

Octobre  
2013

# LES DOSSIERS TECHNIQUES D'INFORMATION

optimind winter

## BIG DATA

- 1/ Introduction
- 2/ Que représente le Big Data?
- 4/ Les aspects clés du Big Data
- 7/ État des lieux de l'industrie de l'Assurance sur le Big Data
- 9/ Le Big Data et l'Assurance
- 13/ Quels chantiers, quelles perspectives demain pour les assureurs et le Big Data?
- 16/ Conclusion

## Opportunités de demain pour les assureurs

À l'heure où certains s'interrogent encore pour savoir si le Big Data est un bon concept, s'il peut être utile pour notre industrie et nos métiers, s'il peut améliorer nos modèles, nous avons décidé de traiter ce sujet au sein d'un *Dossier Technique* complet de manière objective tout en souhaitant mettre en avant les opportunités et les risques sous-jacents. Aujourd'hui, le Big Data est une réalité et non un concept, la vraie question est de savoir comment tirer parti des avantages qu'il procure tout en proposant une démarche raisonnable et progressive pour l'intégrer dans nos processus.

Nous devons admettre que les données volumineuses et non structurées envahissent les réseaux et les systèmes de stockage en tout genre. La première partie de notre *Dossier Technique* cherchera à définir et à illustrer quelques aspects du Big Data.

La responsabilité des assureurs en général et des actuaires en particulier est importante en la matière: nous devons être garants des données que nous produisons et de leur utilisation.

Nous vous proposons ce *Dossier Technique* comme une approche visant à ouvrir une réflexion sur une révolution annoncée, c'est aujourd'hui notre conviction.

Bonne lecture !

Christophe Eberlé, président

Dossier réalisé par Dan Chelly, directeur métier Risk Management, Éric Gaubert, directeur du développement, Bertrand Pitavy, directeur métier Projets & Maîtrise d'Ouvrage, Gildas Robert, directeur métier en Actuariat conseil, Manuel Audrezet, consultant, Frédéric Genet, manager et Vincent Meister, actuaire manager.

# Que représente le Big Data ?

## QUELQUES CHIFFRES ET DÉFINITIONS

En apparence, la définition du Big Data semble évidente et explicite : « grosses données ». En réalité, le concept est bien plus large. Les données de masse existent depuis maintenant déjà bien longtemps et, mises à part les problématiques de stockage, leur traitement tout comme leur utilisation ne posent pas de soucis majeurs.

Une première approche désormais classique est la définition que l'on trouve sur Wikipédia : « Le Big Data [...], désigne des ensembles de données qui deviennent tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l'information. »

Cette définition n'est cependant pas assez explicite selon nous. En effet, le caractère volumineux est une propriété remarquable du Big Data mais pas la seule. Sa deuxième propriété fondamentale est l'absence de structuration des informations sous-jacentes aux données obtenues et stockées. Cette hétérogénéité revêt un caractère novateur au concept de Big Data. Les sources multiples d'origine des données sont une troisième caractéristique du Big Data selon nous. La multiplicité des sources d'origine, particulièrement liées aux nouvelles technologies, ainsi que leur volume très significatif les rendent précieuses. Au-delà des sources classiques de bases de données historiquement constituées, nous avons désormais les sources numériques issues de l'internet et des réseaux sociaux, les sources reliées aux nouveaux et nombreux capteurs qui envahissent notre quotidien, qu'ils soient biométriques ou électromécaniques.

Aujourd'hui, le volume des données stockées est évidemment en pleine expansion et nous vous proposons ci-après quelques éléments de références. Selon une étude IDC, les données numériques créées dans le monde seraient passées de 1,2 zettaoctet par an en 2010 à 1,8 zettaoctet en 2011, puis 2,8 zettaoctets en 2012 et s'élèveront à 40 zettaoctets en 2020. À titre d'exemple, Twitter génère à l'heure actuelle 7 téraoctets de données chaque jour et Facebook 10 téraoctets.

Ce sont pourtant les installations scientifiques qui produisent le plus de données. De nombreux projets

sont ainsi en cours comme par exemple le radiotélescope *Square Kilometre Array* qui produira 50 téraoctets de données analysées par jour, à un rythme de 7 000 téraoctets de données brutes par seconde.

## Quelques ordres de grandeur des volumes de données

Unité	Nombre d'octets	Équivalence
1 yottaoctet (Yo)	$10^{24}$	
1 zettaoctet (Zo)	$10^{21}$	Toutes les données créées en 2012 (2,8 Zo)
1 exaoctet (Eo)	$10^{18}$	Toutes les informations générées jusqu'en 2003 (5 Eo)
1 pétaoctet (Po)	$10^{15}$	Bases de données de Google, Facebook, etc.
1 téraoctet (To)	$10^{12}$	6 millions de livres
1 gigaoctet (Go)	$10^9$	1 film de deux heures
1 mégaoctet (Mo)	$10^6$	1 livre de poche (1 Mo), 1 morceau de musique (5 Mo)
1 kilooctet (Ko)	$10^3$	

## À QUOI CELA SERT-IL ?

Le Big Data sert déjà beaucoup certains acteurs économiques : les professionnels du marketing utilisent ces données en *Datamining* pour des études comportementales. Plus récemment, en politique, les stratégies américains de l'équipe de Barack Obama ont largement utilisé le Big Data pour cibler leurs efforts de campagne en fonction des nombreuses données à leur disposition croisant ainsi des données géographiques avec des données comportementales liées aux opinions politiques.

C'est certainement dans le domaine de la science que le Big Data et les super calculateurs numériques permettent des avancées remarquables. Les séquenceurs d'ADN ont progressé d'un facteur de 10 000 en quelques années : d'environ 10 ans pour décoder l'ADN humain, la durée est aujourd'hui passée à moins d'une semaine. On peut en tirer une leçon simple : le Big Data implique des calculateurs puissants pour être correctement exploité, l'un ne va pas sans l'autre.

“ 90 % des données disponibles aujourd'hui ont été créées au cours de ces deux dernières années. ”

## Des 3V au 2x3V

La littérature évoque généralement les «3V» pour désigner les données de type Big Data, caractérisées par :

- ▶ les Volumes conséquents, de l'ordre de quelques dizaines de Po ( $10^{15}$ ) pour les bases de données de Google ou Amazon ;
- ▶ la Variété significative : fichiers, images, vidéos ou encore des blogs et réseaux sociaux ;
- ▶ la Vitesse : nécessité d'exploiter ces données le plus rapidement possible, voire en temps réel, dans un monde numérique qui exige d'être plus rapide que ses concurrents.

Nous proposons d'ajouter à cette définition quelques V importants pour notre industrie :

- ▶ la Volonté des entreprises de demain d'utiliser le Big Data ;
- ▶ la nécessaire Visualisation des sources existantes ;
- ▶ la Valorisation des données disponibles.

De cette façon, nous définissons le concept nouveau des 2x3V qui autorise, comme souligné dans notre précédent *Dossier Technique* sur les données, une exploitation la plus aboutie possible de cet actif immatériel chargé d'histoire.

### ET DANS LE DOMAINE DE L'ASSURANCE ?

Les domaines d'application futurs du Big Data vont être évidemment très larges et concerneront l'ensemble des secteurs économiques. Notre *Dossier Technique* s'intéresse très logiquement aux opportunités pour le secteur de l'assurance, certainement une

“

Une opportunité pour le secteur de l'assurance, une des industries les plus concernées...”

”

des industries les plus concernées par les perspectives du Big Data. En effet, l'assurance est basée sur des éléments immatériels et sur un cycle de production inversé. Ces deux caractéristiques impliquent une aisance naturelle à utiliser des données numériques ainsi qu'un fort besoin d'éléments prédictifs associés aux sciences probabilistes.

La question aujourd'hui est de savoir comment notre industrie doit se positionner pour bénéficier demain des meilleurs données et outils que nous apportent les nouvelles technologies. L'ambition de la réflexion que nous souhaitons ouvrir avec ce *Dossier Technique* est de permettre au lecteur, décideur ou expert, de se positionner afin de considérer que le Big Data :

- ▶ n'est pas à ce jour une opportunité pour son périmètre d'exercice ;
- ▶ est une opportunité dans une configuration dite « passive » : utiliser progressivement les données disponibles pour des traitements novateurs en termes de gestion des risques ;
- ▶ est une opportunité dans une configuration dite « active » : cette dernière posture vise à faire évoluer ses produits, ses garanties et/ou ses méthodes de gestion des risques pour envisager une meilleure performance et un meilleur service pour ses assurés. L'attitude dite « active » implique également des réflexions sur la mise en place de moyens de collectes comme des capteurs virtuels ou physiques ou encore de partenariats avec des producteurs de données volumineuses, voire une exploitation plus systématique des *open data*<sup>(1)</sup>.

## BFM TV

Dans la séquence d'apparition des phénomènes de société, le bouche à oreille, puis aujourd'hui les réseaux sociaux, sont en général la première occurrence d'une opinion. Fort de ce constat, BFM a construit deux indicateurs pour mesurer l'état d'esprit des Français :

- ▶ le premier relatif à l'état d'esprit des chefs d'entreprise, avec une recherche de mots clés sur les réseaux sociaux comme par exemple rentabilité ou production ;
- ▶ le deuxième relatif à l'état d'esprit des ménages sur la base de mots clés liés à une thématique opposée comme chômage ou logement.

Chacun des indicateurs est le solde de la somme des termes optimistes et pessimistes.

En mai 2013, BFM a pu ainsi annoncer, 15 jours avant tout le monde, la reprise de confiance des ménages et des chefs d'entreprises qui a ensuite été confirmée par des études INSEE.

« Nous avons là un bel exemple d'exploitation de données non structurées, de l'utilisation d'une taxonomie pour classer et trier les termes analysés et surtout une analyse de signaux faibles » explique Emmanuel Lechypre, Éditorialiste et Directeur de l'observatoire économique chez BFM Business.

(1) Voir le site [data.gouv.fr](http://data.gouv.fr).

# Les aspects clés du Big Data

## COMMENT TRAITER LE BIG DATA ?

Le monde de l'assurance est familier des **grilles de calcul** pour modéliser un portefeuille, lancer des milliers de scénarios économiques en mode stochastique pour calculer un SCR, *Capital de Solvabilité Requis*.

Le Big Data est l'avènement des **grilles de stockage**. Les données exploitées sont dispersées au sein d'internet sur une multitude de serveurs, avec un certain niveau de redondance de l'information. À l'échelle mondiale, cela réalise une immense grille de stockage qu'il s'agit d'exploiter en temps réel, avec les problèmes de latence<sup>(1)</sup> des données que l'on imagine aisément.

Une évolution technologique majeure concerne la mise au point de systèmes de fichiers distribués de type NoSQL, pour *Not Only SQL*. Ces patrons d'architecture de données, *frameworks*, répondent aux besoins de performance qui impliquent des traitements distribués massivement parallèles. Les plus connus sont notamment ceux mis au point et/ou utilisés par Google depuis plusieurs années tels que HDFS, *Hadoop Distributed File System*, et GFS, *Google File System*.

Par opposition aux SGBD-R, Système de Gestion de Bases de Données Relationnelles, qui stockent

“

Les données exploitées constituent une immense grille de stockage qu'il s'agit d'exploiter en temps réel.

”

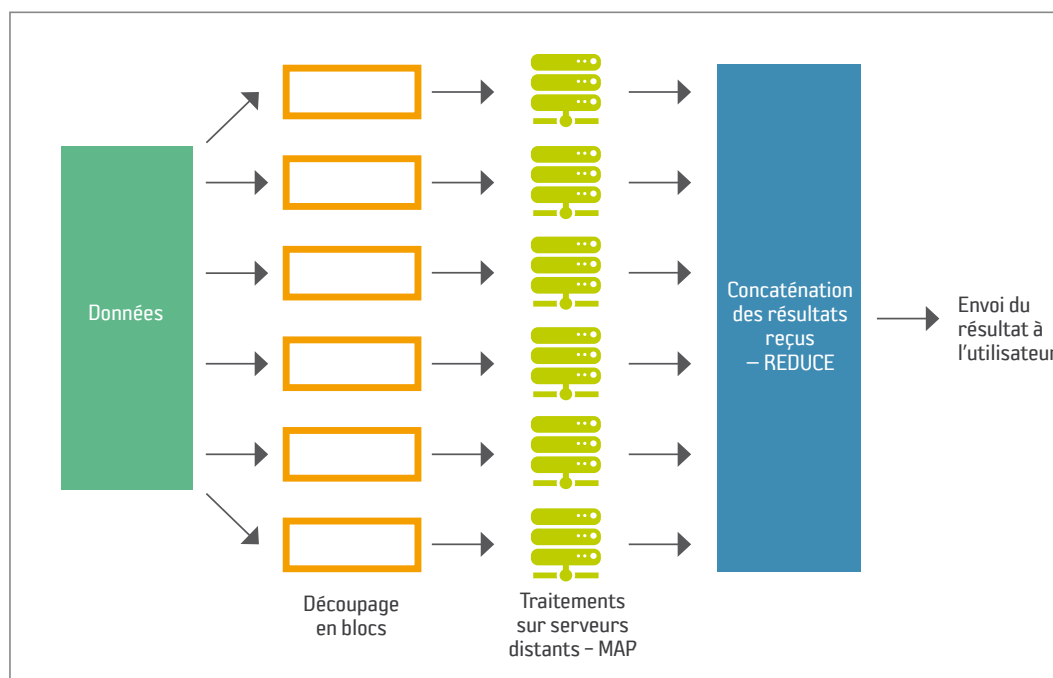
l'information dans des tables indexées, dimensionnées et hiérarchisées, un système de fichiers de type NoSQL peut se représenter par une base de données construite autour d'une unique table qui contient des milliards d'enregistrements de longueur variable et non structurés.

Pour exploiter ces bases, des algorithmes de type MAP REDUCE en quatre étapes sont utilisés, cf. *schéma ci-dessous* :

La table à exploiter est découpée en blocs de mêmes tailles et des serveurs vont chacun appliquer le traitement demandé au bloc à traiter, puis renvoyer leur résultat selon la technique suivante :

1. **Découpage en blocs**
2. **Traitement sur serveurs distants – MAP**
3. **Concaténation des résultats reçus – REDUCE**
4. **Envoi du résultat à l'utilisateur**

## Algorithmes de type MAP REDUCE



Source : Optimind Winter

(1) Temps nécessaire à un paquet de données pour passer de la source à la destination à travers un réseau (définition de Wikipédia).



## QUELLES TECHNOLOGIES POUR LE BIG DATA ?

L'évolution des technologies alliée à la multiplication des serveurs distants permet de gérer des volumes de données conséquents dans des délais particulièrement réduits.

Un premier levier concerne le **matériel** utilisé avec la multiplication des cœurs sur un seul microprocesseur, communément appelé puce. Pour rappel, un cœur contient tous les éléments nécessaires à l'exécution de programmes : unité de calcul, registres, mémoire cache de premier niveau, dit L1, pour bénéficier d'une faible latence.

À titre d'exemple, un seul processeur OCTEON® III MIPS64 contient jusqu'à 48 cœurs. Un cache mémoire, partagé entre les cœurs, permet des échanges des données sans passer par la RAM. Enfin, ces processeurs utilisent un jeu d'instructions RISC<sup>(1)</sup> et sont gravés en 3D et en technologie 28 nm<sup>(2)</sup> pour améliorer davantage encore les performances en réduisant la longueur des connexions entre tous les composants du microprocesseur. Le projet « Angstrom », mené au MIT, vise à concevoir des microprocesseurs à 1 000 cœurs.

Les cœurs sont capables de travailler sur des systèmes d'exploitation différents – Unix –, voire directement en binaire, « langage machine », pour aller encore plus vite. Enfin, quand la charge diminue, on sait diminuer aussi la fréquence d'horloge du microprocesseur, passant par exemple de 1,5 GHz à 750 MHz, pour réduire la consommation électrique. C'est un challenge : l'informatique et les serveurs Web représentent 10 % de la consommation électrique aux États-Unis. Le « *green IT* » devient indispensable.

Le **code informatique** est un deuxième levier pour réduire efficacement les temps de traitements. La programmation récursive permet d'atteindre cet objectif

mais uniquement dans les cas où les calculs peuvent être réalisés en parallèle. Concrètement, cela consiste à développer une procédure informatique faisant appel à elle-même sur un sous-ensemble de données. Ce type d'algorithme permet par exemple de paralléliser les calculs de tri sur plusieurs nœuds distincts.

Cependant, la complexité des algorithmes utilisés et la difficulté à trouver des programmeurs maîtrisant la programmation massivement parallèle rendent ces optimisations délicates à réaliser. On peut aisément anticiper une évolution à court terme d'un certain nombre de ces techniques de programmation, comme on l'a déjà observé lorsque les langages orientés objet sont arrivés à maturité.

Enfin, la **visualisation** est le troisième levier. Connue sous le nom de *Data Visualization* ou *Data Viz*, elle a pour objectifs de **comprendre** et de **faire comprendre**. De multiples techniques existent et les plus performantes sortent des schémas usuels de représentation de l'information.

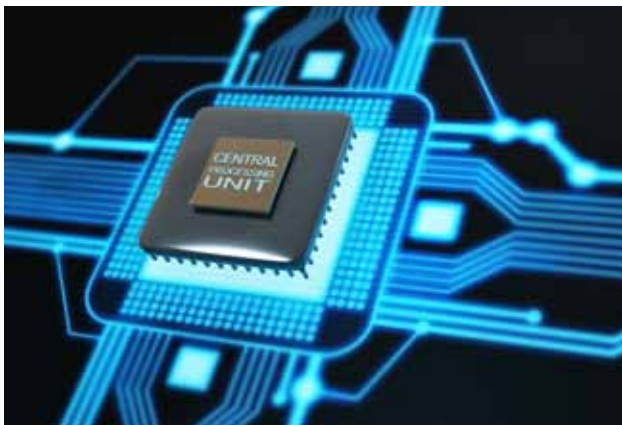
## QUELQUES ENJEUX STATISTIQUES ET MATHÉMATIQUES DU BIG DATA

L'appropriation et l'exploitation des données issues du Big Data représentent un défi nouveau et considérable pour les actuaires, analystes et statisticiens dont le métier est de « faire parler » les données dont ils disposent. L'univers classique actuel, constitué d'éléments normés, explicites et aux formats connus – numériques ou alphanumériques principalement –, est exploré au travers de statistiques et modèles qui s'appuient directement sur ces données. Ces habitudes devront être contournées, voire partiellement oubliées, pour adopter de nouvelles pratiques, en phase avec le fonctionnement intrinsèque des outils de traitement du Big Data.

(1) RISC est l'acronyme de *Reduced Instruction Set Computer*, soit « ordinateur à jeu d'instructions réduit ». Cette technique permet d'exécuter un plus grand nombre d'instructions par unité de temps (cadencement par l'horloge du microprocesseur).

(2) 28 nanomètres soit  $28 \times 10^{-9}$  m. C'est la taille du plus petit élément pouvant être gravé dans le microprocesseur (par photogravure). Cette taille diminue régulièrement grâce à l'amélioration constante des robots de gravure des microprocesseurs.





En fin de processus de l'analyse du Big Data – cf. *algorithme de Map Reduce exposé plus haut* – l'analyse des résultats prend la forme de tableaux de bord, de *reportings* ou encore de graphiques, qui reflètent les interactions ou les corrélations entre les données. L'interprétation de ces sorties passe alors par l'adoption d'un raisonnement non plus cartésien, mais basé sur des réflexions prédictives, voire intuitives.

Il s'agit ici du changement majeur dans l'utilisation d'un environnement Big Data : l'utilisateur devra s'attacher à analyser ces résultats singuliers, les affiner par la création de requêtes complémentaires pour ensuite concaténer cette vision issue des données Big Data avec les données de son portefeuille client. Cette combinaison lui permettra d'explorer de nouveaux indicateurs dont la fiabilité dépendra du degré de confiance accordé aux données issues du Big Data. Les critères classiques de qualité des données n'étant pas applicables, il s'agira surtout pour l'utilisateur de s'approprier progressivement les restitutions issues de ce nouvel univers, en vérifiant notamment la robustesse des résultats dans le temps. La vérification de l'absence de biais lié au processus de collecte des données sera notamment une composante essentielle.

Le déploiement d'un tel entrepôt de données s'accompagne en conséquence d'une augmentation non négligeable des temps d'exploitation et d'analyse. L'expert métier reste néanmoins au cœur du dispositif, en s'appuyant sur la puissance des nouveaux outils à sa disposition qui, par leur intelligence et les bases de données massives disponibles, alimentent et développent ses idées et ses réflexions.

### LES NÉCESSAIRES QUESTIONS LIÉES À LA VIE PRIVÉE

L'approche Big Data amène à utiliser et à croiser différentes natures de données disponibles. Ces données peuvent être publiques, par exemple fournies par des administrations en *open data*<sup>(1)</sup> volontairement disponibles par contrat mais d'accès restreint, comme les

boîtes noires dans les voitures soumises au *Pay as You Drive*, ou encore récupérées sur internet dans un mode, en théorie, plutôt ouvert comme la publication d'un commentaire sur un site.

Toutefois, ces données peuvent faire l'objet de contraintes d'utilisation : contraintes de droit commun sur l'utilisation des données à caractère personnel, CNIL, mais aussi contraintes plus spécifiques liées au secret médical, au secret bancaire ou encore à la propriété intellectuelle. La réglementation française actuelle reconnaît à toute personne des droits sur ses données et définit des règles sur les interconnexions et traitements de ces données. Elle interdit également, sauf exception, la collecte et le traitement des données sensibles, notamment origines ethniques, opinions politiques, philosophiques ou religieuses ou relatives à la santé. La CNIL, autorité administrative indépendante, dispose en parallèle d'un régime de sanctions si ces règles ne sont pas respectées.

L'utilisation des traitements en Big Data en matière d'éthique et de vie privée pose de nombreuses questions alors que l'on parle de plus en plus de patrimoine numérique personnel :

- ▶ la difficulté de qualifier la donnée traitée en temps réel pour savoir si elle est sensible ou soumise à des règles particulières ;
- ▶ la perte potentielle d'anonymat des données dès lors qu'elles sont croisées entre elles ;
- ▶ les difficultés à appliquer le droit à l'information des intéressés, d'obtenir leur consentement, d'exercer leur droit à s'opposer comme leur droit d'accès et de rectification ou de s'opposer à la logique d'une décision liée à un traitement de données ;
- ▶ la marchandisation des données personnelles.

En terme d'éthique et de perception de l'acceptable, il existe une question de génération ou de culture numérique. En effet, il est clair que les nouvelles générations sont bien plus ouvertes à fournir sur internet leurs données afin d'obtenir des offres personnalisées et géolocalisées.

Une future réglementation européenne se construit actuellement progressivement. Elle conserve les principes clés de la réglementation française mais des débats importants subsistent sur le consentement explicite de l'utilisateur – *opt-in* – et sur une forme de droit à l'oubli qui pourraient tout deux contraindre bien plus fortement l'utilisation actuelle des données. Face à ces enjeux, les acteurs de l'économie numérique, qui y voient un gisement de croissance, s'opposent aux défenseurs des droits individuels. Or, le non-respect des règles européennes en la matière est passible d'amendes pouvant atteindre un million d'euros ou 2 % du chiffre d'affaires annuel global d'une entreprise. De quoi être vigilant sur l'utilisation de ces données en assurance...

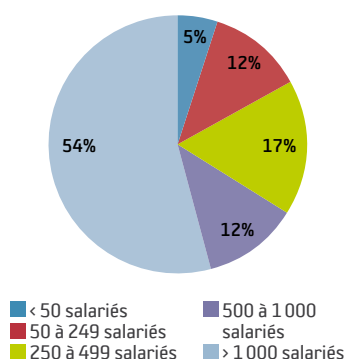
(1) Voir le site [data.gouv.fr](http://data.gouv.fr).

# État des lieux de l'industrie de l'Assurance sur le Big Data

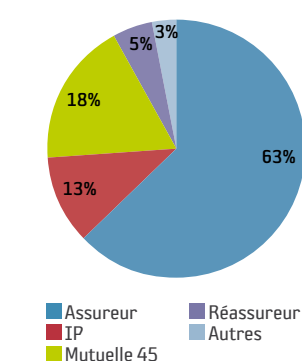
Nous avons souhaité faire un état des lieux de la connaissance et de l'anticipation des assureurs sur ce sujet et avons commandé un sondage à OpinionWay. Volontairement restreint, il porte sur les points cruciaux pour obtenir une vision à la fois globale et instantanée.

L'enquête<sup>(1)</sup> a été réalisée auprès d'un échantillon de 103 répondants qui représentent 48 acteurs de l'assurance. Les principales directions interrogées sont celles qui utilisent les données: directions techniques et actuariat (tarification, provisions), marketing (produits, client), risques (fraude, pertes) et enfin la direction informatique (stockage, mise à disposition).

## Taille entreprise



## Type entreprise



## POSITIONNEMENT DU SUJET : RÉFLEXION STRATÉGIQUE ?

Nous l'indiquions dans l'édition, le thème du Big Data n'est plus un sujet à venir. Les personnes interrogées sont près de 60% à répondre que ce sujet est bien identifié comme étant stratégique. Pour seulement 23 %, il ne l'est pas encore. Enfin, 18 % n'ont pas d'information sur le sujet, sans doute dû à un manque de communication en interne ou bien en raison de son caractère novateur pour certaines personnes.

## DIRECTIONS MÉTIERS IMPLIQUÉES DANS LE PROJET BIG DATA

Pour trois quarts des répondants, les deux directions principalement impliquées par ce projet sont les directions informatiques (75 %) et les directions techniques (73 %). On retrouve bien les contraintes techniques liées à la mise à disposition et au stockage des données. Ces thèmes sont souvent gérés par les directions informatiques. Les directions techniques sont historiquement habituées à utiliser les données pour des besoins de tarification, provisionnement et simulations.

Dans le futur, ils sont 55 % à penser que la direction des risques va être impliquée notamment sur les sujets réglementaires dans le cadre de Solvabilité II mais également pour lutter efficacement contre la fraude et le blanchiment.

En troisième position, on retrouve les directions historiquement « consommatrices » des analyses de données : à 54 %, les directions marketing et à 37 %, les directions commerciales. Nous sommes là en phase avec les besoins liés aux attentes des clients en termes de complétude de la gamme de produits et d'une parfaite connaissance de son portefeuille clients sur les critères discriminants permettant de les fidéliser.

## ÉTAT D'AVANCEMENT PAR RAPPORT AU MARCHÉ, AVANTAGE CONCURRENTIEL ?

Ce sujet intéresse le marché puisqu'ils sont près d'un tiers, soit 32 %, à considérer qu'ils sont en phase avec les évolutions du marché et un autre tiers 35 %, *a contrario* qu'ils sont en retard. Ils sont seulement 2 % à considérer être en avance. Preuve en est que les acteurs du marché surveillent l'état d'avancement de leurs concurrents.

Enfin, 31 % ne connaissent pas leur positionnement, ce qui semble signifier, soit qu'ils n'ont pas démarré, soit qu'ils ne font pas de veille concurrentielle, soit encore que l'absence de mesure claire sur l'adoption du Big Data ne leur permet pas de se positionner aisément.

## UTILISATIONS FUTURES DU BIG DATA ?

Dans un contexte de crise, d'amélioration des marges et de concurrence accrue, tout naturellement, ils sont 45 % à considérer que la première

(1) Enquête réalisée entre le 27 juin et le 16 juillet 2013 auprès d'un panel représentatif du marché français. Utilisant la méthodologie très normée d'un mode d'interrogation en ligne CAWI, *Computer Assisted for Web Interview*, l'étude a été réalisée dans le respect des procédures et règles de la norme ISO 20252.

utilisation sera l'amélioration de l'offre et des produits puis 34 % pour construire un avantage concurrentiel. On retrouve bien là les besoins des directions commerciales/marketing et celles des directions techniques recevant les décisions des directions générales. Information intéressante, ils sont respectivement 38 % et 37 % à penser que le Big Data sera utilisé pour améliorer les processus de l'entreprise et optimiser l'utilisation des ressources. On peut imaginer que les nombreuses données seront analysées pour améliorer la performance de l'entreprise, l'adéquation entre les processus internes et l'allocation des ressources. Ce qui traduit un fort niveau de maturité des réflexions dans les entreprises.

Environ un tiers, respectivement 34 % et 32 %, estime que les données serviront pour identifier les usages ou des comportements anormaux (principalement fraude) et anticiper les demandes de demain. Résolument tournées vers une meilleure connaissance client à travers la systématisation d'études approfondies des données de sinistres et de prestations, il devient important pour les entreprises de mettre tout en œuvre pour réduire les pertes

liées à la fraude. La dernière étude publiée en 2011 par l'ALFA, Agence pour la lutte contre la fraude à l'assurance, indique que les pourcentages de fraude sont estimés à 4 à 8 % sur les montants des sinistres en IARD. Le sujet fraude devient une réelle priorité dans les directions des risques afin de répondre à des besoins réglementaires mais également internes en vue d'améliorer les marges en réduisant les pertes financières liées à la fraude, avec la diffusion de la culture risque dans les entités. En dernière priorité, l'utilisation des données va servir dans un cadre offensif afin de mieux connaître la concurrence (15 %) et d'identifier les signaux faibles sur le marché (12 %).

Enfin, environ 20 % n'ont pas connaissance de l'utilisation possible, traduisant sans doute que le chantier Big Data n'est pas encore lancé ou bien n'est pas encore très mûr concernant ses objectifs.

#### LES FREINS À L'UTILISATION DU BIG DATA

Les deux premières sources sont techniques et financières. Ils sont un tiers (32 %) à considérer que les freins à l'utilisation du Big Data

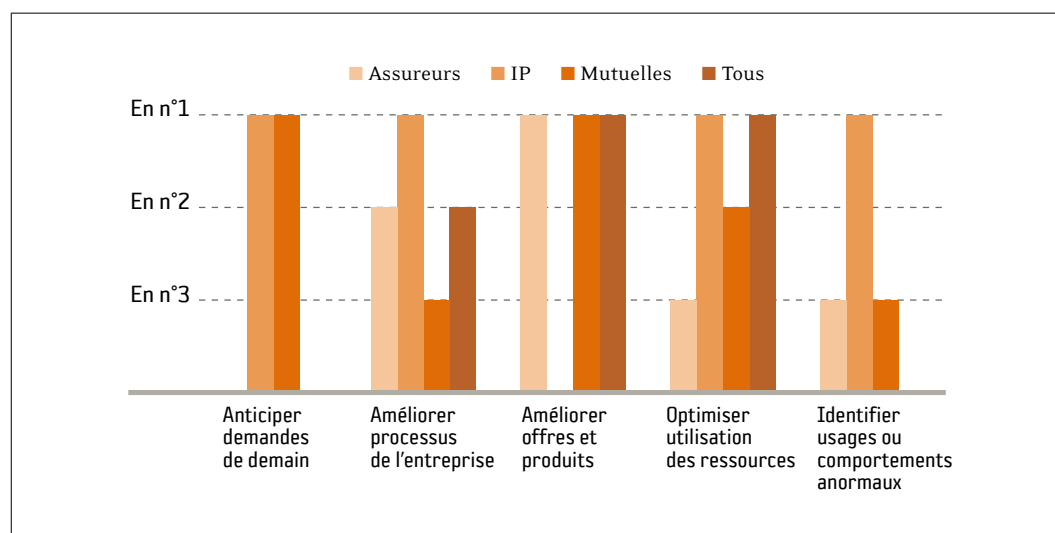
sont dus à un manque d'outils techniques pour gérer un tel volume de données et dans la même proportion, un manque de moyens financiers à investir.

Une limite liée aux ressources humaines apparaît ensuite en deuxième position puisque 24 % considèrent que les freins sont liés à un manque de compétences internes pour le traitement de ce type de données. On va clairement voir naître de nouvelles fonctions, voire métiers, liées aux Big Data qui nécessiteront des compétences transverses et métier à l'entreprise mais aussi très techniques pour appréhender ces nouvelles technologies qu'il faudra parfaitement maîtriser.

Par ordre décroissant, on retrouve ensuite des limites liées à un manque d'intérêt de la direction (22 %) qui ne doit pas considérer ce sujet comme stratégique, ne voyant pas bien comment cela pourrait l'aider au quotidien. D'un point de vue réglementaire, 14 % considèrent que les limitations d'utilisation, par exemple liées à la CNIL, sont trop fortes.

Enfin 29 % ne savent pas, traduisant leur manque d'implication dans le projet interne.

#### Le TOP 3 : dans l'avenir, quelles utilisations comptez-vous faire du « Big Data » ?



Source : Optimind Winter



# Le Big Data et l'Assurance

## POURQUOI LE BIG DATA EST UNE RÉVOLUTION POUR L'INDUSTRIE DE L'ASSURANCE ?

Au regard des différentes présentations du concept même de Big Data proposées dans la première partie de ce *Dossier Technique* et de nos convictions profondes sur l'avenir du secteur de l'assurance, nous sommes convaincus que les nouvelles données disponibles pour les industries du risque doivent leur permettre de mieux répondre aux attentes croissantes des assurés en matière de protection financière face aux risques auxquels ils sont exposés.

Comme dans toutes les révolutions, un certain nombre d'étapes sont à prévoir. Il existera des assureurs pionniers qui anticiperont tôt les échéances nécessaires à l'utilisation du Big Data, ils seront les premiers à essuyer des échecs mais aussi à goûter aux succès et ainsi à pérenniser leur avenir. Il y aura ensuite une étape de normalisation qui permettra à une majeure partie du marché de s'adapter et de suivre les meilleures pratiques en la matière. Dans ce dernier groupe, certains réussiront mieux que d'autres et en définitive, ceux-là, ainsi que les pionniers, se réserveront la possibilité de dominer les futurs marchés de l'assurance liés aux nouvelles technologies des données. Pour les autres, une lente mais certaine décroissance de leur part de marché est à prévoir, le temps qu'ils prennent conscience des efforts à produire pour se mettre à niveau : de ces derniers, certains réussiront peut-être. À l'inverse des précédentes révolutions industrielles de notre secteur, le Big Data que nous pouvons associer à la révolution numérique post-internet, de type 2.0 pour aller vite, est une (r)évolution qui se déploie sur de nouvelles bases temporelles. En effet, on constate une accélération importante du progrès et une nécessité de suivre le rythme.

En définitive, la question centrale est la suivante : pourquoi le Big Data va autant reconfigurer les marchés ? Pourquoi nous pensons que le Big Data est comparable à la révolution informatique des années 70 ?

La réponse tient en trois arguments clés qui consolident l'idée même que l'évolution du marché est structurellement irréversible et donc que c'est une révolution en soit :

- ▶ les sources disponibles pour tarifier et maîtriser les risques d'un produit et/ou d'une garantie deviennent multiples et leurs volumétries exponentielles : la matière première de l'assureur évolue en profondeur ;
- ▶ les risques à assurer peuvent être suivis pour partie en temps réel et ce quel que soit le sous-jacent porteur du risque – personne, automobile, entreprise, etc. : la matière assurable n'est plus la même que par le passé ;



- ▶ l'environnement temporel de l'assurance tend à devenir instantané : le principe du cycle de production inversé est conservé mais son amplitude se rétrécit drastiquement.

Ces trois constats obligent l'assureur à envisager *a minima* trois chantiers concrets dans un avenir proche pour maintenir sa compétitivité concernant :

- ▶ sa matière première : veiller à pouvoir obtenir les données nécessaires à la fabrication et à l'exploitation de ses produits ;
- ▶ sa matière assurable : veiller à construire des produits adaptés aux besoins du marché, c'est-à-dire intégrant les meilleures pratiques en termes de fabrication et d'utilisation des données ;
- ▶ sa réactivité : veiller à ce que son outil industriel, son système d'information, puisse récolter, analyser et utiliser les informations nécessaires pour piloter les produits en temps réel.

## QUELQUES EXEMPLES CONCRETS / LES DOMAINES D'APPLICATION

Les perspectives d'application dans le domaine de l'assurance sont riches et variées : les informations issues du Big Data viendront enrichir les processus et les modèles pour optimiser l'ensemble du cycle de production de l'assureur : de la tarification et la souscription du contrat, à la modélisation, au provisionnement et à la gestion des risques.

L'actuaire, en étant au cœur de ces différents processus, va être très fortement mis à contribution en vue de l'exploitation progressive des informations issues du Big Data. Cela laisse entrevoir de nouvelles perspectives pour cette profession, tout en soulevant de nombreuses interrogations sur les modèles, processus et outils de demain. Un des axes d'investigation sera notamment d'associer, au sein d'un modèle homogène, les données quantitatives utilisées jusqu'à présent avec celles du Big Data, qui se présenteront souvent sous la forme d'informations plus qualitatives.

Le pas à faire est important, mais les perspectives pour le secteur sont décisives. L'actuaire, acteur clé de la solvabilité des assureurs, apportera à ce mouvement sa technicité pour définir de nouvelles pratiques et des modèles toujours plus fiables.

### **Domaines d'application : gestion des risques / fraudes**

La lutte contre la fraude passe nécessairement par la détection de comportements anormaux que l'approche Big Data peut venir nourrir. L'ALFA, l'Agence pour la lutte contre la fraude à l'assurance, propose une définition de la fraude à l'assurance comme un « acte volontaire permettant de tirer un profit illégitime d'un contrat d'assurance ». Elle peut se caractériser lors de la souscription ou encore lors de la déclaration de sinistres. Elle représente pour les acteurs de l'assurance un enjeu estimé à environ 5 % des primes mais cette évaluation, par définition difficile à mesurer, varie fortement en Europe et pèse lourdement sur les marges et donc sur les tarifs proposés.

C'est pourquoi les assureurs prennent très au sérieux ce sujet et s'organisent entre eux pour l'échange d'informations et de bonnes pratiques, objet en France de l'ALFA. En parallèle, chaque acteur consolide son contrôle interne et forme son réseau afin d'assurer une vigilance accrue. La technologie est venue également renforcer le dispositif humain avec la détection systématisée de comportements anormaux.

En effet, en matière de fraude à l'assurance, il est essentiel de pouvoir capter les « signaux faibles » *via* une analyse comportementale efficiente. Or, les limites des approches actuelles sont liées aux périmètres des données sources exploitées, souvent celles fournies et falsifiées par le fraudeur, et aux alertes *a posteriori* laissant une longueur d'avance aux fraudeurs.

C'est là qu'apparaît tout l'intérêt de l'approche Big Data avec des outils ouverts sur les sources d'informations non structurées, notamment les réseaux sociaux et disposants de mesures systématiques des risques générant des alertes en temps réel sur des comportements « anormaux » ou « atypiques ».

Pour ce faire, le préalable est déjà d'identifier les « *red flags* » c'est-à-dire les scénarios qui traduisent une situation anormale et d'identifier les sources d'informations internes et externes, structurées ou non structurées ainsi que les passerelles entre elles.

Les révélations récentes du dispositif PRISM de la NSA, supposé se concentrer sur la lutte contre le terrorisme mais sans nul doute également sur l'intelligence économique, préfigure la puissance d'analyse, en utilisant l'ensemble des sources de données disponibles comme le texte, le son, la photo, la vidéo, la biométrie, ou tout autres capteurs divers, des outils Big Data sur la détection de comportements anormaux.

## **Interview d'Olivier Belma**

*Responsable Fraude & Revenu Assurance  
chez Bouygues Telecom*

### **► Quelles sont les problématiques de fraude rencontrées dans le monde des télécoms ?**

Un fraudeur dispose de deux grandes sources de revenu : le téléphone mobile et les télécommunications. Il doit ouvrir un compte chez l'opérateur ou utiliser le compte d'un vrai client. Les éléments d'identité, physique ou numérique, sont alors primordiaux. Ce sont des « actifs rares » que les fraudeurs exploitent dans tous les secteurs d'activité afin de rentabiliser leur investissement. Nous faisons donc face aux mêmes populations que les assureurs avec des problématiques similaires lors de la vente d'un produit.

### **► En quoi le Big Data vous sert pour votre dispositif de lutte anti-fraude ?**

Notre dispositif de lutte contre la fraude a évolué au fil du temps. Il est positionné à la fois au niveau de l'enrôlement du client et tout au long de sa vie. Concernant les appels, nous analysons les comportements des lignes en temps réel. Le Big Data nous permet aujourd'hui de compléter notre surveillance, de développer le suivi de notre performance avec des outils graphiques et d'être plus pertinents pour les fraudes réalisées sur le Web. Par exemple, pour les renouvellements de téléphone *via* l'espace client Web, nous sommes alertés quand des combinaisons d'actions typiques des fraudeurs se produisent sur un compte. Nous allons développer progressivement d'autres actions en temps réel sur toute la chaîne et aller plus loin dans la modélisation des comportements atypiques.

### **► Quelles ont été les questions rencontrées dans leur mise en place ? Quels sont les avantages indéniables que vous en tirez aujourd'hui ?**

La principale difficulté est de bien définir son besoin afin de disposer de données pertinentes. Le leurre avec le Big Data est de croire que d'avoir toutes les données permet de tout comprendre et tout faire en une fois. Ce n'est qu'un moyen à la disposition d'une stratégie globale de lutte contre la fraude. Quand les besoins sont bien ciblés, les principaux avantages sont la possibilité de mener des actions en quasi-temps réel avec des solutions plus agiles, de diminuer le délai de détection pour agir avant la fraude et d'améliorer la productivité des analystes fraude par un meilleur ciblage.

### Domaines d'application : tarification

Les méthodologies de tarification des produits d'assurance font généralement appel à une segmentation selon des catégories de risques – en fonction des branches d'assurance – telles que les caractéristiques de l'assuré ou celles du bien assuré. Des informations sur la sinistralité passée peuvent également être introduites.

Bien utilisée, l'abondance de données associée à un univers Big Data offre de réelles opportunités pour affiner leurs modèles de tarification et les processus de souscription.

En premier lieu, l'approche Big Data multiplie potentiellement le nombre de variables explicatives de la sinistralité et conduit à revoir en conséquence les opportunités de segmentation tarifaire. De nouvelles pistes seront ainsi explorées pour affiner l'analyse de la sinistralité future probable d'un assuré en fonction de son profil. L'approche Big Data peut par exemple s'avérer particulièrement intéressante depuis la mise en application de la *Gender Directive*, qui interdit depuis fin 2012 la prise en compte du sexe de l'assuré dans les processus de tarification. Des variables corrélées au sexe, non encore exploitées car absentes des dispositifs actuels, pourraient ainsi être mises à profit, comme par exemple la couleur du véhicule assuré. Ou au contraire, de nouvelles variables généralement en lien avec les avancées technologiques pourraient être étudiées pour s'affranchir du caractère prédictif du sexe.

Il est ainsi également parfaitement possible d'imaginer qu'à horizon de 5 ans, les assureurs auto combinent des données « internes », capteurs embarqués, à des données externes traitées en Big Data pour une tarification optimisée des produits en *Pay as You Drive*. Ces données externes pourraient provenir de drones qui analysent le comportement des automobilistes en fonction de la météo, du niveau d'encombrement des axes routiers, et ainsi enrichir les données comportementales dont ils disposent par ailleurs.

L'étude de l'impact météo sur la consommation médicale pour mieux appréhender les comportements de consommation médicale – médecine de ville, pharmacie, etc. – est un autre domaine d'étude. Une régionalisation de cette analyse est d'ailleurs avantageuse pour tarifier au plus près, ce que font déjà certains assureurs à partir du différentiel des coûts moyens.

D'autres applications peuvent être envisagées, par exemple en assurance collective. Partant du principe que le nombre d'arrêts de travail et les dépenses de santé des salariés d'une entreprise sont fortement corrélés à sa situation économique, des informations issues du Big Data sur sa renommée, son secteur et ses concurrents peuvent apporter des informations complémentaires pertinentes pour anticiper une déviation future de la sinistralité.



Les experts métiers devront toutefois veiller à ne pas surexploiter le Big Data par une utilisation abusive du principe de segmentation. En effet, le portefeuille global pourrait alors ne plus profiter de l'effet de mutualisation nécessaire à l'équilibre de l'activité. Dans un environnement où l'utilisation du Big Data serait généralisée, une concurrence accrue menée sur les « bons risques » pourrait se voir proposer des tarifs très bas, au risque de ne plus compenser les « mauvais risques ».

L'approche Big Data ouvre par ailleurs des perspectives plus larges, en permettant aux assureurs d'envisager de disposer d'informations pertinentes sur les enjeux majeurs propres à la souscription d'un risque : les tarifs pratiqués par la concurrence et les critères de choix de l'assuré, en particulier l'importance des aspects tarifaires. Ces informations, associées à une meilleure anticipation de l'espérance de sinistralité, pourraient permettre aux assureurs de considérablement améliorer la maîtrise de leur portefeuille d'assurance, avec la faculté de mieux choisir les risques souscrits.

### Domaines d'application : provisionnement / modélisation

De manière analogue, l'approche Big Data ouvre de réelles perspectives dans le domaine du provisionnement, de la modélisation et de la gestion des risques.

Certains risques, la mortalité par exemple, sont bien maîtrisés par les assureurs : à partir de données historiques, les actuaires sont en mesure de constituer des tables d'expérience, représentatives de la sinistralité passée. Si le caractère prédictif est plus problématique, notamment pour le provisionnement des rentes viagères, le secteur a progressivement appris à utiliser des tables prospectives, qui intègrent des informations externes sur l'évolution future de la mortalité.

D'autres risques en revanche posent plus de difficultés, soulevant de réelles incertitudes sur les évaluations des provisions et des fonds propres des assureurs,

## Interview 01 Business +

### Le big data a séduit Crédit Mutuel Arkéa

**01 Business**

Hier, vingt-quatre heures. Aujourd'hui, quinze minutes. C'est le temps que met désormais Suravenir pour simuler les sommes à provisionner sur trente ans pour ses quelques deux millions d'emprunteurs. Et ce, conformément aux exigences de la directive Solvency 2 garantissant la solvabilité des assureurs. La branche assurance prévoyance de Crédit Mutuel Arkéa exploite Hadoop, une technologie phare du Big Data. *"Nous voulions pouvoir réaliser plusieurs fois par jour des simulations en jouant sur des variables comme les mesures fiscales ou les lois sur le rachat anticipé de prêt, détaille Julien Le Calvez, actuaire et responsable du déploiement de Solvency 2. Notre progiciel n'offrait pas cette souplesse."* Ces calculs étaient limités à 20 000 dossiers à la fois, contre 200 000 avec le nouveau système. Sur le plan technique, Suravenir s'est appuyé sur l'un des meilleurs spécialistes français d'Hadoop, Matthias Herberts. Son équipe (au sein de la DSI de Crédit Mutuel Arkéa) a d'abord monté la plate-forme Big Data. Elle a ensuite cherché à convaincre les opérationnels de l'utiliser. La collaboration avec Suravenir a été facilitée grâce aux compétences de ses actuaires en statistiques, et en développement, qui sont indispensables pour exploiter au quotidien les données dans le système Hadoop.

pour lesquelles l'approche prédictive est de mise. Il s'agit là d'un enjeu majeur pour le secteur: en permettant d'ajuster les niveaux de provisionnement et d'optimiser les dispositifs de couverture et de transfert de risque, les modèles prospectifs constituent un des outils clés de pérennisation de l'activité.

L'anticipation du comportement des assurés - résiliation, rachat et arbitrage - figure parmi les exemples les plus notables. Afin de capter les phénomènes conjoncturels de masse, les modèles actuels intègrent souvent une approche par défaut, forfaitaire, à dire d'expert. Les bases de données historiques ne fournissent en effet pas ou peu d'historique sur ce domaine. L'approche Big Data en revanche pourrait permettre d'enrichir les informations disponibles sur les profils des assurés et de leur appliquer des modèles de comportement prédictifs. Il est ainsi aisé d'imaginer que la propension d'un assuré à racheter son contrat d'épargne, en cas de taux servi inférieur à celui de la concurrence, peut être en partie anticipée par ses habitudes de navigation internet: consultation de sites comparatifs, de boursicotage,

etc. Pour cela, l'assureur doit mettre en œuvre des stratégies et outils de collecte de données sur ses sites et certainement également des partenariats de partage de données avec d'autres plateformes numériques.

Le domaine de la modélisation - qui fait l'objet d'une recherche et développement dynamique ces dernières années - laisse ainsi entrevoir de nombreuses applications du Big Data comme un des axes de perfectionnement des modèles, en vue de pallier à certaines limites bien connues de leurs concepteurs.

#### Domaines d'application: marketing et conception des produits

Les services marketing des compagnies d'assurance peuvent aujourd'hui, avec les bons outils, anticiper les besoins des assurés et proposer au bon moment les bons produits.

#### Collectivement

Sur l'exemple de l'encadré BFM TV sur l'exploitation des bruits de l'internet pour anticiper une tendance, les marketeurs peuvent être à l'écoute des besoins de protections financières et envisager de fabriquer en quasi-temps réel les garanties nécessaires et attendues par le marché. Dans une perspective d'amélioration du cycle de création produit des assureurs à travers leur système d'information, le produit d'assurance peut ainsi être proposé au marché, par une souscription en temps réel *via* internet ou d'autres types de supports connectés.

#### Individuellement

Plus traditionnellement, le Big Data peut fournir des informations comportementales aux professionnels du marketing tout comme aux actuaires de manière à mieux ordonner les besoins et niveaux de couverture attendus par l'individu souscripteur d'une garantie. Mieux encore, les informations spécifiques à cet individu ou objet assuré peuvent également être fournies en temps réel après la souscription de manière à adapter la prime, la provision ou encore réaliser des actions de prévention.





# Quels chantiers, quelles perspectives demain pour les assureurs et le Big Data ?

## METTRE EN PLACE LES INFRASTRUCTURES DE RECUEIL ET DE TRAITEMENT

Un des paradoxes du Big Data provient du fait que la plupart des informations produites l'ont été en dehors d'utilisation de l'information. Par exemple, les réseaux sociaux n'ont pas pour objet initial de fournir d'informations aux assureurs mais simplement d'interconnecter les utilisateurs. Deux types spécifiques d'infrastructure de collecte vont devoir être considérés :

- ▶ des systèmes en charge de collecter des données existantes ou produites par d'autres systèmes : par exemple l'alimentation du système d'information de l'assureur avec des données de téléphonie mobile ;
- ▶ des systèmes propres à l'assureur en charge d'aider à gérer ses risques : par exemple une boîte noire implantée au sein d'un véhicule.

Au-delà de la récupération des informations brutes, l'assureur va devoir investir dans des systèmes de traitement lui permettant de transformer en temps réel les données brutes en système décisionnel à travers de nouvelles modélisations :

- ▶ modélisation de bases de données de type « entonnoir » : à partir d'une base de données à entrée large, transformation par tranche de l'information de manière à pouvoir reconstituer des bases de données structurées exploitables ;
- ▶ modélisation mathématique des processus permettant la création des différentes tranches d'informations progressivement structurées.

## QUELS RISQUES SOUS-JACENTS AUX BIG DATA ?

À la hauteur des opportunités qu'elle suscite, la révolution Big Data entraîne inévitablement un certain nombre de risques que nous devons avoir en tête et que nous listons, de manière non exhaustive ici.

Nous l'avons évoqué plus haut, les éléments liés à la vie privée sont particulièrement sensibles. En effet, l'arbitrage continu entre le droit à la confidentialité et l'utilisation de données privées est déjà une réalité. Les assurés vont certainement accepter de partager certaines informations pour obtenir de meilleures tarifications par exemple, mais un certain nombre de garanties devront être fournies : fixer des limites à l'utilisation des données, assurer la protection des données et éventuellement leur destruction et évidemment montrer l'intérêt réel pour l'assuré à les partager.

À chaque système de tarification existe un système de fraude. L'assureur devra ainsi veiller à ce que



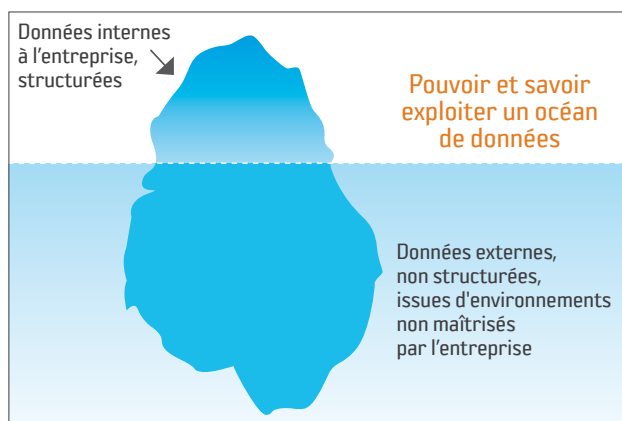
l'information produite et récupérée soit saine et non polluée. En effet, au regard du flou initial des données sources, des systèmes fraudeurs pourraient envisager de produire des informations déviantes de manière à fausser une tarification ou certaines modalités de souscription. Par le passé, certains systèmes connectés aux marchés financiers ont déjà tenté par exemple de soutenir artificiellement des cours de Bourse. Autre exemple plus récent, en utilisant simplement le mode de fonctionnement du moteur de recherche, des épisodes de « google bombing » permettent d'influencer les résultats d'une requête et donc réussissent à générer des résultats orientés.

L'internationalisation désormais instantanée des informations numériques ne permet pas aujourd'hui d'assurer un traitement égal en termes de droit et de confidentialité selon les pays. Ainsi, des opérateurs d'assurance qui exercent dans certains pays pourraient bénéficier d'un avantage concurrentiel évident au détriment des assurés d'un autre pays et de leurs assureurs locaux.



La traçabilité des données constituant le Big Data est quasi inexistante. Ainsi, leur audit est difficile et de nouvelles règles de gouvernance doivent être élaborées pour s'y adapter. De plus, entre l'utilisation des données et l'audit éventuel, les changements de statut de la donnée peuvent être nombreux et ainsi conduire à l'impossibilité de vérifier des informations initiales.

## Bases de données assurancielles versus Big Data



### LE RÔLE DES ANALYSTES PRÉDICTIFS ? L'ÉMERGENCE DE NOUVELLES RESPONSABILITÉS ?

La plupart des autres secteurs d'activité sont également progressivement concernés par le Big Data. Certains, d'ailleurs, depuis assez longtemps du fait de technologies particulières utilisées : la médecine, l'énergie ou encore la téléphonie. Avec le développement de la manipulation des données massives et les fortes attentes sous-jacentes en termes de modélisation prédictive, une nouvelle profession s'est peu à peu formalisée : celle des analystes prédictifs.

La question qui se pose aujourd'hui est de savoir si cette spécialité est concurrente ou complémentaire à la gestion du risque, aux actuaires donc pour la partie quantitative. Notre conviction est forte en la matière : absolument pas. Ces professionnels de la donnée sont évidemment complémentaires aux professionnels du risque, ces derniers étant les seuls garants des équilibres techniques de l'assureur. La responsabilité des actuaires est même plus large selon nous : au-delà des prérogatives actuelles, l'actuaire doit également se porter garant de l'utilisation des données et du principe même de mutualisation assurantielle.

L'actuaire doit œuvrer dans le respect de normes éthiques et du service aux assurés, notamment sur l'anonymat des données, au contraire des analystes prédictifs et des spécialistes du marketing.

Il appartient aux actuaires de continuer à rechercher la meilleure mutualisation possible du risque et de gérer l'équilibre des risques dans les mutualités constituées par la souscription. C'est le retour sur la « ligne de crête » entre segmentation et mutualisation des risques sur laquelle doit cheminer la science actuarielle malgré les fortes tentations offertes par le Big Data d'individualiser et de segmenter.

Bref, l'actuaire de demain, sous Solvabilité II ou non d'ailleurs, apparaît non plus comme un producteur de données, mais comme un certificateur de la qualité, de la pertinence des données et un vérificateur de leur usage. À cet égard, il pourrait devenir le correspondant du régulateur sur les questions de conformité dans le recueil et l'usage du Big Data. Il devient un « commissaire aux données », comme il existe des « commissaires aux comptes ».

L'analyste prédictif sera, lui, positionné en amont de la démarche d'utilisation du Big Data. Il servira à collecter et à assister les professionnels du risque dans l'organisation des données.

Autre nouvelle responsabilité majeure que nous pensons voir rapidement émerger : le *Chief Data Officer*, CDO. En effet, les données issues du Big Data deviennent tellement stratégiques et centrales à l'exercice de la profession d'assureur, qu'il semble opportun qu'une responsabilité centrale dans la gouvernance de l'entreprise émerge.

La mise en place puis l'exploitation d'un dispositif Big Data solliciteront ainsi l'expertise d'un nombre d'acteurs important, partageant tous l'objectif à atteindre : disposer de nouveaux outils permettant d'optimiser l'offre et le développement de l'entreprise. Sous la coordination du *Chief Data Officer*, les équipes d'actuaires, de juristes, de commerciaux, du marketing, soutenus par les analystes prédictifs, seront amenées à travailler ensemble en mode projet, en lien étroit avec les spécialistes du système d'information, afin de mettre en place un écosystème Big Data qui réponde aux besoins des métiers susceptibles d'en bénéficier : délivrer de l'information pertinente pour alimenter les réflexions stratégiques, optimiser le positionnement commercial et renforcer le suivi du risque.

Aujourd'hui, l'industrie de l'assurance est basée sur deux métiers régaliens : la gestion des contrats et celle du risque. Demain, nous avons la conviction que l'assureur aura toujours deux métiers régaliens : la gestion

“ Une nouvelle responsabilité majeure devrait rapidement émerger : le *Chief Data Officer*. ”

## Les 5 erreurs à éviter

- ▶ Limiter le Big Data aux bases de données massives et structurées, déjà dans le périmètre de l'assureur.
- ▶ Impliquer uniquement les directions informatiques : un projet Big Data est d'abord un projet métier.
- ▶ Définir à l'avance ce que l'on attend du Big Data : la puissance des informations que l'on peut en tirer se découvre au fur et à mesure.
- ▶ Stocker d'énormes volumes de données sans en avoir défini l'usage au préalable : intérêt nul et risque de saturation.
- ▶ Définir d'emblée une architecture cible : privilégier un prototype.

des données et celle du risque. En effet, à terme, il semble que la gestion des contrats disparaisse progressivement au sein de systèmes d'informations hybrides orientés sur le traitement des données, la tarification instantanée des produits et l'encaissement tout aussi instantané des primes. Cette vision futuriste, certes simplifiée pour grossir le trait, est sur quelques nouveaux produits déjà vérifiée : par exemple, sur certaines compagnies aériennes en ligne, lors de l'achat du billet, en fonction du profil de l'acheteur, une assurance décès en un clic est proposée, consommée le temps du vol, puis disparaît. Évidemment, d'autres couvertures à développement long subsisteront fort heureusement, mais leur gestion sera différente.

### LE MONDE DU BIG DATA DANS 5 ANS POUR L'ASSUREUR CONNECTÉ

L'exercice de prospective et au regard des différents sujets relatifs au Big Data traités dans ce *Dossier Technique*, nous pouvons aujourd'hui apprécier l'impact, les avantages et les risques du Big Data pour l'assureur. Nous proposons ici une première esquisse d'une vision cible dans quelques années.

#### Le système d'information

Tout d'abord, le système d'information de l'assureur devra évoluer sur la trame proposée ci-après : les aspects Big Data en orange sur le graphique viennent compléter le système d'information existant, illustré en bleu.

Une nouvelle fonction apparaît dans l'entreprise : le *Data Scientist* aura en charge l'analyse des données et l'aide à leur exploitation.

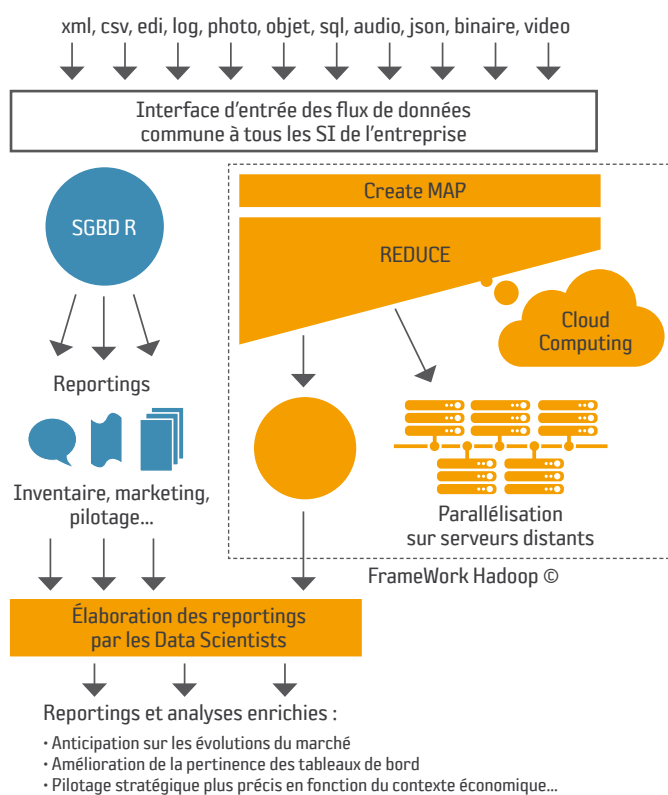
#### Les données

Selon une étude publiée par IDC<sup>(1)</sup>, les volumes de données auxquels nous serons confrontés en 2020 seraient les suivants :

- ▶ Entre 2005 et 2020, l'univers digital va croître d'un facteur 300, passant de 130 exaoctets ( $130 \times 10^{18}$ ) à 40 zettaoctets ( $40 \times 10^{21}$ ), ce qui représentera 5 200 Go de données par personne dans le monde. Entre maintenant et 2020, l'univers digital va doubler tous les ans.

- ▶ À ce jour, seule une petite fraction de l'univers digital a été analysée pour y rechercher des informations de valeur. IDC estime qu'à l'horizon 2020, un tiers de l'univers digital contiendra des informations qui pourraient avoir de la valeur si elles étaient analysées.
- ▶ En 2020, près de 40 % des informations de l'univers digital seront au contact des fournisseurs de *cloud computing*, signifiant que l'octet sera stocké ou traité dans le *cloud* quelque part entre le système d'origine et l'utilisation finale.

#### Le système d'information



Source : Optimind Winter

(1) *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East* (December 2012).



## Conclusion

---

Nous avons pu apprécier dans les précédentes pages les opportunités offertes par le Big Data pour l'assureur. Ses caractéristiques industrielles nous confortent dans l'idée qu'il doit aujourd'hui être appréhendé comme un projet majeur et stratégique à initier au plus vite. Évidemment, les bénéfices réels de tels investissements risquent de prendre quelques années à produire leurs effets.

À terme, les changements dans l'approche comportementale et marketing sont majeurs. La forme même des produits de couverture va être modifiée. Paradoxalement, les techniques mathématiques et actuarielles de tarification vont peu évoluer. En effet,

les techniques destinées à assurer les équilibres techniques sont invariables.

Tout comme la médecine prédictive s'impose de plus en plus comme un outil de notre quotidien, les technologies et les masses d'informations aujourd'hui disponibles peuvent demain nous apporter la capacité à encore mieux répondre aux besoins de protection des assurés, tout en envisageant des tarifs plus adaptés dans un contexte de rentabilité accru pour l'assureur. C'est en tout cas les vœux que nous formulons et les travaux que nous avons désormais décidé d'initier chez Optimind Winter.

---

## Qui sommes-nous ?

---

Leader de l'actuariat conseil et de la gestion des risques en France, OPTIMIND WINTER constitue l'interlocuteur de référence pour les assureurs, mutuelles, administrations, banques et grandes entreprises qui souhaitent un partenaire métier de haut niveau les accompagnant dans leurs projets stratégiques. Éthique, déontologie, expertise, méthode, pragmatisme et investissement sont les valeurs clefs qui animent nos 180 collaborateurs, consultants experts pour la plupart, dont plus de 70 actuaire diplômés membres de l'Institut des Actuaire. Nos clients bénéficient ainsi des plus hautes expertises en gestion du risque associées à la qualité d'une signature de référence d'un des leaders européens en gestion des risques. Notre indépendance, garantie par un capital détenu uniquement par nos salariés et dirigeants, offre à nos clients la perspective d'une collaboration pérenne et engagée.

OPTIMIND WINTER vous apporte son expertise sur les métiers suivants :

- Actuariat Conseil
- Protection Sociale
- Risk Management
- Projets & Maîtrise d'Ouvrage

Pour plus d'informations :

[www.optimindwinter.com](http://www.optimindwinter.com)

**optimind winter** ::

LOCAL **OPTIMIZATION** EUROPEAN **MINDED**