



WAVESTONE

BIG DATA :

UNE MINE D'OR POUR L'ASSURANCE



Déjà, en 1941, on parlait d'« Information explosion » dans l'Oxford English Dictionary pour évoquer le développement exponentiel des données, prémisse du Big data. Mais c'est à l'ère du digital que le terme prend tout son sens : d'après l'ONU et le rapport de l'Institut Montaigne d'avril 2015 « Big data et Objets connectés », plus de données ont été créées en 2011 que dans toute l'histoire de l'Humanité. Selon les estimations, entre 30 et 200 milliards d'objets seront connectés à horizon 2020. Des objets qui amènent une multiplication des données et, avec elle, une révolution des usages.

Et l'Assurance est au premier rang des secteurs impactés par cette révolution !

D'abord parce que la donnée a toujours été la matière première dans ce métier basé sur la prévision et l'estimation des risques. Mais surtout parce que cette nouvelle mine d'or est découverte à l'heure même où les assureurs doivent se réinventer pour continuer à se développer, créer de nouveaux services et repenser leur business model.

Quelles sont les données exploitables ? Comment reconsidérer les métiers de l'Assurance à l'ère du Big data ? Est-ce si simple et sans risque ? Quelle est la meilleure voie pour se lancer et pérenniser la transformation ?

Lors de deux ateliers clients, nous avons réuni autour de ces questions une trentaine de professionnels de l'Assurance en charge de l'innovation, du marketing, de l'actuariat, de l'architecture SI et plusieurs intervenants extérieurs : Stéphane Chappellier, fondateur de la start-up SolvINS, Nicolas Blandel de Témelio et Cyrille Vincey de Qunb.

Cette publication est le fruit de ces échanges et des travaux menés par Wavestone pour concrétiser la promesse du Big data comme levier de transformation.

Bonne lecture !



LAURENT BELLEFIN,
Directeur associé

WAVESTONE

www.wavestone.com

Wavestone est un cabinet de conseil, issu du rapprochement, début 2016, de Solucom et des activités européennes de Kurt Salmon (hors consulting dans les secteurs retail & consumer goods).

Dans un monde où savoir se transformer est la clé du succès, l'ambition de Wavestone est d'apporter à ses clients des réponses uniques sur le marché, en les éclairant et les guidant dans leurs décisions les plus stratégiques.

Wavestone rassemble 2 500 collaborateurs présents sur 4 continents. Il figure parmi les leaders indépendants du conseil en Europe, et constitue le 1er cabinet de conseil indépendant en France.

AUTEURS



Laurence Al Neimi

Laurence Al Neimi est senior manager chez Wavestone, en charge du pôle d'expertise Assurance avec plus de 15 ans d'expérience sur ce secteur. Elle intervient sur des programmes de transformation autour des problématiques d'organisation et systèmes d'information.
laurence.alneimi@wavestone.com



Clément Morizot

Fort de 6 ans d'expérience dans le domaine de l'architecture, Clément Morizot est responsable du développement de l'offre technique Big data chez Wavestone.
clement.morizot@wavestone.com



Stéphane Chappellier

Associé au sein de la *start-up* solvINS, Stéphane Chappellier est diplômé de l'Institut des Actuaire. Avec 25 années d'expérience en assurance en tant que praticien puis consultant, il intervient en tant qu'expert dans le domaine des objets connectés et l'exploitation de leurs données pour l'Assurance.



Laurent Renaudot

Laurent Renaudot est manager chez Wavestone. Depuis plus de 10 ans, il intervient sur les sujets de sécurité financière (lutte contre la fraude et le blanchiment) dans les secteurs de l'Assurance, de la Banque et de la Finance.
laurent.renaudot@wavestone.com

06		La donnée : matière première de l'Assureur
10		L'Assurance à l'ère du Big data : quelles perspectives ?
20		Démythifier la magie du Big data : le processus d'exploitation de la donnée
24		Comprendre la jungle technologique
30		Gérer les risques liés aux données manipulées
36		Comment se lancer ?
43		Big data : une révolution qui passe par la confiance
46		Glossaire

Nous remercions l'ensemble de nos clients et les start-up Témélio, Qunb et SolvINS qui ont enrichi notre réflexion sur le sujet.

LA DONNÉE : MATIÈRE PREMIÈRE DE L'ASSUREUR

Le marché de l'Assurance s'inscrit dans un contexte en pleine mutation. L'enjeu majeur pour les prochaines années est d'identifier un ou plusieurs axes de différenciation bénéficiaire. Cette mutation est telle qu'elle implique une évolution du business model, aujourd'hui fondé sur un calcul de risques établi sur des données historiques. Cette évolution, voire rupture, va pouvoir s'appuyer sur de nouvelles technologies, rendant possibles l'accès et la compréhension d'un volume de données en croissance exponentielle.

UNE TRANSFORMATION EN MARCHÉ

L'Assureur fait face à des défis réglementaires, sociétaux et technologiques. Autant de pressions qui nécessitent de repenser son positionnement.

Côté réglementaire d'abord, les enjeux sont bien réels : selon la Fédération Française des Sociétés d'Assurances, plus de 30 textes ont été étudiés en 2013 ! Au travers de ces textes, la volonté du législateur est :

- / d'accentuer la protection des consommateurs avec par exemple la Loi Hamon (loi sur la consommation) ou la Loi Eckert sur les contrats en déshérence,
- / d'en tirer une manne financière comme avec les contrats responsables et l'augmentation de la taxation de 7 à 14% si les contrats ne respectent pas les conditions dites « responsables »,
- / de garantir la solidité financière des acteurs de l'Assurance avec Solvabilité II.

Ces poussées réglementaires ont pour effet direct de globaliser et tendre le marché.

Sur le plan sociétal ensuite, les principes de solidarité nationale mis en place après-guerre pour la Santé, la Retraite et la Dépendance sont aujourd'hui en très fort déséquilibre compte-tenu de l'accroissement des frais de santé et du vieillissement de la population.

Si l'on prend l'exemple de la Santé, les dépenses représentaient 4% du PIB dans les années 60 contre près du triple aujourd'hui (11,7% en 2013). En sachant que les efforts de prévention ne représentent que 2% des dépenses de Santé, une tendance haussière de la consommation de soins est inévitable. Le Centre européen du diabète note ainsi une prévalence de cette maladie à 6% de la population dont 500 000 à 800 000 diabétiques qui s'ignorent, laissant présager une prise en charge plus lourde et complexe des futurs soins. Sur les dépenses de Santé en général, le reste à charge des Français est en hausse régulière alors que l'accès gratuit à la médecine est devenu un acquis dans l'inconscient collectif (source INSEE).

Ce secteur est donc pris en étau entre un service public dont le budget ne permet plus d'assumer le niveau de consommation et les Français pour qui la Santé est un acquis et n'a pas de prix.

Par ailleurs, dans le contexte d'une économie ouverte, les attentes de l'assuré évoluent. Qu'il soit particulier ou entreprise, ce dernier est aujourd'hui plus averti, plus exigeant et toujours en attente de nouveaux services technologiques. De consommateur, il devient « consom'acteur » de l'économie collaborative. Il est soucieux des principes de durabilité et a toutefois des préoccupations bien individuelles qui s'inscrivent dans un nouveau référentiel de valeur, celui de l'usage et non plus de la possession.

Or, le *business model* de l'Assurance repose sur la promesse d'indemnité ou prestation en cas de survenance d'un risque préalablement mutualisé. Ces risques portent sur des biens ou des personnes mais pas ou peu sur l'usage qui en est fait. Quelles que soient les branches d'Assurance, le modèle est arrivé à sa limite, compte-tenu des nouveaux paradigmes auxquels il doit s'adapter.

Des tentatives d'évolution ont été mises en place avec des services en complément des produits d'assurance (analyse de devis des professionnels de Santé, accès à un réseau de soins en Santé, de garagistes en Auto). Toutefois, ces compléments ne sont pas réellement valorisés et restent essentiellement curatifs.

Une autre voie est donc à inventer pour évaluer différemment les risques et élaborer de nouveaux concepts. Si historiquement, les assureurs sont des consommateurs de données et ont développé une expertise sur la collecte et le stockage, celle-ci permet de calculer les risques et mesurer l'impact de la mutualisation. Désormais, de nouvelles sources sont accessibles et de nouvelles logiques sont à mettre en lumière.

DE NOUVELLES VOIES OUVERTES

Plusieurs nouveaux usages peuvent être imaginés en exploitant le potentiel des données. En premier lieu, mieux exploiter les données internes en repoussant les capacités techniques d'exploitation permettra de valoriser son patrimoine data. Ensuite, élargir à de nouvelles sources de

données apportera des perspectives plus en rupture.

Pour chacune des sources, la qualité de la donnée doit être analysée en suivant deux critères principaux : le degré d'exploitabilité et le degré d'intérêt. La complexité vise à qualifier chaque donnée unitairement mais également en corrélation avec d'autres. Une donnée a priori sans intérêt, comme l'heure de réalisation d'un devis sur internet, pourra acquérir du sens en corrélation avec d'autres.

Et point particulier concernant l'*open data* : l'engagement de service du détenteur d'une source de données est un critère structurant pour garantir l'utilisation pérenne de cette dernière.

Dans le contexte d'une économie ouverte, les attentes de l'assuré évoluent. Qu'il soit particulier ou entreprise, ce dernier est aujourd'hui plus averti, plus exigeant et toujours en attente de nouveaux services technologiques.

DONNÉES INTERNES DU PORTEFEUILLE

- Caractéristiques client et données personnelles
- Contacts, Demandes, Actions
- Contrats, Risques, Devis, Démographies entreprises
- Cotisations émises et payées
- Prestations : Déclaration de sinistres, PJ, Constats

DONNÉES PARTENAIRES

- Réseaux de prestataires : Experts, Carrossier...
- Acteur de l'achat média : Qualificateur d'audience, Annonceurs...
- Réseaux PS Tiers Payant
- PS plateformes de soins
- Rapport d'expertise
- Comparateurs
- Décomptes Sécurité Sociale
- Déclaration sociale des entreprises

DONNÉES SEMI-PUBLIQUES

- Référentiels postaux
- Base Argos (véhicules volés)
- Référentiels AGIRA (victimes, PV, résiliations...)
- SIV • SNIIRAM • ORIAS
- INSEE, Data.gouv.fr
- Base criminalité
- Agenda scolaire
- Programme TV

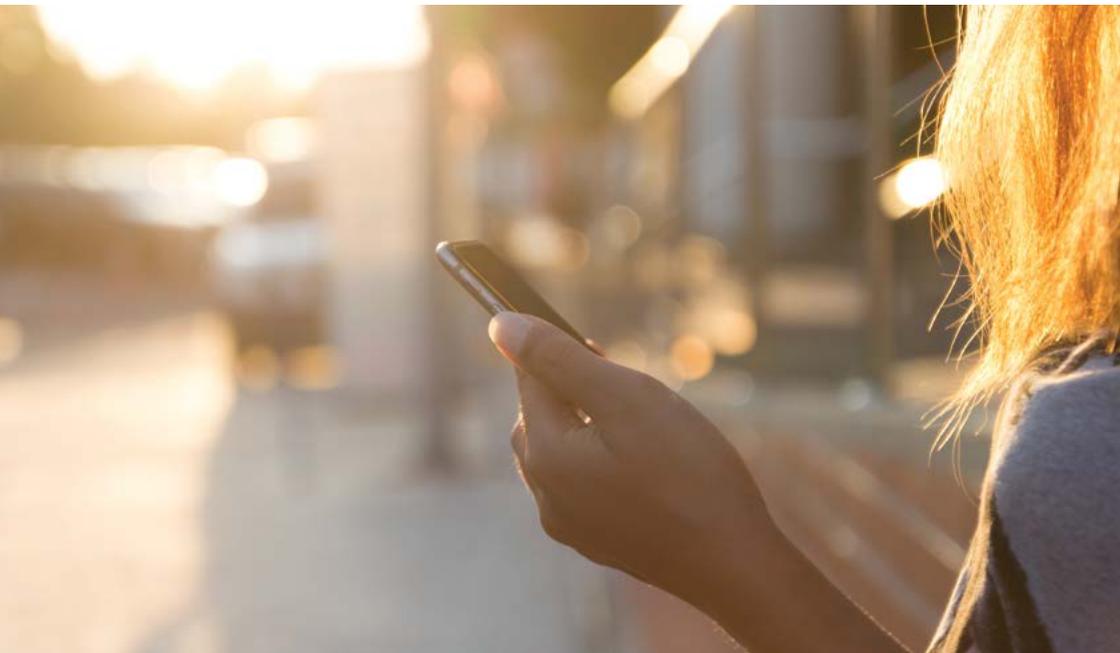
DONNÉES COLLECTÉES

- Compteur/Box intelligent(e)
- Capteurs de la Smart City : Pollution, Traffic, Météo
- Boîtier Habitation (domotique)
- Bracelet / Balance / Tensiomètre /...
- Réseaux sociaux
- Données de navigation
- Google Maps

PS : Professionnels de santé

L'ASSURANCE À L'ÈRE DU BIG DATA : QUELLES PERSPECTIVES ?

Valoriser et exploiter le capital data pour personnaliser la proposition de valeur auprès de ses assurés, en concevoir de nouvelles, mieux gérer le risque grâce à des analyses prédictives, développer une image moderne, renforcer la lutte contre la fraude : telles sont les perspectives des usages du Big data par les assureurs. La puissance du Big data peut être mise à disposition de toutes les fonctions Métier sur la chaîne de valeur de l'Assurance, et au-delà, des fonctions transverses comme la Direction RH pour le recrutement ou la prévention des risques pour ses propres salariés mais aussi la Direction Générale dans la construction de sa stratégie.



LA CHAÎNE DE VALEUR AU REGARD DU BIG DATA

La chaîne de valeur assurantielle peut être améliorée en s'appuyant sur les apports technologiques du Big data. Ces optimisations peuvent être graduelles dans la rupture à opérer. Il est donc possible de construire des stratégies d'évolution à différents horizons allant de l'optimisation du process existant à une toute nouvelle proposition de valeur.

L'attente des assurés a aussi évolué et il est nécessaire de considérer le produit d'assurance sous l'angle du besoin global du client. En Santé par exemple, il conviendrait non seulement de répondre à son attente de remboursement de frais de soins mais aussi de considérer le bien-être, la nutrition, l'activité physique et sportive, l'accompagnement du vieillissement et la protection du cercle familial.

Pour élargir le champ d'action de l'Assureur et construire de nouveaux services apportant plus de proximité avec les assurés, nouer des partenariats apparaît comme une solution pertinente. Solution permettant de partager les informations, les données et amortir les coûts.

Ce fonctionnement en partenariat est idéal afin d'intégrer des objets connectés pour lesquels la modélisation, la construction et le service après-vente restent à la charge de l'industriel. La domotique, les compteurs

électriques, les objets de mesure médicaux ou les assistants de vie pour la dépendance en sont autant d'exemples.

En allant plus loin, une nouvelle voie pourrait consister à concevoir des produits tirant parti des nouveaux aspects prescriptifs. Il conviendrait de proposer des protocoles de prévention et démarches à suivre pour atteindre un résultat calculé et garanti par l'Assureur. Ce dernier s'engagerait alors à indemniser en cas de non atteinte du résultat. L'Assureur ne couvrirait plus un bien ou une personne, mais garantirait un résultat, ouvrant ainsi la possibilité d'une assurance à l'usage.

Cette voie est d'ores et déjà initiée pour le secteur agricole, notamment entre la start-up The Climate Corporation et le géant de la biotechnologie agricole Monsanto. Ce partenariat consiste à proposer un protocole, ajusté en fonction de nombreux paramètres, qui garantit le rendement agricole. Si le protocole n'aboutit pas au résultat escompté, une indemnité est versée à l'agriculteur.

Cette logique prescriptive ne repose donc plus sur les données historiques de sinistralité. Elle permet ainsi à de nouveaux acteurs comme les GAFAM ou mêmes des industriels de se positionner, faisant planer une véritable menace sur les acteurs historiques de l'Assurance.



« Trois fonctions de la chaîne de valeur sont particulièrement impactées par le Big data : marketing, lutte anti-fraude, pricing »

Trois fonctions de la chaîne de valeur sont particulièrement impactées par le Big data. Le marketing et la lutte anti-fraude ont des niveaux de maturité plutôt élevés et peuvent avoir une application proche des process existants. Le Pricing est davantage en rupture par rapport au business model en place.

Le marketing devient data driven. Côté marketing, les enjeux majeurs sont de pouvoir qualifier en temps réel les cibles, de transformer un prospect en client ou de détecter, à la moindre alerte, un risque d'attrition. Pour ce faire, il faut exploiter la quantité de données disponibles en adéquation avec une stratégie pertinente. Le marketing est ainsi devenu data driven, plaçant les données au cœur de sa démarche.

Alors qu'un foyer sur deux résilie son contrat habitation au moment de déménager, l'enjeu pour les assureurs est de pouvoir déceler les signaux précurseurs de ce changement. Comment, dès lors, détecter le bon signal, au bon moment ?

A l'ère du tout connecté, les clients sèment des informations, parfois même sans le savoir, créant de la donnée on-line captée puis collectée par des spécialistes du marketing digital.

En exploitant par exemple les données de navigation d'un client souhaitant déménager, telles que les cookies et autres traces digitales (visite de sites de déménagement, de travaux...), il devient possible pour un assureur de détecter à temps l'intention de déménagement et d'agir en fonction. Une offre, adaptée aux besoins, sur la bonne temporalité, peut alors être proposée au client. Le bénéfice sera pour ce dernier une simplification de ses démarches et dans certains cas une adaptation tarifaire. Le client est satisfait et le taux de rétention chez l'assureur amélioré.

Ces évolutions sont rendues possibles au travers d'une Data Management Platform (DMP), plateforme technologique qui réalise la convergence des données internes (off-line) de l'assureur aux données on-line. Ainsi croisée, la connaissance client est largement enrichie.

Ces solutions permettent de mieux cibler les efforts avec une meilleure qualification des cibles et de la temporalité associée. Il est également possible pour les assureurs de mieux comprendre leur audience.



TÉMOIGNAGE : Nicolas Bandel, CEO & Co-Founder de la start-up Temelio, solution de CRM* Onboarding

Qu'est ce que le CRM Onboarding ?

Les assureurs sont aujourd'hui confrontés à une véritable transformation digitale de la relation client. Or leur richesse réside dans leur outil de gestion de la relation client, le CRM, où se trouvent les informations collectées depuis plusieurs décennies. Le CRM Onboarding est une technologie qui permet d'embarquer la connaissance client CRM off-line des annonceurs dans l'écosystème digital pour mettre en œuvre des actions marketing totalement personnalisées.

Comment le respect de la vie privée est-il assuré ?

La sécurisation des données client doit être garantie lors du process d'Onboarding en anonymisant et supprimant l'ensemble des informations personnelles. Les partenaires, sources de données, s'engagent contractuellement à effectuer une collecte respectant l'ensemble des lois en vigueur et des recommandations de la CNIL. Toute notre technologie est hébergée en France.

Quels ont été les ROI observés chez vos clients assureurs avec cette technologie ?

Nous mettons actuellement en place les premiers pilotes. Nous obtenons déjà des métriques de digitalisation de segments CRM compris entre 25% et 45% et les résultats des campagnes de CRM display sont très encourageants. Suite à un test d'analyse des ventes entre un segment de clients exposé et un segment non exposé, on constate +75% de ventes sur les personnes exposées à la campagne.

Les cas d'usage identifiés chez nos clients assureurs sont principalement :

- / **Le CRM Display** pour un meilleur ciblage lors des campagnes publicitaires. L'assureur peut exposer les segments CRM de son choix à ses publicités dans une logique de multi-détention, ou exclure ses clients actuels pour une approche 100% acquisition.
- / **La personnalisation de site** : l'assureur peut reconnaître un client qui arrive

« les assureurs sont aujourd'hui confrontés à une véritable transformation digitale de la relation client »

* CRM : Customer Relationship Management

sur son site sans login. Grâce à un outil de personnalisation, il est possible de proposer une page web personnalisée selon le client pour lui proposer des offres complémentaires.

/ **L'attribution** : 5% des souscriptions de contrat d'assurance ont lieu en ligne, et pourtant le web influence le consommateur tout au long de son parcours client. Pour aider les Directions digitales à justifier le retour sur investissement de leurs actions, les données de campagnes publicitaires sont croisées avec celles du CRM afin de savoir quel profil a été exposé à une publicité digitale avant de contractualiser en off-line.

« Le CRM Onboarding est une technologie qui permet d'embarquer la connaissance client CRM off-line des annonceurs dans l'écosystème digital pour mettre en œuvre des actions marketing totalement personnalisées »

LA LUTTE CONTRE LA FRAUDE OU LA DÉLINQUANCE

Avec environ 20% des cas de fraude qui ne sont pas détectés en IARD* et un impact évalué à 2,5 milliards d'euros par an (soit 5 à 7 points sur le ratio sinistres / primes), la lutte contre la fraude est devenue un élément fondamental de la maîtrise des risques pour les assureurs. Le renforcement des équipes est donc plus que jamais une nécessité (source ALFA, Association de Lutte contre la Fraude à l'Assurance, rapport annuel 2014).

La professionnalisation de la sécurité financière

Aujourd'hui, la lutte contre la fraude s'est professionnalisée avec le déploiement d'équipes centralisées, spécialisées et dotées d'outils informatiques dédiés, travaillant en lien avec les Métiers. Cela permet d'avoir d'une vision globale, transverse, multi-métiers des clients et des tiers.

Les contrôles sont plus élaborés : blocage a priori des transactions, filtrage et scoring des clients, corrélation d'événements multiples, analyses comportementales et prédictives, mise en évidence de réseaux...

Toutefois, les solutions actuelles restent dépendantes de la qualité des données, coûteuses, complexes et longues à déployer.

L'aboutissement : une baisse de la fraude au bénéfice du client

Le Big data apporte de nouvelles solutions permettant l'agrégation d'une plus grande variété de données, tout en étant moins sensible à leur qualité.

Cette avancée va dans le sens d'un outil unique et adapté traitant l'ensemble des modes de surveillance, tout en respectant la ségrégation nécessaire des données et traitements.

Ces solutions innovantes favorisent la réactivité des fonctions de sécurité financière et une meilleure qualité de la relation client :



* IARD : Incendies, Accidents et Risques Divers

- / En introduisant des corrélations nouvelles, l'approche Big data permet un pilotage plus souple, plus évolutif et en temps réel. La centralisation de la surveillance favorise la cohésion de l'ensemble des acteurs (métier, MOA et informatique) avec une souplesse des analyses, une réactivité des contrôles face à de nouvelles menaces.
- / Grâce à la baisse du coût de la fraude, un retour sur investissement est directement perceptible par la Direction Générale. Une hausse des

garanties et / ou une baisse des cotisations redeviennent possibles, impactant positivement la relation client. L'amélioration du process par l'automatisation des contrôles génère une accélération des délais de traitement des sinistres, et une meilleure qualité des prestations... Ces actions ont également un effet dissuasif envers les fraudeurs et induisent par voie de conséquence une meilleure qualité du portefeuille clients.

LA TRANSFORMATION DES FONCTIONS DE PRICING

Au lieu de passer par le truchement de variables permettant d'expliquer un comportement, les modèles de tarification ont maintenant une lecture directe de ce dernier.

Initialement, la tarification se basait sur les variables âge et sexe ou couleur du véhicule pour contourner la loi du genre, pour catégoriser un assuré dans un segment stéréotypé. Désormais, un jeune conducteur masculin précédemment supposé impulsif peut-être qualifié en fonction de sa conduite réelle grâce aux données issues des boitiers connectés (ou port embarqué type OBD-II).

En conséquence, les techniques qui s'étaient adaptées à un environnement à dimensions restreintes (nombre de cases tarifaires limité) vont devoir s'adapter pour pouvoir fournir une tarification individualisée, fondée sur une masse insoupçonnée de nouvelles données.

Traditionnellement, la tarification repose sur des données statiques. Ces données ne subissent que très peu de modifications d'année en année. La dynamique apportée par les nouvelles données, accessibles en

« La dynamique apportée par les nouvelles données, accessibles en temps réel, va obliger à revoir ces modèles »

temps réel, oblige à revoir ces modèles. Le constat de l'évolution du mode de vie de l'assuré pourra être pris en compte avec, par exemple, un ajustement de la prime suivant les observations, à une périodicité plus fréquente qu'actuellement.

C'est cette nouvelle dimension dynamique qui constituera le moteur de la permanence de la relation client.

« Pour mettre en œuvre une démarche de lutte anti-fraude, il n'est plus indispensable de se lancer dans un grand projet nécessitant une mobilisation d'envergure. Profiter des solutions existantes en interne, développées par exemple par le marketing ou l'actuariat, permet de tester des dispositifs novateurs avec des résultats instantanés »

DÉMYTHIFIER LA MAGIE DU BIG DATA : LE PROCESSUS D'EXPLOITATION DE LA DONNÉE

Les réflexions autour des cas d'usage sont en pleine ébullition pour nourrir la nécessaire transformation du secteur de l'Assurance. Les sources de données sont pour certaines aisément accessibles. Par exemple, les objets connectés domotiques permettent de connaître l'activité d'une maisonnée : état allumé / éteint des lampes, niveau de consommation électrique, de CO2... Mais ces données brutes, sans transformation, ne sont pas très utiles. Elles sont un peu comme le bois pour celui qui souhaite construire un chalet. Un processus de transformation est indispensable pour que le bois devienne chalet. De même, un processus Big data est indispensable pour que ces données deviennent significatives pour un cas d'usage.

DONNER UN SENS AUX DONNÉES

Attention ! Il ne suffit pas d'appliquer un processus prétendument magique sur un ensemble de données hétéroclites pour qu'un enseignement pertinent émerge. Un processus Big data doit passer par différentes phases pour faire ressortir le pouvoir prédictif inclus dans les données manipulées.

Nettoyer les données

Les données brutes doivent d'abord être rendues valides : une ampoule indiquant qu'elle est passée deux fois de suite en mode « allumé » sans être passée par le mode « éteint » constitue une anomalie qui doit être corrigée. Une étape préliminaire de nettoyage des données doit être pratiquée avant toute exploitation pour corriger / valider les données.

Interpréter et contextualiser

Il faut ensuite donner un sens à ces données épurées de toutes anomalies. En effet, savoir qu'une lampe est allumée ou éteinte n'apporte pas d'information en elle-même si ce n'est par son interprétation qu'entre ces deux états, le logement est présumé occupé. Par ailleurs, si l'on recoupe cela avec les données provenant du détecteur de CO2, du détecteur d'ouverture / fermeture de porte, voire de la caméra avec reconnaissance faciale, il est alors

possible de pouvoir conclure plus certainement que le logement est occupé (et par qui). Cette tâche de recouplement

permet d'éliminer le cas où l'occupant serait simplement parti en laissant la lumière allumée (ce qu'on appelle un faux positif).

Il est alors possible, à partir d'un recouplement d'informations et d'une contextualisation à bon escient, de déduire d'une donnée brute plusieurs dizaines de données interprétées ayant acquis un sens. Cela permet d'obtenir des données caractérisant l'intensité, la fréquence, la régularité et plus généralement le mode d'utilisation du bien assuré. Cette étape est cruciale car c'est celle qui permet de dégager les données d'intérêt ayant une signification par rapport à l'objectif « métier » qui aura été assigné au processus Big data. En effet, une mauvaise interprétation à cette étape peut invalider une donnée pertinente et potentiellement aboutir par la suite à des conclusions erronées ou moins précises (laissant passer de faux positifs). Il faut donc allouer un temps suffisant aux indispensables phases de préparation afin de qualifier les données avant de se lancer dans leur analyse.

Analyser

L'analyse consiste à identifier les populations qui ont les caractéristiques recherchées et celles qui en sont divergentes. C'est à ce stade qu'intervient la puissance des traitements massivement parallèles autorisant l'exploitation d'algorithmes machine learning*. Ces algorithmes permettent de distinguer, dans la myriade d'informations, les signaux faibles traduisant un comportement ou



une situation qui ne peut être découverte autrement. Cette approche sera par exemple particulièrement utile pour révéler, parmi les populations actuellement difficilement cernées par les modèles traditionnels, les jeunes automobilistes masculins qui ont des comportements moins à risques.

Le concept du machine learning est totalement nouveau dans le secteur de l'Assurance, dirigé par des processus linéaires et industriels. Associé aux bases de données Big data, le machine learning va permettre d'enrichir tout au long du cycle de vie des produits d'assurance, des analyses de données et des nouvelles corrélations.

Cette technologie n'est pas en opposition avec la Business Intelligence mais bien en complément pour des usages nouveaux. Elle permet d'avoir une vision prescriptive, alors que la Business Intelligence fournit une vision descriptive.

Calibrer

Une fois ces populations différenciées, il reste à distinguer l'intensité de l'effet de cette différenciation sur la variable d'intérêt objet de l'étude. Cette phase de calibrage est importante car il se peut que des caractéristiques très divergentes n'aient que très peu d'impact et vice-versa. C'est ce qui fait que ce processus exploratoire est un processus itératif. En effet, quand la pertinence du modèle n'est pas au rendez-vous, c'est-à-dire que la réponse n'est pas suffisamment

prédictive, il faut recommencer. Il est alors nécessaire d'identifier si une meilleure interprétation / contextualisation des données pourrait apporter plus d'enseignements et/ou si d'autres sources d'information pourraient apporter des éclairages nouveaux.

PROCESSUS ITÉRATIF

L'itération du processus Big data est l'aspect le plus déconcertant pour un assureur habitué au processus linéaire R&D traditionnel. Néanmoins, il s'agit bel et bien de l'ADN de tout processus Big data où le cycle d'accroissement de la connaissance par amélioration des modèles est un processus non seulement itératif, mais continu. À tel point que dans un rythme de croisière, on ne saura plus distinguer l'élément premier dans la démarche :

- / Est-ce la recherche d'événements permettant de caractériser un objectif métier, puis la recherche de données permettant d'identifier cet événement ?
- / Est-ce le processus de qualification permettant de valider l'intérêt de ces données par rapport à l'objectif métier ?

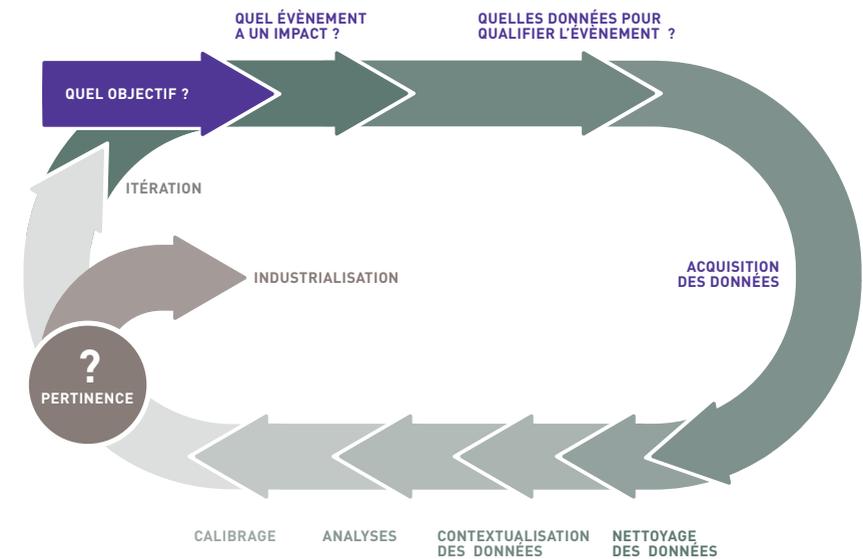
L'ILLUSION DE LA MAGIE : PRÉDICTION DES EFFETS

Le processus d'exploitation des données ne relève pas de la magie. Il s'agit d'un

processus qui tend à reproduire les décisions humaines dont une part est liée à la combinaison d'éléments très diffus (signaux faibles) que la prolifération de nouvelles données rend lisibles et interprétables. Le constat par les algorithmes que les mêmes circonstances produisent les mêmes effets permet d'en prédire la survenance (voire prescrire des mesures de prévention). Et ce, sans qu'il ne soit nécessaire d'établir une relation de cause à effet entre ces circonstances et ces

effets. C'est probablement ce qui donne l'illusion de la magie à un processus tout à fait rationnel.

Stop ou encore : le Big data introduit l'itératif dans le processus R&D



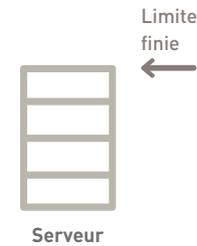
COMPRENDRE LA JUNGLE TECHNOLOGIQUE

Le Big data est né d'une rupture technologique : la distribution des traitements et du stockage. Là où, en 2012, ce concept était novateur, il devient une offre à part entière de la plupart des éditeurs du marché, créant ainsi une jungle technologique dans laquelle il faut savoir se repérer. On compte aujourd'hui plus d'une centaine de solutions packagées sous différentes formes.



La distribution : moteur du Big data

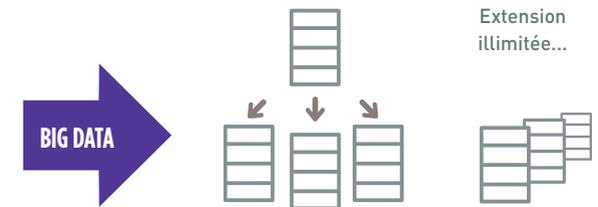
ARCHITECTURE TRADITIONNELLE



MAXIMISER LA PUISSANCE UNITAIRE

La performance est fonction de la puissance unitaire de l'infrastructure sous-jacente

ARCHITECTURE DISTRIBUÉE



EXPLOITER LA PUISSANCE DU NOMBRE

La performance est fonction du nombre d'infrastructures sous-jacentes

Le principe clé du Big data est la distribution des données et des traitements. Il consiste à morceler l'information sur N supports de stockage et à décomposer un traitement applicatif en tâches élémentaires. Ce procédé apporte plus de rapidité dans l'exécution permettant ainsi de traiter davantage d'informations.

NAVIGUER DANS LA DIVERSITÉ DE SOLUTIONS

Quelles solutions répondront aux besoins présents et futurs ?

Trouver la solution ou la combinaison de solutions techniques est une phase critique d'un projet Big data. Or quel que soit le secteur, il n'y a pas de solution Big data universelle.

Le porte-étendard technologique du Big data est Hadoop* - un écosystème open source de distribution du stockage et des traitements sous l'égide de la fondation Apache. L'ensemble des grands éditeurs et constructeurs ont investi dans cette technologie. Toutefois, d'autres types de solutions sont toutes aussi légitimes suivant les usages. Ainsi, dans le cas où la rapidité d'exécution est recherchée, les technologies InMemory sont à privilégier en raison de leur capacité à stocker les données en mémoire et non sur disque. Le NoSQL (Not Only SQL) a quant à lui un avantage dans le monde « pseudo-transactionnel » où il sera un accélérateur dans le stockage et la manipulation de données hors d'une structure relationnelle. La variété des solutions implique une connaissance de l'écosystème technologique et ce dès la phase « test and try ».

En 2015, Hadoop possède plus d'une

vingtaine de modules formant un socle riche. Ces modules sont créés au fur et à mesure par une importante, et dynamique, communauté d'acteurs et développeurs open source. Ils apportent constamment de nouvelles fonctionnalités tout en renforçant l'environnement existant.

Une illustration simple : historiquement Hadoop ne savait traiter que les données massives par des traitements de type batch. Face à ce constat, la communauté a enrichi l'écosystème par la création de modules de type Storm ou Spark permettant de traiter des données unitaires en temps réel et démocratisant encore plus l'usage de plateformes Hadoop.

Cette diversité de fonctions peut conduire à de mauvais choix et par conséquent certaines initiatives à l'échec. Pour tirer le maximum de valeur des données, il faut choisir les modules les plus adaptés aux traitements désirés et aux données manipulées. L'expertise et la connaissance fine sont le prérequis à la création de valeur en vue d'une industrialisation.

LE BIG DATA SOUS DIFFÉRENTES FORMES

Une fois la ou les solutions techniques choisies, il convient de définir le mode de distribution (même si certaines solutions peuvent imposer le leur). Quatre formes sont alors possibles :

1 La distribution construite et gérée par la DSI

L'entreprise construit une infrastructure Big data en agrégeant ses infrastructures conventionnelles via l'intégration d'une distribution. Ce type de solution est principalement destiné aux entreprises ayant un besoin d'évolutivité et d'appropriation fort.

2 L'« appliance » ou solution clé en main

Les grands éditeurs comme Oracle, Microsoft, IBM et les constructeurs proposent une infrastructure sur étagère où matériel et logiciels sont intégrés, préconfigurés et optimisés. Ce modèle a l'avantage d'une efficacité opérationnelle immédiate en contrepartie d'un coût d'investissement qui peut s'avérer élevé.

3 La « plateforme as a Service » ou BDaaS

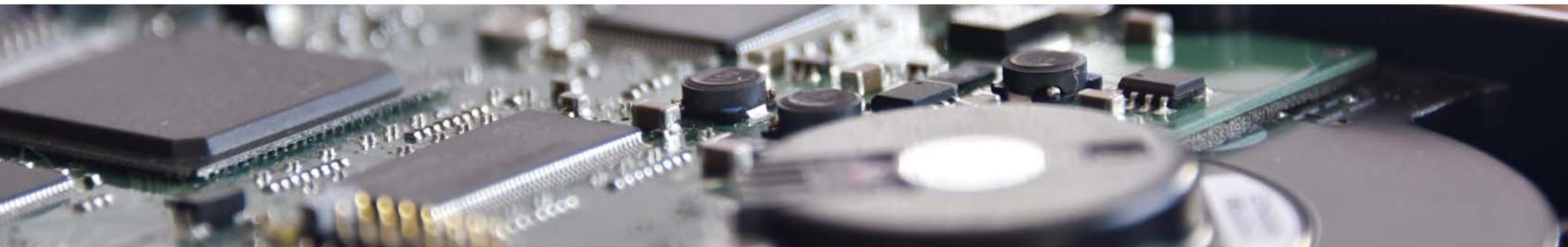
Cette formule connue basée sur le Cloud provisionne un environnement

opérationnel et élastique comprenant l'ensemble des composants Big data pré configurés. L'intérêt de ces offres Cloud est le coût (à l'usage) et la rapidité de mise en place, qui en font une cible parfaite pour des usages de type POC (Proof of concept) afin de tester la valeur du Big data pour l'entreprise. En production, le gain sera plus limité du fait des contraintes liées à l'expatriation des données à l'extérieur de l'entreprise.

4 Le « Software as a Service » ou SaaS

Le SaaS, une formule connue depuis l'arrivée du Cloud, permet la mise à disposition de services métiers sur le web. Le Big data donne un accélérateur à ce modèle en permettant la création de nouveaux services comme l'analyse des réseaux sociaux, l'e-réputation, le marketing ciblé... Ces services reposent sur des moteurs Big data néanmoins invisibles par les utilisateurs finaux. Dans le secteur de l'Assurance, les 4 formules sont légitimes et dépendent réellement du souhait d'investissement et d'appropriation de la technologie. La complexité de ces solutions laisse à penser que le marché va tendre vers une externalisation des problématiques techniques favorisant les offres de type appliance ou Cloud.

* Hadoop : voir glossaire en fin de publication





TÉMOIGNAGE :

Cyrille Vincey, Chief Data Officer de VE Interactive et fondateur de la start-up QUNB

Cyrille Vincey a développé une solution de datavisualisation à partir d'outils tels que Google Insights et Google Analytics pour rendre toutes ces données plus lisibles.

Pourriez-vous nous indiquer l'intérêt que vous avez pour ce domaine ?

Je viens du monde de l'analyse de données notamment sur la fonction du marketing. Faire du Big data signifie disposer d'une forte expertise technique, de beaucoup de curiosité et d'appétence pour la communication. Cette richesse de talents me motive et nous fait avancer dans la compréhension des usages possibles. Le Big data combine à la fois une dimension humaine et technophile. Je suis convaincu du fait que la donnée va transformer la stratégie de l'entreprise de demain.

Quel est votre point de vue sur les technologies ?

On remet au goût du jour des paradigmes techniques imaginés dans les années 60 / 70. À l'époque, des contraintes physiques fortes n'ont pas permis leur essor. La plateforme Open Source Hadoop a été construite au milieu des années 2000 et fait déjà figure d'ancienne. L'écosystème se multiplie de manière pléthorique et rapide. Il est possible d'intégrer soi-même

tous ces outils, sous réserve d'en avoir la capacité. L'avantage est une maîtrise de son infrastructure. L'inconvénient est la nécessité de gérer l'upgrade et la compatibilité de toutes ces solutions. Des distributeurs se sont positionnés sur ce créneau et apportent la garantie de cohérence.

Quelles sont les activités structurantes pour atteindre un résultat ?

Il ne s'agit pas seulement de capturer les données. Il faut les mettre en valeur. Les phases clés sont les phases de nettoyage et de dépollution. Pour cela, une connaissance très fine des contextes Métier et Technique de l'entreprise est primordiale. Nettoyer les données correspond à 80% du travail du Data scientist.

Enfin, même si les infrastructures techniques offrent une puissance élevée, il n'en reste pas moins nécessaire de réaliser un plan de requête adapté pour éviter des temps d'exécution inacceptables.

Quels sont les facteurs clés de succès que vous retenez ?

Tout d'abord, nous sommes tous de grands débutants vis-à-vis de cette matière. Ensuite, la valorisation de la donnée est supérieure à la préoccupation de collecte.

Pour finir, la combinaison de compétences diverses et leur renouvellement régulier permet de garder un œil neuf sur les données et les résultats.

L'objectif est d'être en situation d'étonnement permanent. Les technologies évoluent de manière très rapide. Il est indispensable d'être au fait de celles-ci.

Pour cela, le partage d'expérience est le meilleur moyen de faire de la veille. Il est aussi plus adapté que la documentation.

« Les technologies évoluent de manière très rapide. Il est indispensable d'être au fait de celles-ci. Pour cela, le partage d'expérience est le meilleur moyen de faire de la veille. Il est aussi plus adapté que la documentation »

GÉRER LES RISQUES LIÉS AUX DONNÉES MANIPULÉES

Initier un projet Big data, c'est se poser la question de nouvelles données à collecter, stocker et manipuler pour de nouveaux usages. Autant de besoins qui font peser des risques sur les données concernées et la conformité vis-à-vis des lois de protection des données à caractère personnel.

DONNÉES PERSONNELLES : LE CADRE RÉGLEMENTAIRE

Un cadre légal existe déjà et repose sur la Loi informatique et libertés du 6 janvier 1978, plus particulièrement l'article 34 qui impose aux responsables de traitement de « prendre toutes les précautions utiles, au regard de la nature des données et des risques présentés par le traitement, pour préserver la sécurité des données ».

Ce cadre est en évolution. Le règlement européen sur la protection des données personnelles (GDPR) est applicable le 25/05/2018. A ce règlement, s'ajoute la loi pour la République numérique adoptée le 28/09/2016.

Ces textes sont complétés par des réflexions du G29, groupe de travail des CNIL européennes émettant des opinions sur l'approche par les risques (Article 29 Data Protection Working Party).

Les principes de collecte, transmission, hébergement et traitement des données personnelles doivent suivre des exigences de protection légale très strictes :

- / Recueillir le consentement explicite,
- / Informer sur la finalité,
- / Collecter uniquement les données liées et nécessaires à cette finalité,
- / Détruire les données une fois la finalité atteinte,

- / Respecter les droits à l'accès, la rectification et la suppression de ces données,
- / Sécuriser la collecte, le transfert, l'hébergement et le traitement des données,
- / Effectuer le traitement de façon loyale et licite.

Ceci donne un cadre légal aux démarches Big data avec un processus d'amélioration continue qui requiert parfois plusieurs itérations pour parvenir à un dispositif de protection de la vie privée acceptable. Il exige en outre une surveillance des évolutions dans le temps et des mises à jour régulières.

Pour faciliter la mise en œuvre concrète de ce cadre juridique, la CNIL a publié le 2 juillet 2015 une méthode pour mener une étude d'impacts sur la vie privée (EIVP). L'objectif est d'aider les responsables de traitement dans leur démarche de conformité, de leur permettre de justifier les mesures choisies et de montrer que les solutions ne portent pas atteinte à la vie privée.

Cette méthode est complétée par un guide d'outillage et par un guide de bonnes pratiques.

Cette démarche de conformité se déroule en 4 étapes : délimitation et description du contexte du traitement considéré et de ses enjeux ; identification des mesures existantes ou prévues pour respecter les exigences



légal et traiter les risques sur la vie privée de manière proportionnée ; appréciation des risques sur la vie privée pour vérifier qu'ils sont convenablement traités ; décision de valider la manière dont il est prévu de respecter les

principes de protection de la vie privée et de traiter les risques, ou bien de réviser les étapes précédentes.

LES DIFFÉRENTS TYPES DE DONNÉES MANIPULÉES PAR LE BIG DATA

- **Données personnelles** : toute information, quelle qu'en soit la nature, afférente à une personne identifiée.
- **Données anonymisées** : données qui ont perdu le caractère de « données personnelles » suite à une transformation rendant impossible le rattachement de l'information à une personne identifiée mais uniquement à un groupe de personnes.
- **Données de vie privée** : données personnelles ayant un caractère pouvant remettre en cause le respect du droit à la vie privée (par exemple : géolocalisation).
- **Données sensibles** : données personnelles appartenant à une catégorie particulière dont les traitements sont encadrés. Les données de santé sont des données sensibles. La définition la plus généralement acceptée est celle qui définit la donnée de santé comme caractérisant une pathologie. Les données de bien-être sont des données à la frontière de la santé (poids, nombre de pas, nutrition...) qui, selon certains acteurs, pourraient ne pas relever du caractère « sensible ». Parmi les données sensibles figurent aussi les données relatives à la vie sexuelle, la race, l'ethnie, les convictions politiques, religieuses, philosophiques ou syndicales.

L'ACCÈS AUX DONNÉES DE SANTÉ

L'article 47 du projet de loi de Santé de Marisol Touraine a été adopté par les députés le 11 septembre 2015. Il prévoit la création d'un système national des données de santé (SNDS), grande base de données médicales centralisées accessible au public, afin de faciliter l'accès aux données médicales issues des divers organismes du secteur.

Suivant l'avis favorable de la CNIL à la création d'un numéro d'identification unique, condition sine qua non de la création d'une plateforme numérique unique, le Numéro d'Inscription au Répertoire National (NIR), plus communément appelé numéro de sécurité sociale, deviendra l'identifiant unique de santé.

Pour protéger la vie privée des patients et le secret médical, qui reste la principale critique apportée par le Syndicat des Médecins Libéraux à cet article, de nombreuses garanties ont été apportées par le législateur :

Open data : les données agrégées et anonymisées ne contenant « ni les noms et prénoms des personnes, ni leur numéro d'inscription au répertoire national d'identification des personnes physiques, ni leur adresse » seront accessibles à tous, gratuitement et sans restriction. La réutilisation de ces données sera autorisée si elle n'a « ni pour objet, ni pour effet, d'identifier les personnes concernées ».

Données personnelles : les données rendant l'identification possible seront détenues par une structure distincte et ne pourront être utilisées que sur autorisation de la CNIL et de l'Institut National des Données de Santé (INDS) à des fins de recherche ou d'étude pour l'accomplissement de missions poursuivant un motif d'intérêt public uniquement. L'article 47 interdit expressément l'utilisation de ces données à des fins commerciales ou d'évolution des primes et contrats d'assurance.



CE QU'EN PENSENT LES MÉDECINS

Interview du Docteur Alain Durand, Médecin généraliste

« Les objets connectés pourraient nous aider à détecter les risques de pathologies encourus par nos patients, liés à leurs habitudes comportementales, et à établir un plan de prévention.

Grâce à la connaissance qu'il a de son patient (sociale, familiale et professionnelle), le médecin est à même de l'aider à utiliser des moyens connectés, par la motivation, le suivi et au final l'amélioration de sa qualité de vie. Mais il faut compter avec une population encore hétérogène en termes d'utilisation de l'informatique. Il faudra de nombreuses années pour toucher l'ensemble de la population.

Un échange des données concernant le suivi d'un plan de prévention entre l'Assureur (ou le réseau de soin) et le médecin pourrait probablement nous aider dans notre pratique.

Toutefois, la base inébranlable de notre métier est la confiance que le patient nous accorde et rien ne pourra déroger au secret professionnel. La transmission d'informations devra se faire à l'initiative du patient qui donnera à qui il souhaite les informations qu'il veut »

En leur permettant d'analyser une population donnée, le Big data dans le domaine de la Santé permettrait aux médecins d'améliorer leurs pratiques de prévention et de soin en matière de maladies chroniques, ainsi que leur connaissance de maladies rares et / ou orphelines. Cependant, selon un sondage Odexa paru le 19 janvier 2015, un médecin sur deux considère que l'utilisation de la santé connectée menace le secret médical et 1/4 à 1/3 d'entre eux qu'elle porte atteinte à la liberté des patients. En effet, à l'heure

actuelle, la donnée de santé est qualifiée de « sensible » par la loi « Informatique et libertés » et est par conséquent très protégée. Mais, bien qu'un projet de règlement européen prenant position sur le sujet soit en cours de rédaction, les données de santé et les données de bien-être ne sont toujours pas juridiquement différenciables. Or, si l'on prend l'exemple du « quantified self* », les agissements de l'utilisateur sont enregistrés et peuvent être utilisés. Cette limite doit donc être clarifiée pour lever les réticences des médecins.

* Hadoop : voir glossaire en fin de Synthèse

UN CONTRAT DE CONFIANCE POUR ANTICIPER LES CRAINTES

Citoyens et autorités de régulation des données à caractère personnel expriment des craintes régulièrement relayées par les médias. Maîtriser les risques autour du Big data revient à anticiper ces craintes en communiquant : sur les usages envisagés, la proposition de valeur au regard de ce partage de données, les compétences des équipes en charge des technologies concernées et les mesures de sécurité permettant de garantir la protection des données manipulées.

Il faut donner à l'assuré la maîtrise de ses données personnelles, l'informer sur celles qui seront collectées et l'usage qui en sera fait. Il convient ensuite de lui laisser la possibilité d'arbitrer en permanence sur le partage de celles-ci. Les informations demandées doivent être nécessaires à la réalisation de la finalité. L'objet est d'obtenir un consentement, puis détruire les données brutes pour ne conserver que la finalité. Il

est également envisageable de se limiter aux données anonymisées qui s'avèrent suffisantes dans de nombreux cas. Enfin, un rapport de proportionnalité est impératif : le service proposé grâce à ces données partagées doit être à la hauteur du niveau d'intrusion ressenti.

Ce contrat de confiance pourrait être « géré » via l'espace client / adhérent, donnant à ce dernier la possibilité d'exprimer explicitement son accord ou désaccord avec un historique des décisions.

Tout cela permettra de gagner la confiance du client, de mieux le connaître et de passer du diagnostic à une position prescriptive.

La transparence sur les données utilisées ne fera pas entrave aux secrets de fabrication qui se logent de manière plus structurante sur les algorithmes et sur le volume de données manipulées.

COMMENT SE LANCER ?

Le challenge du Big data est de donner du sens à une masse de données pour aboutir à des cas d'usage générateurs de valeur. Pour se lancer, le meilleur costume est celui de la start-up, dans une dynamique « New Way of Working » : faire émerger les idées, incuber, mobiliser et démultiplier. Ce modèle est le meilleur vecteur pour apporter flexibilité, échange, interaction avec une mixité de profils. S'il fallait quelques mots pour résumer l'esprit dans lequel mener les projets Big data et plus largement les projets de valorisation des données, nous choisirions : créativité, agilité, rythme, expérimentation, simplicité, pluridisciplinarité.



1 - L'IDÉATION

L'émergence des idées est une étape essentielle de prise de recul ayant pour objectif de porter un regard nouveau sur le marché et sur ses propres activités. C'est au travers d'ateliers de créativité que des idées en rupture peuvent émerger. Pour être efficace, cette étape doit être structurée et encadrée : définir les domaines à balayer, identifier les acteurs à réunir par domaine métier, les associer à des acteurs de la DSI et à des spécialistes de la donnée (Data scientists). L'animation est structurante, elle suit des méthodes éprouvées de dynamisation de la créativité. Cette première phase permet de faire émerger les idées les plus prometteuses pour les confronter à l'incubation. Elle peut être renouvelée périodiquement.



2 - L'INCUBATION

L'incubation a pour objectif de concrétiser rapidement les idées sélectionnées. C'est le moment d'affiner ces idées, de vérifier leur faisabilité, de les mettre en œuvre sur de premiers périmètres à coût maîtrisé et de vérifier leur intérêt.

Il s'agit ici de combiner l'approche agile avec le mode projet Test & Learn en privilégiant l'obtention de résultats tangibles sur ces petits périmètres d'expérimentation avant un déploiement à grande échelle. Pour

s'assurer de la dynamique d'ensemble et de l'aboutissement des résultats qu'ils soient positifs ou négatifs, la démarche doit suivre un rythme très cadencé et favoriser le collectif. Il faut s'autoriser à réajuster autant que nécessaire l'idée et sa concrétisation jusqu'à être convaincu de leur intérêt et de leur capacité à être industrialisées.

L'expérimentation permet de prendre concrètement la mesure de l'impact de ces nouvelles technologies et usages sur

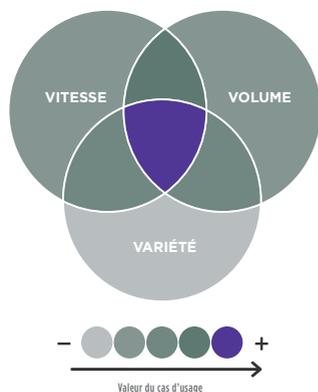
* Le terme anglais churn désigne la perte de clientèle ou d'abonnés.

les équipes informatiques et métiers. Elle permet aussi de capitaliser sur les erreurs commises, normales à ce stade de maturité dans l'exploitation des données et étant donné le foisonnement des technologies.

Les premières expérimentations et premiers projets doivent privilégier la simplicité. Cette simplicité s'entend d'abord par un nombre limité de directions consommatrices. Le résultat attendu doit également être sans impact direct sur le marché et les clients, pour limiter l'effet d'un éventuel échec. Pour finir, se focaliser en priorité sur une seule dimension de la data permet de minimiser la complexité de l'architecture à mettre en œuvre.

La distribution : moteur du Big data

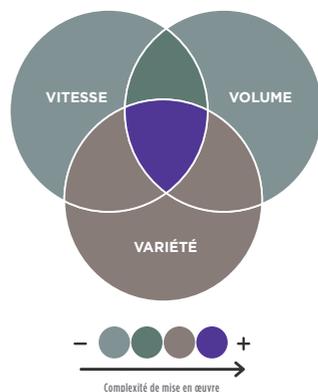
LA COMBINAISON DE LA VITESSE ET DU VOLUME EST GÉNÉRATEUR DE VALEUR, L'UTILISATION DE LA VARIÉTÉ AMÈNE IMMÉDIATEMENT UNE COMPLEXITÉ DE MISE EN ŒUVRE



Aujourd'hui, dans la pratique, nous constatons que nos clients se focalisent principalement sur la volumétrie du Big data. L'aspect variété des données est souvent limité en raison d'une plus grande complexité d'intégration. Un bon niveau de maturité est nécessaire pour explorer toutes les dimensions du Big data et faire valoir toutes ses promesses.

Le diagramme de Venn ci-dessous illustre ce constat et décrit une démarche progressive pour complexifier peu à peu les cas d'usage et leur donner de la valeur.

En un mot, la stratégie gagnante consiste à définir des petits périmètres et à les conquérir pas à pas.



3 - LA MOBILISATION

La réussite de la démarche repose sur la mobilisation des bonnes personnes au bon moment.

L'organisation la plus évidente consiste à mettre en place une équipe cœur, rassemblant et impliquant une large diversité de profils et compétences clés (« sachant » à connaissance métier large, Data scientist, architecte SI, spécialiste des démarches de créativité, juriste...).

Cette équipe cœur vient animer plusieurs équipes pluridisciplinaires mises en place sur des durées plus courtes pour participer à l'idéation et à l'incubation sur des thèmes ciblés.

L'équipe cœur est là pour dynamiser l'ensemble et capitaliser progressivement, tout en diffusant les nouveaux modes de travail et les nouvelles compétences. Elle fournit les méthodes et les outils.

La compétence clé d'analyse des mégadonnées doit au départ être centralisée, l'équipe cœur formant ainsi un guichet unique accélérant l'émergence et le test des idées.

L'équipe cœur aura aussi pour but de préparer la phase suivante, c'est-à-dire l'intégration de la culture de la donnée dans les gènes et dans le fonctionnement courant de l'entreprise. Elle doit identifier les moments clés où les solutions et plateformes techniques associées doivent être

industrialisées.

Elle va pour cela aussi mener des actions plus largement visibles comme des concours d'innovation, des actions de communication, la mise en place de démonstrateurs, avec deux objectifs : légitimer et acculturer.

Dans le cas présent, la difficulté va être d'obtenir et de maintenir cette mobilisation de ressources clés alors que les bénéfices directs sont au départ incertains. Un sponsorship fort est nécessaire, de même que le recours à des accélérateurs externes pour parvenir rapidement à des résultats tangibles.

« L'organisation la plus évidente consiste à mettre en place une équipe cœur, rassemblant et impliquant une large diversité de profils et compétences clés »

DES RÔLES SONT À INCARNER POUR METTRE LA DONNÉE AU CŒUR DU DÉVELOPPEMENT DE L'ENTREPRISE.

Le Data scientist est un nouveau profil clé à faire émerger pour faire le liant entre technologie et valeur de la donnée. Il est à la fois statisticien et expert métier. Il a la capacité à comprendre les enjeux business, les méthodes de traitements de la donnée et les nouvelles technologies.

Les autres profils ciblés sont notamment :

- / les développeurs par la venue des nouveaux paradigmes de développement (nouveaux langages et framework),
- / les architectes par l'usage de nouveaux patterns d'architectures distribuées,
- / les exploitants / techniciens par l'exploitation de nouveaux standards de production,
- / les pilotes des projets, mettant en œuvre de nouveaux modes de travail plus collaboratifs, et qui sauront insuffler le rythme, le bon niveau de communication et compréhension, ainsi que l'osmose nécessaire à la progression de chacune des itérations.

Enfin, une nouvelle fonction s'impose dans le cadre du futur règlement européen pour la protection des données à caractère personnel devant être adopté fin 2015. Il s'agit de la fonction de Data Protection Officer ou responsable de la protection des données, qui remplacera le CIL (Correspondant Informatique et Liberté). Il deviendra incontournable dans toute entreprise concernée par le traitement de données personnelles. Le DPO sera personnellement responsable de la manière dont sont gérées les données personnelles dans l'ensemble des processus de l'entreprise. Il sera également garant du reporting au Contrôleur Européen de la Protection des données (CEPD).

4 - LA DÉMULTIPLICATION

Une fois les premiers retours d'expérience obtenus, il est indispensable de transformer très rapidement ces expérimentations en projets pour accompagner la transformation et éviter l'effet soufflet. Pour aller de l'incubation à la démultiplication, un socle est à construire. Ce socle a 3 piliers : les compétences, le Data Management, le système d'information.

Pilier 1 : les compétences

Pour être en mesure de tirer tout le potentiel des technologies et des données à disposition, une filière Data est à construire pour réorienter une part des profils en place et s'assurer d'une parfaite cohérence avec la transformation engagée. Cette réorientation va passer par une étape d'acculturation puis de formation complémentaire.

Pilier 2 : le Data Management

Tirer la valeur des données est réalisable seulement si l'on connaît le sens des données et que celles-ci sont fiables. Nombre de données internes sont détournées de leur usage premier du fait de systèmes d'information trop permissifs, rendant impossible toute analyse fiable. Pour mener à bien cette mise en valeur de la donnée, des principes d'architecture doivent être intégrés aux processus de gestion et d'évolution des systèmes d'information. Par ailleurs, l'application de règles de sécurité,

d'accès, de ségrégation et d'anonymisation doivent garantir la confidentialité.

Ces règles et principes sont pilotés à travers une gouvernance du Data Management qui se décline comme suit :

- / Sur l'axe métier, il convient de construire un dictionnaire des données, d'identifier de nouvelles sources de données, de veiller au respect de la réglementation, de valoriser et même éventuellement d'identifier les données « monétisables ». La monétisation est encore peu évoquée par les assureurs mais elle pourrait permettre de pérenniser les démarches de Data Management.
- / Sur l'axe fonctionnel, des propriétaires de données doivent être investis

« LES ÉQUIPES MANAGÉ-
RIALES SONT LES PREMIERS
AMBASSADEURS DU PROJET
ET LES RELAIS INDIS-
PENSABLES AUPRÈS DES
ÉQUIPES »

d'un rôle de garant de la fiabilité des données et de leur cycle de vie.

- / Sur l'axe applicatif, un effort doit être consacré à la modélisation des formats canoniques et à la construction des architectures favorisant l'unicité et le partage des données.
- / Enfin, sur l'axe technique, il est nécessaire de construire des offres de services adaptées, optimiser la sauvegarde et l'archivage, sans oublier de suivre et anticiper les augmentations de volumes de données.

La mise en place d'un Data Management structuré est un accélérateur indispensable pour gagner en visibilité sur son patrimoine Data. Elle est à faire vivre et à enrichir avec les données externes qui seront jugées utiles mais également avec la mise en commun de l'ensemble des sources, notamment internes. Pour franchir cette étape, la confiance des différentes entités sur l'utilisation des données est requise. Ce sujet peut être un frein dans certaines organisations. Il est à ce titre nécessaire que chacun prenne conscience de l'augmentation de la valeur des données grâce à la mutualisation de celles-ci.

Pilier 3 : le système d'information

En phase d'industrialisation, l'agilité qui était requise en phase d'incubation va se confronter à la roadmap des projets pluri-annuels et aux maintenances des systèmes existants.

L'organisation en place va devoir s'adapter aux besoins de réactivité des projets Big data. Cette exigence est déroutante pour une majorité de structures organisées autour des méthodologies de cycle en V où l'anticipation est la règle. De nouveaux process industriels techniques sont donc à adapter voire à inventer.

À titre d'exemple, il est nécessaire d'être en capacité de mettre rapidement en place de nouveaux flux de données entre systèmes, sans pour autant dégrader les exigences de sécurité, ni la capacité à absorber une masse d'activité importante par ailleurs.

Le SI doit se doter de plateformes de management de la donnée adaptées aux différents usages. Ces plateformes sont à intégrer dans l'architecture du SI, qui doit elle-même progressivement s'urbaniser, au travers d'une rationalisation des référentiels et des données.

BIG DATA : UNE REVOLUTION QUI PASSE PAR LA CONFIANCE



Le Big data est bien une formidable opportunité et un levier clé pour accélérer et concrétiser la transformation du modèle des assureurs à l'heure du digital. Une véritable mine d'or pour imaginer de nouveaux services, optimiser les modèles opérationnels et trouver de nouveaux positionnements sur la chaîne de valeur.

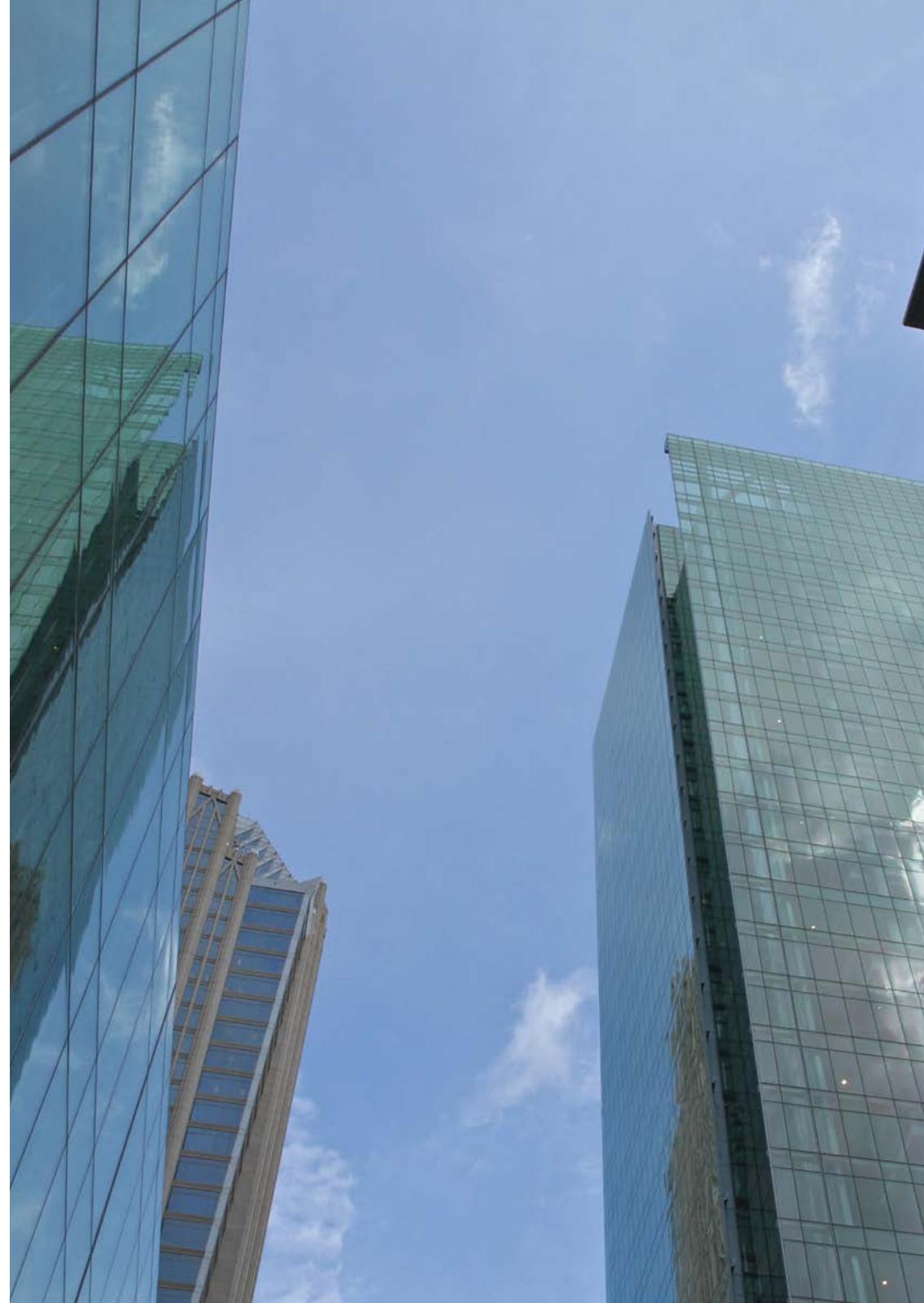
Pour se lancer et réussir, quelques facteurs clés de succès majeurs sont à garder en tête tout au long de cette transformation :

- / Tout d'abord, le sujet est à aborder en partant des enjeux business et en explorant assez largement les cas d'usages potentiels. Cet enjeu business ne doit pas être perdu de vue dans une quête technologique qui aurait pour seul intérêt l'image furtive d'être en ligne avec son temps.
- / Ensuite, il faut rapidement décliner « en vraie grandeur » les idées les plus prometteuses sur la chaîne industrielle de l'Assureur. Pour cela, l'intransigeance sur la qualité des données manipulées est de rigueur. Il convient aussi d'anticiper la mise en place des équipes et des compétences en charge de leur analyse et de leur exploitation.
- / Enfin, il est nécessaire d'apprendre et fiabiliser en avançant. L'analyse des données se fonde sur des corrélations parfois basées sur des « signaux faibles » difficiles à détecter et à interpréter. Des erreurs d'analyse en amont peuvent conduire

à des conclusions erronées ou faire émerger de nouvelles discriminations particulièrement préjudiciables. L'approche Test & Learn en équipe pluridisciplinaire doit être privilégiée, en affinant progressivement les modèles d'analyse. C'est une nouvelle culture et de nouveaux modes de travail qu'il faut s'approprier et diffuser à cette occasion.

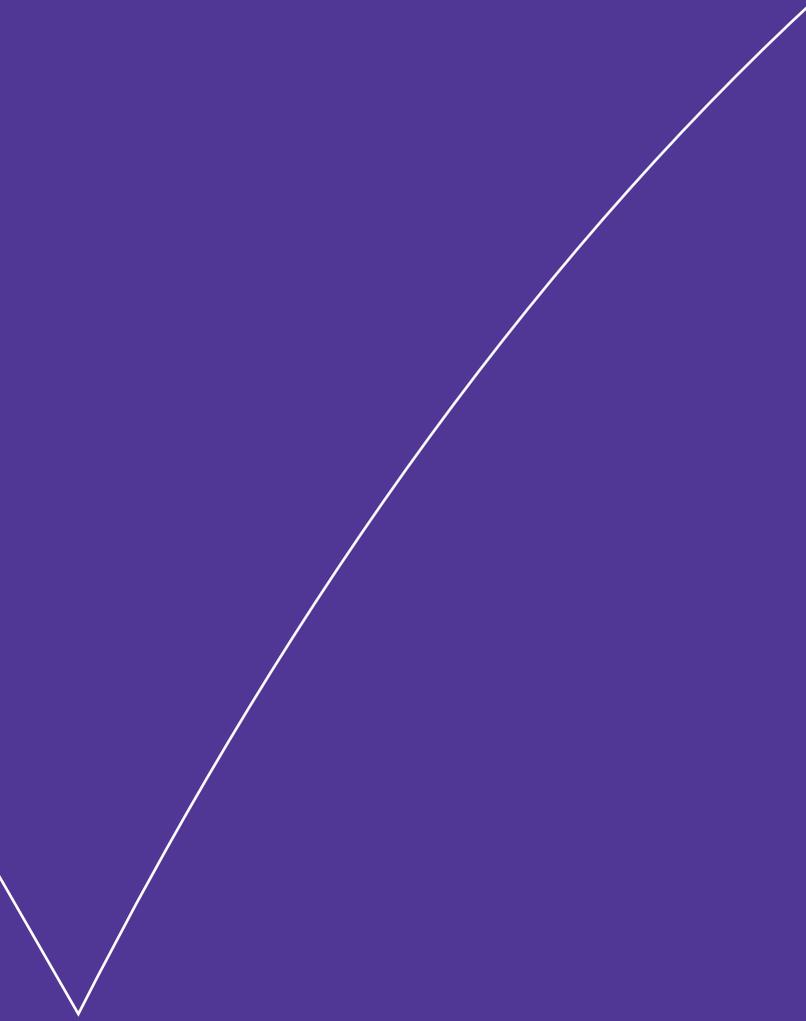
Reste la question du regard que les clients, les assurés, vont porter sur ces nouvelles approches : quelles contraintes, quelles intrusions dans la vie privée et pour quels bénéfices ? Il nous semble que cette question doit être placée au cœur des réflexions des assureurs et doit être abordée dès les premières expérimentations.

Le succès de la révolution du Big data repose en effet largement sur la qualité du pacte de confiance entre assureurs et assurés. Transparence, clarté dans les objectifs et les bénéfices, association des clients dans la démarche sont autant de bonnes pratiques qui pourront donner une longueur d'avance à ceux qui traiteront cet enjeu très en amont.



GLOSSAIRE

- / **3 V** : volume, vitesse et variété désignent originellement les caractéristiques de données Big data. D'autres V ont été rajoutés par différents acteurs, parmi lesquels véracité, valeur...
- / **CEPD** : Contrôleur Européen de la Protection des Données (CEPD). Autorité de contrôle indépendante dont l'objectif principal est de veiller à ce que les institutions européennes respectent et fassent respecter le droit à la vie privée. Le CEPD est assisté d'agents de la protection des données.
- / **Cluster** : grappes de serveurs.
- / **Cookie** : codes enregistrés sur le disque dur de l'ordinateur d'un internaute à la demande du serveur gérant le site web visité. Il permet de recueillir et de stocker des informations sur la navigation effectuée sur les pages de ce site.
- / **CRM** : acronyme de « Customer Relationship Management » ou « Gestion de la Relation client ». Le CRM peut désigner l'ensemble du dispositif informatique consacré à la gestion de la relation client.
- / **CRM Display** : fait de réaliser des campagnes publicitaires display en utilisant des données CRM pour le ciblage et la personnalisation des créations publicitaires.
- / **CRM Onboarding** : technologie permettant de connecter les données off-line d'un CRM à ses différentes plateformes de marketing et de publicités digitales. Cela permet aux entreprises de reconnaître et cibler leurs clients qui naviguent anonymement sur internet.
- / **Datamart** : entrepôt de données ciblées, organisées et regroupées, spécialisé sur un usage métier.
- / **Data Legacy** : donnée stockée dans un format ou système informatique obsolète et dont l'accès ou l'utilisation est difficile.
- / **Data scientist** : le Data scientist a pour mission le traitement et la valorisation de données massives (Big data). Il doit établir une stratégie cohérente pour gérer les données entrantes et les utiliser à des fins commerciales. Le diagramme de Venn décrit les principales compétences qu'il doit avoir : informatique, statistiques et analyse, métier (hacking skills / maths & statistics knowledge / substantive expertise).
- / **Etalab** : service officiel de l'open data français rattaché au Premier Ministre.
- / **GAFAM / GAFAM** : Google, Apple, Facebook, Amazon + Microsoft. Désigne les grandes entreprises américaines qui dominent le marché du numérique.
- / **Hadoop** : framework Open Source écrit en Java, conçu pour réaliser des traitements sur des volumes de données en tirant parti des capacités de calcul massivement parallèles.
- / **Machine Learning** : algorithmes fournissant une analyse prédictive issue d'un auto-apprentissage à partir de l'analyse de données empiriques.
- / **MapReduce** : modèle de programmation et de développement informatique mis en place par Google, dans lequel sont effectués des calculs parallèles de données volumineuses.
- / **NoSQL (Not Only SQL)** : système de gestion de base de données qui ne repose plus sur l'architecture traditionnelle des bases relationnelles.
- / **OBD** : On Board Diagnostic (Diagnostic embarqué à bord). OBD-II est le dernier système électronique embarqué permettant de diagnostiquer le matériel dans la plupart des véhicules à moteur thermique produits dans les années 2000.
- / **Open Data** : données mises à disposition de tous, facilement et gratuitement.
- / **Privacy By Design** : les dispositifs de capture de données doivent intégrer et faire preuve du respect de la vie privée dès leur conception, dans leur fonctionnement en vue de l'élaboration de pratiques responsables.
- / **Quantified self** : applicatif sur smartphone permettant l'apprentissage de soi grâce à l'agrégation de ses données physiologiques. Sans avoir recours à un médecin, l'utilisateur d'objets connectés va pouvoir collecter des mesures sur son état de santé (poids, tension, température, taux d'oxygène dans le sang...).
- / **SNIIRAM** : Système National d'Information Inter-Régimes de l'Assurance Maladie. Il tend à améliorer la qualité des soins, contribuer à une meilleure gestion de l'Assurance Maladie et des politiques de santé, ainsi qu'à transmettre aux prestataires de soins les informations relatives à leur activité.



WAVESTONE

www.wavestone.com