la situation à améliorer et concrètement tous les aspects du problème ou du dysfonctionnement. Pour cela il existe des méthodes par exemple l'analyse QQOQCP, l'analyse modulaire des systèmes, le classement Pareto des données, le diagramme "causes effets" d'Ishikawa.

En entreprise, la description traditionnelle d'un problème est faite à l'occasion d'études du type bilan, audit ou diagnostic. Le bilan est principalement centré sur le présent (points forts, points faibles), mais fait aussi référence au passé, l'audit est principalement centré sur le passé et le présent, mais fait aussi des recommandations pour le futur. Le diagnostic est principalement centré sur le futur, tourné vers l'action en mettant en lumière problèmes et solutions, c'est la première étape d'une action de changement. Au-delà de ces approches classiques, qui visent à opérer des changements ponctuels, il existe une méthode qui vise le changement continu, le Kaizen des Japonais qui encourage à chaque niveau des

organisations des petites améliorations quotidiennes sans gros investissement, contrairement à l'approche occidentale qui fonctionne davantage par grandes avancées souvent coûteuses et peu concertées.

Avec l'internet des objets qui nous génèrent des Big Data, le monde est de plus en plus enregistré & mesuré, il le sera en permanence et dans beaucoup de ses dimensions demain. L'établissement des faits en sera facilité, on le voit dès aujourd'hui avec les faits divers qui sont relatés dans la presse et sont largement alimentés par des données en tous genres (vidéos, réseaux sociaux, antennes relais ...). Dans les entreprises, dans les matériels, dans les produits, les capteurs se généralisent et mettant tout sous contrôle, les approches Kaizen ou classiques en sont facilitées. Mais si les données servent à profiler les problèmes, l'identification d'un problème ne se résume pas, comme on l'a vu plus haut, à la compilation de quelques chiffres et comme l'a dit Albert (Einstein) : « un problème sans solution est un problème mal posé ».

### **13 - De l'analyse à l'action : 50 nuances à prendre en compte**

Ces dernières années, à chaque fois que je tombais sur une citation qui m'inspirait, je la notais. Au final j'ai constitué une collection de plus de deux cents que j'ai souvent utilisée pour mes différentes contributions pour Decideo. Je vous en propose ici une cinquantaine qui concerne le processus qui va de l'analyse à l'action, en passant par la caractérisation d'une situation, la prise en compte du passé, du futur, des risques, la fixation d'un objectif et la mobilisation des hommes. J'ai ci-dessous regroupé grosso modo les propositions par sujet, mais cet ensemble n'a pas bien entendu l'ambition d'épuiser les sujets.

#### **Analyse, bilan, problématique :**

- Le premier problème c'est le problème. Un problème bien posé est un problème à moitié résolu. Il n'y a pas de bonne réponse à une mauvaise question.
- L'œil regarde, mais seule la sensibilité voit.
- L'intérêt particulier fascine les yeux et rétrécit l'esprit.
- Ce sont nos sentiments qui s'interposent entre la réalité et nous.
- Les faits ne pénètrent pas dans le monde des croyances.
- Un oui, un non, une ligne droite.
- L'abus des mots nous trompe sur les choses.
- Les choix les plus essentiels que nous ayons à faire dans notre vie, professionnelle ou privée, consistent à décider de ce qui est important.
- Yvetot vaut Constantinople.
- Ce qui ne me fait pas mourir me rend plus fort.

#### **Passé, futur :**

- Qui ne se préoccupe pas de l'avenir se condamne aux soucis immédiats.
- L'avenir vient de loin.
- À la lumière du passé, le présent s'éclaire.
- Regarder l'avenir bouleverse le présent.
- On ne peut pas prédire le futur à partir du passé.
- L'expérience est une lanterne que l'on porte dans le dos et qui éclaire le chemin parcouru.
- Ce qui est passionnant avec le futur ce n'est pas de le prévoir, mais de le construire.
- La seule chose qui ne changera jamais est que tout change toujours tout le temps.
- Quand tu ne sais plus où tu vas, retourne-toi et regarde d'où tu viens.
- On a deux vies et la deuxième commence le jour où l'on se rend compte qu'on n'en a qu'une.

#### **Fixer des objectifs :**

- Il n'y a pas de vent favorable pour celui qui ne sait où aller.
- L'imagination ne sait pas inventer, c'est l'action qui invente.
- Le hasard ne favorise que les esprits bien préparés.
- Ce n'est pas en optimisant la bougie qu'on a inventé l'ampoule.
- Rien ne se crée avec rien. Ce serait comme dire qu'un enfant naît d'un enfant.

- Ce n'est pas parce que les choses sont difficiles que nous n'osons pas, c'est parce que nous n'osons pas qu'elles sont difficiles.
- Le plus grand danger pour la plupart d'entre nous n'est pas que nos objectifs soient trop hauts et que nous les rations, mais qu'ils soient trop bas et que nous les atteignons.
- Un travail infatigable vient à bout de tout.
- Il y a bien des manières de ne pas réussir, mais la plus sûre est de ne jamais prendre de risques.
- Il faut vouloir fortement une chose pour qu'elle arrive, et vouloir est insuffisant, pour obtenir le succès il faut aussi du talent et de la chance.

#### **Définir des actions :**

- Toute action ne nous engage-t-elle pas dans un jeu que nous ne pouvons entièrement contrôler ?
- Les bateaux sont en sécurité dans les ports, mais ils ne sont pas faits pour cela.
- La guerre n'est pas si onéreuse que la servitude.
- Le jour où nous perdons notre envie de nous battre est le jour où nous perdons notre liberté.
- La gestion ne doit pas être confondue avec la stratégie : la première minimise les risques, la seconde les optimise.
- Un bon plan aujourd'hui est préférable à un plan parfait demain.
- Aucun plan ne survit au contact avec l'ennemi.
- Organiser, ce n'est pas mettre de l'ordre, c'est susciter la vie.
- Attention, chercher à combler ses lacunes, c'est investir là où on a échoué.
- La clé universelle de l'excellence n'existe pas, elle doit être remodelée en permanence au gré de l'évolution de l'environnement.
- Il faut attendre que le sucre fonde (on ne peut pas toujours accélérer l'histoire).

#### **Mobiliser les hommes :**

- Rien de ce qui est grand dans ce monde ne s'est fait sans passion.
- Prenez-moi tout, mais laissez-moi les hommes et je recommencerai tout.
- Le pouvoir n'est rien ... seule compte l'influence.
- Il y a des gens qui disent qu'ils peuvent, d'autres qu'ils ne peuvent pas. En général ils ont tous raison.
- Le plus souple en ce monde domine le plus rigide.
- On ne gère pas les gens, il s'agit de les guider. L'objectif est de rendre productifs les points forts et le savoir de chaque individu.
- La logique est statique : une argumentation logique satisfait l'esprit, elle ne l'entraîne pas.
- Il ne suffit pas d'avoir des idées, il faut que les acteurs concernés se les approprient.
- La reconnaissance silencieuse ne sert à personne.

#### **14 - Seul l'apprentissage automatique permettra à de nombreuses entreprises de valoriser leurs mégadonnées**

Depuis quelques décennies les technologies numériques (texte, image, son, vidéo) envahissent et transforment la société. Presque tout est reconfiguré par le numérique : travail, commerce, administration, santé, loisirs, éducation, culture ... Le numérique redistribue les rôles entre les grands acteurs, publics et privés, entre des anciens et des nouveaux, et fait une très grande place aux usagers. Le développement du numérique crée un monde dématérialisé de données qui ouvre des possibilités d'analyse et d'intelligence économique qui semblent illimitées.

Les grandes entreprises peuvent aujourd'hui se doter de tous les moyens nécessaires pour appréhender leur environnement, toutes les technologies pour intégrer, stocker et analyser les données ont atteint leur niveau de maturité, ou presque pour ce qui concerne le traitement des mégadonnées. Le seul gros bémol en matière d'infrastructure se situe au niveau du nombre de matériels et de logiciels nécessaires, encore aujourd'hui il faut recourir à beaucoup de moyens et de fournisseurs différents selon son niveau d'ambition. Par exemple, une base de données relationnelle classique est rapidement insuffisante pour aller au-delà du traitement des données structurées et commencer à traiter des données semi-structurées, il faut pour cela la plupart du temps mettre en œuvre des moyens spécialisés plus performants de type

MPP. Enfin pour traiter des mégadonnées non structurées, on ne peut échapper à l'utilisation de nouveaux outils comme Hadoop par exemple.

Utiliser des outils comme Hadoop, c'est pour beaucoup d'entreprises sortir de leur petit monde d'informatiques propriétaires, où des fournisseurs, des analystes vous cornaquent, pour aborder le monde libre et extrêmement foisonnant de l'open source. Au final, le système d'information à mettre en œuvre pour traiter des mégadonnées est très important, complexe, onéreux et de plus nécessite de faire appel à des spécialistes pointus en informatique, en analyse de données ou mégadonnées, et dans les métiers de l'entreprise. Il faut non seulement les mobiliser, mais arriver à les faire travailler ensemble. Lors de la création de son équipe, il faut particulièrement faire attention à ne pas succomber au mythe du « data scientist », mouton à cinq pattes du monde analytique, qui serait à même de répondre à lui seul, ou plus ou moins, aux besoins des entreprises en matière de fouille de mégadonnées.

Depuis déjà quelques années de nombreuses entreprises ont su tirer de la valeur de leurs données, et pour certaines construire ex nihilo des activités très rentables. Ces pionniers ont montré la voie, mais l'avenir n'est certainement pas dans l'imitation servile de leurs approches. En effet, toutes les entreprises ne pourront pas constituer des équipes pléthoriques d'analystes pour explorer systématiquement leurs données et tenter de comprendre les tendances, de créer des catégories, de détecter des similitudes ou de faire des prédictions à partir du passé ... ; l'avenir est dans l'apprentissage automatique pour mettre en œuvre des milliers de fonctionnalités sur des milliards d'enregistrements.

L'apprentissage automatique est fondé sur un ensemble de techniques pour tirer parti des données et découvrir automatiquement les relations entre les variables d'un système (entrées, sorties et cachées) à partir d'échantillons directs du système. La valorisation des mégadonnées par les entreprises, se fera demain à partir d'un ensemble d'algorithmes d'apprentissage supervisé et non supervisé (corrélation, classification, statistiques, manipulation de document, N-gramme, calcul d'histogramme, traitement du langage naturel...), permettant d'automatiser leur exploration, d'éviter les réductions de dimension qui peuvent conduire à la perte d'information, pour au final fournir un aperçu complet des données, et par exemple repérer les valeurs aberrantes qui sont souvent la source de fructueuses découvertes.

## **15 - De la difficulté, des échecs et de l'importance des projets Big Data/Intelligence Artificielle**

Les facteurs clés de toute réussite sont la volonté, le talent et la chance. La volonté se matérialise par des actes, elle ne se contente pas de paroles, d'espérances, elle effectue des choix, elle s'engage. Le talent est une capacité à créer qui n'est pas innée, mais qui a été acquise, même si les dons facilitent les apprentissages et la créativité. Quant à la chance, elle n'est qu'un hasard, mais on notera que le hasard favorise les esprits bien préparés qui savent saisir les opportunités.

Le philosophe Alain dit « l'imagination ne sait pas inventer, c'est l'action qui invente ». En fait lorsqu'on aborde des domaines nouveaux, les réalisations se font le plus souvent par des avances pas-à-pas dont on ne sait anticiper toutes les étapes. Dans ce cadre, les premiers objectifs concrets que l'on peut se fixer ne peuvent être que limités, mais pas forcément sans risque, car ce n'est pas en reconduisant exclusivement les objectifs, méthodes, moyens éprouvés du passé que l'on peut être novateur.

Toute action est plus ou moins risquée, il ne faut pas craindre d'aller de l'avant ou d'expérimenter des choses nouvelles pour trouver des solutions. Il y a bien des manières de ne pas réussir, mais les plus sûres sont de ne pas essayer et de ne jamais prendre de risques. Il s'agit cependant de ne pas faire table rase de toutes les expériences passées et de chercher à tout réinventer. On se méfiera aussi de l'idée qu'un travail infatigable vient à bout de tout.

Les projets Big Data/IA n'échappent pas aux impératifs ci-dessus alors que plus de la moitié de ces projets ces dernières années ne sont pas des réussites. Au-delà de la trop fréquente mollesse du support du management et des métiers, les retours d'expériences pointent un manque de compétence, de pragmatisme, d'approche agile, d'itération, de montée progressive en complexité au profit d'une planification d'ensemble intenable, d'un pilotage technique, d'une vision industrielle hâtive, d'une gouvernance inappropriée des données ...

En fait, on ne doit lancer un projet Big Data/IA, que pour répondre à un problème identifié et non pas pour de mauvaises raisons du type : faire comme le concurrent, tester un outil, comprendre ce nouveau domaine qui est annoncé comme prometteur, faire plaisir à une partie de ses équipes, répondre aux critiques des métiers qui se plaignent ... Au-delà de l'engagement sur de mauvais objectifs, c'est le manque de compétence qui explique les échecs. La compétence

correspond à la mobilisation dans l'action d'un certain nombre de savoirs, or aujourd'hui en matière de Big Data/IA, le savoir est globalement limité, pas assez de personnes ont vraiment les connaissances, les années d'expérience et maîtrisent l'état de l'art, pour mettre en œuvre le savoir-être, le savoir-faire et les ressources (hommes/matériel) nécessaires pour résoudre des problèmes.

Cependant, on notera que la réussite est le résultat d'une action décrétée comme un succès, en fonction de plusieurs variables jugées à l'aune des objectifs de départ, et qu'elle est donc relative. À ce sujet, le plus grand danger pour la plupart de ces projets ce n'est pas que les objectifs soient trop hauts, et qu'ils soient ratés, mais qu'ils soient trop bas, et que leurs atteintes soient sans intérêt.

Dans le cadre de la transformation numérique de la société en général et des entreprises en particulier, la convergence des technologies du Big Data et de l'IA va apporter les prochaines disruptions qui vont impacter de nombreux métiers en améliorant significativement : l'exploitation de plus de données, les capacités d'apprentissage, la production de connaissances, la prise de décisions, la réalisation de robots. Comme souvent face aux nouveautés et aux évolutions, en France les entreprises ou les particuliers sont prudents, réticents, réfractaires ... et globalement, nous ne faisons pas la course en tête.

Une clé du succès est probablement à être recherchée dans la formation ; là aussi, la transformation numérique des activités de formation pourrait faciliter et renouveler profondément les pédagogies et les parcours.

## **16 – Nous sommes tous impactés par nos propres biais cognitifs**

Notre perception de la réalité qui nous entoure, de l'information que nous collectons ou recevons est conditionnée par de nombreux facteurs comme de façon non limitative ni ordonnée, nos connaissances, notre mémoire, notre langage, nos facultés de raisonnement, nos expériences, notre attention ... ; on appelle biais cognitifs les distorsions engendrées dans la façon dont nous percevons.

Les chercheurs ont identifié un très grand nombre de biais différents liés à l'émotion, à l'envie de croire ce qui nous arrange, de s'appuyer sur des modèles connus, de sur réagir au risque, d'être en excès de confiance, ou à l'inverse de se conformer à l'avis du groupe, d'imiter, de suivre des stéréotypes, des préjugés, de voir la confirmation de ce que l'on a déjà conclu, de surestimer ou sous-estimer la possibilité d'un événement, de s'ancrer dans des croyances, de projeter son espérance dans un résultat donné, de se croire plus malin ...

Le biais n'est pas que dans la perception il peut être aussi dans la collecte, l'enquête, avec des libellés des questions qui conditionnent, orientent, induisent les réponses. Le biais de réponse peut également se produire dans les échelles d'évaluation, et par-dessus tout venir du questionneur ou de l'évaluateur. Dans cette même ligne, on notera que la plupart des raisonnements en économie comportementale sont fondés sur de vrais sophismes : « le consommateur est pleinement informé, strictement rationnel, complètement intéressé ».

Il existe aussi des biais comportementaux lors de la prise de décision. À ce sujet, il convient de se méfier par exemple et de façon non limitative : des raisonnements par analogie qui pour être souvent intéressants, peuvent être piégeants ; des références au passé, car après coup les choses paraissent toujours plus simples ; de l'aversion très répandue à l'incertitude ; de la facilité de l'inaction qui est souvent gagnante à court terme, de la non-prise en compte que tout change tout le temps ...

Toute la chaîne qui va de la donnée, à l'information, à la décision et à l'action est impactée par des biais de toutes sortes. Les biais nous influencent plus fortement encore lorsqu'il faut gérer une quantité d'informations importantes ou que le temps est limité. Cependant, il convient de noter aussi que les biais s'ils peuvent altérer nos jugements, nos décisions, nos actions, apportent souvent des interprétations simplifiées, des actions réflexes qui dans certains contextes, en temps d'urgence par exemple, améliorent nos capacités d'analyses et de réactions.

Si les biais sont inconscients et conduisent à des erreurs (attention, perception, raisonnement, évaluation, interprétation, jugement, décision, comportement), on peut en limiter les impacts par de la formation, de l'entraînement, le recours à des méthodologies, des techniques, de l'organisation ... Par-dessus tout, il faut prendre conscience du problème, développer un esprit critique sur ses habitudes, croire en la possibilité de changer, développer de nouvelles pratiques, mettre en œuvre des approches, des méthodologies adaptées en fonction des situations.



Le combat contre les biais n'est pas qu'individuel et dans les entreprises par exemple il peut s'appuyer aussi sur de l'organisation comme sur les navires de guerre, par exemple, où il existe d'un côté un officier spécifiquement chargé de recueillir et de traiter toutes les informations, et de l'autre des "metteurs en œuvre". Leurs fonctions respectives sont tout à fait exclusives, et cette dissociation est essentielle pour garantir une objectivité.

### **17 - Big Data : à propos des études bidon qui rapportent gros notamment en pharmacie**

Dans les entreprises, des études quantitatives ou qualitatives sont réalisées pour éclairer les dirigeants, pour fonder les décisions sur des faits observés et des théories vérifiées, pour éviter les spéculations ou les croyances invérifiables. La maîtrise des études est donc un enjeu stratégique et les différentes parties prenantes sont tentées de faire chacune les leurs pour orienter les décisions, ce qui aboutit à une multiplication d'études qui se recoupent, mais aboutissent rarement à des réponses partagées.

Durant mon activité de consultant j'ai pu rapidement constater ce phénomène d'accumulation d'études, à tel point qu'au début de chaque mission je me fixais toujours, comme premier l'objectif, d'identifier toutes les études déjà réalisées concernant le domaine que j'avais à travailler. La compilation des études antérieures est toujours un exercice très profitable, mais il convient de faire attention à ne pas se faire enfermer dans leurs visions, car beaucoup sont souvent très instrumentalisées.

Ce phénomène touche toutes les fonctions de l'entreprise, mais le sujet est généralement tabou. Cependant, il y a quelques années à l'occasion d'une session du SEMO (Salon des prestataires des études et de l'analyse marketing), une conférence avait été consacrée à ce sujet et avait proposé une typologie des études instrumentalisées, dont les noms parlent d'eux-mêmes, par exemple : études-faire valoir, études blanchissantes, études-béquilles, études-alibi, études-agit-prop ... études-pavé de chiffres ... études-destination tiroirs ... Attention, ces études ne sont jamais sans effet, même la destination tiroirs, car si beaucoup vous noient plus ou moins volontairement dans des détails illisibles, elles mettent généralement en exergue des arguments de synthèse, pas toujours tirés de leur contenu, qui sont utilisés par leur promoteur pour atteindre leur but.

La manipulation des études est fréquente, elle se pratique dans tous les domaines et notamment dans la recherche. Tous les chercheurs, les étudiants, ne mènent pas que des recherches éthiques, ne font pas toujours preuve d'honnêteté et de véracité. Les mystifications actives ou passives sont nombreuses. Les chercheurs sont plus susceptibles de se livrer intentionnellement à des actes malhonnêtes suivant les enjeux bénéfiques pour eux-mêmes, pour leur laboratoire ... s'ils considèrent que les chances de détection des manipulations sont faibles, si leur engagement éthique n'est pas ferme ou est compromis.

Les trois grandes manipulations sont : la fabrication, la falsification et le plagiat. La fabrication se présente généralement sous la forme de « données de laboratoire » qui sont simplement inventées. La falsification peut prendre plusieurs formes, les données réelles peuvent être « lissées » ou « arrangées » pour se rapprocher plus étroitement du résultat souhaité ou attendu.

Pour être complet, indépendamment des manipulations, il y a aussi l'incompétence qui apporte son lot d'études douteuses. L'incompétence peut entraîner une conception inappropriée, des procédures d'échantillonnage médiocres ou biaisées, des tests statistiques mal utilisés ou mal appliqués, une tenue de registres inadéquate et tout simplement une négligence. Même s'il n'y a aucune intention de tromper, des informations inexactes peuvent également endommager gravement les résultats.

Dans certaines industries, les enjeux financiers peuvent être colossaux, c'est par exemple le cas dans l'industrie pharmaceutique. Nous le voyons bien, en ce moment avec la pandémie du covid-19, les montants astronomiques en dizaine, centaines de millions, qui sont investis dans la recherche et dans les stocks de certains médicaments qui n'ont pourtant pas forcément fait réellement leur preuve, mais qui ont des études favorables, et peut-être les bons réseaux d'experts, de lobbyistes. C'est d'ailleurs dans cette industrie que les scandales d'études trafiquées ont été les plus médiatisés. Personne ne sait avec certitude si le nombre d'études trafiquées est plus élevé en sciences biomédicales qu'en sciences sociales et comportementales, ou s'il est simplement plus facile de détecter la fraude en biomédecine.

Conscient de cette situation, la recherche médicale a inventé les tests en double aveugle, où ni les médecins ni les patients ne savent qui prend quoi en matière de traitement. Mais si la protection contre les manipulations ou les biais cognitifs est ainsi largement améliorée, elle ne peut jamais être totale, et d'un point de vue éthique cette pratique est douteuse

dans les cas où l'on administre un placebo à un malade qui ainsi ne bénéficie pas des meilleures possibilités de soin. D'autre part, il faut compter aussi sur le biais placebo, des changements dans la maladie du patient attribuable au traitement imaginaire constitué par la prise d'un placebo. Les chercheurs pensent que les placebos agissent en réduisant la tension, la détresse, en créant de puissantes prophéties autoréalisatrices. Les individus pensent et se comportent comme si le médicament, en fait un placebo, était efficace.

## **2 - Quelques aspects des big data et de l'intelligence artificielle**

Les Big Data et l'intelligence artificielle, offrent des perspectives intéressantes, qui intéressent les entreprises, les scientifiques en général et les économistes en particulier, en permettant d'augmenter le nombre de variables par observation, le nombre d'observations, la granularité (des indicateurs macro jusqu'aux individus) ; de multiplier les analyses (périodicité, angles, granularité ...) ; d'améliorer les mesures, les suivis, les prévisions de l'activité économique ; de mieux percevoir les effets des politiques et des chocs économiques ; de rendre possibles des suivis fins des expérimentations ; de développer de nouvelles opportunités grâce aux de nouvelles techniques, par exemple le machine learning ...

Cependant, même si ces nouvelles bases de données et techniques statistiques ouvrent de nombreuses opportunités, elles représentent également de nombreux défis pour la société, notamment au niveau des emplois.

## **21 - Big Data : bientôt les algorithmes vous connaîtront mieux que vous-même**

Depuis plusieurs années, le Big Data occupe le devant de la scène médiatique avec son avalanche de données et la difficulté de les traiter. Depuis peu, les propos se tournent vers les algorithmes qui envahissent notre vie quotidienne. Presque tous les domaines utilisent des algorithmes aujourd'hui : ingénierie, sciences physiques, sciences sociales, médecine, commerce (compteurs, listes, palmarès, classements, recommandations, prévisions ...). D'un côté les avantages sont mis en avant (meilleures précisions, nouveaux services ...) et d'un autre certains pointent le spectre de Big Brother, l'exploitation mal intentionnée de données individuelles et l'enfermement possible dans des bulles personnalisées manipulatrices. En réponse à cette situation et pour lutter contre l'emprise des algorithmes sur le web par exemple, il est proposé de développer la « culture algorithmique » du grand public afin qu'il reste le maître de sa place dans le jeu.

Si le mot algorithme vient du nom d'un mathématicien persan Muhammad Ibn Mūsā al-Khwarizmi (780-850 apr. J.-C.), le premier algorithme « inventé » dans l'histoire aurait été l'algorithme d'Euclide (~300 av. J.-C.) pour trouver le plus grand commun diviseur de deux nombres naturels. La définition informelle d'un algorithme : c'est la méthode pas à pas pour résoudre un problème ou effectuer une tâche. Une recette de cuisine par exemple est un algorithme. En informatique, la différence entre un algorithme et un programme tient au fait qu'un algorithme est exprimé avec une notation indépendante de tout langage de programmation. En analyse de données, l'algorithme est donc une spécification exacte de la façon de résoudre un problème de calcul. Il doit spécifier chaque étape de manière complète, afin qu'un opérateur puisse mettre en œuvre chacune d'elle sans autre "compréhension" et fonctionner pour tous les cas possibles du problème. On notera qu'il peut y avoir plusieurs algorithmes différents pour un même problème.

Dans le monde de l'information d'aujourd'hui, l'avenir appartient aux informaticiens qui sont capables de comprendre les problèmes fondamentaux que nous rencontrons, qui maîtrisent les principes et les techniques pour développer des solutions pour les résoudre. Le processus quant à lui est simple : analyser le problème ; concevoir l'algorithme pour résoudre le problème ; implémenter l'algorithme en code et vérifier que l'algorithme fonctionne ; maintenir l'algorithme (faire évoluer le programme pour répondre aux « évolutions » de la problématique). Mine de rien, la première étape comprendre le problème et ses exigences est souvent très difficile, la conception de l'algorithme quant à elle nécessite une stratégie de résolution (rétro-ingénierie, recherche de problème connexe déjà résolu, raffinement pas à pas, casse du problème en plusieurs sous problème ...).

Alors que l'on assiste à la mise en données de notre société en général et des personnes en particulier, qu'elle doit être la culture algorithmique du grand public qui lui permette de ne pas se faire manipuler ? Apprendre ce qu'est un algorithme, à faire des algorithmes, connaître la liste des algorithmes susceptibles d'attenter à ses libertés ? La CNIL quant à elle, réfléchit sur des thèmes tels que « Quelle éthique pour les algorithmes dans l'IA ? » avec la recommandation de

former à l'éthique tous les maillons de la chaîne algorithmique (concepteur, professeurs, citoyens). On imagine sans peine l'impact qu'une telle formation peut avoir sur les grands fauves de la manipulation.

Le RGPD et la loi pour une République numérique ont introduit le droit à la portabilité des données, c'est-à-dire pour un individu de connaître et de récupérer sous un format standardisé, ses données possédées par un tiers. Ce droit s'inscrit dans la philosophie de l'autodétermination informationnelle du grand public, mais va-t-il vraiment nous aider face aux algorithmes trop envahissants ? D'un autre côté, on recommande au grand public d'être prudent dans la diffusion de ses données, on a mis des contraintes réglementaires sur certaines données et notamment sur le traitement des données nominatives, mais l'expérience montre que les algorithmes savent parfaitement déduire de nombreuses caractéristiques spécifiques d'une personne, même si elle ne les a pas partagées. Les commerçants rêvent d'ailleurs de trouver l'algorithme qui saura proposer les produits auxquels leurs clients n'ont pas encore pensé, mais qu'ils vont adorer. Les manipulateurs en tout genre cherchent activement l'algorithme qui vous connaîtra mieux que vous-même.

## 22 - Le Big data et l'Intelligence économique

Maîtriser l'information est plus que jamais un moteur de compétitivité pour l'entreprise, mais aussi pour les États. Dans le cadre de la mondialisation et du passage d'une logique de production à une logique d'innovation, la puissance d'une entreprise ou d'un État est liée à la maîtrise de toute la chaîne d'information, pour trois raisons principales : nous sommes passés d'un marché de masse à une production très diversifiée, au plus proche de la demande ; le marché est devenu hyper concurrentiel ; le progrès technique s'accélère et justifie une meilleure maîtrise de l'information ; le client n'est plus face à l'entreprise : il pénètre dans l'entreprise, hier la clientèle était disponible, en position d'attente, la concurrence étant limitée, voire nulle, aujourd'hui on écoute le client dans une posture proactive.

Face à l'augmentation du volume d'informations lié à ce phénomène d'écoute proactive et au développement des nouveaux modes de communication, notamment des réseaux sociaux, nous disposons d'un outil : l'intelligence économique, à savoir la maîtrise de l'information pour une plus grande efficacité. Il s'agit à la fois d'un mode de pensée – l'information doit être interprétée – et d'un mode d'action – l'information doit être partagée à bon escient. La démarche s'appuie sur la maîtrise de l'information par l'implication et la coopération de chacun dans le système d'information, à partir d'outils existants : la veille technologique, le benchmarking, le renseignement (dont les informations grises ...) et la gestion des connaissances.

La pratique de l'intelligence économique s'appuie sur quatre idées fondamentales : la maîtrise des connaissances qu'il faut identifier et protéger (il est important de hiérarchiser le niveau de confidentialité afin de ne pas étouffer la communication indispensable au développement de l'entreprise) ; la compréhension des opportunités, des risques et des menaces (la corruption, le piratage, la copie, l'espionnage ...) ; la coordination des acteurs et des activités (par la constitution d'un réseau d'experts internes, de fournisseurs, de clients, de centres de recherche ...) ; la stratégie d'influence (il s'agit de la capacité à maîtriser son environnement : relationnel, lobbying ...).

Avec le développement des technologies du Big Data (cloud computing, parallel computing, NoSQL databases, general programming, machine learning, data visualization ...) une masse extraordinaire de données est aujourd'hui exploitable (enregistrements administratifs et transactionnels ; médias sociaux : messages, photos et vidéos ; capteurs collectant des informations : par ex. climat, trafic ... ; données publiques en open data ; images satellites numériques ; signaux GPS mobiles ...) et sans aucun doute l'intelligence économique a un bel avenir devant elle.

Cependant, le développement de l'intelligence économique, qui connaît une nouvelle phase dans sa dimension connaissance avec le big data, ne convainc pas tout le monde et 53% des personnes interrogées attendent un impact positif alors que 39% pensent que cela va causer de nouveaux problèmes. Côté positif : l'exploitation des informations contenues dans ces ensembles de données a déjà mené à des percées majeures dans des domaines allant de la génomique à l'astronomie, à la physique des hautes énergies et au développement de nouvelles industries fondées sur l'information. Côté négatif : les possibilités d'atteinte à la vie privée sont décuplées et par une sélection judicieuse des données il est possible de construire des théories abracadabrantesques et nuisibles, parfaitement illustrées et plausibles.

En conclusion, l'intelligence économique se développe, car les décideurs ont compris que la capacité de prendre des données - de les comprendre, de les traiter, d'en extraire de la valeur, de les visualiser, de les communiquer – est une compétence extrêmement importante quel que soit le domaine aujourd'hui, que l'information a une immense valeur qu'il faut bien l'utiliser, la protéger, et qu'ils ont lu Sun Tzu qui a dit « Connaissez l'ennemi et connaissez-vous-même ; en cent batailles vous ne courrez jamais aucun danger ».

### **23 - Big data et Prospective : « Il n'y a pas de vent favorable pour celui qui ne sait où aller » Sénèque**

Les spécialistes de l'analyse des données n'ont pas besoin de mainframe décisionnel hors de prix, pour constater que le monde change. Face aux évolutions, les responsables ont le choix entre quatre attitudes : l'autruche (une solution confortable à court terme, souvent choisie par les hommes politiques), le pompier (réactif : attendre que le feu soit déclaré pour le combattre), l'assureur (préactif : se préparer aux changements prévisibles), le conspirateur (proactif : provoquer les changements souhaités).

Il n'y a pas de statistiques du futur, l'histoire ne se répète pas, mais les comportements se reproduisent, et les entreprises ont toujours intérêt à utiliser des données ou des mégadonnées du passé pour établir un diagnostic complet de leur position face à leur environnement, face aux meilleurs (benchmarking), intérêt à se poser les « bonnes questions », à identifier les variables clés, à analyser les stratégies des acteurs, à balayer le champ des possibles, à réduire l'incertitude, à évaluer les choix & les options stratégiques.

L'action sans but n'ayant pas de sens, la définition d'une stratégie s'impose. La stratégie parle de clairvoyance et d'innovation, un problème bien posé est presque résolu : le problème de la stratégie, c'est la définition du problème. Il n'y a pas de bonne réponse à une mauvaise question. Mais attention, certains problèmes sont dans la lumière d'autres dans l'ombre, on se méfiera aussi de la tyrannie des idées dominantes.

Regarder l'avenir bouleverse le présent et l'anticipation invite à l'action. Les concepts de prospective, de stratégie et de planification sont intimement liés. La prospective répond à la question : Que peut-il advenir ? Mais attention toute sorte de prédiction est une imposture. Il faut distinguer les scénarios d'environnement général, des stratégies d'acteurs. Identifier l'éventail des futurs possibles par des scénarios, mettre en lumière des paramètres de bifurcations qui sont des variables clés de l'analyse prospective. Un scénario est un ensemble formé par la description d'une situation future et du cheminement des événements qui permettent de passer de la situation origine à la situation future.

Un scénario n'est pas la réalité future, mais un moyen de se la représenter en vue d'éclairer l'action présente à la lumière des futurs possibles et souhaitables : analyse qualitative et quantitative des tendances lourdes, rétrospectives, jeux d'acteurs, mise en évidence des germes de changement, des tensions et des conflits, construction de scénarios cohérents et complets. De deux modèles, le meilleur sera toujours celui qui pour une approximation donnée représentera le plus simplement les données de l'observation. Attention à ne pas confondre compliqué et complexe, ni simplicité et simplisme. Après les explorations du passé, du présent et du futur, il convient de répondre aux questions : Que puis-je faire ? Que vais-je faire ? Comment le faire ?

En matière de prospective, le recours à un consultant est hautement recommandé. Le consultant ne doit pas être un expert du domaine (rejet par le groupe et frein à l'appropriation), il doit apporter l'ouverture sur l'extérieur, l'impertinence de l'œil neuf et surtout la méthode, car sans elle, pas de langage commun, d'échange, de cohérence, ni de structuration des idées. Dans une étude, quelle que soit sa qualité, le rapport final compte moins que le processus qui y conduit. Mieux vaut une imperfection opérationnelle qu'une perfection qui ne l'est pas.

Le management, c'est l'art de mettre l'organisation au service de la stratégie. Point commun à toutes les méthodes de management : motiver les hommes. Une analyse stratégique partagée peut produire un engagement collectif. Motivation interne et stratégie externe sont donc deux objectifs indissociables qui ne peuvent être atteints séparément. C'est par l'appropriation que passe la réussite du projet. Cependant, la mobilisation collective ne peut porter directement sur les choix stratégiques, par nature confidentiels. En effet pour des raisons de confidentialité et de responsabilité, les choix stratégiques sont du ressort d'un nombre limité de personnes.

Pour aller plus loin, vous pouvez utilement consulter le site de France Stratégie ci-dessous. Cet organisme de réflexion, d'expertise et de concertation, rattaché au Premier ministre, a pour mission d'évaluer les politiques publiques, de

débattre avec une pluralité de parties prenantes, d'anticiper les défis et les mutations et de proposer de nouvelles solutions.

#### **24 - Big data et Lobbying : « Le pouvoir n'est rien ... seule compte l'influence » Hérodote**

Aujourd'hui grâce aux Big data, sans mainframe décisionnel hors de prix, il est possible dans une démarche d'intelligence économique d'appréhender finement, le contexte économique, social & politique d'une organisation. Dans le prolongement de cette approche de veille, les organisations sont naturellement amenées à développer des stratégies de défense de leurs intérêts (communications, constitution de réseaux, de groupes d'intérêts ...).

Les groupes d'intérêts sont nécessaires aux systèmes politiques et ce de façon complémentaire aux partis. Les groupes d'intérêts ont une double fonction : informer les agents du système politique des exigences ou du soutien des agents économiques ou sociaux, et communiquer à ces derniers les possibilités et les volontés des agents du système politique. Il est dans la logique de la démocratie gouvernante de favoriser la formation des groupes d'intérêts.

Pour une entreprise, le lobbying a pour objectif de maîtriser son environnement réglementaire. Il coordonne toutes les actions vers le monde politique et en particulier vers les décideurs : gestion des relations, réalisation de campagnes d'information et de publicité axées sur des positions politiques de l'entreprise, réalisation de dossier de presse et d'argumentation, liaison avec l'univers des études et des sondages politiques. Le lobbying peut être vu comme une branche du marketing (influence des parties prenantes : marketing des idées, communication d'influence, contrôle de la pensée par le contrôle du langage ...). On notera que la plupart du temps, dans les entreprises, le lobbying est géré directement au plus haut niveau (PDG, DG).

Aux USA, le lobbying est organisé, 21 500 lobbyistes sont officiellement enregistrés. Au niveau du parlement européen, on en dénombre 15 000. En France, le lobbying a toujours été considéré comme une pratique grise, même si on constate des interventions de syndicats professionnels, d'organisations patronales, d'associations et qu'il existe ~30 cabinets spécialisés, et des structures spéciales dans les grandes entreprises appelées, affaires publiques. Cependant depuis la loi Sapin II de décembre 2016 relative à la transparence, les acteurs sociaux économiques qui exercent des actions d'influence sur les pouvoirs publics (ministres, parlementaires, etc.) doivent se faire enregistrer sur un répertoire public, et en janvier 2018, 816 représentants de groupements d'intérêts sont enregistrés.

Keynes voyait l'économie comme une suite de rapports de force. Le pouvoir des salariés, qui est capable de bloquer toute baisse nominale des salaires, même en période de chômage. Le pouvoir des entrepreneurs, qui détermine l'investissement, l'emploi effectif, les prix et les revenus distribués. Ce pouvoir considérable des entreprises n'est pas autonome, il doit prendre en compte les niveaux de salaires définis par les conventions collectives et la politique de l'État. Le pouvoir des banques, pour les keynésiens, il est réduit, les banques ne pouvant ajuster la masse monétaire que pour permettre aux entreprises de faire face à l'accroissement des salaires et de leurs autres charges. De plus, elles sont étroitement contrôlées par l'État, qui a le monopole de battre la monnaie et qui contrôle la création monétaire.

Dans ce contexte, le lobbying a de beaux jours devant lui, d'autant plus si on pense comme Coluche que « la moitié des hommes politiques sont des bons à rien. Les autres étant prêts à tout ». Dans ces approches l'information qui alimente le sujet et légitime les positions (études, expertises ...), et les communications politiques, dont l'objectif est d'exposer, médiatiser le point de vue du groupe d'intérêts (position papers) sont une des clés d'influence. Mais l'influence repose aussi sur le recours à des récompenses en faveur de celui qui se soumet à l'influence (persuasion), sur l'utilisation de l'information (manipulation), sur des rémunérations matérielles ou symboliques de celui qui se soumet.

Enfin pour ceux que cela intéresse, il existe des écoles pour apprendre le lobbying, pour apprendre à : accélérer ou retarder l'inscription d'un problème dans l'agenda public ; influencer, légitimer le choix de solution ; faire pression pour l'adoption d'une décision ou bloquer la décision ; retarder-bloquer la mise en œuvre, limiter l'impact de la décision... On n'oubliera pas cependant que les lobbyistes sont organisés en réseau, et que suite à la déculottée subie par les grands partis traditionnels en 2017, des légions de politiques, conseillers et hommes de cabinets ont rejoint les grands groupes privés pour vendre leur carnet d'adresses et leur capacité d'influence.

## **25 - L'intelligence artificielle nous conduit-elle vers une prolétarianisation de la société ?**

L'intelligence artificielle, c'est la science des programmes intelligents. Depuis des décennies des programmes intelligents sont développés pour traités des cas « rudimentaires » régis par des règles, par exemple le grand public a pu découvrir les jeux d'échecs électroniques : un programme gagne contre un humain en 1967, le premier jeu est commercialisé ~50 \$ en 1977, l'ordinateur Deep Blue bat le champion du monde d'échec en 1997.

Cependant, le développement de programmes intelligents a été limité parce qu'il est extrêmement complexe de programmer à la main des algorithmes, notamment dès que le problème est sophistiqué, par exemple établir des diagnostics médicaux ou faire des prévisions de pannes, des prévisions de cours de bourse, parce qu'il s'agit dans ces cas de prendre en compte de très nombreux jeux de données et leurs relations le plus souvent non linéaires. Dans d'autres cas comme la traduction de langues, la reconnaissance d'objets dans des images par exemple, il est très difficile d'élaborer des règles.

Depuis le début les chercheurs ont eu conscience de leurs limites en matière de programmation des algorithmes, ils ont alors eu l'idée de construire des programmes qui sachent programmer à leur place. Pour cela la première capacité qu'un tel programme devait avoir, était celle d'apprendre. C'est ainsi qu'est née l'une des premières machines apprenantes, le Perceptron en 1950 (un réseau de neurones artificiels). Depuis le Machine Learning s'est beaucoup développé et il y existe près d'une vingtaine méthodes d'apprentissage différentes dont le Deep Learning (apprentissage profond) qui a permis dans les années 2000 des progrès importants et rapides dans les domaines de la reconnaissance faciale, de la reconnaissance vocale, de la vision par ordinateur, du traitement automatisé du langage.

Par exemple, l'apprentissage profond a permis : au grand public de disposer de l'assistant personnel virtuel Siri d'Apple, de l'application Google Street View ; aux entreprises de disposer de systèmes de reconnaissance d'écriture manuelle, d'analyse des sentiments pour évaluer les clients et à Facebook, Tweeter ou Yahoo de proposer différents services ; aux médecins de détecter des tumeurs dans des images médicales. Dans le domaine de la recherche scientifique, l'apprentissage profond a des applications en biologie (analyse du génome, structure des protéines), en chimie (prédiction de réactions chimiques), en physique (détection de particules exotiques). Pour ceux qui veulent mettre le nez sous le capot, les principaux outils d'apprentissage profond sont : Pylearn2, Theano, Caffe, Torch, Cuda-convnet, Deeplearning4j.

Le domaine de l'intelligence artificielle est en plein développement, les approches, les méthodes sont nombreuses, variées, en constante évolution, elles doivent le plus souvent être combinées entre elles, chacune ayant ses avantages et ses inconvénients en général ou vis-à-vis d'un problème particulier, l'apprentissage profond par exemple en général nécessite beaucoup de données et de temps. En conséquence ce domaine passionnant de l'intelligence artificielle nécessite et nécessitera dans les vingt ans à venir, de très nombreux spécialistes de l'analyse de données, ce qui créera beaucoup d'emplois qualifiés bien payés. L'avenir pour ces spécialistes est particulièrement radieux.

Cependant, l'Intelligence artificielle qui débute est une des phases de la révolution numérique, et les experts en économie sont pessimistes au sujet de son impact sur l'emploi, à l'instar de Jean Tirole, Prix Nobel d'Économie qui a dit récemment « Il faut se préparer à la révolution numérique qui arrive, sans quoi le choc pourrait être très violent. Ce qui est clair, c'est que les destructions d'emplois vont s'accélérer, même si personne n'a vraiment idée à quelle vitesse exactement ». En parallèle de ces destructions, de nouveaux emplois seront créés, mais la question est de savoir s'ils compenseront les pertes et de quels niveaux de qualification ils seront. Le risque, largement évoqué par de nombreux experts, étant que la société se mette de plus en plus à offrir surtout des emplois peu qualifiés, voire très peu, à une population de mieux en mieux formée.

### **3 - De la maîtrise des données dans la société d'aujourd'hui**

Les données permettent de refléter la réalité, et leur production, leur maîtrise, est un enjeu majeur de nos sociétés, notamment en matière d'intelligence économique, de prospective ...

Dans une démocratie, le gouvernement est le gardien des renseignements qui intéressent la population, et les citoyens devraient avoir accès à la plus grande partie possible des données, le plus rapidement possible. L'information est une

base nécessaire de la démocratie dans un esprit de connaissance civique des affaires. L'information, sans être une forme précise de participation, est un préalable nécessaire à toute participation.

Cependant, il n'y a pas de construction ni de développement de la démocratie sans mise à niveau des mentalités et des comportements, sans civisme et sans participation citoyenne. La démocratie suppose un niveau de conscience individuelle et collective qui n'est ni inné ni donné, mais qui s'acquiert par l'éducation. On ne naît pas citoyen, on apprend à le devenir.

En démocratie, l'accès à l'information est essentiel et les technologies de l'information ont un rôle important, comme l'a souligné Al Gore il y a de cela plus de vingt ans « Je vois un nouvel âge athénien de la démocratie forgée dans les forums créés par l'infrastructure globale de l'information ». L'Internet est un instrument permettant de stimuler et d'enrichir la discussion entre citoyens.

### **31 - De l'analyse de données aux choix politiques**

Les théories et discours économiques sont de plus en plus fondés sur des analyses de données, comme le montre actuellement le dernier best-seller international « Le capital au XXIe siècle » de l'économiste français Thomas Piketty. Cependant, que ce soit pour des démonstrations économiques ou pour des approches financières et marketing, les résultats dépendent d'abord des données disponibles et de leur qualité.

Partant d'une base de données compilant deux siècles d'événements, l'économiste montre que le taux de rendement du capital dépasse le taux de croissance de l'économie, ce qui concrètement signifie que l'enrichissement de la société profite plus aux porteurs de capitaux, et que donc les inégalités s'accroissent inexorablement. Fort de cette analyse et d'autres à découvrir dans le livre, Piketty, en bon français normal, propose de créer un impôt mondial sur le capital, afin d'arrêter l'explosion des inégalités.

Sans entrer dans le détail des polémiques que ce livre a généré, des erreurs relevées par certains experts, il convient de noter que les données utilisées dans les calculs sur les inégalités de revenus sont celles des revenus avant impôts et avant transferts sociaux. Il en ressort que Piketty ne fait qu'enfoncer une porte ouverte sur la nécessité, pour une société démocratique, de disposer d'un système de régulation des inégalités (impôts et transferts).

Partant de données plus ou moins bonnes et en tout cas incomplètes, la démonstration débouche sur des conclusions mal fondées. Les recommandations dès lors ont peu de chance d'être les bonnes, vu qu'elles cherchent à résoudre, via de nouveaux impôts, un problème qui n'existe pas, à savoir l'accroissement inexorable des inégalités.

Comme en France actuellement, le monde a besoin d'une croissance solide, de forts investissements dans les hommes, dans leur formation, mais surtout pas de nouvelles taxes et de nouveaux fonctionnaires pour les collecter. Pour mémoire, le coût des fonctionnaires en France représente 12% du PIB contre seulement 7% en Allemagne.

### **32 - Les données sont le carburant indispensable de la dynamique démocratique**

Dèmos = le peuple, kratos = le pouvoir, la démocratie, c'est le régime ou le peuple est au pouvoir. Cela ne veut pas dire qu'il gouverne ou qu'il vote les lois, mais que rien ne se fait sans son accord ni hors de son contrôle ; mais dans un grand pays encore faut-il que le peuple soit informé. De la même manière que nul n'est censé ignorer la loi (par exemple les 3600 pages et plus de 10 000 articles du Code du travail), tout se passe comme si le peuple était informé de tout, et pouvait tout contrôler.

Pris dans le rythme de ses activités, le citoyen le plus souvent ne cherche pas à comprendre par lui-même, et se fie à la presse, à des organismes plus ou moins indépendants, à des associations de toute nature, aux hommes politiques qu'il estime et se contente d'adhérer à des éléments de langage, des opinions en prêt à consommer. Sans compréhension, nous avons une démagogie, pas une démocratie.

Bien sûr, des organismes comme la cour des comptes ou des think tanks, des experts, des journalistes ... révèlent des faits sur notre environnement, mais outre que les faits ne déplacent pas les gens, ils sont noyés parmi l'avalanche des

communications quotidiennes plus ou moins téléguidées par des milliers de lobbyistes, d'affidés de toute nature et par-dessus tout d'hommes politiques champions toutes catégories de la désinformation.

La désinformation consiste à volontairement fausser la compréhension d'une situation. Désinformer, c'est fournir de fausses indications afin d'orienter dans une mauvaise direction. Il s'agit le plus souvent de présenter de façon tendancieuse un fait avéré, de fournir une information biaisée, de brouiller les pistes, de noyer l'information vraie dans une avalanche de considérations et de détails. Les expressions politiquement correctes sont une forme de désinformation, la langue de bois une autre forme.

Pour s'informer aujourd'hui il y a internet et c'est sans conteste un grand progrès, mais chacun sait que l'on trouve tout et son contraire sur internet. Wikipedia est probablement précis à 99,9% (?), mais il faut absolument s'en méfier dès que le sujet à une dimension politique. Il devrait y avoir les sites officiels des administrations centrales, locales, etc., mais force est de constater que l'Open Data avance assez lentement en France.

Il faut cependant noter que le gouvernement vient de nommer (septembre 2014) un administrateur général des données (chief data officer), autorisé à connaître les données détenues par l'administration de l'État et ses opérateurs. Cet administrateur a pour mission d'organiser une meilleure circulation des données dans l'économie comme au sein de l'administration. Cependant, la lecture détaillée de la mission montre que l'objectif principal vise à mieux partager les données entre les administrations, plus qu'à les mettre rapidement à disposition du grand public.

Nous pouvons compter sur les combats d'arrière-garde des administrations pour freiner le partage des données entre elles, et le réflexe de défense collective des fonctionnaires pour freiner l'ouverture au public, sans compter les tentatives de recul comme au ministère de la Santé.

### **33 - L'open data au service des citoyens pour contrôler les politiques et l'administration**

Les élus et les administrations publiques sont censés travailler au nom des citoyens, et ces derniers les contrôler, les soutenir ou pour les politiques les sanctionner éventuellement via les élections. Tout cela serait parfait si le citoyen était bien informé ; or ce n'est pas le cas. Il faut reconnaître que cela est difficile, mais je ne suis pas sûr de la constante bonne volonté de ceux qui ont les informations.

Quant aux personnes susceptibles de nous éclairer, les journalistes sont trop souvent uniquement le nez dans le guidon de l'actualité, et les analystes toujours très politisés font plus de la communication pour leur chapelle que de l'information. Les données communiquées sont sélectionnées pour appuyer les messages, et de communication en communication les angles sont différents, les données aussi, elles ne peuvent pas ou très peu être recoupées, ni comparées dans le temps.

D'autre part si les grandes notions qui intéressent le citoyen ont des définitions bien arrêtées, les chiffres les concernant sont la plupart du temps des regroupements de comptes comptables qui sont bien difficiles à tous cerner. Finalement, seuls les spécialistes peuvent s'y retrouver. C'est le cas par exemple des prestations sociales qui représentaient, en 2013, 33 % du produit intérieur brut. Elles comprennent, les prestations sociales qui sont constituées des versements en espèces (pensions de retraite, allocations familiales, RMI, allocations chômage, etc.) ou en nature (remboursements de soins ou de médicaments, etc.) et compte satellite de la protection sociale qui introduit la notion de « prestation de protection sociale », qui comprend en sus les prestations de services sociaux : accès gratuit ou à tarif réduit à des services comme l'hospitalisation publique, l'hébergement des personnes handicapées, la réduction de tarif des transports en commun pour les familles nombreuses ou les personnes âgées.

Les chiffres de toutes les dépenses sont publiés annuellement, ils permettent à tout un chacun un suivi par grandes catégories par exemple la répartition, des prestations d'un montant global de 597, 57 milliards d'euros en 2009, était : vieillesse-survie 45,5%, santé 35 %, maternité-famille 9 %, emploi 6 %, logement 2,7%, exclusion sociale 1,8 %. Mais ne permet pas de suivre à qu'elle logique ils correspondent. En effet, les prestations sociales peuvent répondre à trois logiques : une logique d'assurance sociale financée par des cotisations sur les salaires (comme dans une assurance privée) et sont donc réservées à ceux qui cotisent ; une logique d'assistance, versée sous condition de ressources, mais non de



cotisations préalables ; une logique de protection universelle, accordée sans condition de cotisations ni de ressources, mais sont les mêmes pour tous.

Il faut non seulement des chiffres qui soient compréhensibles, mais qui soient aussi internationalement comparables, pour nous situer, par exemple la France c'est 1% de la population mondiale, 3,7% de la production mondiale, et 15% des transferts sociaux.

Il faut reconnaître que l'administration met actuellement beaucoup de données à disposition, mais il reste beaucoup à faire avant que le citoyen puisse s'y retrouver facilement dans le dédale des dépenses, des déficits et des dettes qu'elle génère. Seules des données et une implication citoyenne très forte, peut nous permettre d'empêcher l'administration et les politiques de continuer à par exemple : augmenter de façon continue les effectifs de la fonction publique, voter de nouveaux impôts (4,5 Milliards supplémentaires pour les entreprises en 2015), imposer le kafkaïen compte pénibilité, continuer à creuser notre dette qui atteindra 100% du PIB en 2015.

### **34 - Les données quantitatives sont des arguments de poids**

Un argument c'est une proposition qui tend à justifier une idée, ce n'est pas une preuve, mais peut en tenir lieu. Une donnée peut à elle seule caractériser une situation et ne nécessiter aucun autre argument pour faire passer une idée. C'est pourquoi les données chiffrées sont très appréciées dans notre société hyper médiatisée où il faut savoir délivrer un message en quelques secondes, et exprimer synthétiquement un point de vue sur des réalités complexes.

Quelques exemples :

En 2014, la croissance mondiale devrait être de 3,4%, celle de la zone euro 1,2, celle de la France ne pas dépasser 0,7%.

On compte 5,3 millions de demandeurs d'emploi en France toutes catégories confondues (DOM compris), au total la population active est de l'ordre de 30 millions (chômeurs compris).

Au sortir de l'enseignement primaire, 15% des élèves sont en grande difficulté, 60 000 jeunes sortent chaque année du système éducatif sans aucune qualification, seulement 40% des diplômés ont un niveau égal ou supérieur au Bac ou au Brevet professionnel ou équivalent.

La dette publique qui représentait 35% du PIB en 1991, qui est passée à 67% en 2006, est actuellement à 93% et continue à se creuser un peu plus tous les jours (en fonction des rêves de gagner plus en travaillant moins, de réduire les déficits en augmentant les dépenses, et de tout changer en Europe, sauf en France).

Caractériser globalement une situation c'est bien, mais la comprendre en détail c'est mieux. Or pour cela l'usage de données quantitatives n'est pas toujours facile. Certaines données administratives ne mesurent pas aussi exactement ce que l'on souhaite (exemple les données sur l'emploi qui ne disent rien sur les embauches et les ruptures) ; de plus beaucoup sont publiées souvent tardivement, pas assez fréquemment, et les communications dont elles font l'objet ne portent souvent que sur ce qui arrange le communicant ; ou elles utilisent un indicateur-écran (exemple : l'indicateur des prix qui est une moyenne et n'est l'indicateur de personne) ; elles peuvent aussi changer au fil du temps (exemple les dépenses publiques qui deviennent dépenses publiques primaires qui excluent le remboursement de la dette, qui ne fait que croître) ; ...etc.

La maîtrise de l'information est source de pouvoir, c'est pourquoi il est stratégique d'obtenir de nos gouvernements et surtout de l'administration (des fonctionnaires) : la production de données comparables à celles des autres pays ; des facilités pour que des universitaires, des experts du privé, des journalistes puissent fonder leurs recherches sur des bases de données de qualité ; un accès facile aux études produites et aux principales données pour le plus grand nombre.

Le mouvement international de l'Open Data va dans ce sens. Le Parlement européen a émis une directive en 2003 sur la réutilisation des informations du secteur public, qui a été transposée en France en 2005. En 2011 il a été créé au niveau national une mission pour promouvoir l'ouverture et le partage des données publiques (cf. [www.etalab.gouv.fr](http://www.etalab.gouv.fr)), et en parallèle quelques communes se sont lancées à publier des chiffres (Bordeaux par exemple), mais globalement nous ne

voyons pas encore beaucoup de résultats probants. Il faudrait sans doute un second souffle. C'est un enjeu majeur de démocratie.

### **35 - Big Data, statistiques et infographie : les chiffres disent-ils toujours la vérité ?**

Une opinion largement répandue veut que les données parlent d'elles-mêmes, que leur simple présentation suffise et nous évite ainsi de longs discours. Ceci est totalement faux, les chiffres ne parlent vraiment qu'après un long interrogatoire pendant le lequel on leur fait décliner leur identité, où on vérifie tous les détails, où l'on recoupe les informations, où l'on traque les contradictions, où l'on compare avec des résultats obtenus en d'autres occasions, etc. ...

Un des grands classiques de données qui peuvent induire en erreur est la présentation de deux séries de données qui à première vue semblent porteuses de sens, mais qui sont biaisées, car en fait dépendantes d'une variable intermédiaire non apparente qui est la clé explicative principale. Par exemple, le niveau de revenu induit des facilités qui vont impacter des résultats indirectement, par exemple en matière d'instruction, d'hygiène, d'implantations géographiques ...

Compter et mesurer sont historiquement devenues nécessaires aux hommes pour maîtriser leurs activités, dès les premières civilisations. Il existe des preuves archéologiques de compteurs et de systèmes de comptage dans certaines des plus anciennes cultures humaines (cf. les os d'Ishango, datant de 18000 av. J.-C.). Dans l'Égypte des pharaons, pour garder une trace de la quantité de terre allouée à un agriculteur, de la quantité de récolte, ainsi que de tout impôt ou taxe à payer, un système de mesures et de comptage avait été développé.

Si les recensements et les comptes de gouvernement remontent à l'antiquité, la statistique qui date de 1800 est venue avec l'ambition de transformer l'artisanat de l'état en une science. La statistique se veut au début une science sociale qui permet d'identifier rapidement les causes de problèmes sociaux tels que le crime, le suicide, l'épidémie ... Mais en fait, les gouvernements n'ont jamais vraiment écouté, car le message statistique n'est souvent pas aussi simple que l'espèrent les statisticiens, les graphiques ne montrant pas de lien simple entre instruction et crime, par exemple.

Les statisticiens estiment que les statistiques sont une forme d'information, que les chiffres s'expliquent d'eux-mêmes et sont facilement accessibles à quiconque. Mais les statistiques ne sont que très rarement vraiment transparentes et pas seulement pour le grand public, même si certaines statistiques sont assez lisibles pour décrire les résultats de la mesure de variables uniques et pour construire et évaluer des échelles multi-items. Ces statistiques comprennent des distributions de fréquences, des mesures de la tendance, de la variation centrale ...

D'autres statistiques sont utiles pour atteindre une validité causale en nous aidant à décrire l'association entre variables et à prendre en compte d'autres variables ou à en tenir compte. (Statistiques bivariées et multivariées). La tabulation croisée est la technique permettant de mesurer l'association et de contrôler d'autres variables. Il est aussi possible d'estimer le degré de confiance que l'on peut accorder à la généralisation d'un échantillon à la population à partir de laquelle l'échantillon a été sélectionné. Mais, arrivé à ce niveau d'information on commence à perdre beaucoup d'écoute et l'idée que l'on peut faire dire ce que l'on veut à des statistiques est très répandue. Pour Mark Twain par exemple, il y a trois sortes de mensonges : les mensonges, les sacrés mensonges et les statistiques.

Avec les Big Data une autre difficulté se présente, celle de l'avalanche de chiffres qui sont à communiquer. Mais, heureusement nous avons l'infographie pour nous aider qui permet de présenter des informations complexes rapidement et clairement, avec des éléments visuels (couleurs, graphiques, icônes, panneaux, cartes ...), des éléments de contenu (texte, faits, statistiques, délais, références ...), des faits et la conclusion pour transmettre le message ou l'histoire globale. L'infographie attire l'œil, l'attention des gens, d'autant plus que 65% de la population apprend plus facilement visuellement.

Cependant, comme nous sommes dans un monde de post-vérité, où le politiquement correct assume de mentir pour le bien de tous, comme les régimes totalitaires, il convient de regarder n'importe quelles données chiffrées avec beaucoup de prudence, y compris et peut être même surtout celles présentées par les gouvernements et la presse. Ici on notera qu'au jeu de la vérité, internet est à la fois une aide et une source d'erreurs.

### **4 - De la nécessité de disposer d'hommes compétents pour tirer parti des mégadonnées (big data)**

La numérisation de la société provoque de grands chambardements (penser à la numérisation de la photo, numérisation des textes, des livres, de la presse, de la musique, du cinéma, des télécommunications, des mesures ; penser à la

multiplication des capteurs, à l'impact sur les entreprises, les clients mieux informés, les citoyens, les gouvernements ... ), la numérisation génère donc beaucoup de données que l'on peut analyser (big data) et en outre partager facilement (open data)

Le big data pénètre désormais de nombreux secteurs d'activité. Des données bien exploitées dotent les entreprises d'un avantage concurrentiel difficile à battre. Le temps des affaires menées à l'instinct est révolu, l'ère de la décision par les données a sonné. La plupart des décideurs dans le monde ont adopté les outils analytiques pour les aider à prendre des décisions. À noter que pour les pionniers le big data ne booste pas l'avantage concurrentiel, il est au cœur même de leur modèle d'affaires.

Les moyens matériels ou logiciels nécessaires ont vu le jour et continuent à être développés pour lancer, accompagner cette vague qui ne fait que commencer. Tout va bien, mais un domaine mérite cependant beaucoup d'attention, c'est celui de la mobilisation des compétences indispensables pour profiter des opportunités offertes par le big data.

#### **41 - Propos sur les conseils en Big Data (Méga Données)**

Face à un nouveau sujet, les entreprises ont généralement le réflexe de chercher à se faire conseiller sur la manière de l'aborder. Mais que peut-on attendre d'un conseil et qu'est-ce que le conseil ? Un conseil est un spécialiste extérieur à une organisation à qui l'on fait appel, afin d'obtenir un avis au sujet d'une question ou de l'aide pour résoudre un problème précis.

La prestation d'un conseil suit un processus dont il est difficile de caractériser l'essence en une phrase. Le processus de conseil est indépendant du contenu, c'est-à-dire que les conseils suivent les mêmes étapes, quel que soit leur domaine d'expertise. Quoique le processus de conseil soit indépendant du contenu, c'est une activité très dépendante du client. Il s'agit donc pour le conseil de mettre en œuvre ses compétences pour résoudre le problème de son client. Le conseil ne s'accommode pas de solutions standardisées. Le conseil doit être adapté spécifiquement aux besoins particuliers du client. Le conseil doit avoir des compétences c'est-à-dire une combinaison de connaissances, d'expériences et de comportements qui le qualifie pour conseiller sur un sujet.

Le processus du conseil comprend différentes dimensions, toutes liées entre elles. Il s'agit d'influencer sans pouvoir hiérarchique direct ; d'apporter son expertise et la faire appliquer ; d'établir une relation de confiance avec le client ; de s'impliquer tout en restant impartial ; de travailler sous pression et dans le conflit ; de fournir un cadre pour la recherche de solution ; et finalement d'avoir un impact sur l'entreprise.

Un conseil s'efforce toujours de répondre aux besoins spécifiques de son client, mais son rôle évolue avec les circonstances. Au cours de la même mission peut-être même dans le même jour il est amené à jouer des rôles bien différents : médecin, facilitateur, ambassadeur, gourou, catalyseur, arbitre, animateur, détective, chef d'orchestre, confident, entremetteur et malheureusement souvent bouc émissaire.

Cette liste n'est bien sûr pas exhaustive. On peut trouver sans doute d'autres rôles à y ajouter. Cette liste est une simple description des rôles les plus courants joués par les conseils. Suivant la mission, le conseil joue tous ces rôles à la fois ou seulement quelques-uns d'entre eux. En améliorant sa connaissance de son métier, le conseil peut se rendre compte qu'il devient plus facile d'évaluer les situations et de passer d'un rôle à l'autre. De plus, il est appelé à jouer des rôles plus variés au fur et à mesure que ses responsabilités, son expérience et ses rapports avec ses clients augmentent.

Il est important de reconnaître que des contextes différents appellent des traitements différents. L'efficacité du conseil dépend souvent de son habilité à évaluer les situations et à choisir les rôles qui sont les mieux adaptés. Pour évaluer et choisir un conseil, il faut considérer les compétences, les connaissances et surtout les références spécifiques relatives au problème à l'étude. Le plus important est d'engager un conseil qui peut apporter une valeur ajoutée à l'entreprise. Soyons clair, aujourd'hui en France pour le thème du Big Data (Méga Données) les personnes à même de conseiller sont assez rares.

#### **42 - Mégadonnées : la construction d'algorithme nécessite toujours du jus de cervelle**

À l'origine, un algorithme désignait une suite de calculs pour effectuer une opération mathématique complexe. Aujourd'hui, on s'en sert pour caractériser un processus, c'est-à-dire un enchaînement de tâches, pour obtenir un

résultat. Un algorithme se définit donc par sa finalité. Ainsi sur internet, nous utilisons ou nous subissons de nombreux algorithmes : de recherche, de recommandation, de suggestion, de pistage ...

Un algorithme de recommandation de livres par exemple, va manier des notions telles que la similarité (ouvrages sur le même thème ou d'un genre similaire), la complémentarité, la diversité et tenir compte des caractéristiques de la personne à traiter (âge, localisation, achats précédents, notations d'autres ouvrages). On remarquera que si le système manipule beaucoup de données individuelles ou non (c'est le côté mégadonnées), il est fondamentalement construit sur des classifications humaines des ouvrages et des personnes.

Pour un algorithme de prévision, il faut traditionnellement identifier les variables et rassembler les données pertinentes par rapport au sujet, il faut aussi généralement nettoyer ces données (corriger les erreurs, colmater les trous) et assurer leur interopérabilité. Ensuite il faut affecter des pondérations, selon les analyses escomptées, toutes les données croisées dans une même base ne recouvrent pas le même intérêt, d'où la nécessité de les pondérer. Les pondérations sont réglées à partir d'échantillons de cas réels jusqu'à ce que l'algorithme soit à même de reproduire la réalité. On remarquera ici encore la main prépondérante du pilote.

Avec l'apprentissage automatique, ou machine learning, qui vise à développer la capacité des machines et des logiciels à apprendre de leurs résultats, nous entrons dans un autre monde. Les algorithmes utilisés permettent à un système d'adapter ses comportements et réponses de façon autonome, en fonction d'une base de données qui a été conçue et constituée avec un peu de jus de cervelle. La donnée brute est un mythe, elle est toujours construite, elle ne vient pas de la nature, mais du système qui la mesure.

Pour reprendre l'exemple précédent de la prévision, de vente par exemple, on parle de machine learning dans le cas où l'algorithme rectifie tout seul les pondérations des données en fonction du résultat obtenu lors du dernier cycle de vente, et rectifie sa formule pour ne pas répéter les inexactitudes repérées dans le cycle suivant. L'algorithme apprend et se corrige de façon autonome, mais il reste fondé sur le postulat que les corrélations entre les jeux de données suffisent pour prévoir les nouveaux modèles à appliquer. Or si le futur vient de loin, il n'est jamais totalement contenu dans le passé.

Bien sûr avec des algorithmes appropriés on peut optimiser les ventes d'une entreprise ou les flux des transports urbains, faire baisser la criminalité ou la pollution, mais tout dépend de la volonté, des stratégies, des politiques et des organisations mises en place, les mégadonnées dans tout cela ne sont jamais qu'une technologie à éthiquement mettre en œuvre. On cite toujours en exemple les réalisations de Google, Yahoo, ..., mais le plus gros utilisateur de mégadonnées et d'algorithmes sophistiqués est quand même la NSA, dont on a pu apprécier une toute petite partie des activités, avec l'affaire des écoutes des dirigeants européens.

L'un des plus grands problèmes avec les algorithmes de mégadonnées, c'est leur coût, car contrairement à ce que l'on dit, ils sont onéreux à mettre en œuvre et donc seuls les grands organismes peuvent aujourd'hui en tirer vraiment profit. Certes on n'a pas besoin comme pour les entrepôts données traditionnels de mainframes décisionnels hors de prix, des solutions open source et des matériels lambda conviennent, mais pour gérer un gros volume, une grande variété de données, de la vitesse ..., il faut mettre en œuvre beaucoup de moyens matériels, logiciels, beaucoup de spécialistes, il faut du temps, beaucoup de jus de cervelle, et in fine cela est onéreux.

### **43 - L'analyste de « Mégadonnées ou Big Data » doit se méfier de son enthousiasme**

Le traitement des mégadonnées d'une entreprise nécessite de mettre en œuvre un processus itératif de construction, d'exploration et d'exploitation d'information. Il s'agit pour l'analyste avant de commencer, de fixer des objectifs qui tiennent compte surtout du contexte, de la situation de l'entreprise, et pas seulement de l'audience que constituent les dirigeants concernés.

Comme la pertinence des analyses dépend moins de la quantité que de la qualité des données, le contrôle très approfondi de leur origine est fondamental. Sélectionner des données, pour construire des métriques, des algorithmes qui permettront de répondre à des questions métier nécessite des compétences multiples, une très grande rigueur, pour

bien poser le problème et définir l'ensemble des objets à manipuler pour obtenir un résultat. Attention à ne pas admettre au départ dans son raisonnement ce que l'on veut prouver.

Même si une analyse globale ou statistique est souvent concluante, il faut toujours accorder toute son attention à l'incertitude des résultats. Il faut surtout se montrer prudent dans les interprétations, ne pas oublier qu'un scénario n'est pas une loi, l'analogie n'est pas une preuve, le bizarre & l'improbable sont probables, le possible n'est pas toujours possible, les anomalies, les mystères, ne constituent pas un fondement, l'inexistence de la preuve n'est pas la preuve de l'inexistence. La communication des résultats est toujours un moment extrêmement délicat. Il convient de ne pas confondre corrélation et causalité, prendre l'effet pour la cause, par de petits oublis ouvrir la voie à ses conclusions, se méfier du poids des mots, et des discours creux, mais efficaces qui multiplient les connotations permettant d'obtenir une adhésion subjective.

Finalement des décisions seront prises et des actions menées. Il faut ici encore être méthodique pour rester objectif et ne pas privilégier, au moment de fixer les orientations, une perception intuitive de la situation. Il convient de bien prendre en compte le contexte, les objectifs et les capacités d'actions des différentes parties prenantes, même si la revue systématique de la stratégie de chaque acteur est fastidieuse (finalités, objectifs, motivations, contraintes, moyens d'action ...). Comme dans ce processus on a recours à des données externes issues de l'intelligence économique, il faut se méfier de l'intox et se rappeler l'une des remarques principales de Sun Tzu « tout l'art de la guerre est fondé sur la duperie ».

Et les technologies dans tout cela ? Elles doivent être sélectionnées pour répondre aux besoins de l'entreprise, mais comme ils sont très divers, et qu'aucune solution ne peut répondre à tous les besoins, il faut donc multiplier les solutions, les fournisseurs, intégrer tant bien que mal l'ensemble et surtout le maintenir dans le temps (bonjour Sisyphe). Dans le monde changeant d'aujourd'hui, c'est un grand défi que seul le recours ponctuel, mais régulier à des experts indépendants de toute solution peut aider à relever.

#### **44 - Big Data : au-delà des incontournables solutions open source, il faut savoir mobiliser les bonnes compétences**

Depuis le début des années 2000, les ordinateurs sont devenus beaucoup plus puissants et des algorithmes ont été développés qui permettent d'exploiter des ensembles de données pour réaliser des analyses plus larges et plus profondes. Les grandes entreprises ont alors massivement investi dans leurs infrastructures de gestion des données. Aujourd'hui, toutes les fonctions des entreprises font l'objet de collecte, de conservation et de traitement de tout ou partie de leurs données : les ventes, le comportement des clients, la performance des campagnes de marketing, la gestion de la chaîne logistique, la fabrication, etc. Dans le même temps, les données sur des événements extérieurs tels que les tendances du marché, les nouvelles techniques de l'industrie, et les mouvements des concurrents sont maintenant plus largement accessibles.

Cette large disponibilité des données a conduit à un intérêt croissant pour les méthodes d'extraction, de représentation synthétique des informations et des connaissances. Au début, les entreprises employaient des équipes de statisticiens et d'analystes pour explorer leurs ensembles de données, mais les volumes et la variété des données dépassaient largement les capacités d'analyse dont une partie du processus restait en partie manuelle. Au mieux, des analyses au champ limité permettaient d'éclairer certains aspects ponctuellement. Mais depuis l'apparition des solutions open source (Hadoop, NoSQL, ...) les capacités et coûts ont été drastiquement changés et on peut constater que les entreprises, dans tous les secteurs, cherchent maintenant avec ces nouveaux moyens à exploiter leurs données pour obtenir un avantage concurrentiel.

Par exemple de façon générale, l'exploration de données est utilisée pour comprendre les préférences de la clientèle, pour améliorer la gestion des relations, pour analyser le comportement des clients afin de gérer l'attrition et de maximiser la valeur de la clientèle attendue. Le secteur de la finance utilise l'extraction de données pour la notation de crédit et de négociation, et pour la détection des fraudes ou la gestion des risques. Les principaux distributeurs de Walmart à Amazon appliquent l'extraction de données à travers toute leur entreprise, des achats à la vente en passant par le marketing et la chaîne d'approvisionnement. Finalement de nombreuses entreprises se sont différenciées de manière stratégique grâce à la science des données, parfois au point d'évoluer vers des sociétés d'exploration de données.

Par exemple de façon plus détaillée, dans le monde des télécommunications, les opérateurs ont des problèmes avec la fidélisation de leurs clients de téléphonie mobile. On constate habituellement que 20% des utilisateurs de téléphones cellulaires quittent leur opérateur lorsque leurs contrats expirent, alors qu'en parallèle il devient de plus en plus difficile et cher d'acquérir de nouveaux clients, étant donné que le marché du téléphone cellulaire est maintenant saturé. Les entreprises de télécommunications sont donc maintenant engagées dans des batailles sans fin pour attirer les clients des uns des autres, tout en conservant les leurs propres. Ici les vastes ressources de données des opérateurs peuvent permettre de découvrir quels clients devraient avoir une formule spéciale de rétention avant l'expiration de leur contrat. Il s'agit donc après analyse pour un opérateur de choisir un ensemble de clients à qui il proposera une offre adaptée afin de réduire le taux de désabonnement.

Cependant, mettre au point des programmes de fidélisation comme celui évoqué ci-dessus est beaucoup plus compliqué que cela puisse paraître au départ. Certes, les gisements de données peuvent aider, certes les technologies open source rendent cela plus accessible et moins onéreux qu'avec les mainframes décisionnels propriétaires, mais la grande difficulté est dans la mobilisation des compétences adéquates en matière d'informatique, d'analyse de données, de métier (marketing, finance, logistique, production ...), de management ... De management pour les budgets, de métier pour l'élaboration d'initiatives créatives, d'analyse pour savoir faire parler les données, et d'informatique pour savoir assurer l'approvisionnement en données de qualité. Or constituer des équipes mixtes de ce type et arriver à les faire travailler ensemble est très difficile dans les entreprises françaises où l'esprit de chapelle règne toujours.

Le sujet de la constitution d'équipes mixtes n'est pas nouveau, mais avec le big data on rajoute le besoin de nouvelles compétences en matière d'analyse de données, ce qui complique encore la situation, d'autant plus si à tort, on voit le data scientist comme un mouton à cinq pattes autonome, ce qui est le plus sûr moyen d'aller droit dans le mur.

#### **45 - Big Data : Il faut un bon équipage pour ne pas se noyer dans le data lake**

Le battage publicitaire promotionnel au sujet des lacs de données n'aide pas les entreprises à se faire une idée exacte de ce que cette approche peut leur apporter. De nombreux fournisseurs commercialisent des solutions pour réaliser des lacs de données et les présentent comme un élément essentiel pour capitaliser sur les opportunités Big Data, mais ils sont peu disert sur la façon de les exploiter et de tirer des avantages métiers significatifs.

Le plus souvent les lacs de données sont commercialisés en tant que plates-formes qui permettent de gérer et d'analyser les différents gisements de données de l'entreprise dans leur format brut d'origine. Ainsi au lieu de la traditionnelle approche de sélection, de transformation et d'organisation des données dans des entrepôts d'entreprises ou des data marts, il s'agit de simplement déverser, dans leur format original, des données dans le lac. Hélas, sans grand effort les données sont alors disponibles pour être analysées, croisées et en tirer toute la valeur qu'elles renferment au grand bénéfice des différentes fonctions, métiers de l'entreprise.

Les lacs permettent de gérer en un seul lieu toutes les données utiles à l'entreprise, quels que soient leurs sources ou leurs formats, permettant ainsi de dépasser les problèmes de silos constitués indépendamment les uns des autres. Cependant, alors que le battage publicitaire promotionnel suggère au public que toute l'entreprise pourra tirer profit des lacs de données, ce qui dans l'absolu est vrai, aujourd'hui très peu de personnes, d'entités, d'entreprises pourront en fait y arriver. Le processus de mise en valeur des données d'un lac est plein de difficultés, d'embûches, et ce qui est incontournable c'est que sans une bonne gouvernance des métadonnées il est impossible de se sortir des pièges sémantiques des données brutes. Sans métadonnées descriptives, le lac de données risque de se transformer en un marécage de données.

De par sa conception, le lac contient toutes sortes de données dont il convient de déterminer la qualité, d'assumer la sécurité et de contrôler l'accès en respect des exigences réglementaires, en particulier de la vie privée des personnes concernées par toutes ces informations. Sa valorisation suppose aussi que les utilisateurs connaissent et comprennent le biais contextuel de la façon dont les données ont été originellement saisies, qu'ils appréhendent la nature des données et leur structure, qu'ils savent comment concilier et fusionner les différentes sources de données disparates.

Les utilisateurs finaux à qui on donne accès au lac de données, n'ont sans de bons budgets et une aide très conséquente d'informaticiens, de data scientists, de statisticiens, aucune chance de tirer une valeur significative de leurs données. De plus il est évident qu'aujourd'hui la majorité des équipes mixtes de professionnels (métier, IT, analyste ...) dans les grandes entreprises n'ont pas beaucoup d'expérience en matière d'analyse de mégadonnées et leur montée en compétence sera longue.

Face à cette situation, les entreprises n'ont pas intérêt à essayer de tout expérimenter, de tout inventer, seules. Il convient sans aucun doute de se faire conseiller, encore faut-il éviter de s'appuyer uniquement sur des consultants avant-vente, spécialistes auto proclamée qui cherchent avant tout à vous vendre leur solution, ou plutôt la nouvelle solution de leur maison mère, le plus souvent américaine, qu'ils ont découverte lors de leur dernier kick off, en même temps que leurs objectifs de vente.

#### **46 - L'avenir est radieux pour les spécialistes de l'analyse des données**

Le travail des spécialistes de l'analyse des données, aujourd'hui le plus souvent, vise à permettre au plus grand nombre d'accéder à des éléments chiffrés en les rendant disponibles à travers l'utilisation de catalogues sur internet ou intranet (entreprises, universités ...). Il a pour but de faciliter l'accès et la compréhension des données, de façon que même les gens qui ne sont pas des experts dans l'analyse des données puissent en tirer des bénéfices. En raison de leur format standard, les données sont bien comprises par les humains et à l'avenir elles seront encore bien plus indispensables notamment pour les très nombreux robots qui nous entoureront.

Les spécialistes de l'analyse des données sont indispensables à tous ceux qui veulent appréhender le monde tel qu'il est, pour comprendre et agir. Les spécialistes stockent et soignent les données comme des conservateurs ou bibliothécaires et peuvent être considérés comme des facilitateurs. Ils aident aussi les gens à accéder aux données et les conseillent comment les utiliser dans leur domaine. Ils font en sorte que les données soient mises dans un format qui peut être lu facilement, non seulement par des experts dans ce domaine, mais aussi par le plus grand nombre. Les spécialistes de l'analyse des données doivent avoir de bonnes aptitudes à la communication, car ils interagissent avec beaucoup de personnes différentes dans le cycle qui va de la création de la donnée à son utilisation.

Tous les spécialistes de l'analyse des données n'ont pas les mêmes compétences ni les mêmes rôles, le monde des données offre de nombreux emplois différents. Il y a les emplois de direction centrés sur la gestion des données et l'organisation. Il y a les emplois de diffusion et de formation qui consistent à faire des publications, des présentations et à enseigner aux gens comment utiliser les données dans les différentes fonctions de l'entreprise ou dans différents métiers autres qu'en entreprise (gouvernement, police, santé, armée ...). Il y a tous les emplois liés au cycle de création et d'exploitation des données, de leur transformation en information (informaticiens, data-scientists, statisticiens, analystes métier (actuaire ...), contrôleurs qualité, archivistes, data journalistes ...). Il y a aussi les chercheurs qui mettent au point de nouveaux algorithmes, de nouvelles approches scientifiques, qui travaillent sur les outils de demain.

La plupart des spécialistes de l'analyse des données n'ont pas originellement étudié pour devenir experts des données, mais proviennent d'une profession liée à la création, la gestion ou l'utilisation des données. Les spécialistes de l'analyse des données viennent d'horizons différents, mais beaucoup ont un bon niveau en mathématiques et en statistiques. De plus il est certain qu'une connaissance de l'informatique est importante dans la science des données, la maîtrise de certains logiciels est indispensable selon les fonctions (base de données, outils de traitement, d'analyse, de restitution ...), mais de purs littéraires, des historiens, des juristes, des psychologues et bien d'autres profils encore, trouvent très bien leur place dans les équipes de spécialistes de l'analyse des données qui sont toujours multidisciplinaires.

Enfin ce qu'il faut prendre en compte, c'est que le monde des données est passionnant, parce qu'il permet d'être utile à la société, qu'il est en constante évolution et que l'avenir à moyen/long terme est particulièrement prometteur, notamment avec le développement de la technologie quantique. En attendant la révolution quantique, on peut noter que les spécialistes de l'analyse des données sont aussi nécessaires au bon fonctionnement de la démocratie.

## 5 - La multiplication des technologies du Big Data amène son lot d'inconvénients

La multiplication des technologies du Big Data nous a déjà apporté beaucoup de possibilités et nous ne sommes qu'au début du phénomène qui va encore s'amplifier avec l'internet des objets. Cependant, tout cela n'est pas sans inconvénient, il y a des constantes qui se retrouvent toujours avec les nouvelles technologies de l'information et des communications, des questions comme celle relative à leur impact sur la vie en société, les droits individuels ou celles relatives aux hommes et à leur capacité à intégrer, utiliser, développer ces nouvelles techniques.

### 51 - L'internet des objets ou la consécration du Big Data

Depuis longtemps nous utilisons des dispositifs pour mesurer, le monde qui nous entoure, pour mesurer nos actions, nos résultats, avec à l'esprit que l'on ne peut pas maîtriser ce que l'on ne peut pas mesurer. L'internet des objets c'est la connexion à l'internet de ces dispositifs qui créent ou manipulent les données dont nous avons besoin au quotidien pour prendre des décisions intelligentes. L'internet des objets se fonde sur les technologies et les solutions qui permettent l'intégration des données et des services du monde réel dans les technologies actuelles du réseau d'information, de façon à délivrer la bonne information à la bonne personne (ou à la machine) au bon moment.

L'internet des objets ne se contente pas de relier les dispositifs existants depuis longtemps, il offre l'opportunité de gérer de nouvelles données du fait d'un développement important de nouveaux dispositifs de mesure et de communication. C'est dans les années 80 avec les radio-étiquettes (RFID) que le nombre des dispositifs de ce type a explosé et c'est diversifié (capteur de proximité, onde acoustique, gyroscope, accéléromètre, réseau de capteurs sans fil, GPS...), jusqu'à commencer à s'intégrer à internet il y a un peu moins de dix ans.

Ainsi un nombre sans précédent de périphériques devrait être connecté à internet, et selon différents experts 50 milliards d'objets devraient être connectés d'ici 2020 et permettre de transformer les informations en actions en créant de nouvelles capacités, des expériences plus riches et des opportunités économiques sans précédent pour les entreprises, les particuliers et les pays (des moniteurs de santé et de performance portables, des véhicules connectés, "smart grids" pour l'énergie, production manufacturière connectée ...). Cet internet des objets révolutionnera notre façon de travailler, de vivre, de jouer et d'apprendre.

De façon plus détaillée par exemple dans le monde ferroviaire : un réseau connecté de dispositifs divers relié à l'internet des objets peut aider à améliorer la sécurité des passagers en analysant et en corrélant les événements dans différentes stations, ainsi que sur les trains. L'intelligence obtenue à partir de systèmes connectés multiples permet d'identifier les goulets d'étranglement, d'autoriser l'optimisation des itinéraires et des horaires pour une plus grande efficacité et même d'éviter les collisions. De même, les capteurs attachés aux pièces critiques telles que les roues peuvent déterminer de manière proactive si une pièce doit être remplacée - avant qu'elle ne puisse provoquer un accident dévastateur. Les capteurs peuvent également alerter les superviseurs si le train est utilisé de manière dangereuse ou alerter l'opérateur de changements importants dans les conditions de circulation.

Autre exemple : la voiture connectée. Les voitures d'aujourd'hui sont hautement informatisées, avec des centaines de capteurs pour évaluer à peu près tout, mais aujourd'hui, ce sont des centaines de capteurs individuels qui peuvent pour certains allumer une lumière au tableau de bord, et c'est à peu près tout ce qu'ils peuvent faire. En connectant tous ces capteurs, de sorte que les données peuvent être communiquées et centralisées on peut produire des informations plus utiles. Par exemple pour une perte de pression sur un pneu, tenir compte non seulement de la pression, mais aussi de la vitesse, des conditions routières, de la température extérieure, de la pression atmosphérique et d'autres données pertinentes, pour un contrôleur centralisé qui peut les analyser pour les transformer en intelligence utile ... le pneu doit-il être réparé immédiatement (le GPS peut dire où) ? Peut-il attendre ?

Avec l'internet des objets, il y a un champ immense de travail pour les vingt ans à venir, et les besoins en spécialistes de la mesure, des réseaux, des données, de leur gestion (qualité, sécurité...), de leur analyse, de leur usage (individuel, industriel...) sont immenses.



## 52 - Le monde merveilleux de l'internet et des big data : GAFAM, HACKERS, NSA ... et usurpation d'identité

Si vous êtes mal intentionné et à la recherche de forfaits à faire, internet est un paradis vous offrant 3 milliards de personnes, des millions d'entreprises ou d'organismes privés ou publics à potentiellement ciblés depuis votre canapé. Dans ce contexte, la cyberdélinquance se développe de façon exponentielle. Les médias parlent souvent des hackers, des crackers, des hacktivistes, parfois des terroristes, des criminels financiers, des employés mécontents, jamais des sociétés d'acquisition & de fusion, des entreprises ciblant un concurrent, et en ce qui concerne les pays qui font de l'espionnage, de la déstabilisation, de la manipulation, seuls quelques-uns sont dénoncés alors que la pratique est très largement répandue.

Si l'on imagine bien les motivations des états, des entreprises vis-à-vis de leurs concurrents, des criminels financiers, des éditeurs de logiciels de sécurité ... pour les autres les spécialistes nous évoquent des besoins psychologiques : reconnaissance, envie d'apprendre, curiosité, vengeance, malice, expérimentation, mentalité de gang, désir d'embarrasser et même des raisons altruistes de défense de la population contre Big Brother et autres institutions ou pratiques portant atteinte aux libertés fondamentales et à la vie privée.

Via internet, les acteurs mal intentionnés peuvent s'attaquer à des infrastructures (électricité, télécommunications, systèmes bancaires ...), engendrer des perturbations des services, des dégradations, des pertes d'argent et de crédibilité, ou opérer des extorsions, des vols (propriété intellectuelle, données confidentielles, identité ...) ou réaliser des manipulations (propagande, fake news, désinformation politiquement correcte ...).

En matière de cyber délinquance, nous sommes passés en trente ans du geek isolé, au gang d'hacktivistes ... et aujourd'hui il y a surtout des armées (privées ou publiques) de cyber acteurs aux intentions variées. Parmi les nations qui ont officiellement reconnu avoir des cyber capacités offensives on trouve notamment les pays suivants listés sans ordre particulier : Iran, Inde, Grande-Bretagne, Chine, Russie, USA, Australie, Italie, Israël, Syrie, Allemagne, France ... et tout le monde se souvient des écoutes illégales de dirigeants européens par la NSA. Concernant le vol de données, d'images, de sons privés ... certains grands fournisseurs de services, de logiciels, de matériels (GAFAM, etc. ...) se sont fait prendre la main dans le sac, et pour certains plusieurs fois.

Le plus étonnant dans ce domaine est l'inefficacité des organismes de contrôle type CNIL et la grande tolérance du grand public, qui pour l'essentiel continue de déverser ses données privées sur les réseaux sociaux, sur les sites de sauvegarde ... Le grand public néglige les risques, et agit aussi dans ce domaine comme si cela n'arrive qu'aux autres. Bien sur les attaques engendrant un déni de service, les divulgations d'informations privées, voir les pertes de données, les attaques calomnieuses ... ont la plupart du temps que des effets limités, mais le vol de données privées peut amener des usurpations d'identité et là l'impact peut non seulement être très conséquent (financièrement, judiciairement ...), mais durer pendant des années.

Le vol d'identité, cela arrive quand quelqu'un vole vos informations personnelles, les utilise sans permission et ce faisant passer pour vous, finit par endommager votre réputation, vos finances ... En fait, on distingue différents types de vol d'identité suivant l'usage qui en est fait : l'identité peut être utilisée pour obtenir des biens ou des services (ouvrir de nouveaux comptes bancaires, demander des cartes de crédit ou des comptes de crédit en magasin, multiplier les dettes ...), se faire passer pour un autre lorsqu'on est appréhendé pour un crime, on peut utiliser les informations d'un autre pour assumer son identité dans la vie quotidienne (clonage : obtenir un emploi, louer un appartement, souscrire des prêts ...), ou utiliser le nom commercial d'une autre entreprise pour obtenir du crédit ....

Il faut savoir que deux millions de Français ont déjà été victimes d'usurpation d'identité. Si l'on est victime d'usurpation d'identité, il faut immédiatement déposer plainte au commissariat de police ou à la brigade de gendarmerie, mais cela ne sera pas suffisant, il faudra en fait déposer plainte à chaque fois que l'on a connaissance d'une fraude, c'est le seul moyen de s'opposer aux créanciers. On notera qu'il n'existe pas de fichier central des usurpations d'identités, ce qui ne facilite pas la vie ni des victimes ni de la police, mais facilite bien celles des usurpateurs.

## 53 - Big Data et protection de la vie privée : mégadonnées => méga problèmes

Tout le monde (entreprise, gouvernement, université) s'accorde pour reconnaître aux mégadonnées (Big Data) un potentiel remarquable pour stimuler les affaires, les innovations, et les progrès de la société. En effet, avec une utilisation adéquate des mégadonnées on peut améliorer l'intelligence économique et l'efficacité des secteurs privé et public, en

développant de nouvelles applications qui permettent une exploitation de nouvelles opportunités sociales et économiques qui transforment la société.

Tout le monde souhaite protéger sa vie privée, et la loi en Europe prévoit que toute personne a droit au respect de son intimité et protège les données à caractère personnel telles que définies dans la directive 95/46 / CE, à savoir toute information relative à une personne physique identifiée ou identifiable. Cela comprend toute information qui se réfère à l'identité, aux caractéristiques ou au comportement d'une personne ou qui est utilisée pour déterminer ou influencer la façon dont cette personne est traitée ou évaluée.

Les mégadonnées peuvent être utilisées pour identifier des corrélations et définir des tendances générales permettant de mieux comprendre un marché et ses évolutions, par exemple, mais elles peuvent également être traitées afin de décortiquer très exhaustivement le comportement des individus. Les mêmes algorithmes et outils analytiques utilisés par un distributeur ou un banquier, par exemple, pour comprendre les intérêts, les désirs des individus, et déterminer ce qu'ils peuvent lui vendre, peuvent être utilisés par un gouvernement, des partis politiques, des entreprises privées de sécurité pour calculer (et souvent mal calculer) si vous pouvez être une menace (exemple les fiches S), maintenant ou dans l'avenir.

Même si ce n'est pas les mégadonnées (3, 4, 5, 6 V) qui sont en elles-mêmes inquiétantes, mais bien les utilisations des informations qui peuvent être faites, la multiplication des capteurs qui a lieu via le développement de l'internet des objets (GPS, montres connectées, traceurs d'activité, domotique ...), augmente considérablement les flux de données et les possibilités de percer de nombreux espaces qui étaient auparavant privés. Des exemples précis de collecte de données peuvent ne pas sembler importants par eux-mêmes, mais lorsqu'ils sont agrégés, ils peuvent créer une image complète d'une personne qui peut être extrêmement nuisible aux personnes concernées, en particulier dans les mains de tiers non autorisés. Aujourd'hui, le risque que les gens perdent le contrôle de leurs propres données devient maximum.

La protection de la vie privée est un sujet que les vendeurs de solutions décisionnelles éludent le plus souvent, le voyant principalement comme un frein à leur développement. La législation en Europe s'appuie principalement sur des directives qui ont plus de vingt ans et recommandent des solutions comme l'anonymisation qui sont de moins en moins efficaces au fur et à mesure que plus de données sont collectées et échangées. La réforme amorcée en 2012 et adoptée en 2016, même si elle a essayé de renforcer les possibilités pour le citoyen de contrôler ses données, ne règle absolument pas le sujet.

À noter cependant que des communautés de développeurs proposent des solutions comme : le « Privacy by Design » (construire la protection de la vie privée dans les dispositifs dès le début) ; le « Data Minimization » (selon le principe que les données qui ne sont pas recueillies, ne peuvent pas tomber entre de mauvaises mains) ; le « Notice and choice for unexpected uses » les personnes devraient recevoir des avis clairs et simples sur la manière dont leurs données seront utilisées avec un mécanisme de consentement. Ces communautés sont motivées et leur slogan est : « Big Data and Big Privacy : Yes, we can ».

#### **54 - Lutte antiterroriste et mégadonnées vont-elles conduire à des atteintes à la vie privée ?**

Compte tenu du contexte actuel, les dernières mesures gouvernementales pour lutter contre le terrorisme, apparaissent spontanément souhaitables : budgets, effectifs, moyens supplémentaires, notamment concernant l'acquisition et le traitement des données. Une grande liberté va donc être accordée aux services spécialisés pour utiliser toutes sortes de techniques liées aux mégadonnées, qui permettront de suivre au plus près les personnes ciblées.

En effet avec les technologies actuelles il est possible de traquer en même temps un très grand nombre de personnes. Consultation des mails et logiciels espions dans les ordinateurs des personnes ciblées (dont l'objectif est d'analyser le contenu d'un disque dur ou encore de connaître à distance les informations qui apparaissent sur l'écran de la personne ciblée), ou des dispositifs mobiles de proximité, qui permettent d'intercepter des communications dans un périmètre défini. Élargissement des écoutes administratives à l'entourage des personnes ciblées. Surveillance des déplacements des personnes ciblées par les services de sécurité via une balise GPS posée sur leur véhicule, et pose de micros ou de caméras dans un lieu privé. Sans compter les opérateurs de communications électroniques qui peuvent, et devront sur ordre, détecter par un traitement automatique, une succession suspecte de données de connexion.

Tout serait bien dans le meilleur des mondes, si l'on était sûr que les services ne vont s'intéresser qu'aux terroristes actifs ou potentiels. Mais les récents débordements en matière d'écoute d'avocats à des fins politiques ne nous inspirent pas confiance. Certes, lutter contre le terrorisme est un objectif prioritaire, mais les hommes politiques aiment par-dessus tout se piéger entre eux, y compris entre eux du même bord, et l'administration qui soupçonne de fraude tout le monde, ne se prive jamais d'utiliser tous les moyens techniques disponibles pour poursuivre les entreprises et les contribuables.

Certes la législation actuelle ne permet pas tout, et il existe théoriquement des risques en cas de dérive pour les contrevenants : amendes et peines de prison sont prévues par la loi, notamment pour des atteintes à la vie privée. Mais avons-nous déjà vu l'administration ou des politiques condamnés ? De plus malgré toutes les précautions que prendront les contrevenants, pour tenir caché les découvertes qui ne les intéressent pas, personne (cf. Wikileaks) n'est à l'abri de fuites d'informations, de piratages, de pannes, d'erreurs, de malveillances. L'industrie du chantage a de beaux jours devant elle.

Depuis de nombreuses années, les États-Unis espionnent sans vergogne leurs alliés, et tous les grands fournisseurs de solutions décisionnelles et de mégadonnées collaborent très profitablement avec la NSA. Il se prétend même que la NSA dispose, grâce aux fournisseurs, de moyens discrets pour accéder aux mégabases de données des entreprises pouvant disposer de données intéressantes, comme les banques ou les opérateurs de télécommunications. En tout cas selon les dernières révélations de Snowden (cf. le New York Times), la NSA collecte plusieurs millions de photos d'internautes par jour pour se constituer une base de données pour la reconnaissance faciale.

La CNIL toujours très présente pour tancer les entreprises, est actuellement quasi muette sur tous ces sujets concernant les états et les administrations (excepté un avis non publié en mars 2015), mais communique sur la surveillance des employés par leur entreprise ou donne des conseils pour maîtriser les « réglages vie privée » des téléphones. La presse s'est fait l'écho de quelques remarques, mais l'actualité l'amène à surtout regarder ailleurs. Il ressort de tout cela, que les technologies des mégadonnées en général et de surveillance de masse en particulier, vont donner naissance à une nouvelle ère de notre histoire, qui risque fort d'être peu démocratique.

Dans son livre « L'Ère numérique, un nouvel âge de l'humanité », Gilles Babinet identifie cinq domaines, intrinsèquement liés aux mégadonnées, dont l'évolution en cours va changer le monde : connaissance, éducation, santé, production et État. Il dit aussi que la société doit réfléchir collectivement à l'utilisation démocratique de ces technologies, si l'on ne veut pas créer un monde orwellien.

## **55 - Vers une révolution de la cryptographie avec le quantique**

Le besoin d'assurer la confidentialité des écrits a dû probablement suivre de peu l'invention de l'écriture. Garantir des secrets militaires, diplomatiques, commerciaux, personnels ... est l'objectif de la cryptographie qui transforme du texte ordinaire (texte en clair) en texte chiffré illisible pour qui n'a pas la méthode pour récupérer les informations. Crypter un texte peut être facile, il suffit d'utiliser un alphabet ou des caractères étrangers, par exemple le texte qui suit entre crochets (bravo vous avez trouvé) au lieu d'être en « Calibri » est en « MTextra » (en standard dans Word).

Le système de chiffrement utilisé pour les affaires militaires, le plus ancien connu, est celui de César qui en 40 av. J.-C. remplaçait chaque lettre selon la méthode d'un décalage de l'alphabet, par exemple : C remplaçait A ... P remplaçait N ... Z remplaçait X. Cette méthode est facilement cassable, si l'on subodore qu'il y a un décalage de lettre il n'y a que 25 clés possibles à tester ; on peut aussi faire une analyse de fréquence d'apparition des lettres (les plus fréquentes en français : e, s, u, n, i, o, a ...) et avec des textes longs, il est très probable qu'il n'y ait qu'une seule possibilité de déchiffrement. À noter que dès le XVIe siècle on s'est mis à utiliser plusieurs alphabets pour masquer les lettres fréquentes.

Au XXe siècle l'utilisation de machine pour crypter s'est beaucoup développée, la plus célèbre est l'ENIGMA, le système de code principal des nazis était capable de générer 159 ZETA clés différentes (ZETA = 10 puissance 21). Turing a réussi à casser le code, mais pour cela il a dû inventer un ordinateur spécifique, c'était un grand mathématicien, ce qu'il a fait pour craquer l'ENIGMA a été classé secret jusqu'en 1975 et certains de ses papiers sont toujours classés.

Disposer de la capacité de casser le code de son ennemi donne un très grand avantage. Par exemple les Américains qui avaient percé le code japonais ont éliminé l'Amiral Yamamoto en abattant son avion ayant eu connaissance de son déplacement (avril 1943). En revanche, les Japonais se sont cassé les dents sur les communications américaines

effectuées en langage Navajo. La population parlant Navajo étant inexistante dans le camp de l'Axe, la langue étant très complexe, les communications étaient indécryptables.

De nos jours, l'utilisation de la cryptographie s'est beaucoup développée pour assurer différents services : la confidentialité qui permet de garder secrètes des informations, sauf pour ceux autorisés à les avoir ; l'intégrité des données qui traite de la modification non autorisée des données ; l'authentification qui s'applique à la fois aux entités émettrices ou réceptrices et aux informations elles-mêmes ; la non-répudiation qui empêche une entité de refuser des engagements ou des actions antérieurs.

De nombreux systèmes de cryptographie civils ont été créés, ils sont à la fois hautement sécurisés et largement disponibles, comme DES d'IBM (Data Encryption Standard) ou RSA créé par Rivest, Shamir et Adleman ou ... Compte tenu de la place de l'ordinateur et de l'internet dans notre société, tous les moyens de sécurité et de confidentialité ont dû être réévalués. Ceci a amené le progrès probablement le plus important dans la longue histoire de la cryptographie, l'utilisation de deux clés (une clé publique et une clé privée) qui utilise une application intelligente des concepts de la théorie des nombres pour fonctionner et complète plutôt que remplace la cryptographie à clé privée.

La cryptographie à clé publique implique l'utilisation de deux clés, une clé publique, qui peut être connue de n'importe qui, et qui peut être utilisée pour crypter des messages et vérifier des signatures, une clé privée, connue uniquement du destinataire, utilisée pour déchiffrer les messages et signer (créer) des signatures. La cryptographie à clé publique est asymétrique, car ceux qui chiffrent les messages ou vérifient les signatures ne peuvent pas déchiffrer les messages ou créer des signatures.

Les systèmes cryptographiques modernes sont à la fois hautement sécurisés et largement disponibles, cependant l'attaque de recherche exhaustive par force brute est toujours théoriquement possible même si les clés publiques sont grandes (>512bits) et nécessitent beaucoup de temps pour être cassées. Une attaque par force brute implique d'essayer toutes les clés possibles jusqu'à ce qu'une traduction intelligible du texte chiffré soit obtenue. En moyenne, la moitié de toutes les clés possibles doivent être essayées pour réussir. Autant dire que c'est juste trop difficile à faire dans la pratique aujourd'hui.

Cependant, les choses vont être remises en question avec l'arrivée des calculateurs quantiques qui s'appuient sur les propriétés quantiques de la matière, et utilisent des techniques de calcul différentes de la programmation, comme l'algèbre linéaire classique pour traiter simultanément des ensembles de données. L'avenir commercial de ces calculateurs ne sera pas dans le grand public, en revanche ils sont très bien pour des calculs dont la complexité réside dans la combinatoire. Les calculateurs quantiques ont un bel avenir en cryptanalyse, car bien plus rapides qu'un ordinateur classique. En effet avec un classique le nombre de calculs augmente de façon exponentielle ( $2^N$  puissance N) en fonction de la taille de la clé à casser, alors qu'avec le quantique il n'augmente que de façon linéaire (N).

Les premiers petits calculateurs quantiques ont été construits dans les années 90, et un très grand nombre d'organismes publics ou d'entreprises privées travaillent à créer les premiers gros calculateurs quantiques, mais les experts invitent à éviter tout enthousiasme prématuré pour le moment, même si différentes sociétés annoncent des succès : IBM, Google, Honeywell ... On notera cependant que des moyens de chiffrement quantique existent déjà, ils ne demandent pas de calculateur quantique, et apportent la facilité de rendre toute interception de message immédiatement détectable. Enfin, les experts pensent que si les transmissions quantiques se généralisaient à l'avenir, elles pourraient assurer une confidentialité totale.

## **6 - Propos sur les entreprises qui développent les technologies du Big Data**

Les nouvelles technologies sont en évolution constante, et il y a régulièrement des révolutions techniques plus ou moins importantes qui remettent en cause la place de certains fournisseurs.

### **61 - Big Data & Open Source**

Les principales technologies Big Data ont été initiées par de grands acteurs du Web, comme Google & Yahoo, puis passées sous licence libre ce qui leur a permis un développement et une diffusion rapide. Ces logiciels open source, autorisent une liberté d'utilisation et des développements collaboratifs. « Libre » ne signifie pas toujours gratuit et de nombreux projets open source sont clairement développés dans une logique marchande. La diffusion de logiciels libres peut être faite selon différents modèles économiques : communautés, fondations, éditeurs, intégrateurs et prestataires de

services. Pour mieux connaître ce domaine, vous pouvez utilement lire le livret de Systematic, consacré aux Modèles Economiques du Logiciel Libre.

Cela dit face à un projet Big Data, il est possible de s'équiper en logiciels libres que cela soit pour l'intégration, les traitements, le stockage, l'analyse et la restitution, comme le montre la petite revue ci-dessous de solutions libres :

Solutions d'intégration et de traitement de mégas données : Hadoop est un ensemble de projets et d'outils de la fondation Apache permettant de stocker et traiter massivement des données, (voir aussi Storm) ; Apache Pig est un framework de requêtage et de traitement, (voir aussi Cloudera Impala, Flume, Hive) ; Talend offre un ETL for Big Data, (voir aussi Pentaho Data) ; Mule est un ESB.

Solutions de stockage de mégas données : Apache Cassandra est une base de données NoSQL Colonne répartie en Peer to Peer (voir aussi Apache HBase, la base de données du framework Hadoop) ; MongoDB est une base de données NoSQL Document, (voir aussi Elasticsearch) ; Neo4j est une base de données NoSQL Graph ; il est aussi possible d'intégrer des données NoSQL dans des bases de données relationnelles (solutions de PostgreSQL ou de Cassandra) au sein d'un ODS ou un entrepôt de données de type base de données relationnelle et ainsi d'y accéder avec un langage SQL classique.

Solutions d'analyse et de restitution de mégas de données : Pentaho Business Analytics est un portail décisionnel complet, (voir aussi JasperSoft BI Suite, Vanilla Platform) ; Elasticsearch Kibana est lui un portail de tableaux de bord web.

Pour une entreprise qui se lance pour la première fois dans un projet Big Data/Open Source, il est hautement recommandé de se faire accompagner par un conseil, qui devra d'abord aider à évaluer la faisabilité du projet dans le contexte spécifique de l'entreprise. Il existe de nombreux intégrateurs spécialisés dans les solutions open source comme : Altic, Audaxis, Core-Techs, Opensyscom, Progilone, Smile, sans oublier les départements spécialisés des grands intégrateurs classiques comme Capgemini ou Atos Consulting.

## **62 - Big Data : Hadoop doit être le cœur de votre système d'information décisionnel**

Hadoop a totalement changé le jeu en matière d'analyse de données. Les responsables dans les entreprises ont à leur disposition ce nouveau moyen et peuvent choisir de l'utiliser et de faire la course en tête dans leur activité, ou non. Hadoop présente de nombreux avantages, maintes fois démontrés en matière de coûts et de performance, notamment en permettant de s'affranchir des solutions ETL coûteuses et complexes, ou en accueillant des traitements de données antérieurement effectués sur des mainframes et entrepôts de données onéreux.

Hadoop permet de relever les défis classiques auxquels les entreprises ont aujourd'hui à faire face : forte croissance du volume des données à traiter, augmentation continue des demandes des métiers, complexité des cahiers des charges en matière d'ETL, limite au niveau des fenêtres de traitement, escalade des coûts face à des budgets de plus en plus serrés.

L'approche aujourd'hui recommandée par les conseils indépendants et sérieux, c'est-à-dire pas ceux qui sont juges & parties, consiste à mettre en œuvre une architecture de référence centrée sur Hadoop, de déplacer les traitements batch de l'entreprise sur Hadoop, de faire des données gérées par Hadoop le point de vérité unique de l'entreprise, de massivement réduire le recours aux solutions ETL en assumant notamment la partie transformation avec Hadoop, de conserver sur Hadoop les données à leur niveau de granularité le plus fin pour d'autres traitements ultérieurs, et de renvoyer dans les systèmes en place les résultats pour exploitations par les métiers.

Les grandes entreprises ont besoin d'un système d'information décisionnel complet et les solutions aujourd'hui en place sont complexes, avec des entrepôts de données d'entreprise, des data marts, de nombreux outils d'analyse, de reporting et une intégration poussée avec les systèmes opérationnels qui gèrent les transactions. Les solutions d'entreprise qui utilisent Hadoop doivent être conçues comme des écosystèmes globaux ayant pour objectifs de : réduire les coûts, diminuer considérablement le temps de traitement par lots pour les mainframes et les entrepôts de données ; de conserver et d'analyser les données à un niveau beaucoup plus granulaire, et avec une plus grande profondeur d'historique ; de délivrer de façon fiable la production dans les fenêtres de temps ; d'éliminer en grande partie l'utilisation des outils ETL traditionnels ; de permettre aux utilisateurs métier d'améliorer leurs analyses sur de très grands ensembles de données.

Les distributeurs ont par exemple besoin de s'améliorer dans les calculs de l'élasticité des prix de leurs articles en magasin. C'est-à-dire de répondre plus fréquemment à la question " qu'entraîne une évolution du prix sur la consommation de mes produits ? ". Plus l'élasticité est forte, plus la variation du prix a un effet important sur la consommation. Pour cela il faut prendre en considération les données de ventes par article, ce qui pour un grand distributeur peut représenter plusieurs milliards de lignes, ce qui nécessite de grandes capacités de stockage et de traitement. Avec des moyens traditionnels, la plupart des distributeurs font ce type de calcul trimestriellement, ce qui limite la possibilité de suivre au plus près les évolutions des marchés et de piloter le lancement de nouveaux produits.

Les distributeurs qui aujourd'hui chargent les données nécessaires dans Hadoop (données concernant les offres, les ventes, les prix, les stocks, les magasins, les paramètres d'élasticité ...), peuvent maintenant calculer l'élasticité des prix chaque semaine et l'ensemble de données constitué pour cette application permet en outre de développer d'autres analyses particulièrement utiles pour les analystes métiers. Au-delà des cas métiers d'utilisation d'Hadoop dont on pourrait multiplier les exemples dans de nombreuses industries (banque, télécommunication, distribution, assurances ...), nous pouvons constater qu'Hadoop sert principalement à : être le point central de collecte directe de données en provenance des systèmes opérationnels, être le lieu de transformation des données, gérer les données au niveau de granularité le plus fin, éliminer les stockages redondants, et au final être la source unique de la vérité d'entreprise.

Mais attention introduire Hadoop dans l'architecture de son système d'information décisionnel n'est pas simple, l'aide d'un conseil indépendant est recommandée, la bonne nouvelle étant qu'il y a déjà beaucoup de grandes entreprises qui ont fait le chemin et dont on peut s'inspirer.

### **63 - Installez votre système décisionnel dans les nuages à vos risques et périls**

Avec le cloud computing, les experts nous disent qu'une entreprise peut avoir un accès commode & à la demande à des ressources informatiques (réseaux, serveurs, stockage, applications et services), sans investissement, avec un minimum d'effort de gestion ou d'interaction avec le fournisseur. Il permet d'éviter d'avoir à assumer l'investissement, la gestion, l'exploitation de matériels aux capacités fixes spécifiques à l'entreprise et de profiter de capacité à la demande sans limites, de ne payer qu'en fonction de l'utilisation.

Des responsables informatiques qui cherchent à permettre à leur entreprise, d'innover, de faire plus à moindre coût, sont séduits pas cette informatique en nuage qui semble être une opportunité. Cependant, beaucoup sont préoccupés par la visibilité réduite dans les centres de données cloud, par le moindre contrôle sur les politiques de sécurité, par les menaces encore mal connues quant aux environnements partagés et à la complexité de la démonstration de la conformité. Finalement, la sécurité reste la plus grande objection à l'informatique en nuage, et l'inhibiteur numéro un à l'adoption à grande échelle.

Dans ce contexte, il est surprenant de voir se développer de plus en plus d'offres de systèmes décisionnels en mode cloud computing. En effet pour une entreprise, en matière d'information, quoi de plus précieux que les données de son système décisionnel, à part les secrets de fabrication non brevetés ; dès lors il est curieux de confier la gestion de ses données historiques clés à un tiers, même de confiance, car on s'expose à des indiscretions, fuites d'information, piratages, malveillances, usurpateurs, Big Brother ... Confier à un tiers la gestion de ses secrets qui ont une valeur marchande certaine, c'est à coup sûr tenter le diable, c'est inviter bien des personnes à des indiscretions rémunératrices.

L'actualité démontre régulièrement la faiblesse des systèmes d'information des entreprises, des administrations gouvernementales et la force, la créativité des personnes mal intentionnées. Des experts en analyses de données, en sécurité sont régulièrement pointés du doigt dans de nombreuses affaires, pour n'en citer que deux par exemple : Palantir et Kaspersky. Des pays sont connus pour leurs activités d'espionnage (USA, Russie) ou de cyber attaques (Corée du Nord, Chinois). En France, il est clair sur la place publique que les services de renseignements français sont en plein développement et recrute à grande échelle, montrant l'importance grandissante de la guerre de l'information dans notre société actuelle.

Les systèmes d'information sont des mille feuilles de logiciels composés de millions de lignes de codes qui sont vulnérables & pénétrables discrètement. Les gouvernements ou les entreprises ne sont pas peuplés de bisounours, encore une fois l'actualité nous le montre régulièrement (élection américaine, diesel gate) et très récemment dans le monde décisionnel la plainte portée par Teradata contre SAP, qui dans le cadre d'une coentreprise a détourné à son

profit des secrets d'un des leaders des solutions d'entrepôts de données. En matière de sécurité le maillon faible c'est le plus souvent l'homme, ses défaillances bien sûr, mais par-dessus tout sa volonté de s'enrichir.

Si vous confiez vos données historiques à une société tierce, les candidats à l'indiscrétion rémunératrice sont nombreux, du PDG au balayeur intérimaire, en passant par le responsable des sauvegardes. Si en tant qu'entreprise vous êtes prêt à courir le risque, il faudra informer vos clients, dont vous détenez des informations nominatives, à minima les rassurer, voire obtenir leur accord. À noter que l'entrée en vigueur en mai 2018 du RGPD (règlement général sur la protection des données en Europe) qui laisse aux utilisateurs le contrôle sur leurs données personnelles, renforce la transparence et la confidentialité, ne tranquillise pas 70% des Français d'après un récent sondage.

#### **64 - Big data, analyse de données et modèles métiers**

L'analyse de données permet d'appréhender le monde qui nous entoure et aujourd'hui de façon de plus en plus fine avec le Big data. Dans le cadre de l'utilisation de données pour guider leurs actions, les entreprises réalisent des modèles, élaborent des connaissances et ceux dans toutes les fonctions (finance, marketing, logistique ...). Un des modèles marketing les plus connus est celui du cycle de vie des produits.

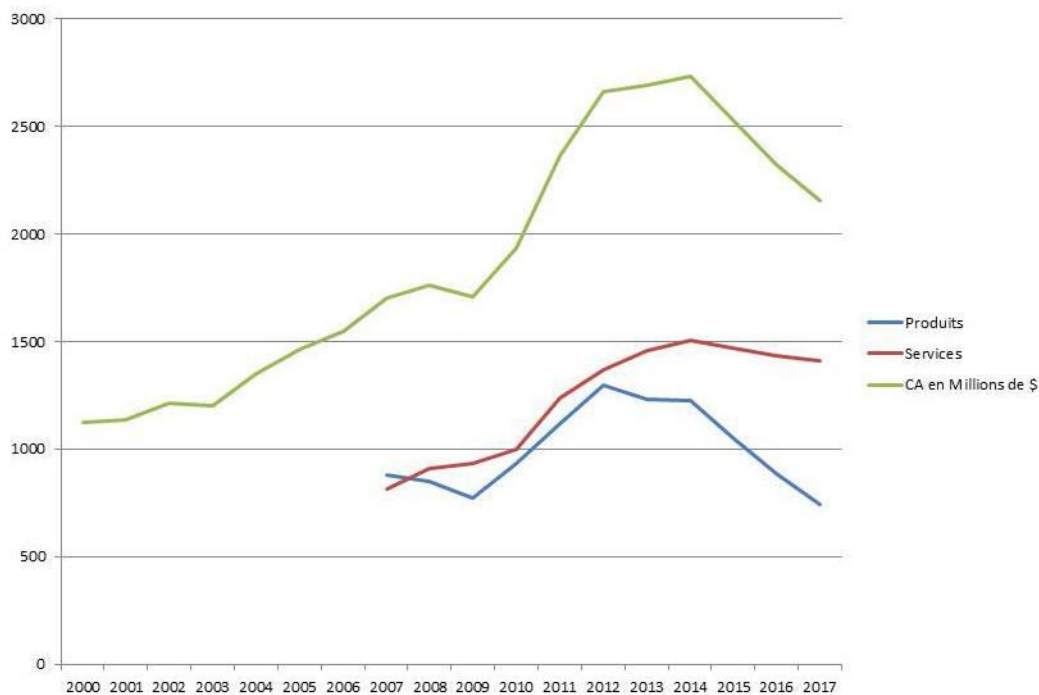
Le cycle de vie d'un produit, pour ceux qui réussissent, se décompose en cinq phases : recherche & développement, lancement, croissance, maturité, déclin. Il n'y a pas de chiffre d'affaires avant la phase de lancement, celui-ci croît jusqu'à la phase de maturité, durant laquelle il se met à décroître. En termes de profit, il faut généralement attendre la fin de la période de lancement pour voir les premiers bénéfices et parfois jusqu'à la phase de maturité pour couvrir les frais de recherche, de développement et de lancement.

Chaque phase nécessite de mettre en œuvre des stratégies spécifiques en matière de ressources à mobiliser et de clientèles à cibler. Par exemple, il faut successivement toucher les clients-testeurs d'innovation, les adoptants précoces, la majorité précoce, la majorité tardive et enfin les retardataires. Le produit lui-même évolue durant toutes ces phases, on l'associe souvent à d'autres produits ou services dans des packages et le chiffre d'affaires et les bénéfices s'en ressentent dans le temps.

Deux phases sont particulièrement difficiles à gérer, le lancement et le déclin. Dans les deux cas, la grande question est la même : continuer ou arrêter. Des phases de lancement s'éternisent et il faut souvent savoir s'arrêter à temps. Des phases de déclin peuvent être retardées avec des stratégies d'extension ou écourtées avec le retrait du produit au profit d'un remplaçant.

La plupart des entreprises ont de multiples produits et gèrent des portefeuilles de produits et ne voient pas leur viabilité à terme en péril quand un de leurs produits arrive en phase de déclin. Il n'en va pas de même pour les entreprises qui n'ont qu'un produit ou dont l'un des produits est indispensable pour leur viabilité. C'est souvent le cas de PME qui n'ont pas trouvé de successeur à leur produit phare, qui n'ont pas su se diversifier, qui n'ont pas, etc.... Cela peut aussi arriver à des entreprises plus importantes, dans le monde du Big data, Teradata est dans ce cas par exemple.

Le succès de Teradata a été fondé sur son offre de base de données qui lors de son lancement durant les années 80 était très innovante. En 2017, dans un marché en progression de plus de 10% par an ces dernières années, l'entreprise a fait un chiffre d'affaires de 2156 millions de dollars en baisse pour la troisième année consécutive (-7 à -8% chaque année). D'un point de vue global, l'historique de ses résultats, classe l'entreprise en phase avancée de maturité. D'un point de vue plus détaillé, on constate que le CA global est maintenu à un bon niveau grâce aux activités de services et qu'en fait la baisse des résultats au niveau des produits (base de données, serveurs ...) est amorcée depuis 5 ans : CA produits 2012 = 1297 millions de dollars vs CA produits 2017 = 747 millions de dollars. La ligne de produits Teradata est donc peut-être déjà en phase de déclin (cf. tableau du CA de Teradata ci-dessous)



Les modèles que l'on élabore, comme celui du cycle vie de produit, sont d'une grande utilité pour les entreprises, dans le cas du cycle de vie de produit cela permet de se situer sur le marché et de définir les stratégies qui s'imposent. En parallèle de la gestion des produits, les entreprises doivent assurer leur survie dans le long terme, et il existe différentes voies possibles comme l'innovation ou la diversification. Ce sont des stratégies à mettre en place avec des investissements appropriés (Recherche & Développement, rachat d'entreprises innovantes ...). Les entreprises qui n'ont pas de vision, pas de stratégie de marché, se contentent d'une gestion financière avec le rachat de leurs propres actions.

### 65 - Fournisseurs de solutions Big Data : la vie est un « long » torrent déchaîné

Pour un entrepreneur, le jeu consiste à développer une idée d'affaires, de prouver qu'elle fonctionne en lançant ses produits et/ou services, de devenir bénéficiaire, de faire croître son activité, de durer en s'adaptant à son environnement en constant changement, de compléter son offre, d'aller vers de nouveaux marchés, d'ancrer sa marque, sa réputation, de se diversifier, d'acquérir de nouvelles compétences, de capitaliser sur sa marque ...

Pour réussir, il faut bien sûr avoir une première idée créatrice de valeur, mais aussi savoir définir un modèle d'affaires adéquat qui permet d'être maître d'un avantage compétitif. Il s'agit d'organiser les rôles, les relations entre les clients, les alliés, les fournisseurs, et de calibrer les principaux flux de produits/services, d'informations, d'argent entre eux. Concrètement, il faut fixer l'offre de valeur pour le ou les types de clients ciblés, la tarification, la répartition des rôles entre partenaires (fournisseur, producteur, distributeurs ...) pour fournir la valeur, le tout en tenant compte de nombreuses contraintes : capacité de financement, taille des marchés, réglementations, normes, concurrences, évolutions prévisibles (marchés, techniques, ressources humaines ...) ...

Dans le monde des nouvelles technologies de l'information, pour durer une entreprise doit bien sûr comprendre comment les besoins des clients vont évoluer, mais par-dessus tout comment les technologies vont évoluer, à la fois les siennes et celles sur lesquelles elle compte pour s'intégrer dans les environnements parfois très compliqués de ses clients. Pour pouvoir durablement vivre de produits et de services de classe mondiale, il faut disposer au cœur de son offre d'au moins un élément très différenciateur par rapport à la concurrence et il faut aussi avoir le modèle de commercialisation adéquat.

Souvent les start-ups dans l'enthousiasme de leur développement cherchent à couvrir avec des offres spécifiques trop de domaines où elles ne font pas mieux que les autres. Une première stratégie pour durer consiste à adopter le plus de normes possible, de produits ou services du marché pour se concentrer sur ses composants les plus différenciateurs. Cela



été le cas de Teradata par exemple, dont les premières solutions dans les années 80 étaient élaborées avec pratiquement que des composants propriétaires : matériels, os, bus, base de données, outils d'interrogation ... Cela a été un des grands apports de NCR dans les années 90 que de réorienter les investissements de Teradata, après son rachat, sur le cœur différenciateur, à savoir la base de données et d'aller progressivement sur le marché chercher les autres bons composants.

Les start-ups au moment de se lancer ont aussi souvent des approches commerciales souvent frustes se fondant sur la simple démonstration technique de leurs avantages. Très rapidement, ces approches montrent leurs limites. Pour continuer avec le cas Teradata, l'entreprise est passée de la démonstration purement technique (années 80), à la promotion d'exemples d'analyses ponctuelles remarquables (années 90), à la fourniture de catalogue d'opportunités et d'applications (années 2000), à l'offre de services d'accompagnement technique, fonctionnel, de mise en œuvre et plus récemment d'hébergement. Les services qui étaient inexistantes dans les années 80 ont représenté 65% du CA en 2017, selon la présentation des résultats cette année-là.

Tout cela amène l'entreprise à se modifier en profondeur, pour bien des composants au lieu de faire, on achète, au lieu de vendre un produit ou un système interconnecté qui représente un investissement pour le client, on ne vend plus de produits, mais on rend une prestation de service au client ... Au fur et à mesure des évolutions, les différentes fonctions de l'entreprise sont transformées (production, finance, commercialisation, ressources humaines ...). À ce jeu, il est illusoire de croire que l'esprit des créateurs qui ont amorcé l'aventure dans leur garage (c'était la mode à une époque) peut perdurer, ni même celui des différentes équipes aux diverses étapes de la vie de l'entreprise.

Comme pour les produits, le développement d'une entreprise suit une courbe en S qui se termine inéluctablement par un déclin si l'on ne trouve pas le moyen de repartir sur une nouvelle courbe en S, avec une offre renouvelée et/ou un nouveau modèle d'affaires. Teradata était dans cette situation de devoir impérativement se renouveler et tente d'évoluer en développant depuis quelque temps une nouvelle stratégie, avec de nouvelles offres, une nouvelle communication ...

À la vue des résultats de Teradata fin 2018, la partie ne semble pas mal engagée, la chute du chiffre d'affaires est stabilisée (le CA 2018 est à 2164 millions de dollars par rapport à 2017 : 2156), avec selon la présentation des résultats de cette année, des revenus dits récurrents en hausse +10%, des services de consulting en légère baisse -2%, des revenus des licences et des matériels liés à la base installée en forte baisse -21%. Tout va donc se jouer sur la capacité à continuer à augmenter suffisamment le récurrent, alors que les revenus des licences et des matériels (actuellement 340 millions de dollars) vont inexorablement baisser.

En conclusion, il faut prendre conscience que l'entreprise est un des rares corps sociaux dont l'existence ne peut durer qu'en dynamique. Si l'entreprise arrête d'innover, elle perd de sa valeur et se condamne. Le rôle de l'entreprise est de gérer le processus d'innovation. Tout produit commence à être obsolète dès le moment où il est diffusé sur le marché et les entreprises doivent toujours avoir de nouveaux produits dans les tuyaux.

## **66 - Big Data : les Systèmes d'information décisionnels migrent dans les nuages**

C'est à la fin des années cinquante qu'un analyste, l'allemand Hans Peter Luhn, a décrit pour la première fois un BIS (Business Intelligence System) pour recueillir des faits, les analyser, en tirer des modèles et diffuser des informations aux différentes parties prenantes d'une organisation pour mieux gérer les activités. Le mot marketing lui, est apparu au début des années soixante s'appuyant sur des idées & des expériences de la décennie précédente en particulier aux États Unis. Dans les années soixante-dix, les best-sellers des éditions en matière de management concernaient les SIM pour Système d'Information Marketing.

Un SIM de l'époque visait à rassembler toutes les informations possibles internes ou externes nécessaires aux décideurs marketing, en particulier pour réaliser le plan marketing annuel. Concrètement, il s'agissait de centraliser des données qui pour l'essentielle étaient papier, mais on commençait à s'appuyer sur de l'informatique pour les données internes qui en ce temps-là étaient souvent limitées et ne couvraient que quelques fonctions (principalement : comptabilité, facturation, paye). En fin de compte, les analyses, les rapports, les plans ... sortaient tous des machines à écrire utilisées par des secrétaires.

Cependant, cette approche d'analyse et de synthèse étant fructueuse et pas seulement dans le monde marketing, la demande en informations internes s'est accrue en nombre, en fréquence, d'autant plus que les systèmes informatiques

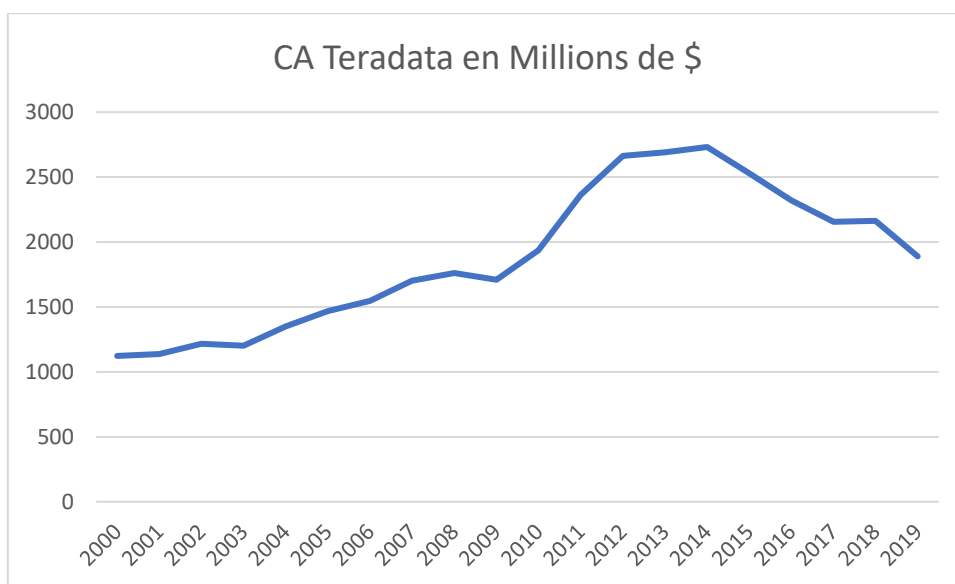
couvraient de plus en plus de fonctions de l'entreprise. Les directions informatiques très occupées à développer et gérer les systèmes de production renâclaient à répondre à toutes les demandes. Alors le concept d'Infocentre a été inventé, qui consistait à mettre à disposition des utilisateurs des données, dans un premier temps sous la forme de simple copie de données de production, mais ainsi les premiers vrais systèmes décisionnels étaient nés.

Le marché du décisionnel s'est développé au fur et à mesure des avancées techniques (PC, base de données relationnelle, MPP, Hadoop ...) et des nouveaux concepts (entrepôts départementaux, d'entreprise, lac de données ...). Les sources de données se sont multipliées, au-delà des applications internes de production, avec les données textuelles, géospatiales, vidéos, des capteurs, des séries temporelles, des réseaux sociaux ... le tout en grand volume et numérisé. Les systèmes décisionnels permettent de mixer et d'explorer tout cela, de faire des descriptions, des découvertes, d'établir des connaissances, des prédictions, des prescriptions.

Aujourd'hui, c'est le temps de l'informatique dans les nuages et les systèmes décisionnels ni échappent pas. Donc les entreprises se demandent si elles doivent cibler d'être dans un nuage privé ou public ou les deux (hybride). La réponse à cette question à court terme n'est pas simple, d'autant moins si vous êtes Directeur des Systèmes d'Information et que vos troupes sont en jeu, d'autant moins si vous êtes une très grande entreprise avec des métiers et des applications très spécifiques. Inversement la plupart des PME ont tout à gagner en cherchant à bénéficier de la qualité industrielle apportée par les nuages publics tant pour les applications de production que les systèmes décisionnels.

Le virage amorcé est en fait le passage d'une informatique trop souvent quasi artisanale, à l'exception des très grandes entreprises, à une informatique industrielle qui va apporter au-delà du coût une qualité des prestations pas souvent atteinte aujourd'hui. Dans ce contexte, les fournisseurs de composants matériels ou logiciels vont voir le nombre de leurs clients se réduire fortement à quelques grands prestataires de services dans les nuages. À ce jeu, pour chaque besoin seules quelques technologies survivront.

Pour une entreprise comme Teradata dont la base installée est quasi exclusivement composée de très grandes entreprises qui vont à court terme opter majoritairement pour des solutions hybrides, ne mettant que peu de choses en dehors de leur nuage privé, l'avenir se jouera probablement sur sa capacité à : acquérir de nouveaux grands clients pour des solutions en mode nuage privé, freiner l'attrition de la base installée, conquérir de nouveaux clients avec ses solutions en mode nuage public. Les résultats de l'année 2019 montrent que Teradata a réussi à doubler son nombre de « Cloud Customer », mais que globalement son chiffre d'affaires de 1889 millions de dollars est en baisse de 12% (-24% pour les services, -69% pour les licences de matériels et logiciels). Pour finir, on a un résultat négatif de 23 millions de dollars.



## 67 - La diversification, une voie pour la pérennité

Les entreprises se diversifient pour accroître leurs activités et mieux répartir leurs risques.

La diversification consiste à élargir la gamme de produits/services ou de marchés desservis par une organisation. La diversification connexe implique la diversification dans des produits ou services ayant des relations avec l'entreprise existante. La diversification non liée (en conglomérat) implique d'aller vers des produits ou services sans relation avec les actuelles activités.

L'une des diversifications connexes les plus courantes est l'intégration verticale qui consiste à entrer dans des activités où l'entreprise devient son propre fournisseur ou client. Le processus d'ajout à l'entreprise de nouvelles activités distinctes de ses activités établies peut se faire par le démarrage d'une nouvelle entreprise à partir de zéro, par acquisition ou par une coentreprise.

Une des grandes questions est de savoir si l'on veut s'étendre au-delà de son industrie d'origine. Les avantages de rester dans une seule industrie sont de concentrer ses ressources et ses capacités dans un domaine connu, sur ce que l'entreprise sait et fait de mieux. Les inconvénients sont liés au possible déclin de l'industrie, au fait de rater l'occasion de tirer parti de ses ressources et de ses capacités pour d'autres industries, de se reposer sur des lauriers et de ne pas chercher à apprendre continuellement.

Un préalable à la diversification est souvent de conceptualiser l'entreprise comme un portefeuille de compétences distinctives plutôt que comme un portefeuille de produits & services. Dans cette optique, il s'agit de recenser les compétences, de mettre en lumière les compétences clés, d'examiner comment ces compétences pourraient être mises à profit pour créer des opportunités dans de nouvelles industries et cela mène à des restructurations, des investissements, des ventes, des spins off, des achats.

Pour une entreprise, savoir se diversifier est un gage de longévité.

La société NCR pour laquelle j'ai travaillé de 1993 à 2007 est un exemple que je connais. Créée en 1884 sur une activité de production de caisses enregistreuse, elle s'est diversifiée au fil du temps dans ce domaine du paiement avec des trieuses de chèques, des distributeurs automatiques de billets par exemple, mais aussi dès les années 50 dans des ordinateurs, des moyens de stockage de données, produisant des mainframes, des mini-ordinateurs et même dans les années 90 des PC.

Partenaire de Teradata dès 1990, NCR finit par racheter cette start-up et l'a développée jusqu'au moment de sa mise sur le marché par un spin off en 2007. Cependant, autant NCR a montré une capacité de diversification, autant Teradata qui n'a jamais su réellement le faire, en subit les conséquences aujourd'hui. En effet, les nombreux achats de sociétés que Teradata a réalisés n'ont jamais débouché sur une réelle diversification, même si plusieurs sociétés de conseils ou de services ont été rachetées et que cela a permis de bien développer ce type d'activité.

De nombreuses sociétés de software dans le domaine de la gestion de la relation client ont aussi été acquises, telles de façon non exhaustive que Ceres, Aprimo, eCircle, Appoxxee ..., mais toujours dans le but de renforcer la vente des mainframes décisionnels et pas dans la perspective de créer des lignes d'activités distinctes, pour finalement pour certaines s'en débarrasser à un prix inférieur à l'acquisition. De nombreuses autres acquisitions ont été réalisées pour mettre la main sur des technologies pour améliorer l'offre de mainframe décisionnel, par exemple Kickfire (SGBD orienté colonne), Aster Data (Big Data), Raisntor (Hadoop)..., cela a certes conforté l'offre, mais n'a permis que de rester dans la course d'un marché en pleine évolution et non pas se diversifier.

Comme pour tout, l'entreprise est soumise à la tyrannie de la réalité

Comme je l'ai déjà montré dans mon article Big data, analyse de données & modèles métiers, les résultats de Teradata, faute de diversification, sont totalement dépendant d'une seule offre technologique et en subissent le cycle de vie. Le marché évolue et les systèmes d'information décisionnels migrent dans les nuages, Teradata tente de suivre, et les résultats de 2020 montrent une évolution forte du modèle d'affaires avec une baisse de 31% du CA réalisé avec les licences logiciels & matériels, une baisse de 28% du CA réalisé avec les prestations de conseil, mais ces baisses sont en

partie compensées puisque finalement le CA global ne baisse que de 3%, ce qui n'est pas si mal dans le contexte de la pandémie.

En l'absence d'une vraie diversification, Teradata est dans un combat de survie, avec une baisse continue de son Chiffre d'Affaires depuis 6 ans. La partie n'est pas perdue, mais il n'est pas sûr que son nouveau modèle d'activité puisse lui assurer son indépendance longtemps. Demain l'entreprise regrettera peut-être de ne pas avoir tenu compte des conseils de Mona Chollet dans son livre La Tyrannie de la réalité : « Point de salut hors de la diversification inlassable : ne jamais se croire arrivé, ne jamais cesser de faire des branches, qui elles-mêmes produiront de minuscules rameaux ».

À suivre !

\*\*\*