Les travaux de l'Enass

Ecole nationale d'assurances





REMERCIEMENTS

Je remercie mon entreprise, la MAAF, de m'avoir permis de suivre ce MBA CNAM-ENASS avec une majeure Assurance, qui aura été une aventure aussi enrichissante que passionnante. Je remercie, également, M. Jean-François VIGNERON, Directeur de la Direction des Systèmes d'Information MAAF, ainsi que M. Bruno LACOSTE-BADIE, Directeur de la Direction Marketing et Communication MAAF, pour leur soutien lors de mon inscription à ce MBA.

Je remercie aussi, tout particulièrement, mon directeur de thèse M. Jean-François ANDRJANCZYK, Directeur de la direction Actuariat et Compte Technique Covéa, pour son accompagnement, ses conseils et sa disponibilité pour me guider dans mes travaux. Merci à M. Olivier TAPIN et M. Etienne ARBOGAST, respectivement, Responsable Domaine Etudes Statistiques à la MAAF et Responsable Etudes Actuarielles Covéa, pour leurs conseils et la documentation qu'ils m'ont transmis. Je remercie M. et Mme HEURTEBIZE pour leur soutien et recommandations avisées. Merci, également, à M. et Mme Marc NABETH pour leurs soutien et conseils, ainsi qu'à Mme Angélique BAUX anciens élèves du MBA. Merci à M. Franck LAURENCE pour sa relecture et ses conseils.

Merci aux personnes qui ont accepté de me recevoir pour me faire partager leurs connaissances, lors d'interviews :

- Mme Marielle VO VAN LIGER Directrice du Département Relation Client Marketing Stratégique MAAF
- M. Charles DUMARTINET Responsable de Pôle Pôle Risques Majeurs Covéa
- M. Thierry BEILLON Responsable Organisation Méthodes AIS MAAF
- M. Rolf NEUERBURG Responsable du Département Business Intelligence ALLIANZ
- M. Xavier DUCURTIL Directeur de Projets de Transformation Stratégique Covéa
- M. Olivier DESBIEY et M. Pascal PALUT CNIL
- M. Philippe RONDEAU Directeur de la Direction Technique Produit MAAF

Je remercie, également, les membres de ma famille pour leur soutien et leur patience et, plus particulièrement, mon épouse Sandra, qui m'a encouragé, tout au long de cette formation et donné de précieux conseils pour ma thèse.

RESUME

Volontairement ou involontairement, l'humanité génère de plus en plus d'informations. Les progrès, en matière de capture et de stockage, autorisent la conservation de cette masse de données. Des outils et des méthodes permettent, aussi, de traiter ces volumes gigantesques d'informations. Ce phénomène est appelé le Big Data.

L'une des grandes promesses du Big Data est de permettre le développement des entreprises, grâce à l'analyse et la transformation de ces informations. Le besoin se recentre alors sur le client, ce qui de fait, le fidélise et l'incite à acheter.

L'enjeu, dans le domaine de l'assurance, sera d'arriver à personnaliser les offres, tout en maintenant le pilier qu'est la mutualisation. Le Big Data permettra des prédictions plus sûres, grâce au volume et à l'utilisation de nouveaux critères d'analyses.

Tout en restant conforme aux recommandations de la CNIL, l'utilisation des données personnelles de nos assurés paraît inévitable. L'objectif est d'offrir des services et un suivi personnalisé des risques de nos clients, en s'appuyant sur les données qui leurs sont propres. Cet échange, pour se faire dans la légalité, devra être consenti et justifié au regard du service à rendre. L'assuré devra être d'accord pour que l'assureur utilise ses données personnelles. L'assureur, devra l'informer de la manière dont elles sont interprétées et à quelles fins. Il sera responsable de la sécurité de ces données et devra s'engager à les utiliser, uniquement, selon les modalités définies avec leur propriétaire. De plus, il devra offrir des services d'accès et de modifications des données et permettre au propriétaire de les récupérer.

Le Big Data est l'opportunité, pour l'assureur, de renforcer son rôle de conseiller grâce à une meilleure connaissance des assurés et de leurs risques.

Mots-clés : assurance – Big Data – marketing – organisation – client – gestion des risques – Quantified Self – services personnalisés - CNIL

ABSTRACT

Voluntarily or involuntarily, humanity generates more and more information. Progresses that were made in capture and storage technology allowed storing this important volume of data. Tools and methods also allow to handle these gigantic volumes. This phenomenon is called Big Data.

The Big Data will be a useful tool to increase company's profits by analysing and transforming information. This phenomenon allows to focus on the customer needs that will permit to develop loyalty and invites him to buy your products.

The stake for insurance will be to manage personalization of its products while maintaining its pillar that is the mutualisation. Big Data will allow more reliable predictions due to the volume of data and the use of new criteria of analyse.

While staying compliant with CNIL the use of personal data of our clients seems to be the only issue to be competitive in insurance tomorrow. Based on these personal data, we will be in capacity to offer personalized services and risks management. To be legal, this data exchange should be approved by the client and justified in regards of the returned service. The client must be agree and informed by the insurer about the use of his personal data and for which purpose. The insurer is also responsible of the data security and is engaged to use them only according to the modalities defined with their owner. Furthermore, he had to give the possibility to the client to access his data for consulting them, for modifying them and even to get them back.

The Big Data is an opportunity for insurer to reinforce their advisor role with a better knowledge of their clients and their risks.

Key words: insurance - Big Data - marketing - organisation - customer - risks management - personalized services - Quantified Self - CNIL

SOMMAIRE

IN	INTRODUCTION7					
1.	QU	'EST CE QUE LE BIG DATA	15			
	1.1	PRÉSENTATION ET APPROCHE DU BIG DATA	15			
	1.2	LES TECHNIQUES D'ANALYSE ET DE VISUALISATION DU BIG DATA	30			
		1.2.1 L'analyse	30			
		1.2.2 La visualisation	39			
CO	NCL	USION PREMIÈRE PARTIE	43			
2.	LE	S ENJEUX DU BIG DATA	44			
	2.1	IMPACTS ET CHANGEMENTS LIÉS AU BIG DATA	47			
		2.1.1 Les villes où nous vivons	47			
		2.1.2 Le domaine de la santé	49			
		2.1.3 Le commerce	51			
		2.1.4 L'Etat et le secteur public	52			
		2.1.5 Le secteur industriel	54			
	2.2	LES ENJEUX DANS L'ASSURANCE	56			
		2.2.1 Les changements dans le métier de l'assurance	58			
		2.2.2 Les impacts organisationnels dans les entreprises d'assurance	63			
		2.2.3 Gouvernance des données	64			
		2.2.4 Evolution des Rôles	68			
	2.3	GOUVERNANCE ET DÉROULEMENT D'UN PROJET BIG DATA	70			
CO	NCL	USION DE LA DEUXIÈME PARTIE	74			
3.	LE	S RISQUES ASSOCIÉS AU BIG DATA	75			
	3.1	CLASSIFICATION DES RISQUES EN FONCTION DE L'IMPACT	76			

Le Big Data dans l'Assurance

3.2	L'UNIVERS DES RISQUES DU BIG DATA POUR UNE SOCIÉTÉ D'ASSURANCE	
	3.2.1 Cartographie des risques externes avant mitigation	79
	3.2.2 Mitigation des risques externes	80
	3.2.3 Cartographie des risques externes après mitigation	81
	3.2.4 Cartographie des risques Entreprise avant mitigation	84
	3.2.5 Mitigation des risques Entreprise	85
	3.2.6 Cartographie des risques Entreprise après mitigation	87
3.3	LA CNIL	
3.4	L'ASSURANCE ET L'UTILISATION DES DONNÉES	92
CONCL	USION DE LA TROISIÈME PARTIE	99
CONCL	USION	100

INTRODUCTION

Depuis ces dix dernières années, les informations disponibles via Internet et les différents objets connectés ne cessent d'augmenter. Les entreprises ont elles aussi accumulées des informations utilisées ou non sur leurs clients. Ces données recueillies sur le web (réseaux sociaux, plateforme téléphoniques, messagerie électronique, ...), open data (données en accès libre), géolocalisation des personnes, ouvrent des possibilités de personnalisation des produits. Les entreprises voient dans ce phénomène l'occasion de cibler très précisément les attentes des clients.

Le terme Big Data désigne ce volume d'informations.

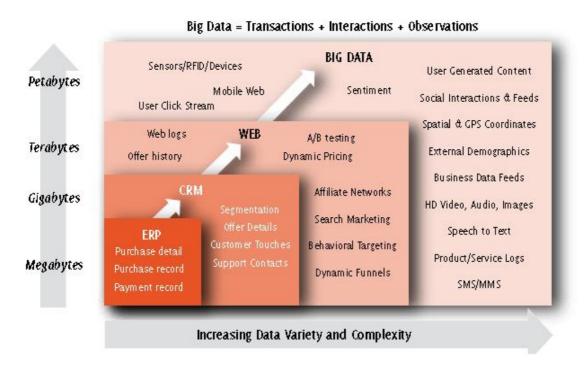
Tout comme les énergies telles que l'électricité et le pétrole ont marqué la révolution industrielle au XXème siècle, le XXIème siècle sera celui de la donnée et de la manière de l'interpréter.

« Nous construisons la France de la souveraineté numérique. Le développement de l'Internet nomade, avec les smart phones et les tablettes, entraîne un véritable déluge de données numériques que renforce encore l'arrivée dans notre quotidien des objets connectés.

Les données deviennent ainsi le carburant de l'économie numérique. Tous les secteurs économiques, du commerce au secteur automobile en passant par le secteur énergétique, tous les domaines de la vie quotidienne (santé, éducation...) sont concernés. Le Big Data regroupe à la fois le traitement de ces grandes masses de données, leur collecte, leur stockage jusqu'à leur visualisation et leur analyse. Comment mettre en relation toutes ces données ? Comment les faire parler ? Comment les mettre en valeur au service des consommateurs, de nos concitoyens ? »¹

.

¹ La Nouvelle France Industrielle: Ce rapport a été conduit par la Direction générale de la Compétitivité, de l'Industrie et des Services (DGCIS) du Ministère du Redressement productif (ministre Arnaud Montebourg), appuyé par l'expertise internationale du cabinet McKinsey. Ce chantier a été mené en lien avec les pôles de compétitivité et les comités stratégiques de filières au sein desquels les chefs d'entreprises, les partenaires sociaux, les administrations concernées et les fédérations professionnelles sont représentés et prennent une part active.



Le principe du « BIG DATA »²

Le schéma, ci-dessus, représente la masse de données et son évolution croissante au travers des différentes technologies interconnectées. Les entreprises au travers des ERP³ et des CRM⁴ disposent de données leur permettant un suivi de leurs activités et de leurs clients. A cela viennent s'ajouter les données disponibles sur le Web au travers, par exemple, des moteurs de recherche comme Google. Les clients se sont appropriés le WEB via les réseaux sociaux, échangeant ainsi des avis, des vidéos et autres médias sur les entreprises dont ils sont clients. Afin de suivre cette activité parallèle, les entreprises ont, à leur tour, investi ces réseaux pour écouter les avis de leurs clients internautes et, ainsi, essayer de s'adapter pour leur donner satisfaction. Cette masse de données accumulées depuis quelques années, contient une matière riche d'enseignements sur la connaissance client.

² VINT Research report 1 of 4 – Creating clarity with Big Data – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren - Sandeep Sachdeva - SOGETI

³ Enterprise Resource Planning : logiciel correspondant pour une organisation, au support de base capable d'assurer une « gestion intégrée », définie comme étant l'interconnexion et l'intégration de l'ensemble des fonctions de l'entreprise dans un système informatique centralisé.

⁴ Customer Relationship Management : outil permettant d'adresser tous les aspects de la relation client –service client, ventes et marketing.

D'après le Gartner⁵, le « Big Data » peut se définir avec le concept des 3V :

- Volume des données disponibles : 2.8 zettaoctets (10²¹ octets) en 2012 pour atteindre 40 zettaoctets en 2020.
- Variété : données semi structurées voire non structurées complexes et de tous types (textes, images, ...)
- Vélocité : fréquence à laquelle les données sont générées, capturées et partagées. (nombre de tweets à la seconde par exemple)

« Le monde traditionnel de la donnée et de la transaction rejoint celui de l'interaction et de l'observation donnant naissance à la formule suivante :

Big Data = Transactions + Interactions + Observations, le but étant de générer plus d'affaires, une productivité accrue et de nouvelles opportunités commerciales. »⁶

Dans un premier temps, il s'agira de bien comprendre ce que les entreprises peuvent mettre derrière le terme Big Data.

Ce volume de données est, aujourd'hui, disponible grâce aux capacités de stockage qui ont augmenté, alors que leur coût a diminué, de manière conséquente, durant ces dix dernières années. De plus, des outils permettent, aujourd'hui, de traiter et d'assimiler plus rapidement cette masse de données.

Une des innovations, rendant possible le stockage et le partage de quantités importantes de données, est le « Cloud ».

Le « Cloud Computing » est un modèle permettant l'accès à des ressources informatiques partagées telles que des réseaux, des serveurs, des capacités de stockage importantes, des applications et des services. ⁷

⁵ Entreprise de recherches et de conseils en technologies dont l'expertise est reconnue dans le monde entier : http://www.gartner.com/technology/home.jsp

⁶ Traduit du VINT Research report 1 of 4 – Creating clarity with Big Data – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren – Sandeep Sachdeva - SOGETI

⁷ Basé sur la definition du <u>www.nist.gov</u>: National Institute of Standards and Technology

Mais comment tirer partie de ces données sans remettre en cause nos fondamentaux et faire de ces informations une plus value pour le métier d'assureur ?

L'enjeu des entreprises d'assurance, face au Big Data, se situe, aujourd'hui, dans l'étude de ces données et la construction de modèles capables de fournir des analyses « Data Mininig », en adéquation avec les enjeux business.

« Le « Data Mining » apporte aux entreprises une connaissance approfondie de leurs clients, en termes de profil, de comportement d'achat et de consommation des produits. C'est un préalable indispensable à l'apport de la meilleure réponse possible aux attentes de chacun de leurs clients. Très concrètement, l'utilisation des techniques de « Data Mining » permet à une entreprise, dans le domaine du marketing de :

- Connaître précisément, sa clientèle pour optimiser son offre (étude de profiling)
- Evaluer le potentiel de chacun de ses clients (life time value)
- Segmenter ses clients, pour optimiser la rentabilité des campagnes (segmentation)
- Cibler, plus efficacement, ses clients (scoring)
- Prévenir le départ de tel ou tel client et minimiser le risque de churn (change and turn : résiliation) (taux d'attrition)
- Proposer le bon produit, au bon client, via le bon canal (appétence)
- Analyser le panier d'achats, pour optimiser son merchandising (cross et up selling)
- Prédire les comportements d'achats futurs et le ROI (modélisation prédictive) »⁸

Quels sont les possibilités et les enjeux de l'interprétation de ces données pour les entreprises d'assurance ?

Comprendre le passé pour anticiper le futur et mettre en place les actions qui en découlent. Les données internes, longtemps exploitées, à posteriori, pour les bilans et les suivis, permettent de comprendre et d'améliorer le fonctionnement d'une entreprise.

⁸ http://www.autarcia.com/data-mining/

En y ajoutant les données issues du web, jusqu'alors impossibles à stocker et à interpréter, nous ouvrons des possibilités sur l'anticipation, plutôt que sur la réaction. Le choix de ces données, les corrélations possibles avec les données existantes, ainsi que les modèles mathématiques à appliquer pour les faire parler, représentent tout l'enjeu du Big Data.

Cette analyse permettra de proposer, une ou plusieurs approches, pour déterminer les attentes des consommateurs d'assurance et, ainsi, ajuster la stratégie de l'entreprise.

Le schéma ci-dessous représente ce principe d'utilisation des données.



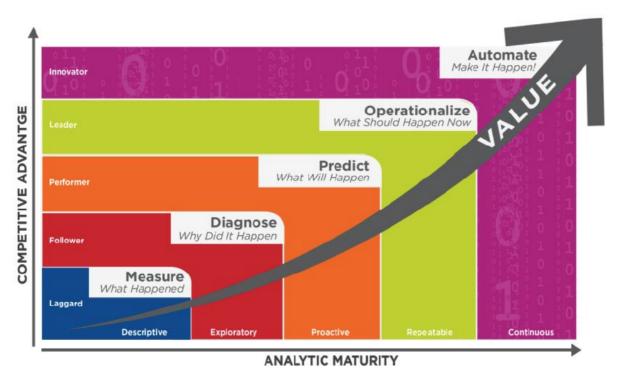
<u>Understand – Anticipate - Act</u>⁹

Une stratégie sur la gouvernance des données doit donc être mise en place dans chaque société, qui souhaite utiliser le Big Data.

⁹ Salon « transform 2013 » : **TIBCO Software**

Au regard de ces évolutions, l'adaptation de nos organisations et de nos méthodes de travail doit être envisagée. La mise en place de ces nouvelles méthodes de travail nécessite de nouvelles compétences pour les entreprises d'assurances. Un de ces profils, le plus central pour le Big Data actuellement, est le « Data Scientist », analyste des données.

Le MIT¹⁰ Technology Review rappelle que le terme « Data Scientist » a été inventé, en 2008, dans la Silicon Valley, par deux analystes employés par LinkedIn¹¹ et Facebook¹². « La mission d'un « Data Scientist » est de récupérer des milliers, voir des millions d'informations, désormais disponibles grâce à Internet sur les consommateurs, et essayer d'en tirer, sinon une stratégie pour son entreprise, au moins quelques actions concrètes à mettre en œuvre. » ¹³Ci-dessous sont représentées les différentes étapes au travers desquelles les données doivent passer pour prendre de la valeur et se transformer en avantage compétitif.



Utilisation du « Big Data » 14

¹⁰ Le Massachusetts Institute of Technology ou MIT, en français Institut de technologie du Massachusetts, est une institution de recherche et une université américaine, spécialisée dans les domaines de la science et de la technologie

¹¹ réseau social permettant de mettre des professionnels en contact

¹² réseau social

¹³ http://lentreprise.lexpress.fr/formation/quand-je-serai-grand-je-serai-data-scientist

¹⁴ Salon « transform 2013 » : **TIBCO Software**

Le Big Data dans l'Assurance

Cette masse de données ouvre aussi la possibilité d'être plus efficace en terme de prévention des sinistres. Le volume de données accessibles laisse entrevoir des perspectives telles qu'une meilleure maîtrise des risques et une meilleure connaissance des usages clients.

Accompagner nos clients, les prévenir, être auprès d'eux pour les aider à anticiper les risques, toutes ces notions seront rendues possibles, grâce au Big Data.

Ce travail sur la prévention des sinistres permettra aussi de lutter plus efficacement contre la fraude.

Cette avalanche de données cependant n'est pas sans risque pour la protection de la vie privée des assurés.

Les autorités Européenne et la CNIL¹⁵, veillent à ce que les entreprises possédant et gérant ces données mettent en œuvre les politiques nécessaires, afin de sécuriser l'accès et la consultation de ces informations. La loi prévoit aussi la possibilité, pour tout individu, d'avoir accès aux données qui le concernent et de pouvoir les modifier.

D'autre part, l'analyse de ces données entraîne des risques auxquels les autorités sont attentives. En effet, le rapprochement de données issues de sources différentes, sur un même individu peut permettre de déduire des informations à caractère privé.

A quel type d'analyse les entreprises d'assurances peuvent-elles prétendre ?

Quelles données vont-elles pouvoir interpréter afin de rester compétitives, sans être mises en cause pour atteinte à la vie privée de leurs clients ?

Dans ce contexte, un nouveau mouvement fait l'objet de réflexions : le « Quantified Self », « il s'agit de capturer, d'analyser et de partager ses données personnels » ¹⁶.

L'objectif est de mettre en commun des bonnes pratiques dans le but d'améliorer la connaissance de soi et de ses propres performances.

¹⁵ Commission Nationale de l'Informatique et des Libertés

¹⁶ Emmanuel Gadenne. Consultant Internet. Manager chez Sopra Consulting. Co-fondateur de QuantifiedSelf Paris

Le Big Data dans l'Assurance

Dans cette nouvelle approche, le fournisseur de service est aussi responsable, de la sécurité et de la protection des données sensibles mises à sa disposition par ses clients.

Dans cet échange consentant, l'assureur a un rôle important à jouer, au niveau professionnel et social, en fournissant les services adaptés à ses clients pour qu'ils puissent mesurer et améliorer leur assurabilité.

1. Qu'est ce que le Big Data

1.1 Presentation et approche du Big Data

Pour rappel, le Big Data se définit selon les trois 'V', volume, variété et vélocité. On ajoute un quatrième 'V' qui est la valeur que l'on pourra attribuer à ce volume et cette variété de données.

Le Volume

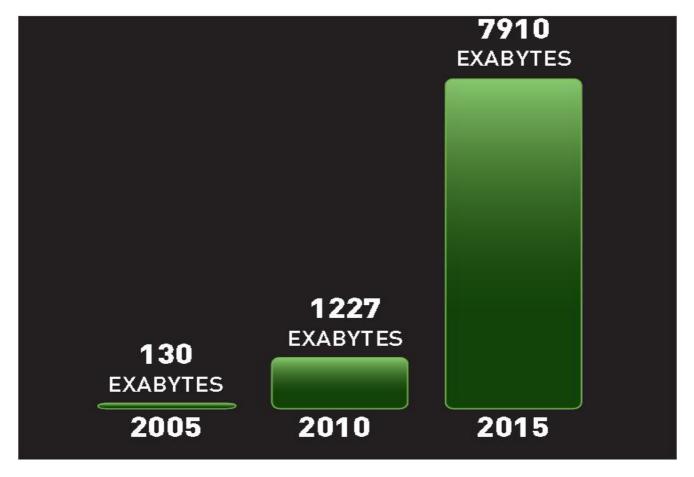
Derrière le terme « Big » se cache un volume de données, jamais atteint jusqu'à aujourd'hui. Le flux de données étant continu, le volume ne fait que croître chaque jour. On connaît aujourd'hui avec le Web 2.0, une avalanche d'informations laissées par les internautes sur les réseaux sociaux, entre autres. A cela, viennent maintenant s'ajouter les objets connectés. On parle de smartphones et de tablettes, géolocalisés, capables de recevoir et d'envoyer des informations de manière continue.

De plus en plus d'objets, deviennent personnalisables et communiquants :

- ° Les voitures avec l'état d'usure de certaines pièces motrices
- Envoi de données liées à la façon de conduire (freinage, temps de pause respecté, distances de sécurité, ...)
- L'e-call qui envoie un signal en cas d'accident
- Des sièges de voitures intelligents capables de détecter les crises cardiaques et d'envoyer un signal
- Des maisons connectées avec des détecteurs d'incendies, des caméras de surveillance capables de communiquer via une connexion internet des informations
- USAA (United Services Automobile Association), un assureur américain, a obtenu une licence pour un appareil pouvant être installé dans les maisons et permettant :
 - o de détecter des différences de températures
 - o de mesurer la vitesse du vent
 - o de détecter des vibrations

o de mesurer le taux d'humidité¹⁷

Tous ces nouveaux appareils fabriquent encore plus de données et viennent, à leur tour, enrichir le volume existant.



Prévisions du volume mondial de données créées (source : EMC)¹⁸

Ce volume devrait atteindre entre 35 et 40 zéttaoctets (1 zéttaoctet = 10^{21} octets) en 2020.

 $^{^{17}}$ basé sur l'interview de Thierry Beillon - Responsable organisation méthode Direction Générale AIS du 02/12/2013 18 www.emc.com

Certaines solutions de stockage et d'utilisation de ressources sont aujourd'hui disponibles via le « Cloud Computing ». « Le « Cloud » permet l'accès via un réseau de télécommunications, à la demande ou en libre service, à des ressources informatiques partagées. » ¹⁹

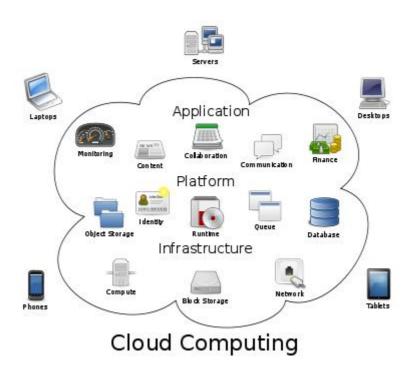


Illustration Wikipedia

On note sur ce schéma l'accès à des ressources de stockage, mais aussi à des infrastructures permettant de traiter ces volumes, ce qui diminue le temps nécessaire à ces opérations. On notera aussi la possibilité d'accéder aux architectures « Cloud », à partir de n'importe quel objet connecté, quel que soit l'endroit où il se trouve.

¹⁹ Traduction de la définition du National Institute of Standards and Technology (NIST)

Le Big Data dans l'Assurance

La variété

Ces données arrivent de différentes sources et sont de nature variées : SMS, Tweets, réseaux sociaux, messageries, vidéo, ...

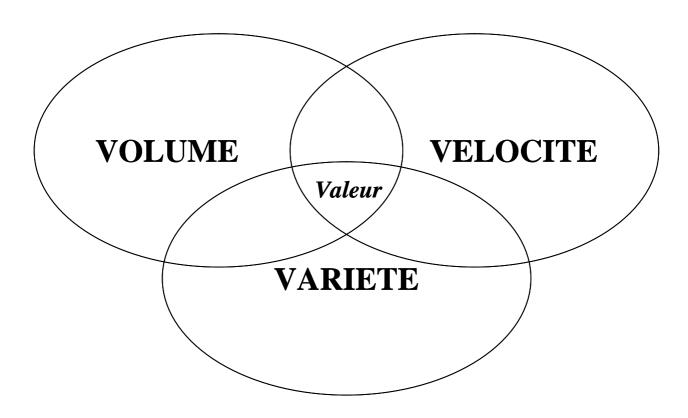
Elles représentent une somme d'informations conséquentes et sont issues de différents domaines.

La vélocité

Il s'agit de la rapidité avec laquelle ces données sont produites, mises à disposition et interprétées.

La valeur

Un quatrième V, la valeur, peut être ajouté. En effet, l'objectif des entreprises est de tirer parti de ce déluge d'informations, notamment en exploitant cette masse de données non structurées pour leur donner du sens.



Caractéristiques du Big Data »²⁰

La somme de ces 4 'V' ouvre des perspectives notamment sur la création de nouveaux produits, sur une meilleure maîtrise des risques et une segmentation plus fine des clients. Des actions pour anticiper et réduire les résiliations, où fidéliser le client, sont aussi envisageables ...

 $^{^{20}}$ Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Le Big Data dans l'Assurance

Plusieurs outils arrivent aujourd'hui à maturité et permettent de traiter ces volumes variés de données.

Dès le début des années 2000, les géants du Web tels que Google, Yahoo, Facebook, ont été les premiers confrontés à cette masse de données à traiter. Pour ce faire, ils ont imaginé des solutions répondant aux exigences suivantes :

- Traitement, stockage et intégration de données volumineuses, structurées ou non, en fonction des diverses sources de données existantes
- O Architecture parallèle, basée sur la technique des clusters (grappe). Il s'agit de mettre ensemble plusieurs serveurs pour répartir la charge d'une tâche ou bien d'utiliser plusieurs serveurs, en même temps, pour effectuer des calculs et ainsi être plus rapide.
- o Analyse, en temps réel, de données non structurées
- Scalabilité: adaptabilité au volume et à la charge, que représentent ces masses de données à traiter.
- o Performance et haute disponibilité des applications et de l'accès aux données.
- o Coût raisonnable, pour ne pas avoir une explosion du coût lié à la scalabilité des architectures.
- Polymorphisme : la plupart des solutions Big Data proposées sont disponibles, via plusieurs langages.²¹

La plupart de ces solutions font appel à un duo « Stockage – Traitement » et font parties, aujourd'hui, du monde open source.

²¹ Basé sur le livre de Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Société	Technologie	Type de Technologie
	Développée	
Google	Google Bigtable	Système de base de données propriétaire reposant sur GFS (Google File System). Technologie non open source, mais qui a inspiré HBase qui est open source.
	MapReduce	Plate-forme de développement pour traitements distribués.
YAHOO!		Plate-forme Java destinée aux applications distribuées et à la gestion intensive des données. Issue à l'origine de Google Big Table, MapReduce et Google File System.
	\$4 distributed stream computing platform	Plate-forme de développement dédiée aux applications de traitement continu des flux de données.
	Cassandra	Base de donnée de type NoSQL et distribuée.
facebook.	HIVE	Logiciel d'analyse de données utilisant Hadoop

	Storm	Plate-forme de traitement de données massives.
twitter	FlockDB	Base de données distribuées de type graphe.
	Apache Kafka A high-throughput distributed messaging system.	Système distribué de gestion de messages
Linked in.	SenseiDB	Base de données temps réel distribuée et semi structurée
	Project Voldemort A distributed database.	Base de données distribuée destinée aux très grosses volumétries

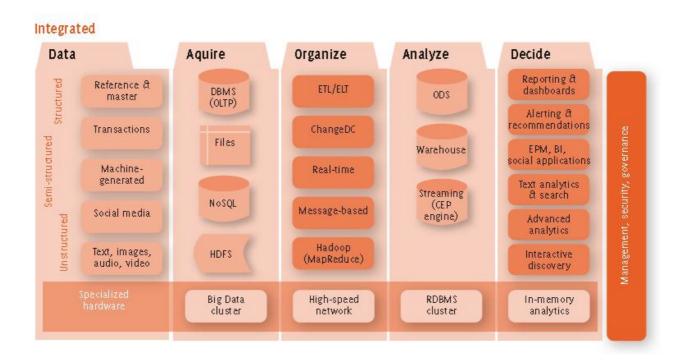
Quelques technologies Big Data chez les sociétés du Web²²

²² Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Le schéma ci-dessous illustre les différentes étapes à mettre en œuvre pour tirer profit du Big Data dans une entreprise et faire cohabiter ces nouvelles données avec l'existant.²³

Dans un premier temps, il s'agit d'acquérir les données et de les charger dans différentes bases de données (relationnelles ou non → NoSQL). Ensuite, on croise ces données pour les organiser et les charger avec différents outils pour les analyser. On combine ensuite les données de ces différentes sources pour les transformer dans un schéma unique et aligner les concepts. On élimine les doublons, on complète les données, on résout les contradictions, en interagissant avec des humains (crowdsourcing). Enfin, on interroge les entrepôts de données ainsi créés, on visualise les résultats sous diverses formes. On analyse les données pour découvrir des corrélations. ²⁴

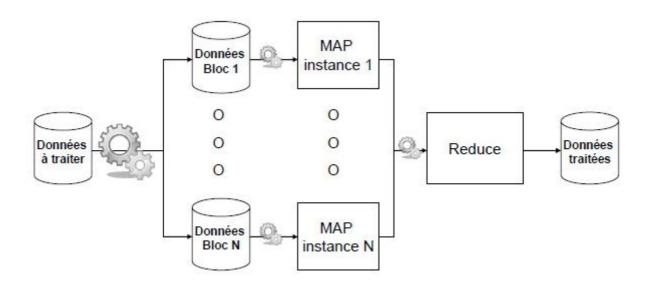
L'objectif final étant de développer de nouveaux services personnalisés pour nos assurés.



²³ VINT Research report – No more secrets with Big Data Analytics – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren - Sandeep Sachdeva - SOGETI

²⁴ Inspiré de Vers une Nouvelle Science des Risques - Serge Abiteboul – Revue Risques n°95 Novembre 2013.

Sans présenter dans le détail une architecture technique Big Data, je pense qu'il est important d'expliquer le fonctionnement du MapReduce qui est une des composantes et un des fondements du Big Data.



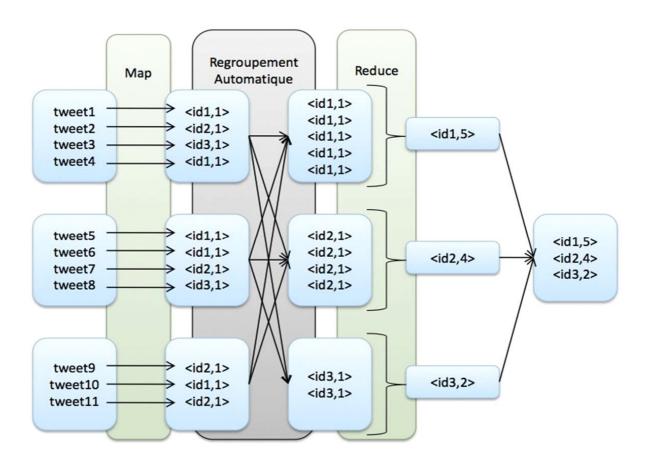
Principe de fonctionnement du MapReduce ²⁵

Comme expliqué précédemment, l'utilisation de plusieurs serveurs, en parallèle, permet le traitement de plusieurs tâches, en même temps (Map).

Les grandes quantités de données traitées sont découpées en plusieurs fichiers, répartis entre les différents serveurs.

 $^{^{25}} http://www.journaldunet.com/developpeur/outils/les-solutions-du-big-data/principe-de-fonctionnement-de-mapreduce.shtml$

« La fonction Map accomplit une opération spécifique sur chaque élément, d'où la parallélisation des traitements afin que ces derniers soient plus rapidement exécutés. L'opération Reduce combine les éléments selon un algorithme particulier, et fournit le résultat. Soulignons que le principe de délégation peut être récursif : les nœuds à qui sont confiées les tâches peuvent aussi déléguer des opérations à d'autres nœuds. »²⁶



<u>Un exemple d'algorithme MapReduce : le calcul de nombre de tweets par utilisateur²⁷</u>

²⁷ http://datascience.bluestone.fr/blog/les-origines-les-grands-principes-du-big-data

²⁶ Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Le principe de fonctionnement est le suivant :

Un découpage des fichiers de données à analyser sur des nœuds de calcul (clusters), extensibles à l'infini, en fonction des besoins (volume de données, impératifs de performance),

Etape Map : La création d'une association <clé, valeur> pour chaque observation des fichiers partitionnés, et la redistribution des clés sur les clusters,

Etape Reduce : La réalisation d'une opération mathématique d'agrégation de valeurs, par clé, et l'envoi des résultats agrégés au client.

Aujourd'hui les technologies offrent une vision des données structurée et, de ce fait, statique. Les limites, en terme de performance et de stockage relationnel, réduisent les possibilités d'analyse à des ensembles explicites de données.

Le décisionnel tel que nous le connaissons aujourd'hui, se base sur une matière cohérente. Les actuaires effectuent leurs études sur des données structurées, bien rangées et des échantillons limités.

Les indicateurs, issus des techniques décisionnelles, sont construits selon des axes d'analyse prédéfinis. Pour un indicateur chiffre d'affaires, par exemple, il sera analysé selon les axes agences (localisation géographique), clients et produits vendus. La combinaison de ces axes pourra permettre d'autres types d'analyses, mais nous évoluons dans un univers limité. Toutes ces données et ces axes sont fournis par le système d'information et saisis par les utilisateurs des applications. Le système décisionnel, ainsi construit, sera à l'image du système d'information. L'obtention de ces données et leur visualisation dans des tableaux de bord peuvent, parfois, prendre jusqu'à une semaine. La plupart du temps, les résultats obtenus ne le sont pas, en temps réel, le temps d'extraction et d'agrégation des données, étant très long.

Ces outils permettent de piloter et d'étudier des portefeuilles de clients importants. Nous sommes dans le « mass marketing ».

A contrario, le Big Data est une petite révolution à plusieurs titres. Tout d'abord, le volume et la possibilité de stocker plus de données personnelles permettent, ainsi, d'entrer dans l'air du « personnal marketing ». Grâce à la mise en œuvre de nouvelles puissances de calcul, nous ne sommes plus limités à quelques données pour effectuer nos évaluations de risques assurantiels.

Cela ne signifie pas que le Big Data va remplacer la « business Intelligence », que nous connaissons. Il s'agit d'un nouvel outil, qui doit s'intégrer et devenir complémentaire du pilotage actuel, en proposant des nouveaux axes d'analyse. Le Big Data va permettre de travailler sur des volumes de données conséquents, non utilisés par le « BI », pour trouver des corrélations et cela plus rapidement.

Par exemple, les résiliations de masse, chez les opérateurs téléphoniques, peuvent, parfois, s'expliquer par un phénomène social qui est le suivant :

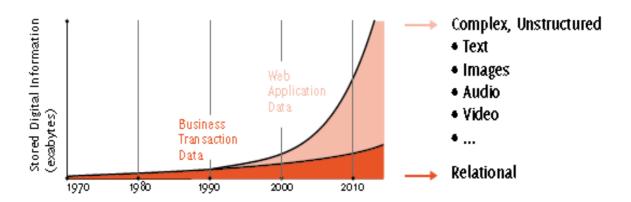
si une personne trouve un opérateur, meilleur marché, à service équivalent, il va en faire part à ses amis et sa famille, qui vont le rejoindre et, ce faisant, résilier leur forfait chez leur ancien opérateur.

Pour identifier ce type de relation et réagir, face à ce genre de phénomène, l'analyse de tout le fichier client est nécessaire, un échantillon ne suffirait pas, car il ne contiendrait pas la complétude des relations possibles.²⁸

Grâce aux nouvelles techniques et outils associés dans le cadre du Big Data, il devient possible de travailler sur des volumes de données importants. Il en découle deux avantages. Premièrement, la possibilité de mettre en évidence des corrélations, jusqu'ici inconnues, avec des nouvelles données. Deuxièmement, asseoir des théories, grâce à la fréquence élevée d'apparition de la donnée dans le volume.

On peut, aussi, envisager d'avoir des analyses dynamiques et, en temps réel, grâce à la vitesse à laquelle les techniques et outils rattachés au Big Data permettent d'obtenir ces résultats.

²⁸ Basé sur l'interview de Marielle Vo-Van-Liger du 04/11/2013



Volume de données structurées et non structurées ²⁹

Ce schéma illustre, parfaitement, le fait que le volume de données exploité par les entreprises est bien, en deçà de ce qui existe. Le rapport entre les données structurées et les données non structurées illustré, ci-dessus, affiche clairement la marche à franchir, en terme d'exploration de données. On notera, aussi, la variété des sources existantes pour les données non structurées, ce qui a longtemps complexifié leur stockage et leur analyse.

Une autre source de données, aujourd'hui mise à disposition et qui a permis la démocratisation du Big Data est l'Open Data. Il s'agit de données anonymisées, mises à disposition du public.

Plusieurs pays ont, par exemple, décidé de mettre à disposition des données publiques, collectées par leurs administrations, pour permettre des utilisations d'intérêt général.

En France, ces données sont consultables sur le site http://www.etalab.gouv.fr/.

²⁹ VINT Research report 1 of 4 – Creating clarity with Big Data – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren - Sandeep Sachdeva - SOGETI

On peut, ainsi, avoir accès à un volume de données important rangées par domaine. Voici quelques exemples :

- ° Agriculture et Alimentation
- ° Données sur l'inventaire forestier
- ° Culture
- ° Liste des immeubles protégés, au titre de monuments historiques
- Economie et emploi
- ° Le nombre de personnes soumises à l'ISF, par commune de plus de 20 000 habitants
- ° Education et recherche
- Statistiques brevet des collèges
- ° La Santé
- 0

L'information est aujourd'hui disponible. Qu'elle provienne de l'extérieure ou bien de l'intérieure de l'entreprise, les outils existent, pour la transformer, et en retirer de la valeur pour les clients et pour l'entreprise.

1.2 <u>LES TECHNIQUES D'ANALYSE ET DE VISUALISATION</u> DU BIG DATA³⁰

1.2.1 *L'ANALYSE*

Il existe plusieurs modèles mathématiques, répondant aux principes des méthodes prédictives. Ces modèles mathématiques ne sont pas nouveaux, même si des recherches sont toujours en cours, pour les améliorer. Ce qui est nouveau, c'est la masse de données mise à disposition de ces modèles pour leur permettre de fonctionner à plein régime. La plupart de ces modèles demandent des mises au point et des réglages, pour obtenir des résultats pertinents. Ce détail, important, fait partie intégrante de l'analyse des données et, encore plus, lorsqu'elles sont volumineuses.

Dans le monde de la grande distribution, certains produits phares sont repérés (type Nutella, Coca-Cola, ...), car ils sont, systématiquement, achetés par les clients. Si toutefois, le client ne trouve pas son produit phare, il peut quitter le magasin pour faire ces courses ailleurs. Ayant analysé et compris ce phénomène, la grande distribution a mis en place un balayage systématique des tickets de caisse afin de repérer ces différents produits. Si un de ces produits n'apparaît plus sur un ticket de caisse, audelà de 3 heures, un responsable de rayon est aussitôt envoyé pour étudier pourquoi (plus de produit, produit trop cher, ...). Ce système de corrélation, en temps réel, avec un flux incessant de données, montre comment utiliser un volume de données important dans la vie pratique.³¹

³⁰ Inspiré du livre "Big Data: The next frontier for innovation, competition, and productivity" de McKinsey Global Institute – Mai 2011.

³¹ Basé sur l'interview d'Olivier Tapin du 15/10/2013

Ce type de corrélation se définit comme de la fouille de données : « Data Mining ». Il s'agit de « l'ensemble des :

- ° Algorithmes et méthodes
- ° ... destinés à l'exploration et l'analyse
- ° ... de (souvent) grandes bases de données informatiques
- ... en vue de détecter dans ces données des règles, des associations, des tendances inconnues (non fixées a priori), des structures particulières restituant de façon concise l'essentiel de l'information utile
- ° ... pour l'aide à la décision »³²

Ce mode d'exploration des données n'est pas une technique nouvelle. Dans les premiers essais de fouille de ce type, on peut citer le célèbre exemple de l'entreprise WAL-MART. Ces derniers, en analysant des millions de tickets de caisse, sont arrivés à une corrélation très forte entre l'achat de couches culottes pour enfants et la bière, le Samedi après-midi. Après analyse, il fut mis, en évidence, que les paquets de couches étant très volumineux, les épouses envoyaient leurs époux acheter les couches ce jour là et ces derniers en profitaient pour acheter de la bière. Fort de ce constat, WAL-MART rapprocha les couches et la bière et vit ses ventes grimper en flèche.

Le terme « Data Mining » est particulièrement adapté à cette masse de données dans laquelle on doit fouiller, pour faire parler et interpréter ces volumes d'informations.

Ceci permet, entre autre, de corréler des données, telles que des couches culottes et de la bière, comme exposé dans l'exemple ci-dessus.

La nouveauté réside dans le fait que des progrès ont été réalisés dans les domaines de l'intelligence artificielle et de la théorie de l'apprentissage. En associant ces progrès avec les capacités de stockage

³² Stéphane Tufféry – Data Mining – http://data.mining.free.fr

offertes, ainsi que les puissances de calculs des machines et des outils devenus, aujourd'hui accessibles, on dispose de formidables opportunités pour fouiller et faire parler les données.

L'interprétation de ce type de données passe par l'adoption d'un raisonnement orienté vers des réflexions prédictives, basé sur des présomptions, voire des intuitions.

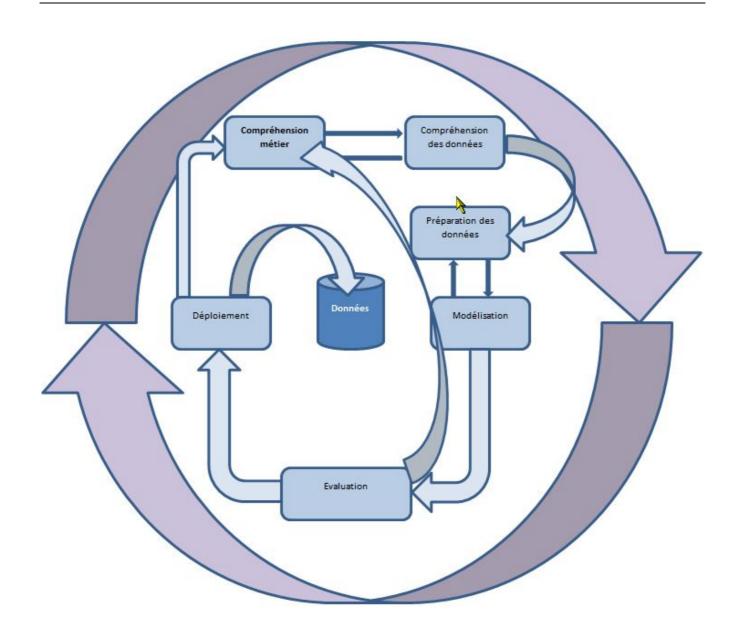
Le principe de l'analyse cartésienne, très utilisé dans les analyses traditionnelles, s'appuie sur les concepts suivants :

- o analyse de tous les éléments et de leurs caractéristiques
- o analyse des relations entre tous les éléments (hiérarchique, appartenance, ...)
- o principe de causalité : « La cause précède l'effet et l'effet reflète l'influence de la cause ».

Du fait du principe de causalité, l'analyse cartésienne est une méthode rétrospective qui part d'une observation, à un instant 't' et examine les circonstances antérieures qui ont conduit à cette situation.

Pour une analyse Big Data, l'analyse systémique sera plus adaptée. Nous partons, en effet, de l'observation d'un cas d'utilisation, qui peut s'apparenter à un système et nous cherchons les éléments, les plus pertinents, qui le composent. Cette approche permet de comprendre et de prévoir le comportement du système.

Un des objets des modèles prédictifs, issus du « Data Mininig », consiste à mettre, en évidence, des corrélations informelles entre les données, qui vont constituer les éléments pertinents du cas étudié. Ces modèles, associés au volume, permettent l'exploration et la mise, en évidence, de pistes, jusqu'à présent inaccessibles, car dissimulées dans cette masse de données et, ainsi, de prédire des nouveaux comportements ou des nouveaux besoins.



Cross Industry Standard Process for Data Mininig 33

³³ WIKIPEDIA : conçu en 1996 et pris en main par l'Union Européenne (ESPRIT : European Strategic Program on Research in Information Technology) via 5 sociétés : SPSS, TERADATA, DAIMLER AG, NCR CORPORATION, OHRA.

Cette figure représente le découpage du processus de "Data Mininig" en 6 phases :

- ° Compréhension Métier : définition du problème et des objectifs
- Compréhension des Données : collecte des données, filtre en fonction des objectifs et détection des données intéressantes pour les premières hypothèses
- ° Préparation des Données : mise en forme des données à utiliser.
- Modélisation : mise au point du ou des modèles à utiliser sur les données sélectionnées. Un retour à l'étape de préparation peut être nécessaire.
- ° Evaluation : étude et vérification que les résultats obtenus sont en adéquation avec les besoins du départ et que le modèle répond aux exigences business attendues. Vérification qu'aucune piste n'a été sous-estimée ou oubliée avant le déploiement final.
- ° Déploiement : diffusion des résultats ou bien déploiement, du modèle mis au point.

Grâce à ce volume de données maintenant disponible, les modèles mathématiques appliqués aux données sont en mesure de fournir des résultats plus précis. Ces techniques d'observation et d'analyse des données sont appelées à se développer et à s'enrichir.

Nous allons exposer ici, quelques techniques existantes, pouvant être utilisées. Ces techniques ne se veulent pas exhaustives, il s'agit d'éclairer, le domaine du possible, avec des méthodes connues et maîtrisées.

« Il existe 2 types de techniques de modélisation pour explorer des données :

Les techniques descriptives:

- ° Analyse factorielle: projection des données, sous forme graphique, pour obtenir une visualisation, de l'ensemble, des liaisons (rapprochements et oppositions) entre les différentes données.
- ° Classification automatique (clustering segmentation): permet de rassembler des groupes homogènes d'individus ou de données, permettant de mettre, en évidence, une segmentation des individus dans des classes.
- ° Recherche d'associations (analyse du ticket de caisse) : il s'agit de détecter des dépendances entre les objets ou individus observés.

Objectif des techniques descriptives :

- Visent à mettre en évidence des informations présentes, mais cachées par le volume des données (c'est le cas des segmentations de clientèle et des recherches d'associations de produits sur les tickets de caisse – Cf WAL-MART-).
- ° Réduisent, résument, synthétisent les données
- ° Il n'y a pas de variable « cible » à prédire

Les techniques prédictives :

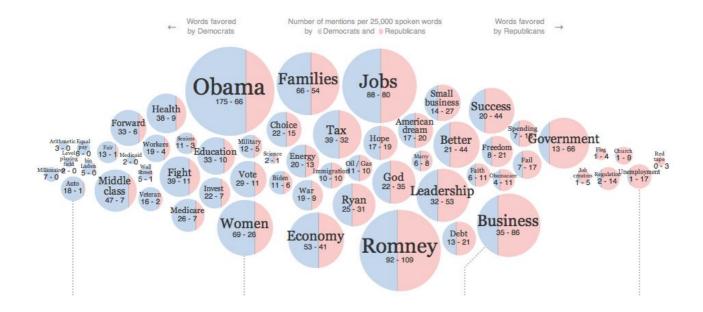
- ° classement/discrimination (la variable à expliquer est qualitative)
- analyse discriminante / régression logistique : trouver des règles d'affectation des individus à leur groupe.
- ° arbres de décisions : permettent de répartir les individus d'une population en n classes, on commence par choisir la variable séparant le mieux les individus de chaque classe en fonction de la variable cible, en sous populations appelées noeuds : le critère précis (C1) de choix de la variable et de sa valeur testée, dépend de chaque type d'arbre (pour déterminer, par exemple, en fonction de critères si un individu est malade ou bien portant).
- réseaux de neurones : issus du domaine de l'intelligence artificielle, c'est un ensemble de nœuds connectés entre eux et possédant des valeurs. L'apprentissage se fait en ajustant le poids de chaque nœud jusqu'à trouver une solution optimale où atteindre le nombre d'itérations fixées. (Deep Blue célèbre jeu d'échecs ayant battu Kasparov en 1997 utilise ce principe de modélisation)
- ° prédiction (la variable à expliquer est continue ou quantitative)
- régression linéaire (simple et multiple) : permet de modéliser la variation d'une variable Y par rapport à une autre variable X et donc de prédire l'évolution de Y en fonction de X. Cette modélisation reste vraie en introduisant d'autres variables explicatives, indépendantes de X et Y.
- Modèle linéaire général, généralise la régression linéaire avec des variables explicatives continues.

Ces méthodes statistiques analysent une relation entre une ou plusieurs variables dépendantes par rapport à des variables indépendantes. Ces méthodes prédictives peuvent être utilisées par exemple pour déterminer l'influence d'une molécule médicamenteuse, sur un échantillon de patients, en fonctions de critères comme le poids et l'âge du patient.

Objectif des techniques prédictives :

- o visent à extrapoler de nouvelles informations, à partir des informations présentes (c'est le cas du scoring)
- o expliquent les données
- o il y a une variable « cible » à prédire »³⁴

L'hétérogénéité des sources de données, ainsi que leurs caractéristiques, font que le « Data Mininig» seul, ne suffit plus. On entre dans le domaine du « Text Mining ». Il s'agit de l'analyse de texte pour évaluer, par exemple, un sentiment positif ou négatif, envers une société ou des opinions sur les réseaux sociaux.



³⁴ Stéphane Tufféry – Data Mining – http://data.mining.free.fr

Ci-dessus, un exemple de démarche de « Text Mining » utilisée pour extraire les concepts-clés des discours présentés, lors des conventions républicaines ou démocrates : sur un axe horizontal, les mots sont classés en fonction de leur utilisation démocrate ou républicaine, dans des bulles, dont la taille est fonction du nombre d'occurrences et dont la couleur dépend de son utilisation par l'un ou l'autre camp. On y observe, clairement, le camp Obama avec les mots « Middle Class », « Forward », « Health » et « Women » face aux mots « Leadership », « Success », « Business » et « Government » de Romney.³⁵

Il existe, aussi, l'« Image Mining » qui permet l'interprétation d'images, notamment, dans le domaine médical ou la biométrie. Plusieurs logiciels d'interprétation d'images existent, depuis longtemps, dans le domaine industriel (les photos satellites, par exemple) et l'imagerie médicale.

Enfin, nous avons, aussi, l' «Audio Mining » qui s'apparente au logiciel de reconnaissance vocale. Ces systèmes sont très utilisés pour les serveurs vocaux, par exemple.

Dans le domaine du Big Data, on parle, de plus en plus, de « Machine Learning ». Ce concept est issu du domaine de l'intelligence artificielle. Il s'agit de modéliser l'intelligence humaine et de permettre à des machines d'apprendre, par déduction, en faisant des comparaisons avec des données prédéfinies ou bien en testant des combinaisons différentes, jusqu'à trouver celle qui répond le mieux à la demande.

« Les algorithmes utilisés permettent, dans une certaine mesure, à un système piloté par ordinateur (un robot éventuellement), ou assisté par ordinateur, d'adapter ses analyses et comportements en réponse, en se fondant sur l'analyse de données empiriques provenant d'une base de données ou de capteurs. » 36

Ces méthodes sont utilisées dans la partie modélisation représentée dans le schéma du process de « Data Mininig ».

Associées ensemble, elles permettent de former des algorithmes utilisés par ces machines auto apprenantes. Ainsi, les résultats interprétés par des analystes permettront des prédictions dans des domaines comme la santé, la météorologie, ...

•

35 Site PEGASUSDATA.COM

³⁶ Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

1.2.2 LA VISUALISATION

Une fois l'analyse terminée, les résultats doivent être présentés de manière lisible. Il s'agit de permettre des prises de décisions, en enlevant la partie complexité et en étant le plus clair possible sur les enjeux.

Au-delà des graphiques classiques tels que nous les connaissons, histogrammes, courbes, camemberts, ... existent d'autres représentations plus lisibles pour le Big Data.

La visualisation est un élément indispensable au Big Data et peut être considérée comme le cinquième 'V'. Il s'agit de trouver de nouvelles formes de visualisation facilement interprétables qui mettent en valeur cette masse de données. L'information est ainsi mise en contexte.

Voyons quelques exemples de matérialisation possible d'une analyse Big Data.

Chaque résultat pourra faire l'objet d'une présentation particulière. Les principes qui doivent guider de telles restitutions restent la lisibilité et la compréhension.

Le « Tag Cloud »:



www.123rf.com

Les mots, les plus souvent rencontrés, apparaissent avec une taille supérieure aux autres.

On peut utiliser ce type de graphique pour représenter des classifications :

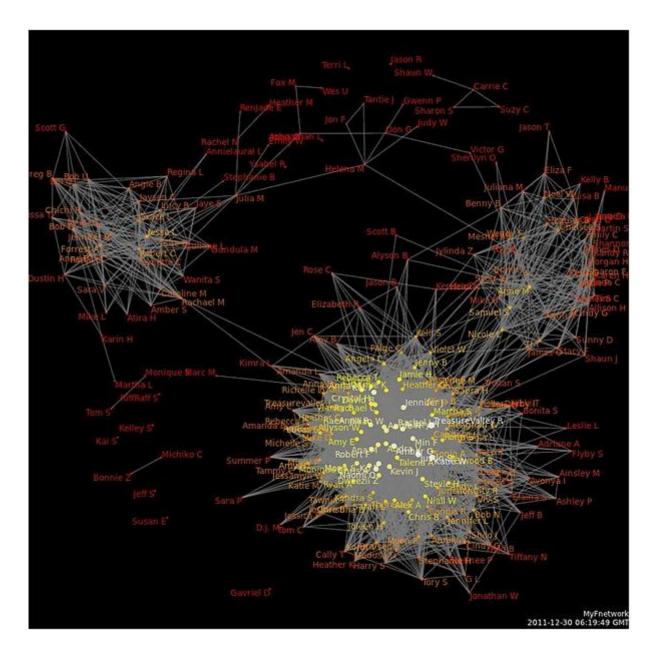
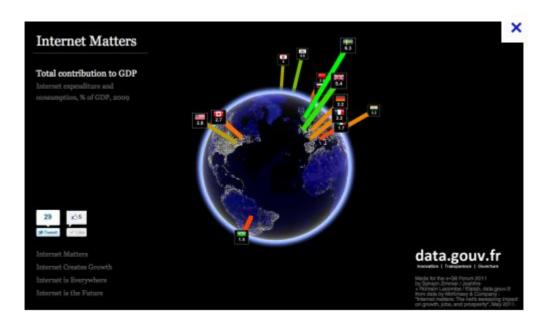


Image Wikipedia représentant les relations Facebook d'une personne

On remarque, très facilement, les relations et les différents groupes créés sur le réseau social Facebook, au travers des relations d'une personne.

Un exemple de représentation animée dans l'espace :

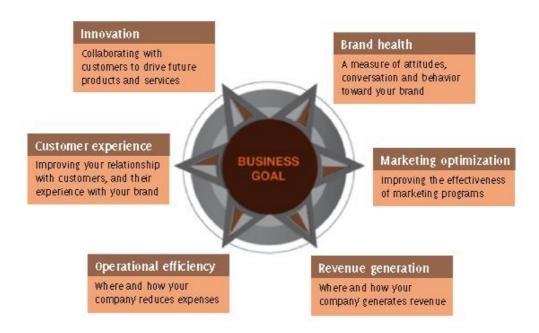


On y distingue la contribution du secteur Internet au PIB des pays du G8, des BRIC (Brésil, Russie, Inde, Chine) et de la Suède, la contribution du secteur Internet à la croissance du PIB de ces pays, la distribution, pays par pays, des revenus des 250 plus grandes entreprises du secteur Internet, et l'investissement, pays par pays, en R&D (Recherche & Développement) liés à Internet.³⁷

 $^{^{37}\} http://www.etalab.gouv.fr/article-data-visualisation-a-l-e-g8-forum-internet-dans-l-economie-mondiale-74865969.html$

CONCLUSION PREMIERE PARTIE

Ce schéma illustre les différents éléments qui doivent être pris en compte, aujourd'hui, pour rester une entreprise compétitive. ³⁸



Le Big Data va devenir un élément important dans la prise de décision des dirigeants d'entreprises. Nous allons devoir apprendre à travailler, ensemble, en innovant avec nos clients, en tenant compte de leurs comportements et en étant à leur écoute pour créer les produits et services de demain qui nous permettront de générer des bénéfices pour atteindre nos objectifs.

³⁸ VINT Research report 2 of 4 – Big Social - Predicting behavior with Big Data – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren - Sandeep Sachdeva - SOGETI

2. LES ENJEUX DU BIG DATA

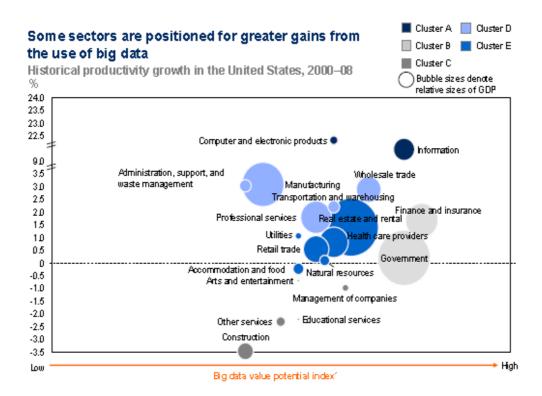
Le Gartner affirme que ceux qui seront capables d'analyser et d'interpréter les données réaliseront, en moyenne, 20% de bénéfices supplémentaires, par rapport à leurs concurrents.

La maîtrise de cette information est bien l'enjeu du siècle au même titre que l'ont été le raffinement du pétrole, la maîtrise de l'électricité et autres innovations des siècles passés.

C'est une des raisons pour lesquelles ce volume de données ne peut plus être ignoré par les entreprises.

Plusieurs entreprises utilisent déjà le Big Data pour cibler les besoins de leurs clients. C'est le cas d'Amazon, par exemple, qui vous suggère, systématiquement, des livres en rapport avec celui que vous avez récemment acheté ou consulté. En fonction de vos consultations et de vos achats, Amazon reconstitue un profil et le compare aux différents contenus des autres ouvrages à sa disposition et peut, ainsi, vous proposer des livres en adéquation avec vos goûts.

Mc Kinsey Global Institute dans son étude « Big Data : The next frontier for innovation, competition, and productivity », illustre dans le schéma, ci-dessous, les gains estimés, par secteur, grâce au Big Data. Tous les secteurs ne sont pas concernés, de la même manière.

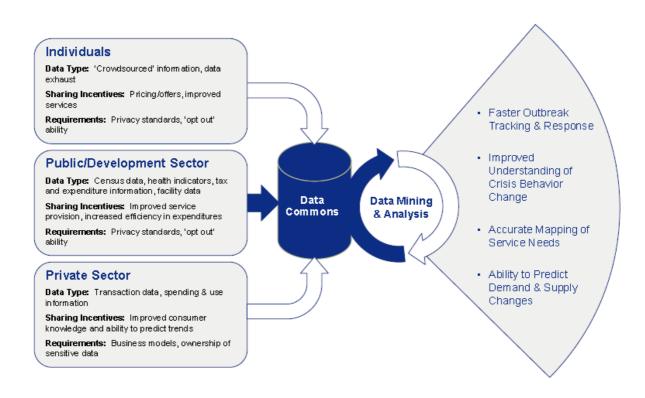


Le groupe A, qui regroupe les ordinateurs et les produits électroniques, ainsi que le secteur de l'information, possède et a accès à des volumes de données conséquents. L'utilisation du Big Data sera, pour eux, une source de revenus importants.

Le groupe B, avec le secteur de la finance, les assurances et le gouvernement est aussi, bien positionné, pour tirer parti du Big Data, si la réglementation ne se durcit pas. Par le biais du Big Data, ces secteurs doivent pouvoir améliorer les performances de leurs produits et de leurs services.

Pour le groupe C, avec l'éducation, par exemple, il s'agira, grâce au Big Data, de faire progresser la performance des professeurs, en mesurant et en étudiant la réussite des élèves aux examens. Ainsi, en fonction des critères et des conclusions, on peut imaginer adapter les programmes scolaires et les outils à la disposition de l'éducation nationale, afin de faire progresser cette performance.

Pour les groupes D et E, comme le commerce et la santé, par exemple, le potentiel à exploiter le Big Data permettra des progrès colossaux. Le frein essentiel reste la réglementation et ce qu'elle permettra de faire.



Exemples de retours, grâce à une meilleure utilisation des données³⁹

Ce schéma illustre, parfaitement, le mélange de données d'origines variées, publiques, privées ou individuelles. Les analyses mettent, en évidence, des changements de comportements client, et font apparaître de nouveaux besoins. Elles permettent, aussi, de prédire des demandes qui n'existent pas encore. Tous ces éléments contribuent, à la prise de décisions stratégiques, pour diriger une société.

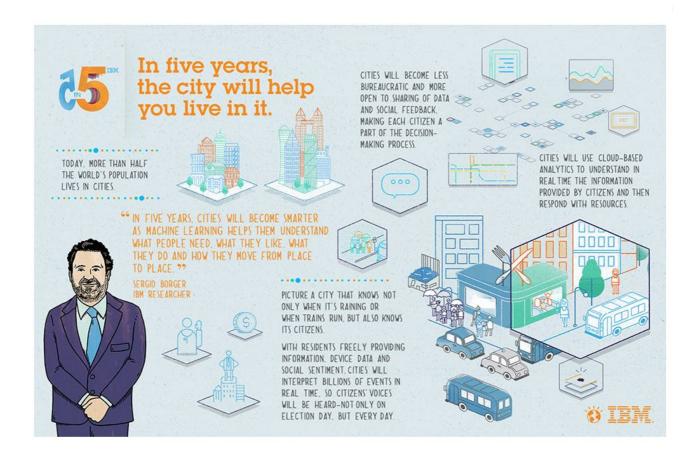
 $^{^{39}}$ World Economic Forum – Big Data, Big Impact : New possibilities for International Development – 2012

2.1 IMPACTS ET CHANGEMENTS LIES AU BIG DATA

Découvrons les domaines dans lesquels les applications du Big Data vont apporter des changements.

Plusieurs illustrations sont tirées du « 5 in 5 report » d'IBM. 40

2.1.1 <u>LES VILLES OU NOUS VIVONS</u>



.

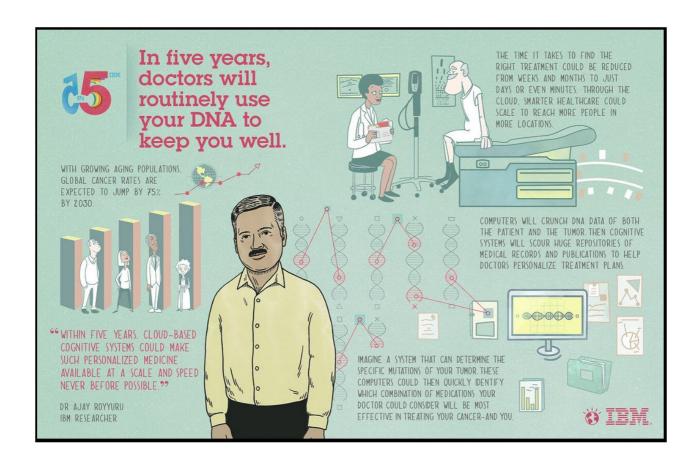
⁴⁰ The 5 in 5 – Innovations that will change our lives in the next five years - http://www.ibm.com/smarterplanet/us/en/ibm_predictions_for_future/ideas/

Pour les villes, on peut imaginer des applications Big Data et des algorithmes capables de calculer, en temps réel, via les informations collectées sur les smartphones des citoyens, les besoins en transport et ainsi, d'augmenter ou diminuer la fréquence et le nombre de bus, de métro et autres moyens de transports publics.

Une autre application pourrait aussi être l'éclairage urbain, la nuit, en fonction de la fréquentation des différentes places en ville. Les endroits peu fréquentés pourraient être moins éclairés ce qui permettrait des économies d'énergie et de frais pour les villes et leurs habitants.

2.1.2 <u>LE DOMAINE DE LA SANTE</u>

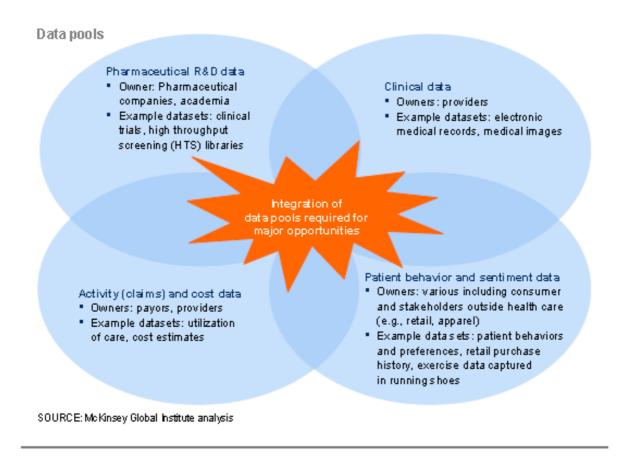
La santé est, sans doute, l'un des secteurs qui présente le plus fort potentiel de progression grâce au Big Data.



Deux axes importants concernent la santé. Le premier axe permet la réduction des dépenses liées aux différents actes médicaux. Pour se faire on peut personnaliser les traitements, grâce aux masses de données dont disposent les hôpitaux, les cliniques, les pharmacies, ..., faire des diagnostics plus justes en évitant de multiples examens, proposer les médicaments et les soins adaptés aux patients (éviter les effets secondaires ou les médicaments qui ne fonctionnent pas).

Grâce à la connaissance du patient et aux données disponibles sur des patients identiques, on peut imaginer faire de la prévention, avant que la maladie ne se déclenche.

Le schéma suivant représente un exemple de données médicales, à mettre en commun, pour obtenir des diagnostics plus efficaces.



Le deuxième axe concerne la recherche médicale. L'accès à des millions de données de patients ayant suivi un même protocole médical permettra, en fonction des données génétiques de ces patients, d'étudier dans quelles conditions le protocole a fonctionné et dans quelles conditions il n'a pas été efficace, afin de l'adapter pour l'améliorer.

Il s'agit de faire avancer la recherche médicale, en permettant des séquencements et des tests de protocoles médicaux plus rapides, plus efficaces et mieux ciblés.

<u>LE COMMERCE</u>

C'est, sans nul doute le domaine, qui possède le plus de données personnelles sur ses clients et, donc le plus d'expérience, sur la segmentation et la personnalisation du besoin.

La conservation des historiques de courses permet de faire des propositions d'articles similaires à ce que le client a l'habitude de consommer. C'est le principe des sites de e-commerce, qui personnalisent leurs propositions en fonction des articles que le client a déjà consultés ou achetés sur leurs sites.



La géolocalisation permet de faire parvenir aux clients les promotions en cours dans leurs magasins favoris. L'écoute des réseaux sociaux permet, également, aux marques de faire des promotions, si un de leurs produits fonctionne, car ainsi, ils peuvent anticiper les ventes et ajuster leurs marges et travailler, au plus juste, l'approvisionnement des stocks.

Les clients sont, aussi, de plus en plus écoutés. Leurs propositions sont prises en compte et deviennent des adaptations des produits. C'est l'innovation participative ou « crowd innovation ». Le principe est simple, on demande des idées nouvelles sur des produits, pour ensuite, retenir les meilleures dans la construction du produit. (Microsoft, Coca Cola, Volvo, ...). Ford va même, jusqu'à permettre, l'accès aux API (Application Programming Interface) embarquées dans ces véhicules et donne la possibilité de les modifier.

Le client s'approprie le produit, en le personnalisant, ce qui permet de le fidéliser.

2.1.4 L'ETAT ET LE SECTEUR PUBLIC

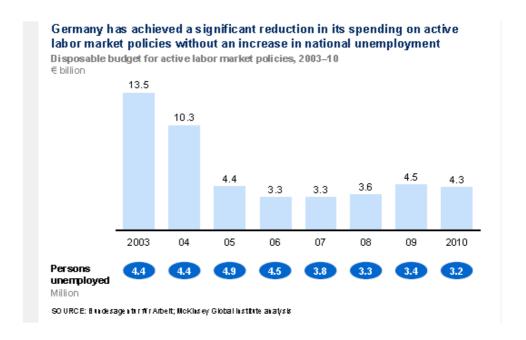
Dans son rapport, le Mc Kinsey Global Institute estime que le secteur public Européen pourrait, grâce au Big Data, réduire ses coûts de 15 à 20%, créant ainsi, l'équivalent d'un gain allant de 150 à 300 billions d'euros. Ce calcul inclut la différence entre les impôts et taxes collectées, actuellement, et le montant des ressources que cela devrait, normalement, rapporter. C'est-à-dire, en supprimant les erreurs et la fraude, par exemple.

Le remplissage automatique des formulaires permet d'éviter les erreurs et, surtout, de passer moins de temps à vérifier les données. De plus, la fraude devient plus compliquée, lorsque les données sont connues des services administratifs.

Ce mode de fonctionnement est possible, lorsque les différents services de l'état communiquent entre eux et partagent les informations qu'ils possèdent sur les citoyens.

L'Agence Nationale pour l'Emploi, en Allemagne, a aussi, mis en œuvre une technique de segmentation des demandeurs d'emploi, comme on le fait dans le domaine commercial, pour cibler ses actions de formation et de placement sur la bonne population.

Et voici les résultats obtenus, illustrés par le graphique ci-dessous.



Le Big Data peut, aussi, être utile pour la surveillance des épidémies et, ainsi, éviter ce que nous avons connus, en 2009, pour le vaccin contre la grippe A (H1N1) :

94 millions de vaccins commandés : 50 millions feront l'objet d'une annulation de commande et, seulement, 6 millions de Français se feront vacciner. 300 000 seront vendus au Qatar : 16 millions seront donnés à l'Organisation Mondiale de la Santé et 2 millions iront aux Ambassades Françaises à l'étranger. 19 millions de vaccins seront détruits, pour un coût total de 400 millions d'euros. 41

⁴¹ Source l'express.fr

En observant les données médicales nationales d'une part, et les prescriptions médicamenteuses d'autre part, les responsables de santé publiques peuvent, rapidement, détecter des épidémies voire, les enrayer. Les avantages sont nombreux, notamment, l'amélioration de la prévention et, en conséquence, la diminution du nombre de maladies et des coûts associés.⁴²

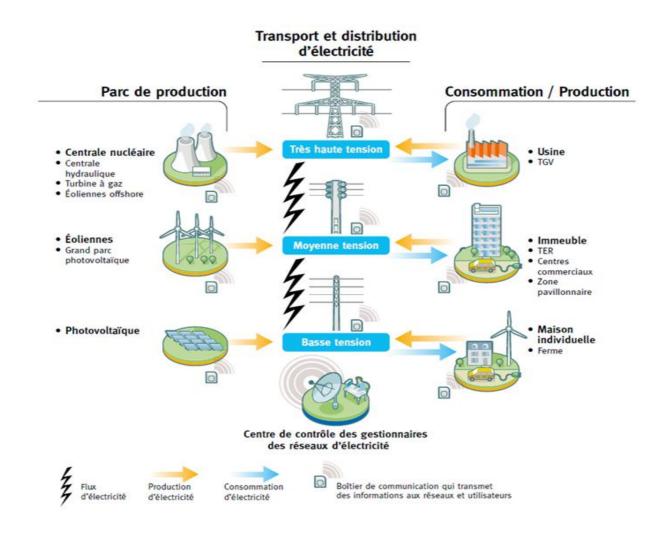
2.1.5 <u>LE SECTEUR INDUSTRIEL</u>

Ce secteur utilise, énormément, de données dans différents domaines.

Le monde du digital et des objets communiquants va accroître ce volume de données et poser la question de leur utilisation.

Un exemple qui prend de l'ampleur, au niveau mondial, est celui des « smart grids ». Il s'agit de réseaux électriques « intelligents » publics, auxquels sont ajoutés des fonctionnalités issues des nouvelles technologies de l'information et de la communication (NTIC). Le but est d'assurer l'équilibre entre l'offre et la demande d'électricité, à tout instant, et de fournir un approvisionnement sûr, durable et compétitif aux consommateurs. La pièce centrale de ce dispositif, étant un compteur, qui envoie des données, pour permettre d'ajuster cette fourniture en électricité (vu sur le site internet de la Commission de Régulation de l'Energie).

⁴² Basé sur le livre de Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »



Commission de Régulation de l'Energie (CRE)

2.2 LES ENJEUX DANS L'ASSURANCE

Les services marketing qui travaillent dans les assurances utilisent, de plus en plus, leur connaissance des clients pour déterminer les comportements et les nouveaux besoins en assurance.

« L'assurance vise à évaluer non pas tant des situations financières que des attitudes à l'égard du risque. » ⁴³ Ce que cherche l'assureur au travers de l'information c'est la connaissance du risque qui lui est proposé. Il a l'habitude de travailler sur des critères ou caractéristiques du risque technique (voiture et conducteur en assurance automobile, type d'habitation et valeur des biens en assurance habitation, âge en assurance santé, …). Les critères liés au comportement sont peu utilisés. La principale difficulté étant de trouver une corrélation entre ce comportement et le risque.

Le grand principe de l'assurance est d'intervenir pour ses clients en cas de sinistres. Le cycle inversé des entreprises d'assurance, qui consiste à encaisser une prime pour couvrir un sinistre qui n'a pas encore eu lieu, a obligé ces dernières à mettre au point des modèles très pointus pour évaluer des tarifs adaptés en regard des risques proposés. Pour se faire on observe ce qui s'est passé en terme de vente et de gestion de contrats, de sinistres et des éléments s'y rapportant, pour en déduire une tarification adaptée.

« Pour cela, prévenir, anticiper et prédire le risque sont à toutes les étapes de la chaîne de valeur de l'assureur, des enjeux. » ⁴⁴ Des enjeux pour :

- Sélectionner les risques à la souscription,
- Appliquer un tarif en adéquation avec le risque sélectionné et des critères de tarifications supplémentaires
- ° Evaluer, correctement, un sinistre lorsque ce dernier se réalise.

-

⁴³ Big Data en assurance clients et souscriptions en question – Romain Durand – Revue Risques n°95 – Novembre 2013.

⁴⁴ En quoi le « Big Data » change-t-il l'évaluation des risques ? – Gontran Peubez – Revue Risques n°95 – Novembre 2013.

Grâce aux possibilités offertes de pouvoir croiser des données d'origines et de sources différentes, le Big Data permet de comprendre le comportement de nos clients de manière très précise et détaillée.

Ces possibilités ouvrent des perspectives sur de nouvelles segmentations et une meilleure maîtrise des risques. Cela va aussi modifier le métier d'assureur en faisant évoluer les stratégies et les organisations des entreprises d'assurance.

Le schéma ci-dessous, représente différentes opportunités, offertes par le Big Data, dans les différents domaines composants une entreprise d'assurance.

Les Produits	Marketing & Ventes	Gestion des contrats	Gestion Sinistres	Gestion des Actifs
Nouveaux marchés	Nouveaux clients	Gestion des paiements	gestion des réseaux de prestataires placem	Stratégie de placement
Nouveaux produits	Vente multi-canal	Tarification temps réel Déte frauc Prév mitig		Gestion de
Connaissance des risques et meilleure	Détection résiliation Optimisation et ciblage des campagnes commerciales « just in time »		Détection des fraudes	Portefeuille
tarification			Prévention et mitigation des risques	Gestion des actifs et passifs
Segmentation plus fine				Modélisation des risques
Optimisation des				
produits Gestion de la performance				

Principales opportunités pour l'assurance⁴⁵

Pour chaque opportunité, un indicateur de mesure du retour sur investissement devra être mis en place. La mesure du taux d'attrition pourra être un indicateur sur la fidélisation de nos clients, la performance d'un nouveau service pourra se mesurer grâce à l'évolution du « Loss Ratio » (Sinistres sur Primes) concernant ce produit, etc. ...

⁴⁵ Basé sur une analyse du Boston Consulting Group (BCG)

2.2.1 LES CHANGEMENTS DANS LE METIER DE L'ASSURANCE

« Cette couche de mathématiques » ⁴⁶ placée sur nos comportements va modifier en profondeur la connaissance client ainsi que la connaissance des risques.

Aujourd'hui, l'assurance repose sur le principe de mutualisation.

« La mutualisation des risques consiste à partager le même risque entre plusieurs personnes. L'organisation de la solidarité réalisée par l'assureur se fait à partir de l'ensemble des cotisations versées pour un même type de risques. Cela permet la couverture du risque et en même temps de plusieurs risques analogues. » ⁴⁷

Or le principe de mutualisation et de segmentation de l'assurance tel que nous le connaissons est en passe d'être modifié par cet accès facilité aux données et aux puissances de calculs.

Nous n'allons pas basculer dans un monde déterministe, ou chacune de nos actions va devenir prévisible, mais un monde basé sur des probabilités de plus en plus justes. La réalisation d'une action aura été déterminée pour nous en corrélant des données. Ce type de déduction permettra de se préparer à rendre un **service personnalisé** aux clients.

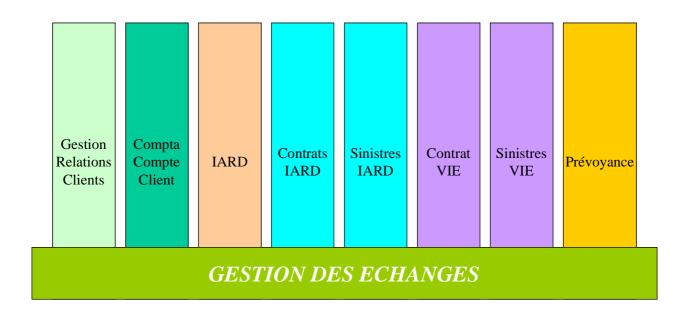
Le métier de l'assurance, a longtemps, été basé sur la capacité des produits à être innovants par rapport aux sinistres auxquels ils sont rattachés. Depuis plusieurs années, les assurances se tournent, de plus en plus, vers leurs clients et se mettent à leur écoute. Le phénomène Big Data est un puissant moteur permettant d'interpréter des informations sur les attentes des clients en matière d'assurance.

⁴⁷ Les fondamentaux de l'assurance – Cours ENASS – MBA – Dominique Laborie

_

⁴⁶ La société du calcul – François-Xavier Albouy – Revue Risques n°95 – Novembre 2013.

La difficulté, aujourd'hui, réside dans le poids de l'histoire de nos produits d'assurance, indépendants les uns des autres et qui utilisent des données communes, en les nommant différemment.



<u>Un ensemble de solutions différentes pour chaque métier qui communiquent via une couche</u> <u>d'échanges standardisée et centralisée ⁴⁸</u>

Ce modèle, en silo, avec une gestion des échanges transverses, illustre la problématique des assureurs pour proposer des solutions personnalisées à leurs clients.

Aujourd'hui, l'accélération des demandes et les frontières entre ces domaines, qui deviennent de plus en plus floues, se retrouvent être des freins à l'innovation et à la production.

⁴⁸ Quel Modèle d'Entreprise pour l'Assurance de demain ? - Jean-René Lyon et Stéphane Arbus

La gestion transverse des échanges montre la volonté de lier les différents produits, entre eux, pour proposer des solutions partagées et complémentaires, en fonction des attentes des clients.

Les données sur lesquelles les assureurs définissent les stratégies futures sont basées sur l'observation du passé. Les changements comportementaux que nous vivons, avec le développement des objets connectés, sont de plus en plus rapides. Il devient donc, de plus en plus compliqué, pour un assureur de faire des prédictions sûres en observant le passé. L'accès à ce volume de données va permettre de trouver de nouveaux critères pour affiner les prédictions des assureurs.

Qui dit volume, dit aussi population importante. Ces nouveaux critères, pour être efficaces, sont liés à leur fréquence d'apparition dans les échantillons observés. Il s'agit de faire de la **personnalisation mutualisée**. De ce fait, l'assurance conserve ce pilier mais doit l'envisager de manière différente. On peut se retrouver sur des mutualités nouvelles et plus étendues (en terme d'éloignement géographique).

Les études et analyses marketing effectuées par les entreprises d'assurance cherchent à déterminer le taux d'attractivité d'un segment de population pour une offre produit. L'objectif étant, que les personnes composant ce segment, soient le plus nombreuses possible, afin de fabriquer un produit qui intéressera le plus grand nombre. Grâce aux volumes proposés et à la diversité des informations le Big Data donne la possibilité de définir, puis de s'adresser, à de nouveaux segments de populations et ouvre des perspectives sur de nouveaux marchés. Ces données analysées permettront la prise en compte de nouveaux critères comportementaux, autorisant des corrélations entre le risque et le comportement. Elles permettront également de déterminer l'équilibre entre tarif et provisions techniques, en regard de ces nouveaux segments d'assurés.

Il conviendra de rester vigilant sur cette nouvelle forme de mutualisation, afin de conserver un équilibre technique répondant à la solvabilité nécessaire d'une société d'assurance. Il s'agira, aussi, de respecter le droit des assurances dans la sélection des risques.

La réflexion à mener, va bien au-delà, de la mutualisation. Le métier d'assureur peut se recentrer sur la prévention du risque pour définir, dans sa stratégie, les services adaptés à associer au produit d'assurance. Grâce au digital, la distribution et la gestion de ces services peuvent se faire à distance. La personnalisation du produit se fera, par l'intermédiaire, du service proposé.

Le modèle assurantiel, tel que nous le connaissons aujourd'hui pourrait donc, d'ici quelques années, devenir obsolète. En effet, grâce à cette connaissance accrue de nos clients, nous devrions être en mesure de développer la prévention des risques voire pour certains, notamment en santé, les supprimer.

De plus en plus d'appareils de notre vie de tous les jours sont capables d'envoyer des informations. Il s'agit, par exemple, d'un détecteur d'incendie intelligent, qui, via un accès internet, sera en mesure de déclencher une alerte pour une intervention, permettant d'éviter que la maison de notre assuré soit entièrement détruite. Ainsi, l'assureur y gagne, car il ne devra pas rembourser le risque dans sa totalité et, à terme, l'assuré abonné à cette catégorie de services verra sa cotisation réduite, puisque le coût des sinistres pourra baisser. Il appartiendra aux assureurs de définir, à partir des données issues de ces capteurs, celles qui seront déterminantes pour créer de nouvelles mutualités et, ainsi, fabriquer les services et tarifs, en adéquation avec les possibilités offertes par ces données.

L'assureur doit ajouter à ses produits ces nouveaux services afin, de permettre la transformation de la mutualisation classique, telle que nous la connaissons aujourd'hui, en mutualisation « responsable ».

Ces services, « gagnant – gagnant » permettent à l'assureur et à l'assuré de se focaliser sur la prévention du risque et de rendre la mutualisation plus sûre. Certains services trouveront leur existence dans le Big Data, d'autres existent déjà. Grâce au Big Data il sera plus facile de repérer les populations en quête de nouveaux services en assurance. Il en résultera une nouvelle segmentation client et une nouvelle mutualisation des assurés.

A titre d'exemple, plusieurs applications de coaching santé sont à disposition, aujourd'hui, via des objets connectés, type smartphones, capteurs, chaussures, ... Un produit d'assurance décès IT, couplé à une telle application et suivi par l'assureur, permettrait de proposer aux clients, avec leur accord, une prime réduite sur ce type de produit d'assurance. Pour cela, il suffira que le client s'engage à suivre les recommandations du coach, pendant toute la durée de la validité du contrat.⁴⁹

⁴⁹ Discussion du 04/12/2013 avec André FRAGNERE directeur gérant ADICAM (Courtier de l'assurance emprunteur du Crédit Agricole)

Ainsi, **la fidélisation du client** se fait par « une personnalisation demandée remplaçant une personnalisation subie » ⁵⁰. Au travers de ces services, on équipe mieux nos clients, afin de les fidéliser.

Les assureurs, depuis plusieurs années vendent le Pay As You Drive (PAYD) pour l'assurance auto, le Big Data va permettre de fabriquer du **Be Assured as You Are (BAYA)**.

Dans cette optique, à côté de la prime qui permet de se prémunir face à un hypothétique sinistre car le risque zéro n'existe pas, on trouvera, au travers d'un service adapté, un suivi, un accompagnement et un coaching personnalisé qui permettront d'éviter un sinistre.

Le service offrira la possibilité à l'assuré de personnaliser son besoin d'assurance en fonction de ses attentes. L'assureur retrouvera, au travers de ce suivi, son rôle de conseiller vis-à-vis de ses clients. Il renforcera sa présence, en amont du sinistre. Les moyens modernes de communications serviront ces stratégies d'accompagnement (SMS, tweets, ...).

Le monde digital offre les outils nécessaires à la personnalisation de l'assurance via des offres de services adaptées à des segments de population plus restreints.

C'est une modification du modèle économique et le développement de stratégies individuelles par client pour travailler la matière assurable de manière personnalisée.⁵¹

⁵⁰ Daniel Kaplan - Cofondateur et Délégué Général de la Fondation Internet Nouvelle Génération (FING) – Vie Privée à l'Horizon 2020 – Cahiers IP N°1

⁵¹ Basé sur l'interview de Xavier Ducurtil – Direction COVEA TS du 16 Janvier 2014

2.2.2 <u>LES IMPACTS ORGANISATIONNELS DANS LES ENTREPRISES</u> D'ASSURANCE

Aujourd'hui, il apparaît de plus en plus nécessaire de « désiloter » nos organisations assurantielles pour rester compétitif et créatif.

Le poids de l'histoire ne permet pas de décider, un matin, de remettre à plat nos organisations. Le Big Data va nous obliger à réfléchir, de manière plus transverse, sur certains domaines et, notamment, de créer des passerelles entre le marketing et la direction des systèmes d'informations. La mise en place d'une gouvernance des données devient obligatoire pour une entreprise qui manipule des volumes de données importants.

Les opportunités proposées par le Big Data ne doivent pas remettre en cause nos fondamentaux techniques, mais bien trouver comment s'y intégrer. Le provisionnement et la tarification des produits d'assurance traditionnels sont maîtrisés par nos organisations et assurent, ainsi, un équilibre vital entre les primes et la sinistralité.

La mise en place de ces nouvelles approches d'analyse vient en complément de nos fondamentaux, pour les enrichir ou innover et mettre en œuvre de nouveaux produits d'assurance.⁵²

Cette approche pourra permettre d'affiner des segmentations client, déterminer des tarifs adaptés, par une analyse plus précise, des risques et des assurés. La prise en compte de critères, aujourd'hui non utilisés comme les émotions ou l'appétence à la conduite, par exemple, peuvent offrir des possibilités de répartitions et de tarifications nouvelles.

⁵² Réflexion issue de l'entretien du 19/12/2013 avec Jean-François Andrjanczyk (Directeur de la direction Actuariat et Compte Technique COVEA)

2.2.3 GOUVERNANCE DES DONNEES

« La notion d'information partagée est extrêmement importante et il faut veiller à organiser les données de façon non redondante. »⁵³

Le volume de données mises à disposition implique la mise en place dans nos organisations d'une structure de gouvernance des données qui saura sélectionner les données utiles en fonction des attentes de l'entreprise. En effet, la célèbre maxime « Garbage in garbage out » prend toute sa signification avec l'arrivée du Big Data. Des filtres sont nécessaires pour retenir les données significatives avec lesquelles travailler et prédire des axes de développement en accord avec les stratégies des directions métiers d'une part et les attentes des clients d'autre part.

Dans cette optique, la mise en place d'une gouvernance des données au sein de l'entreprise est indispensable. Cette gouvernance des données instaure des règles et des procédures à appliquer, afin de permettre à l'entreprise de choisir et de valider la pertinence des données sur lesquelles les analyses doivent être faites, au regard de la stratégie choisie. Cet organe de décision implique donc de nombreux acteurs dans les différents domaines de l'entreprise tels que :

- Direction générale
- Direction des systèmes d'information
- Oirections métiers (Marketing Métiers produits Réseaux Actuariat)
- Contrôle de gestion
- Direction du Management des risques

_

⁵³ Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Il s'agit pour ces acteurs de participer aux groupes de travail et d'arbitrer en fonction de la stratégie à suivre.

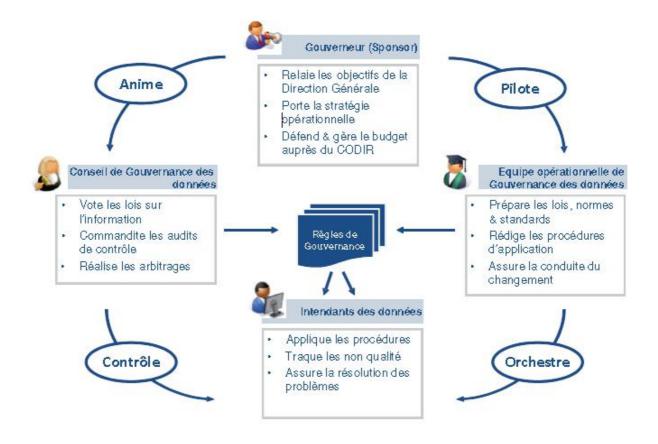


Illustration de la mise en place d'une fonction « gouvernance des données »⁵⁴

En centralisant cette gouvernance, cela permet de contrôler la pertinence des données utilisées, en regard du métier d'assureur. Les données sont définies de la même manière pour tous les domaines. Cela permet de prendre des décisions comprises et applicables par tous les acteurs de l'entreprise.

La gouvernance assure la qualité et l'intégrité des données. Elle maintient également l'équilibre et la cohérence entre les données internes et les données externes, issues du Big Data.

⁵⁴ SOLUCOM - Gouvernance des données : comment mettre durablement les données sous contrôle? - Jean-Nicolas Biarrotte-Sorin et Gilles Baillou

Le Big Data dans l'Assurance

Certaines entreprises d'assurance ont choisi de profiter de solvabilité 2, pour mettre en place une gouvernance des données.

L'exemple d'Allianz Global Corporate & Speciality qui a mis en place une gouvernance des données, illustre ce besoin, qui va s'amplifier avec le Big Data.

Il s'agissait pour Allianz, au travers de cette gouvernance des données, couplée avec SIX SIGMA⁵⁵, d'identifier et de corriger les données de qualité insuffisantes. De plus, celles qui ne répondaient pas aux besoins de solvabilité 2, dès la source, ont été enrichies ou supprimées, avant d'être utilisées dans les traitements de restitutions de tableaux de bords.

L'équipe, formée d'experts informatique et de compétences maîtrisant les données et leur cycle de vie, s'est attachée à faire des recommandations au comité de direction sur les règles à appliquer, pour maintenir la qualité des données de l'entreprise utiles au métier de l'assurance et à la gestion d'une entreprise d'assurances.

En utilisant SIX SIGMA, l'équipe de gouvernance des données peut mesurer l'impact des règles de qualité appliquées sur les données.

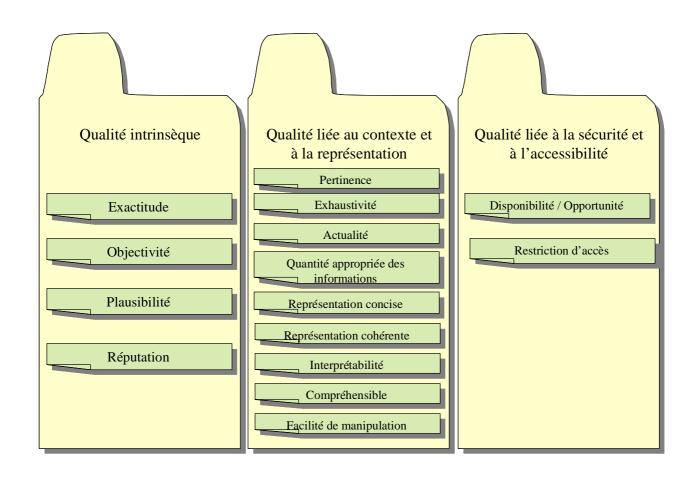
Cette approche a permis d'obtenir un degré de fiabilité des données, très élevé, pour les domaines du marketing, des risques et le métier d'assureur.

Rolf Neuerburg, explique qu'avoir cette structure de gouvernance de données permet de se poser, sereinement, la question de la fiabilité des données sur lesquelles sont prises les décisions stratégiques.⁵⁶

_

⁵⁵ Méthode d'amélioration de processus ayant pour objectif de réduire la variabilité pour accroître la qualité – Cours MBA CNAM-ENASS de Franck Pivert Allianz

⁵⁶ Basé sur l'Interview du 17/12/2013 de Rolf NEUERBURG Responsable du département Business Intelligence Allianz -Unité Opérations - Direction de Solutions IARD



Les dimensions de la qualité des données⁵⁷

Cette illustration représente des indicateurs possibles pour mesurer la qualité des données et, ainsi, s'assurer d'avoir toujours des données fiables sur lesquelles baser des analyses pertinentes pour déterminer nos stratégies.

⁵⁷ Livre blanc ISACA : BIG DATA : Impacts et Avantages

2.2.4 EVOLUTION DES ROLES

Le Big Data est lié à l'essor du digital et entraîne le client dans une attente forte, en terme d'usage et de réactivité. Ces comportements, en pleine mutation, obligent les assureurs à repenser leur modèle pour le recentrer sur le client. Ainsi, le marketing, grâce au Big Data, va pouvoir travailler l'aspect orientation client. Les Directions Techniques Produits vont optimiser les performances des risques assurés, en terme d'appétence et de comportement client. Ainsi, suivant les différentes responsabilités portées par chacune des Directions métiers dans les organisations assurantielles, le besoin va se calquer sur les comportements client. Les Directions Marketing et Directions Techniques Produits deviendront les interlocuteurs principaux de la Direction des Systèmes d'Informations.

L'objectif poursuivi est d'innover pour nos clients et de créer des services associés à nos produits afin de prévenir des risques et, ainsi, de personnaliser nos offres. C'est pour cette raison que le marketing devra s'appuyer sur des Directions Techniques Produits pour déterminer les services adaptés aux clients et en adéquation avec les caractéristiques techniques du produit d'assurance auquel il sera rattaché. On pourrait envisager des directions techniques, qui acquièrent des compétences en marketing, afin de créer des produits accompagnés de services préventifs.

Les directions en charge de la gestion des sinistres ont également un rôle prépondérant à jouer dans l'innovation et la mise en place de services préventifs. Très en avance dans le domaine du préventif, depuis plusieurs années, et très au fait du Big Data dans le domaine du risque, ces services pourront amener une expertise dans le domaine de la corrélation des données, entre le risque et l'assuré. Ces services n'ont de cesse de travailler sur le risque pour le réduire et protéger les assurés, par exemple en croisant les données météorologiques, avec un portefeuille d'assurés.

COVENTEO, est un outil qui permet de renforcer les plateformes de gestion sinistres en cas de tempêtes. Cet outil, couplé avec des données issues de sites météorologiques (données externes) et le portefeuille des assurés permet de prédire les régions où les assurés risquent d'être le plus touchés. De cette manière, les plateaux de gestion sinistres peuvent être renforcés en conséquence.⁵⁸

⁵⁸ Basé sur l'interview de Charles Dumartinet du 07 Novembre 2013

Pour la réalisation de cet outil, des compétences en gestion de sinistres ont été nécessaires pour les aspects organisation des plateaux de gestion, mais aussi des compétences météorologiques pour mettre en œuvre les modèles prédictifs et interpréter les résultats.

Les compétences et les collaborations vont devoir évoluer. Les métiers de l'assurance, tels que nous les connaissons aujourd'hui, quel que soit le domaine, vont devoir s'adapter pour s'enrichir de nouvelles compétences.

L'assurance d'aujourd'hui est un métier d'experts: expertise dans le domaine technique de l'assurance, expertise dans le domaine juridique, lié aux produits d'assurance, expertise dans le domaine du marketing, expertise dans le domaine de l'actuariat, avec un chef d'orchestre à leur tête capable de faire appel à ces expertises pour fabriquer des produits d'assurance. Demain, avec le Big Data, il s'agira pour ces experts d'élargir leur palette de compétences, afin d'apprendre des autres domaines dans le but de trouver des corrélations significatives et pertinentes pour fabriquer les services en adéquation avec les attentes de nos assurés. Ces compétences seront techniques, métiers, mathématiques ou stratégiques, en fonction des compléments pertinents à apporter à son domaine de connaissances.

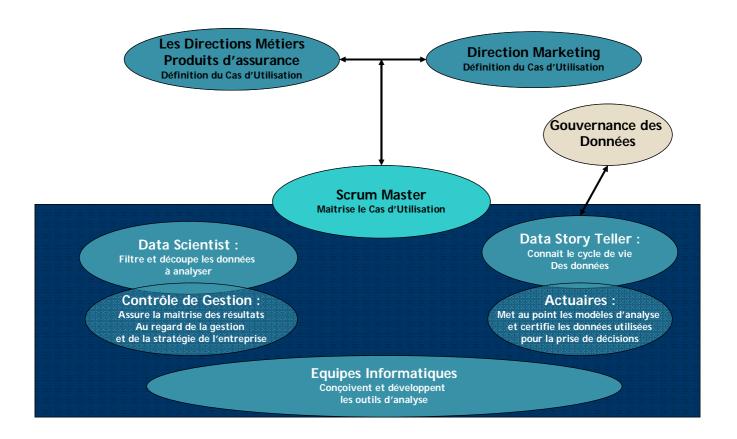
Les méthodes et approches des sujets assurantiels évoluent car il devient impératif de comprendre ce que les assurés souhaitent. Les rôles doivent donc devenir transverses. La base d'une corrélation, élément clé du Big Data, est d'avoir une vision globale du sujet à traiter. C'est pour cette raison que les phases de construction des projets ne peuvent plus se faire les unes après les autres comme aujourd'hui mais bien ensemble.

2.3 GOUVERNANCE ET DEROULEMENT D'UN PROJET BIG DATA

Les méthodes actuelles de conduite de projets sont basées sur le « cycle en V ». Il s'agit de définir, précisément, un besoin pour lequel on sait lister les données à utiliser, pour y répondre. L'approche, par la donnée avec le Big Data, sera différente. On détermine un cas d'utilisation et on cherche avec quelles données on peut y répondre. Il convient, auparavant, de déterminer la ou les sources de données les plus pertinentes, pour effectuer nos fouilles (réseaux sociaux, images, SMS, ...). Les travaux d'analyses consistent, ensuite, à déterminer les données à utiliser et à corréler pour répondre, efficacement, au cas d'utilisation imaginé.

Les méthodes agiles sont tout à fait adaptées à ce type de travaux. Elles permettent l'association de compétences différentes, qui collaborent ensembles pour trouver des réponses. La collaboration de ces profils est basée sur la mise au point de la solution, par paliers successifs, en fonction des résultats observés. Les différentes itérations sont réduites dans le temps et permettent ainsi plusieurs tests avant de trouver la meilleure réponse au cas d'utilisation.

C'est l'étroite collaboration entre les différents profils, qui fera naître des corrélations de données intéressantes pour l'entreprise.



Exemple d'équipe en mode agile pour un projet Big Data⁵⁹

Ce schéma représente la collaboration nécessaire, au sein d'un projet Big Data. Les directions métiers et marketing définissent un cas d'utilisation, pour démarrer les fouilles dans les données :

- ° Pourquoi nos clients résilient ? Existe-t-il un comportement, permettant d'anticiper le départ d'un client ?
- ° Peut-on anticiper les besoins d'assurance de nos clients en IARD ?⁶⁰

 $^{^{59}}$ Basé sur le cours de Management Information Systems du CNAM de M. Franck LAURENCE & M. Jean-Michel RAICOVITCH

⁶⁰ Basé sur l'interview du 04/11/2013 de Marielle Vo-Van-Liger – Directrice du département Relation Client et Marketing Stratégique

Le Big Data dans l'Assurance

Une fois le « Cas d'Utilisation » défini, le rôle du « Scrum Master » est d'animer l'équipe projet qui va, effectuer les recherches dans les données. En fonction des corrélations le « Scrum Master », en relation avec les directions métiers, sera en mesure d'orienter les recherches, en fonction des résultats obtenus.

Dans la littérature dédiée au Big Data, on parle d'un profil particulier, le « Data Scientist », scientifique de la donnée autrement dit, un expert en analyse des données. Les compétences du « Data Scientist » regroupent les domaines suivants :

- ° « Comprendre des problématiques business
- ° Collecter les données
- ° Optimiser la qualité des données
- ° Définir et utiliser des algorithmes et modèles d'analyse
- ° Dégager des tendances et donner du sens à partir des données
- ° Communiquer, de façon intelligible, des résultats d'analyse »⁶¹

En plus de ces compétences métiers, mathématiques et statistiques le « Data Scientist » doit, aussi, connaître le langage informatique utilisé pour mettre au point son algorithme d'analyse.

Toutes ces compétences sont compliquées à trouver dans un seul et même profil. C'est pourquoi le schéma présenté, ci-dessus, illustre une collaboration entre plusieurs profils. Et comme expliqué, plus haut dans ma thèse, cette collaboration permet de monter, en compétence, dans les domaines nécessaires à la réalisation de ces analyses de données.

⁶¹ Christophe Brasseur « Enjeux et usages du Big Data technologies, méthodes et mise en œuvre »

Dans cette articulation, le « Data Scientist » est présent pour apporter son expérience sur l'échantillonnage et l'optimisation de la qualité des données sources. Il a, en charge, la conduite d'ateliers pour déterminer, avec l'équipe, quelles données il convient de sélectionner pour effectuer les fouilles.

Le « Data Story Teller » fait partie du domaine métier, pour lequel le cas d'utilisation a été défini et apporte sa maîtrise du cycle de vie des données de son domaine (« Data Story Telling »). En relation étroite avec la gouvernance des données, il s'assure de la signification et de la pertinence des données utilisées dans l'analyse. Il porte, aussi, via la gouvernance des données, la cohérence avec les autres données de l'entreprise.

Le rôle des actuaires, dans ces nouvelles approches de l'assurance permet de mettre au point et de valider les modèles et algorithmes prédictifs, permettant la mise en œuvre de nouveaux services ou de nouvelles mutualités de risques assurables. Ils sont garants, avec l'organe de gouvernance des données, de la qualité des nouvelles informations, à partir desquelles ces résultats ont été obtenus. Le contrôleur de gestion sera là, pour s'assurer que les nouvelles opportunités trouvées sont maîtrisées et en adéquation avec la stratégie définie par la direction de l'entreprise. De plus, il s'assurera, que ces nouveaux services ne font pas courir à l'entreprise un risque dans sa gestion technique assurantielle. Ainsi, le processus de décision est contrôlé et validé, tout au long de la chaîne de décisions.

Les équipes informatiques exécuteront quant à elle les tâches techniques, comme la collecte des données et la maintenance de l'architecture.

CONCLUSION DE LA DEUXIEME PARTIE

Les possibilités offertes par le Big Data, dans le domaine de l'assurance, ouvrent des perspectives nouvelles pour ce métier traditionnel.

« Quand, auparavant, il s'agissait de vérifier si telle décision entrait bien dans le plan d'une providence divine, il s'agit désormais de mesurer le « risque » que l'on prend à faire tel choix plus qu'un autre. Cela ne réduit pas pour autant l'incertitude (ontologique⁶²) des choses mais fournit une aide utile à la décision. Il va bientôt devenir « rationnel » de « préférer » les hypothèses qui maximisent « l'espérance » de gain, celles qui réduisent le « risque », -ce qui en réalité ne fournit aucune certitude concernant la marche des évènements. »⁶³

Ces nouvelles opportunités, induites par le Big Data, nécessitent une analyse des risques, afin de déterminer les conséquences de ces opportunités vis-à-vis du métier d'assureur et les conséquences pour nos assurés.

Bien entendu, ces décisions ne vont pas sans se poser la question de l'éthique et de la réglementation y faisant référence.

_

⁶² Relatif à l'étude de l'être en tant qu'être

⁶³ François Ewald – Assurance, prévention, prédiction ... dans l'univers du Big Data

3. LES RISQUES ASSOCIES AU BIG DATA

En regard des enjeux présentés dans le chapitre précédent, j'ai souhaité étudier les risques induits par l'arrivée et l'utilisation du Big Data par les assureurs.

Ces risques sont classés en fonction de l'impact financier qu'ils auraient sur l'entreprise.

Une fois l'univers des risques défini et cartographié, selon les impacts pour une société d'assurance et la réglementation, j'ai étudié la manière de les réduire, pour les rendre acceptables et maîtrisables.⁶⁴

⁶⁴ Basé sur les échanges et le cours de M. Stevan P.G. Corbett – Directeur Partner Reinsurance

3.1 <u>CLASSIFICATION DES RISQUES EN FONCTION DE</u> <u>L'IMPACT</u>

Les sommes financières du tableau représentent la classification des risques et sont estimées par rapport :

- au risque d'image,
- aux relations clients et à la perte potentielle d'une partie des assurés
- à l'organisation de l'entreprise et l'opérationnel (panne grave du système d'informations)
- au règlementaire et au juridique.

IMPACT	FINANCIER	Réputation, Clients, Salariés	Conséquences et sanctions
Critique	Plus de 5M€	* Mise en cause de la réputation de la société > à 1 an * Perte des clients (15%) * Suspension de l'activité suite à la mise en cause de la réputation	* Suppression de l'accord de branche assurance. * Sanctions pénales pour la direction.
Elevé	De 1M€ à 5 M€	* Publicité négative avec mise en cause temporaire de la réputation de la société (quelques mois) * Perte de clients (10%)	* Suspension de l'activité pendant 1 à 2 semaines * Sanctions administratives et amende
Moyen	De 250K€ à 1M€	* Publicité négative avec des impacts modérés (quelques semaines) * Perte potentielle de clients (5%)	* Suspension de l'activité pendant quelques jours * Sanctions administratives ou amende (rappel à l'ordre) et révision de la politique commerciale
Bas	De 50K€ à 250 K€	* Faible impact sur la réputation et la perte de clients	* Adaptation mineure de la politique commerciale

3.2 L'UNIVERS DES RISQUES DU BIG DATA POUR UNE SOCIETE D'ASSURANCE

Risques Externes

- 1. Modification de la loi
 - CNIL
 - ACPR
 - Fiscalité
- 2. Comportement des clients Attentes des clients
- 3. Risques Concurrentiels

Risques Entreprise

- 4. Risques Stratégiques
- Réputation avec l'utilisation de données personnelles assurés
 - Gouvernance des Données
- Mauvaise orientation stratégique
 - 5. Risques Financiers
 - Coût de mise en oeuvre
 - Investissement à perte
 - 6. Risques Opérationnels
 - Compétences humaines

Risques Externes:

1 - Modification de la loi :

Regroupe toutes les instances gouvernementales, susceptibles de légiférer sur les données et la protection de la vie privée.

2 – Comportement des clients / Attentes des clients :

La sensibilisation et l'information faites autour du «Big Data» peuvent faire évoluer les comportements des clients. Ils peuvent penser que leur assureur est trop intrusif dans leur vie privée. A contrario, leurs attentes vis-à-vis des fournisseurs de services, vont s'accroître en terme d'exigences.

3 – Risques concurrentiels :

Plusieurs entreprises d'assurance investissent dans le Big Data. Les entreprises pionnières qui mettront en évidence des résultats très pertinents pourront définir une stratégie de conquête de parts de marché sur des segments d'assurés appartenant à des entreprises concurrentes. Le risque réside dans le fait de ne pas s'intéresser à ce que peut apporter le Big Data pour une entreprise d'assurance.

3.2.1 <u>Cartographie des risques externes avant</u> <u>MITIGATION</u>



3.2.2 MITIGATION DES RISQUES EXTERNES

1 – Mitigation du risque « Modification de la loi » :

« Lobbying », auprès des autorités Européennes pour leur faire prendre conscience de la dimension Big Data dans les textes de loi. Autoriser l'utilisation de données, permettant d'anticiper les risques de manière ciblée avec le consentement du client. Trouver, une juste mesure, entre le service apporté et la protection de la vie privée.

2 – Mitigation du risque « Comportement des clients / Attentes des clients » :

Surveiller et anticiper les attentes des clients via les travaux Big Data permanents pour être au rendez vous. La consolidation des données utilisées se fera, au travers, de la gouvernance des données de l'entreprise.

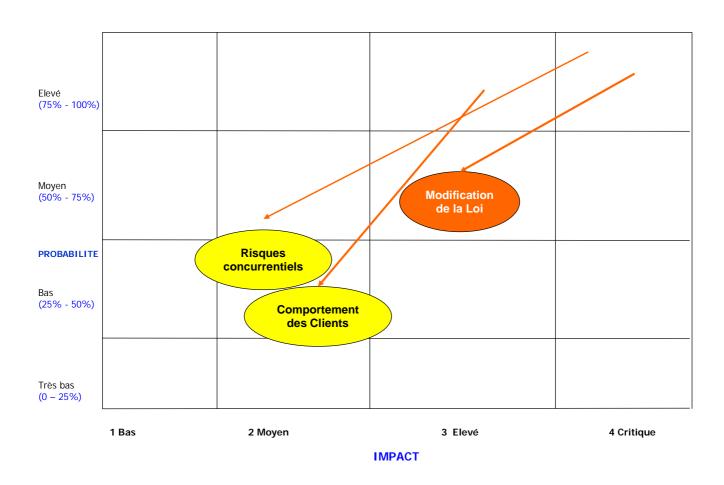
Se faire connaître, pour que le client associe le besoin d'assurance et notre entreprise. Publicité ciblée, vers les clients à atteindre et les médias sur lesquels on les trouve.

Inciter le client à prendre conscience des services et accompagnements proposés par son assureur.

3 – Mitigation du risque « **Risques Concurrentiels** » :

Investir dans le domaine du Big Data pour progresser et acquérir une expérience significative afin de rester compétitif dans le domaine des assurances.

3.2.3 <u>Cartographie des risques externes apres</u> <u>MITIGATION</u>



Risques Entreprise:

4 – Risques stratégiques

En fonction des données utilisées par l'entreprise et des offres, en regard de cette utilisation, certains clients n'apprécieront pas cette intrusion dans leur sphère privée. Le rôle de l'assureur, et l'image qu'il véhicule auprès de ces assurés, n'est pas en phase avec ce type d'intrusion. Il convient donc de porter une attention particulière aux données utilisées et à la manière dont est proposé le service issu de l'interprétation de ces données.

Dans la même catégorie, dite stratégique, une mauvaise interprétation des données peut induire en erreur et donner lieu à la fabrication de produits ou de services non adaptés au besoin du client.

La composition des équipes en charge des analyses Big Data, présentée dans la seconde partie de cette thèse, permettra de surveiller, en permanence, ce risque. Associée à des enquêtes de satisfactions client, l'évolution de ce risque sera contrôlée et le risque d'image diminué.

5 – Risques financiers

Les coûts de mise en œuvre d'une architecture permettant de faire du Big Data sont à étudier, en regard des objectifs à atteindre. Il conviendra d'étudier les solutions du marché et voir celle qui convient le mieux à l'entreprise. Où stocker un tel volume de données pour travailler dessus reste une question fondamentale ? Le « Cloud » avec son offre de stockage est une des réponses possibles.

En ce qui concerne les investissements à perte, il faut savoir que dans le Big Data, on parle souvent de ce qui a fonctionné et très peu, de ce qui ne fonctionne pas. En d'autres termes, il peut arriver de ne pas trouver de corrélations significatives, en fouillant les données. Dans les démarches classiques on établit un besoin, puis un investissement en rapport avec ce besoin, pour permettre de le développer. Dans le cadre du Big Data, on va déterminer un cas d'utilisation, puis on va itérer pour vérifier d'éventuelles corrélations entre des données permettant de reproduire ce cas d'utilisation. Il peut arriver que les données ne correspondent pas à la réalisation du cas d'utilisation et

même en itérant nous n'y arriverons pas. On peut, aussi, découvrir un autre cas d'utilisation, qui n'était pas celui identifié au départ. L'investissement, mis en regard du cas d'utilisation, pourra donc se révéler infructueux ou bien inadapté.

6 – Risques Opérationnels

Cette nouvelle façon de travailler et d'aborder le métier de l'assurance comporte un risque, en terme de compétences opérationnelles. En effet, certains profils essentiels, comme le « Data Scientist », sont difficiles à recruter. Les formations n'existent pas encore.

L'organisation de ce type de projet diffère aussi de l'organisation classique des projets menés dans l'assurance. Le fameux cycle en V n'est plus adapté à ce type de projet où le besoin est exprimé sommairement au démarrage et peut évoluer en cours de projet.

3.2.4 <u>Cartographie des risques Entreprise avant</u> <u>MITIGATION</u>



3.2.5 <u>MITIGATION DES RISQUES ENTREPRISE</u>

4 - Mitigation du « Risques Stratégiques » :

Il faut, aussi, être prudent dans notre approche et convaincre que nous proposons des services pour accompagner nos assurés et diminuer les coûts de leur assurance, par une meilleure prévention des risques. Ces nouveaux types de service, vendus avec le produit d'assurance doivent avoir pour objectif de réduire le risque, voire dans certains cas de le faire disparaître.

Contre le risque d'erreur, dans le process de « Decision Making », la mise en place de pilote pour tester les solutions mises en œuvre sera une réponse efficace, avant un investissement conséquent. De plus, l'équipe en charge des fouilles dans les données, sera constituée d'actuaires et de contrôleurs de gestion, en relation avec la gouvernance des données. Cette composition devient un gage de sécurité, au niveau de l'interprétation et de la sélection des données.

5 - Mitigation du « Risques Financiers » :

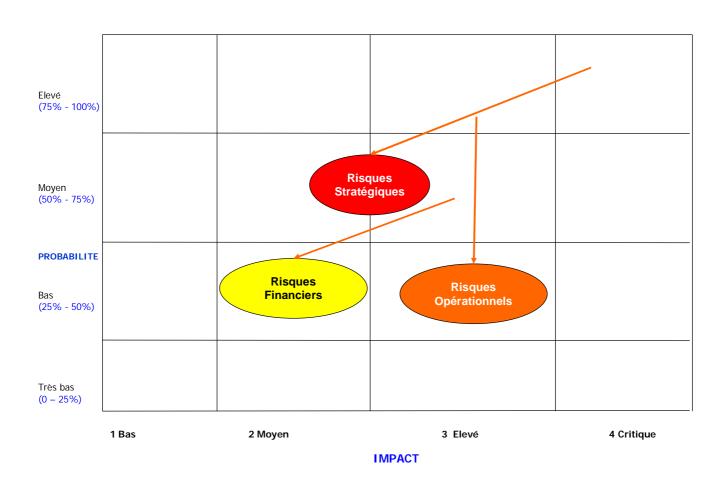
L'architecture informatique, mise en place, pourra être réutilisée. La perte financière, en terme d'investissement matériel, doit être mise en regard des possibilités de réutilisation. En terme d'investissement projet, cela revient à investir sur une équipe entre huit et dix personnes, pendant 1 an, soit un coût moyen de $10*600 \ \text{e}/\ \text{jour} = 6\ 000 \ \text{e}/\ \text{jour} = 220\ \text{jours} = 1\ 320\ 000 \ \text{e}$ par an, hors matériel informatique.

6 – Mitigation du « Risques Opérationnels » :

Il s'agit, de mettre en place un accompagnement efficace de nos équipes, capable de produire ce travail. Faire appel à une entreprise expérimentée, dans ce nouveau domaine. De plus, les méthodes de conduite et d'organisation du projet évoluent aussi. Un accompagnement sur ces nouvelles manières de travailler est indispensable pour faire monter nos équipes internes en compétence, afin qu'elles deviennent autonomes sur ces sujets.

Pour les compétences, qui n'existent pas encore sur notre marché, des partenariats avec des écoles et des organismes de formation sont à lier, afin qu'ils investissent sur ces nouvelles formations.

3.2.6 <u>Cartographie des risques Entreprise apres</u> <u>MITIGATION</u>



3.3 $LA CNIL^{65}$

Dans le cadre du respect et de la protection de la vie privée, la CNIL s'intéresse depuis plusieurs années au Big Data.

« La CNIL est l'autorité en charge de veiller à la protection des données personnelles. A ce titre, elle dispose notamment d'un pouvoir de contrôle et de sanction. Jouant aussi un rôle d'alerte et de conseil, elle a pour mission de veiller à ce que le développement des nouvelles technologies ne porte atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques. De manière plus générale, la CNIL protège la vie privée et les libertés dans le monde numérique.

De plus, le projet de règlement européen devrait consacrer le principe d'un « droit à l'oubli » numérique nous permettant de mieux maîtriser notre vie en ligne. Ce nouveau droit devra s'exercer dans le respect de la liberté d'expression, du droit de la presse et du devoir de mémoire. Dans ce contexte, la CNIL a lancé une consultation auprès des internautes sur ce droit, souvent invoqué mais dont les contours restent encore à préciser. En parallèle, elle consulte également les professionnels concernés.

Dans ce contexte, la CNIL invite l'ensemble des pouvoirs publics nationaux à se mobiliser et à promouvoir une vision humaniste de la protection des données personnelles, qui définissent un juste équilibre entre droits fondamentaux et innovation technologique. »⁶⁶

La CNIL a mis en place des lois et des obligations de déclarations pour contrôler l'utilisation des données à caractère personnel afin de protéger les internautes et les clients présents dans les fichiers informatiques des entreprises.

⁶⁵ Commission Nationale Informatique et Liberté – Paragraphe réalisé avec l'assistance de Messieurs Pascal Palut et Olivier Desbiey de la CNIL.

⁶⁶ Site www.cnil.fr

« Cinq règles d'or sont à respecter :

- ° Finalité du traitement : la CNIL a refusé que des fichiers de caisses de sécurité sociale soient utilisés pour envoyer de la publicité aux assurés ; pas d'utilisation des fichiers administratifs à des fins de prospection politique.
- Pertinence des données : interdiction de collecter les données sensibles qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale ainsi que les données relatives à la santé ou à la vie sexuelle. Sauf exceptions (consentement, intérêt public...)
- ° Conservation limitée des données : les données ne peuvent être conservées dans les fichiers au-delà de la durée nécessaire à la finalité poursuivie qu'à des fins historiques, statistiques ou scientifiques
- Obligation de sécurité: Respect de l'intégrité et de la confidentialité des données: empêcher que les données soient déformées, endommagées ou que des tiers non autorisés y aient accès
- ° Respect des droits des personnes : Les personnes doivent être informées, lors du recueil, de l'enregistrement ou de la première communication des données :
 - o de la finalité du traitement,
 - o du caractère obligatoire ou facultatif des réponses et des conséquences d'un défaut de réponse,
 - o de l'identité du responsable du traitement,
 - o des destinataires des données,
 - o de leurs droits (droit d'accès et de rectification, droit d'opposition),
 - o le cas échéant, des transferts de données vers des pays hors UE. »⁶⁷

_

Aurélie Banck – Ancienne Juriste à la Direction des affaires juridiques de la CNIL Pôle Banque – assurances – finances
 Intervention MBA CNAM – 13 décembre 2013

Le Big Data dans l'Assurance

Une étude a montré aux Etats-Unis, que 87% des personnes étaient identifiables avec uniquement trois données :

- le sexe
- la date de naissance
- le code postal⁶⁸

En France, la CNIL est très vigilante sur l'utilisation de ce type de données à des fins d'identification.

Mais des études menées dans d'autres pays montrent qu'il est possible d'identifier des personnes grâce à des données moins significatives.

⁶⁸ Quantified Self Data & Privacy - Third Wave : http://thirdwaveberlin.com/2011/12/quantified-self-data-privacy/

Yves-Alexandre de Montjoye, doctorant au MIT, illustre dans une étude « Unique dans la foule » avec César A. Hidalgo, Michel Verleysen et Vincent D. Blondel que les traces laissées par nos déplacements sont uniques. Cette étude, effectuée pendant quinze mois, sur les données issues des déplacements d'un million et demi de personnes a permis d'identifier 95% des gens.

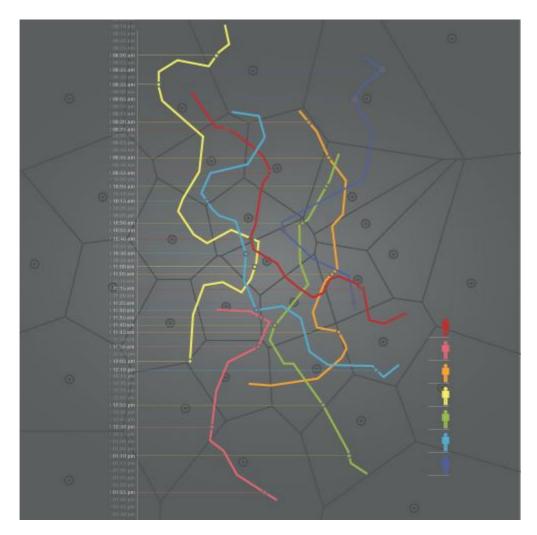


Illustration de l'unicité de nos parcours repérés via des antennes mobiles.⁶⁹

⁶⁹Image tirée de l'étude "Unique dans la foule" - http://www.internetactu.net/2013/12/05/en-quoi-les-big-data-sont-elles-personnelles/

Comme le rappelle François Ewald dans After Risk – Vers un nouveau paradigme de l'assurance

L'Assurance à l'âge de Big Data de Septembre 2013 : « il ne s'agit pas de brandir abstraitement les grands principes de la régulation des données personnelles (protection de la vie privée et discrimination) contre l'assurance. Le risque serait alors de priver les citoyens, les consommateurs des nouveaux services de protection que les assureurs seraient en mesure de proposer grâce à l'usage des nouvelles technologies : nouvelles manières de rendre les services, réduction des coûts, offre de nouveaux services. »

Face à ces défis, l'assureur a un rôle important à jouer et une dimension nouvelle à acquérir. Le dialogue avec les autorités doit être engagé, dès maintenant, afin de leur proposer de nouvelles perspectives, par rapport aux possibilités technologiques qui apparaissent.

3.4 L'ASSURANCE ET L'UTILISATION DES DONNEES

« Enfin, une voie innovante émerge à partir de l'idée d'un partage réciproque des données entre clients et entreprises, selon le principe résumé par Daniel Kaplan : « Si l'entreprise a une information sur le client, le client doit l'avoir aussi » (projets MesInfos et MiData), alors qu'actuellement, les clients connaissent de moins en moins ce que les entreprises savent d'eux. Comme l'explique Daniel Kaplan, « cette idée s'inscrit dans le concept de Vendor Relationship Management (VRM) créé par Doc Searls⁷⁰ pour équilibrer le Customer Relationship Management (CRM) », dont les dérives provoquent, paradoxalement, une chute de plus en plus forte de la fidélisation. Pour Doc Searls, le seul endroit où le 360° est réellement possible, c'est du côté du client et le mouvement VRM veut poser les bases d'une relation client saine dans l'ère de l'industrialisation du numérique, « une personnalisation demandée remplaçant une personnalisation subie » grâce par exemple à des outils émergents visant à « outiller » le client et à responsabiliser sa prise de décisions face aux entreprises, comme les Personal Data Stores (MyDex, Privowny, personal. com...)⁷¹.

⁷¹ Logiciels de gestion et protection de données personnelles sur le web

_

⁷⁰ Journaliste Américain et auteur de *The Intention Economy: When Customers Take Charge*

Mais comment ces données vont-elles pouvoir être « réutilisées » par les individus ? Daniel Kaplan évoque le mouvement de « quantification de soi » (« Quantified Self ») : avec ces données de consommation, l'individu pourrait réfléchir à sa mobilité, à son empreinte carbone, enrichir son bilan de compétence, appliquer des mesures d'éco-responsabilité à sa consommation… »⁷²

Le « Quantified Self » permet de regrouper des communautés qui partagent des données personnelles. « Les domaines principaux sans qu'ils soient exhaustifs sont les suivants :

- o Expérimentation sur soi-même
- o Gestion du comportement
- o Suivi de sa position dans l'espace
- o Information digitale sur le corps
- o Données biométriques
- o Auto-évaluation
- o Auto-diagnostic
- o Préparation physique et sportive
- Données de santé
- o Nutrition et perte de poids
- o Suivi et amélioration de sa productivité
- o Informations génétiques personnelles
- o ... »⁷³

⁷² Cahier Innovation & Prospective n°1 (CNIL) - Vie Privée à l'Horizon 2020

⁷³ http://interestingviews.fr/2011/05/31/le-quantified-self-doit-on-compter-sur-soi : Interview : Emmanuel Gadenne, Cofondateur de QuantifiedSelf Paris

Les objectifs de ce partage d'informations personnelles sont de mieux se connaître, de s'améliorer, en confrontant nos bonnes pratiques, pour progresser. De plus, en s'affichant ainsi, dans une communauté choisie, cela permet de trouver une motivation dans le regard des autres.



Exemple de données personnelles partagées pour progresser 74

Cette application, « BodyMedia », travaille en collaboration avec des coachs sportifs et des professionnels de santé. Quoi de plus naturel pour un assureur avec le consentement de l'assuré, que de

_

⁷⁴ http://www.fredcavazza.net/2011/06/15/le-quantified-self-au-service-de-la-productivite-individuelle-et-collective/

coupler ce type d'application avec un produit santé ou bien une assurance décès, comme évoqué précédemment, dans ma thèse.

Il ne s'agit pas de faire de la discrimination, mais bien de proposer une nouvelle forme d'assurance, où les bons risques aident les mauvais risques à progresser pour devenir, à leur tour, des bons risques. Il s'agit d'ajouter une dimension supplémentaire à la mutualisation, afin de pouvoir en tirer un vrai bénéfice.

Ce partage de bonnes pratiques, animé par l'assureur lui-même, permet de répondre à plusieurs axes :

- ° Fidélisation de l'assuré : via l'utilisation régulière de l'application de « Quantified Self », liée au produit d'assurance, pour saisir ses données et ses progrès, consulter les bonnes pratiques, être en contact avec son assureur pour des points réguliers, ...
- Ajustement du tarif : grâce aux progrès accomplis et, en fonction de paliers définis par avance, le tarif peut être ajusté. On retrouve le principe de la valorisation des données personnelles mises à disposition de son assureur.
- Détection des fraudes : un mauvais risque, qui ne progresse pas, va attirer l'attention et sera repéré plus rapidement par l'assureur.

Cette méthode peut aussi s'appliquer aux produits d'assurance. Un des concepts comportemental issu des réseaux sociaux est celui de la communauté. Il s'agit de relations ayant des goûts et des comportements proches voir similaires aux vôtres. On partage ainsi des loisirs, des produits et surtout on les recommande à sa « communauté » de relations. Lorsqu'on recherche un conseil, il n'est pas rare aujourd'hui de se tourner vers sa « communauté » pour obtenir un avis. L'assurance n'échappe pas à ce type de comportement. L'assureur s'inspirera des communautés pour adapter ses produits aux attentes des assurés. Sur le même principe que le « Quantified Self », la création d'un « Quantified Product » avec les informations issues de commentaires de ces communautés sur nos produits d'assurance pourrait permettre de les améliorer en répondant aux attentes des clients.

⁷⁵ Basé sur la Conférence «Le digital dans l'assurance » Chapuis Halder & Compagnie le 30 janvier 2014

Dans cette perspective, l'assureur va devoir jouer plusieurs rôles :

- o récolter les données nécessaires pour rendre un service,
- o faire connaître à l'assuré la manière dont il va les utiliser,
- o lui permettre de les modifier ou de les retirer à tout moment de son choix,
- o garantir que ces données personnelles seront protégées et exclusivement réservées à l'usage convenu avec l'assuré.

On pourrait par exemple mettre en place un service d'aide à la conduite en utilisant les données issues de boîtiers embarqués dans les véhicules des assurés. Il conviendra au préalable d'informer les personnes souhaitant adhérer à ce service de la manière dont leurs données seront interprétées.

Une plateforme de type « Quantified Self » sur les bonnes pratiques de conduite permettra de réunir une communauté de conducteurs désireux de partager et d'apprendre des autres. L'animation de ces communautés est réalisée par l'assureur avec la mise en place d'observations, permettant de contacter nos assurés pour faire des bilans sur les progrès constatés.

Il existe chez de nombreux assureurs des plateaux téléphoniques dédiés à certaines tâches, comme la vente à distance de contrats ou bien la gestion des sinistres. L'animation de ces nouvelles communautés assurantielles pourra se faire sur le principe du plateau téléphonique. Ce nouveau métier d'assureur « Web 2.0 » est à construire et à développer, selon les axes suivants :

- o animation,
- o conseil en assurance,
- o coaching et suivi,
- o fidélisation, via la personnalisation.

C'est bien, au travers, de ces services, que se fera la personnalisation demandée par le client. Sans remettre en cause les fondamentaux assurantiels, il s'agit d'ajouter une dimension supplémentaire, afin d'optimiser les produits via des services.

Le Big Data dans l'Assurance

En terme stratégique, plusieurs possibilités s'offrent aux assureurs pour créer ces services à plusvalue : créer des partenariats, ou bien acheter et fusionner avec des structures déjà équipées de services pouvant intéresser l'assurance, étudier les données issues de ces services et voir quelles adaptations sont à envisager pour les coupler à des produits d'assurance.

Bien entendu, en tant qu'assureur et pour rester en conformité avec la CNIL, il conviendra de maîtriser le cycle de vie des données utilisées. Dans cette optique, l'assureur devra investir pour créer des partenariats ou acheter ses propres applications. Il sera ainsi en mesure de garantir la sécurité et la protection de ces informations.

Des travaux, à destination des assurés, pour les informer des différentes utilisations possibles de leurs données devront être mis en œuvre. En effet, les données collectées pour répondre aux besoins d'un service pourront aussi, comme déjà expliqué dans ma thèse, être utilisées dans le cadre d'analyses pour effectuer de nouvelles corrélations.

La transparence et la garantie de l'accès à ses propres informations, pour les modifier ou interdire leur usage, sont des devoirs que se doivent de respecter les assureurs.

Le gouvernement britannique est à l'origine du projet «MyData», qui regroupe une vingtaine d'entreprises, qui se sont engagées à partager les données qu'elles possèdent sur leurs clients, avec eux. Ainsi, les dits clients seront en mesure de transférer et de réutiliser ses informations (qui leurs appartiennent), pour d'autres services.

De plus en plus de sociétés, dont les activités sont la protection et la portabilité de vos données personnelles sur le web, voient le jour, actuellement. Ce phénomène trouve son explication dans le fait que, chaque individu « digital », prend conscience de la valeur de ces propres données. Jusqu'à présent, les générations Y ou « digital Natives » ⁷⁶ avaient pour habitude d'exposer au grand jour leurs données et humeurs personnelles. Puis, petit à petit, ces expositions se sont vues restreintes à une population d'intimes. De nos jours, il faut être de plus en plus vigilant sur ce qu'on publie et quels sont les amis qui y ont accès

 $^{^{76}\} Voir\ thèse\ de\ Guillaume\ Billy\ -\ Les\ communautés\ virtuelles\ peuvent-elles\ réinventer\ l'assurance\ ?-ENASS\ 2012$

Le Big Data dans l'Assurance

Chaque individu prend conscience de la valeur de ses données. Grâce aux méthodes et outils induits par le Big Data nous savons comment valoriser toutes ces données.

A l'assureur de proposer des solutions pour protéger les aspects privés de ces données et les utiliser, à

bon escient, pour la prévention des risques, via un service personnalisé.

L'assurance de demain, sera un échange transparent entre un assuré et un assureur, où chacun pourra s'exprimer et accéder, en toute confiance, à ses données pour les optimiser et choisir comment les

utiliser.

La protection des données passe par un ajustement entre le « tout interdit » et le « tout permis ». Comme je l'ai déjà évoqué, plusieurs fois, dans ma thèse, l'interdiction des données identifiantes ne suffit plus aujourd'hui pour anonymiser des données. Les croisements et corrélations de données rendus possibles grâce au Big Data permettent d'étendre la connaissance du client, voire de l'identifier.

Demain, il s'agira de conforter la CNIL dans son rôle, mais de l'appréhender différemment. Il s'agira

de permettre à la CNIL d'éduquer la population aux pratiques numériques, afin de mettre en place

« une pédagogie des usages et des bonnes pratiques. » 77 Cette pédagogie passe par des

recommandations à faire, auprès des entreprises d'assurance, sur la mise en place de notices

d'informations et d'assistance, auprès des assurés. Cette protection passe aussi par une

responsabilisation des entreprises sur les aspects sécuritaires et protection des données à caractères

privés.

« Si la CNIL ne peut pas tout auditer elle-même, elle doit être en mesure de distribuer et de superviser les contrôles. Certains pensent, également, que, à condition de disposer de l'autonomie et des

compétences nécessaires, le Correspondant Informatique et Libertés pourrait devenir, demain, un vrai

« responsable de la conformité » de son organisation. »⁷⁸

Toutes ces mesures auront pour but de rendre aux assurés le pouvoir de contrôler et de valoriser leurs

données personnelles, au travers des opportunités proposées par le Big Data.

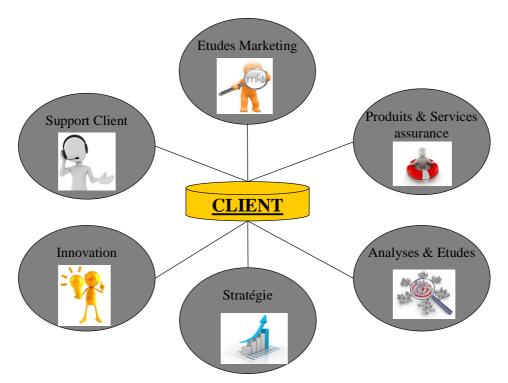
⁷⁷ CNIL – Cahier IP n°1

⁷⁸ CNIL – Cahier IP n°1

CONCLUSION DE LA TROISIEME PARTIE

Le numérique et la révolution digitale sont, depuis plusieurs années, au centre des stratégies de nos entreprises. « Partout le numérique va introduire des ruptures fondamentales, qu'il convient d'anticiper et de maîtriser. Le modèle d'affaires de l'entreprise numérique ne ressemble plus au modèle d'affaires classique, dans ses fondamentaux et sa déclinaison sur le marché » ⁷⁹. La nouveauté, aujourd'hui, dans cette avalanche de données, réside dans le souhait des entreprises de replacer le client au centre des préoccupations. « La personne est désormais l'objet de toutes les attentions. » ⁸⁰

De cette manière, tout va tourner autour du client et des données visibles qu'il souhaitera communiquer.⁸¹



⁷⁹ Bruno Ménard :2010

⁸⁰ Entreprises et Culture numérique - CIGREF

⁸¹ Inspiré du VINT Research report – No more secrets with Big Data Analytics – Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen – Erik Van Ommeren - Sandeep Sachdeva - SOGETI

CONCLUSION

L'assureur a un rôle essentiel à jouer avec l'avènement du Big Data. La barrière entre le fait d'offrir un service à un consommateur et le fait de l'inciter à choisir tel ou tel produit, selon notre convenance, est très mince. Les informations accumulées et mises à disposition, sur chaque individu, ouvrent des perspectives aussi intéressantes que risquées.

Ce paradigme, expliqué en détail par François Ewald dans son rapport pour l'institut Montparnasse⁸², permet de passer d'une moyenne de masse à une individualisation de la prévision.

Pour revenir à un des fondamentaux de l'assurance, qui consiste à protéger les personnes et les biens, l'arrivé du Big Data va renforcer les aspects préventifs de l'assurance vis-à-vis de ses clients.

La dimension sociale de ces aspects n'aura échappé à personne.

Dans le modèle économique de l'assurance, une des principales composantes reste la prévention, afin d'éviter la réalisation du risque. Le Big Data doit servir cette prévention, permettant de confirmer l'assureur dans son rôle d'assistant et renforcer la confiance des consommateurs, à l'égard de ce métier. L'Etat, qui règlemente et demande de plus en plus à l'assurance de le suppléer pourrait y voir une opportunité de soutien dans différents domaines.

Si le Big Data évolue vers une dimension sociale et s'affiche comme une véritable assistance, et non pas comme une machine intrusive permettant de faire des bénéfices à tout prix, alors le « Data Driven » sera réellement bénéfique pour tous. L'échange de données personnelles servira à l'obtention d'un service personnalisé, en toute confiance. Avoir le choix et les recommandations adaptés, en fonction de sa personnalité et de ses comportements, va bouleverser la manière d'assurer un risque.

_

⁸² Assurance, prévention, prédiction ... Dans l'univers du Big Data

Analyse et présentation des données et de leur cycle de vie pour des prises de décisions pertinentes

En effet, cette nouvelle répartition des forces entre pouvoir et savoir, telle que décrite par François Ewald, va transformer l'assurance. Ce nouvel équilibre, provoqué par la révolution digitale, implique, de la part des assureurs, d'accentuer les aspects personnalisations des offres, afin d'être en phase avec les attendus des clients.

Cela ne va pas à l'encontre de la mutualisation, mais implique de la repenser. Le développement des réseaux sociaux et la création de communautés nouvelles pourraient être la réponse à une nouvelle forme de mutualisation.

La dimension personnelle de ces données implique, aussi, des responsabilités importantes vis-à-vis des assurés.

Les aspects sécurité sont aujourd'hui très peu étudiés, car il n'existe pas de véritable preuve d'attaque d'objets connectés. Toutefois, cela reste envisageable et il faudra s'en préoccuper, rapidement. Les conséquences d'une voiture autonome, piratée et devenue de ce fait incontrôlable, seraient dramatiques. L'assureur de la voiture doit aussi se préoccuper des nouveaux risques qui peuvent naître des objets connectés. Les flux de données, la manière dont ils circulent entre les objets connectés et celui qui les utilise, devront être totalement sécurisés pour protéger l'assuré et lui garantir des services de qualité.

L'enjeu principal, pour les assureurs, est de trouver un équilibre entre mutualisation et personnalisation des offres proposées, mais aussi entre progrès et vie privée.

Arriver à convaincre, en éludant les aspects intrusifs, tout en maintenant des services de qualité basés sur les fondamentaux assurantiels, sera l'objectif à atteindre pour l'assurance de demain avec le Big Data.

SOURCES

INTERVIEWS

10 Septembre 2013 : Etienne Arbogast

Responsable Etudes Actuarielles - MAAF

15 Octobre 2013 : Olivier Tapin

Responsable Domaine Etudes Statistiques - MAAF

04 Novembre 2013 : Marielle Vo Van Liger

Directrice du Département Relation Client Marketing Stratégique - MAAF

07 Novembre 2013 : Charles Dumartinet

Responsable de Pôle - Pôle Risques Majeurs - Covéa

02 Décembre 2013 : Thierry Beillon

Responsable Organisation Méthodes - AIS Covéa

17 Décembre 2013 : Rolf Neuerburg

Responsable du Département Business Intelligence - Unité Opérations - Direction de Solutions IARD - ALLIANZ

16 Janvier 2014 : Xavier Ducurtil

Directeur de Projets de Transformation Stratégique - Covéa

17 Janvier 2014 : Olivier Desbiey

Pôle Innovation - CNIL

06 Février 2014 - Philippe Rondeau

Directeur de la Direction Technique Produit – MAAF

BIBLIOGRAPHIE

Talend

Les quatre piliers d'une solution de gestion des Big Data – White Paper - 2013

IBM

Harnessing the power of Big Data for insurance – White Paper – Mars 2013

ISACA

Big Data & Protection de la Vie Privée – White Paper – Août 2013

The World Economic Forum

Big Data, Big Impact: New Possibilities for International Development - 2012

Economist Intelligence Unit

Big Data Lessons From The Leaders - 2012

Le Comité stratégique de la « filière Assurance » au pôle FINANCE INNOVATION

Le Livre Blanc 2 de l'Innovation dans l'Assurance

Viktor Mayer-Schönberger and Kenneth Cukier

Big Data: A Revolution That Will Transform How We Live, Work, and Think - 2013

Christophe Brasseur

Enjeux et usages du Big Data – technologies, méthodes et mise en œuvre – 2013

Jean-René Lyon et Stéphane Arbus

Quel Modèle d'Entreprise pour l'Assurance de demain ? – 2011

McKinsey Global Institute

Big Data: The next frontier for innovation, competition, and productivity" - Mai 2011

SITES INTERNET

http://www.decideo.fr

Les ECHOS

 http://lecercle.lesechos.fr/entrepreneur/tendances-innovation/221166793/data-scientist-jobplus-sexy-21eme-siecle

La CNIL

www.cnil.fr

Stéphane Tufféry

- http://data.mining.free.fr/
- www.nist.gov
- http://www.autarcia.com/data-mining/
- www.emc.com
- http://fr.wikipedia.org/wiki/Big_data
- http://www.clientaucoeur.com/2013/05/23/data-mining-analyse-semantique-et-predictive-text-mining-et-big-data-ces-technologies-au-service-de-lintelligence-client/
- http://www.journaldunet.com/developpeur/outils/les-solutions-du-big-data/principe-defonctionnement-de-mapreduce.shtml
- http://datascience.bluestone.fr/blog/les-origines-les-grands-principes-du-big-data
- http://www.pegasusdata.com/
- http://www.etalab.gouv.fr/article-data-visualisation-a-l-e-g8-forum-internet-dans-l-economie-mondiale-74865969.html
- http://lentreprise.lexpress.fr/formation/quand-je-serai-grand-je-serai-data-scientist
- http://thirdwaveberlin.com/2011/12/quantified-self-data-privacy

- http://www.internetactu.net/2013/12/05/en-quoi-les-big-data-sont-elles-personnelles/
- http://www.fredcavazza.net/2011/06/15/le-quantified-self-au-service-de-la-productiviteindividuelle-et-collective/
- http://www.lexpress.fr/
- http://www.gartner.com/technology/home.jsp

SALONS ET CONFERENCES

• Le LAB: 19 Février 2013

Conférences Big Data

EURAPCO : 29 Mai 2013

Intervention sur le Big Data du Boston Consulting Group

Salon « transform 2013 » :11 Juin 2013

TIBCO Software

SCOR: 26 Mars 2013

Conférences Big Data

Chapuis Halder & Compagnie : 30 janvier 2014

Conférence «Le digital dans l'assurance »

RAPPORTS ET THESES

SOLUCOM - Jean-Nicolas Biarrotte-Sorin et Gilles Baillou

Gouvernance des données : comment mettre durablement les données sous contrôle?

CIGREF

Entreprise et Culture numérique

IBM

The 5 in 5 – Innovations that will change our lives in the next five years - http://www.ibm.com/smarterplanet/us/en/ibm_predictions_for_future/ideas/

Direction générale de la Compétitivité, de l'Industrie et des Services (DGCIS) du Ministère du Redressement productif (ministre Arnaud Montebourg), appuyé par l'expertise internationale du cabinet McKinsey. Ce chantier a été mené en lien avec les pôles de compétitivité et les comités stratégiques de filières au sein desquels les chefs d'entreprises, les partenaires sociaux, les administrations concernées et les fédérations professionnelles sont représentés et prennent une part active. – 20 Novembre 2013

La Nouvelle France Industrielle

SOGETI: Jaap Bloem, Menno Van Doorn – Sander Duivestein – Thomas Van Manen –
 Erik Van Ommeren - Sandeep Sachdeva- 2013

VINT Research report – No more secrets with Big Data Analytics

François Ewald – Rapport pour l'institut Montparnasse : Octobre 2012

Assurance, prévention, prédiction ... dans l'univers du Big Data

Guillaume Billy MBA CNAM - ENASS: Thèse soutenue en 2012

Les communautés virtuelles peuvent-elles réinventer l'assurance ?

- Revue Risques n°95 : Novembre 2013
- Cahier Innovation & Prospective n°1 (CNIL) Vie Privée à l'Horizon 2020

Thèse soutenue en 2014

pour l'obtention du **MBA majeure assurance** du Cnam Enass

Président du Jury : Alexis Collomb

Sous la direction de Jean-François Andrjanczyk

Une école est un lieu de production et de diffusion de connaissances.

L'Ecole nationale d'assurances s'organise pour répondre le mieux possible à cette mission en direction de ses élèves d'abord, mais aussi de la profession de l'assurance et de ses partenaires :

- les « dialogues de l'Enass » éclairent l'actualité par le débat avec une personnalité remarquable ;
- « les travaux de l'Enass » sont destinés à faire bénéficier la profession des travaux menés au sein de l'Enass par ses professeurs et ses élèves, à tous les niveaux, dans la mesure où les jurys qui les ont évalués ont noté leur qualité et leur originalité. Ces travaux vous seront adressés par Internet, certains d'entre eux pouvant faire l'objet d'un tirage sur papier ou même, être édités.

Nous souhaitons que toutes ces initiatives vous soient profitables.

François Ewald

Président du Conseil scientifique et pédagogique de l'Université de l'Assurance