

## Les définitions et l'écosystème du Big Data

Le Big Data tantôt fait rêver, tantôt inquiète. La multiplicité des possibilités de connections électroniques (les « clics ») depuis le début d'Internet laisse imaginer des solutions commerciales innombrables se substituant au commerce de proximité. Mais beaucoup de consommateurs (clients) se sentent « espionnés ». La Carte Orange défunte devenue le Passe Navigo permet de connaître, comme le GPS, les itinéraires de chacun dans et hors la ville. Le contenu du caddy de supermarché est un outil de marketing qui pourrait bien modifier les politiques de prix de la grande surface, mais peut-être pas au profit du consommateur. Mon banquier connaît aujourd'hui beaucoup d'éléments de mon comportement d'achat et de mon patrimoine. Orange dispose d'énormes bases de données sur les communications téléphoniques et Internet de ses clients. La génération 2.0 se félicite, au moins dans la publicité télévisée, de communiquer avec sa banque *via* le *smartphone*. Les « applis » se multiplient, dans une ambiance de gratuité un peu suspecte. La marionnette de Tim Cook, dans « Les Guignols de l'Info » sur Canal+, présente chacune des innovations d'Apple comme une « révolution », sous l'œil sarcastique et dubitatif de PPDA. Beaucoup d'entre nous racontons notre vie sur Facebook, et/ou participons au réseau LinkedIn. Nous vivons, comme le dit la marionnette de Tim Cook, une « révolution » de la connaissance, fondée sur la prolifération de « données » (*data*) qui ne cesse de croître. Twitter génère 7 téraoctets de données et Facebook 10 par jour<sup>1</sup> et rien ne semble indiquer un ralentissement de croissance. Le Big Data c'est d'abord cette prolifération de données transmises sous forme

---

1. Source : *Le Big Data dans l'assurance*, Marc Dupuis et Emmanuel Berthelé, L'Argus de l'Assurance, collection Les Essentiels, décembre 2014 : 1 téraoctet = 10<sup>12</sup> octets = 6 millions d'ouvrages de librairie.

électronique, dans l'ensemble de la société, dans un cadre très peu structuré, et qui laisse à penser que tout un chacun fournit et reçoit, mais aussi peut utiliser toute cette masse d'informations disponibles pour peu qu'il sache la canaliser et la traiter. Des événements récents (et parfois délictueux) montrent qu'il y a en réalité peu de limites à l'obtention et à l'utilisation des informations (WikiLeaks).

Le monde de l'assurance est, curieusement, pris à revers par cette « Révolution ». Curieusement, car l'assurance comme la banque sont, depuis un demi-siècle, des gestionnaires de données de masse : mouvements des comptes courants, ordres de Bourse, émissions de cotisations, gestion de contrats Vie et Non-Vie. Tous les ménages ou presque sont bancarisés et la très grande majorité d'entre eux sont assurés pour leurs biens, leurs responsabilités, leur santé, leur épargne. Mais les données recueillies sont très cloisonnées, sur un compte (nonobstant le fait que le client peut en ouvrir plusieurs auprès de la même banque), sur un contrat d'assurance Auto, MRH<sup>2</sup>, Vie, Santé, Entreprise, sans que le rapprochement de ces « couloirs » d'information sur la gestion d'un même client soit toujours possible (voire tout simplement impossible). Le Big Data ouvre une ère d'informations multiples et sans destination spécifique, des mouvements bancaires aux photos de famille sur Facebook, en passant par les itinéraires suivis sur le GPS du véhicule et tant d'autres informations sur notre comportement, assuré le plus souvent. Tim Cook a raison (au moins sa marionnette dans « Les Guignols »), c'est pour l'assurance aussi « une Révolution ».

C'est pourquoi, avant de décrire comment l'industrie de l'assurance peut (et sans doute **doit**) transformer ses *business models* en utilisant cette « Révolution », il est utile d'esquisser à grands traits (pardon pour les spécialistes des systèmes d'information) la définition du Big Data, la nature des outils qui créent son environnement (Dupuis et

---

2. Multirisques Habitation.

Berthelé<sup>3</sup> parlent de « l'écosystème »), les acteurs du Big Data, et montrer comment se construit un « projet Big Data » pour l'assurance au sens large, sans prendre parti sur les acteurs de l'Industrie les mieux placés pour agir, ce qui sera fait dans un autre chapitre.

\*  
\*   \*   \*

## 1. Big Data : essai de définition

Le Gartner Group<sup>4</sup> a publié la définition de référence : « *Big Data is high volume, high velocity and high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making* ». Tout est dit dans cette excellente définition.

- **Les 3V = volume, velocity, variety** caractérisent la « révolution numérique ».

Des **volumes** considérables, sans commune mesure avec l'information que l'assureur reçoit (ou obtient) de son client, particulier ou industriel, sur le risque qu'il lui est demandé d'assurer. Nous dirons que le Big Data pourrait produire une inversion de l'asymétrie d'information entre client et assureur, et bousculer sévèrement le risque d'anti-sélection (le vieux *moral hazard* qui pousse l'assureur ignorant à souscrire des risques à probabilités fortes de réalisation).

La **Velocity** est la rapidité d'obtention d'une information abondante sur les clients. Le contrat annuel traditionnel français « à tacite reconduction » d'un risque évalué et réexaminé à date

---

3. M. Dupuis, E. Berthelé, *Le Big Data dans l'assurance*, Éditions de l'Argus de l'Assurance, 2014.

4. In Dupuis et Berthelé, *op. cit.*

fixe est submergé par les informations sur le comportement du client et l'évolution quotidienne (ou quasi) du risque qui arrivent à l'assureur.

La **Variety** est peut-être le plus important. L'information est multiple, chiffrée ou non, vérifiée ou pas, pertinente ou inutile. Elle n'est plus codée et gérée en silo : à l'assurance Auto, l'âge du permis de conduire ; à l'assurance Santé ou emprunteur, le questionnaire « de santé » ; à l'assurance Vie, des informations narratives sur le patrimoine et les compétences financières du prospect. Le Big Data est aussi, ou d'abord, une stratégie de **tri de l'information pertinente**, fût-elle « à bas bruit<sup>5</sup> ». La fréquence n'est plus nécessairement un critère de pertinence de l'information.

- L'information est un **actif** (*asset*). Pour l'assureur, c'est un moyen de souscription. Pour le *data scientist*, c'est un trésor, qu'il faut déterrer pour concevoir des produits et des tarifs adaptés au plus près des besoins du client. Pour l'actuaire, voici venu le temps du « prix actuariel » (donc exact !) du risque présenté par le prospect. Nous dirons plus tard les ravages que cette idée peut causer sur les théories classiques de l'assurabilité du risque. Au diable la mutualisation ! Nous voici à l'ère de l'évaluation du risque que chacun payera son prix, sous réserve – évidemment – que celui-ci soit financièrement et socialement acceptable. Quoi qu'il en soit, la donnée a désormais de la **valeur**, pour peu que l'on sache la reconnaître et l'extraire. Nous verrons qu'il s'agit de la question cruciale. Qui crée de la valeur et qui est le mieux placé pour l'extraire ? Qui sait faire les « barrages filtrants » pour chercher les « pépites » marketing, tarifaires et de gestion qui amélioreront substantiellement la « proposition de valeur » (la célèbre *value proposition* des consultants de la décennie 90) au prospect ? Qui assurera l'avantage compétitif de l'un ou l'autre

---

5. Les données « à bas bruit » désignent les signaux peu nombreux mais révélateurs de l'évolution des comportements.

des maillons de la chaîne de valeur de l'assurance ? Qui, finalement, s'appropriera la marge générée par la gestion efficace du Big Data ?

C'est à ce point qu'interviennent les notions de « monétisation » et de « marchandisation » des données, et le célèbre adage : « Si c'est gratuit, c'est vous le produit ». Facebook et Twitter ne connaîtraient pas leur actuel succès en Bourse s'ils n'étaient que de sympathiques lieux d'échanges d'informations et de retrouvailles entre anciens du Lycée de Romorantin. Il s'agit bien de céder des informations sur les clients et prospects et de donner aux vendeurs de services l'accès à d'immenses bases de données clients disponibles. L'assureur est aujourd'hui peu familier de ces procédures de démarchage commercial, même si, depuis une dizaine d'années, la présence de certains sur les réseaux sociaux est avérée et efficace.

- Des formes nouvelles de gestion (*processing*) de l'information pour une meilleure gestion des entreprises. C'est là que la technologie intervient. Il faut collecter cette information, la stocker, la trier, la valider, etc. pour la rendre utilisable, donc pertinente, pour les usagers assureurs. Nous verrons comment chacun des métiers de l'assurance peut ainsi extraire de la valeur d'une vague d'informations non codifiées et non orientées « métier » au profit de la compétitivité de son activité.

Il manque à cette définition générale deux caractéristiques majeures du Big Data qui en font une révolution dans l'approche même du risque pour l'assureur.

- Aujourd'hui, les informations de gestion du risque sont **internes à l'entreprise**, ou ont été internalisées (questionnaires, déclarations, constat amiable, éléments chiffrés sur les transactions du compte en banque). Le Big Data ouvre d'innombrables voies de renseignement externes sur le ou les clients ou prospects à travers Google, Twitter, Facebook et les objets connectés.