

DÉCEMBRE 2015 - N°27

Le mariage de l'homme et des Big Data, c'est presque maintenant...

par Philippe Rouger, *European Institute for Systems Biology & Medicine.*

Après le haut débit pour les réseaux et le cloud pour les traitements, les *Big Data* sont le nouveau produit d'appel du numérique. Les chercheurs, les météorologues ou les hommes de marketing les collectent, les analysent et les exploitent depuis bien longtemps. Le fait nouveau, c'est la prise de conscience par l'opinion et la classe politique des perspectives ouvertes par la mise en données de la vie intime des individus par-delà les frontières. Une révolution qui attend d'être encadrée.

2015 restera-t-elle vraiment comme l'année du Big Data ? Dans les faits, non, car ces « grosses données » existent depuis de nombreuses années dans le domaine des sciences (CERN), du marketing, de la finance et dans notre prédictif journalier tels que les sondages et surtout la météo. Par contre, pour le grand public et les politiques, oui, car cela sonne l'entrée dans une nouvelle époque de l'ère numérique comme si ce seul terme permettait de trouver des solutions à tous les maux de l'économie française et de relancer la croissance...

Pas une conférence parlant d'innovation sans que les Big Data soient au centre des débats ! Et plusieurs dizaines de livres ont été publiés ces deux dernières années pour décrire, expliquer les changements, que dis-je, la révolution qu'annonce cette nouvelle technologie !

Mais en fait, les Big Data, qu'est-ce que c'est ?

C'est à la fois une réalité technique, un produit marketing et un débat de société (en particulier sur la mise en données du corps humain)

Ces « grosses data » répondent à la règle des 5V. Les trois premiers V (Volume, Variété et Vélocité) apportent les éléments objectifs à la qualification de Big Data, c'est-à-dire un très grand nombre de données, d'une importante variété et arrivant par flots très rapides. Les deux derniers V (Véracité et Valeur) donnent une information qualitative à ces données par l'analyse de la qualité, de la fiabilité, de l'intérêt ou de l'« actionnabilité » des données collectées.

Pour se donner une idée des données stockées en 2010, on les estimait à 1.2 zeta (1000 milliards de milliards d'octets - 10 puissance 21) dans le monde. Les prévisions pour 2014 avaient bondi à 4.5 zeta. On en sera certainement à 40 zeta en 2020.

Et le Big Data s'accompagne du développement d'applications ou plateformes (multi applications), d'IA (intelligence artificielle) qui visent à donner un sens aux données traitées.

Humainement difficile voire impossible de traiter un tel volume d'informations. Nous devons nous reposer sur des plateformes qui regroupent infrastructure de stockage, calculateur hyper puissant, systèmes de gestion de données particuliers et algorithmes pour les traiter.

Au départ, les Big Data étaient principalement liées à la science (recherche en particulier nucléaire) qui pouvait disposer de super calculateur et de stockage presque illimité. L'exemple le plus proche de nous qui se perfectionne depuis des dizaines d'années est la météo dont le dernier calculateur comporte

60000 cœurs Intel Xeon et un stockage de 45 peta-octets (million de milliards d'octets). C'était le temps des « data science ».

Mais depuis 20 ans, les capacités de traitements et de stockage se sont démocratisées et sont à la portée de toutes les grandes entreprises.

Voilà l'arrivée des *Big Data*. Le premier secteur à se lancer a été le marketing (data mining) ou les sociétés de sondage. La finance a suivi avec ses plateformes décisionnelles (haute fréquence ou autres). Aujourd'hui, plus de 60 % des transactions financières sont traitées par des plateformes numériques.

A suivi l'industrie du numérique, portée par les GAFAMT (Google, Amazon, Facebook, Apple, Microsoft, Twitter), qui a vite compris l'intérêt marketing et commercial qu'elle pouvait tirer de cette technologie dont le nom sonne si bien aux oreilles, qui se traduit et s'explique si facilement.

Cette démocratisation technologique et la simplicité du terme font qu'aujourd'hui les Big Data c'est un peu tout, et surtout n'importe quoi si on lit ce qui est proposé ! Statistique, analyse de données, open data, outil de diagnostic... Bref, dès que le nombre de données est important, nous sommes dans les Big Data qui représentent le nouveau produit d'appel du numérique (après le haut débit pour les réseaux et le cloud pour les traitements).

Le sujet qui pose question aujourd'hui, ce sont les applications et les algorithmes qui doivent traiter les Big Data pour en faire une sélection, un regroupement ou un traitement pour transformer une série de données en une métadonnée qui sera accessible et compréhensible par les personnes qui vont les utiliser dans leur métier, et plus généralement dans leur vie.

Pourquoi ? Tout simplement parce que les Big Data s'attaquent maintenant à l'homme.

Les Big Data sont en train de s'immiscer de plus en plus intimement dans la vie de l'individu, que ce soit dans sa consommation ou sa vie, dans son être biologique (génomique, biochimie, micro biome), dans la collecte d'informations liées à notre environnement, à nos activités physiques ou mentales avec les objets connectés.

Et encore une fois, les milliards de données que porte chacun devront être traités par des plateformes, des applications et des algorithmes.

Heureusement, du côté de la maladie et globalement de la santé, les données sont bien gardées. De nombreux projets de recherche sont lancés et protégés dans le cadre académique. Ces nouvelles capacités technologiques sont notamment utilisées pour des analyses de données personnalisées en comparaison aux référentiels complexes des maladies chroniques dans le cadre de la médecine 4P - Prédictive, Prédictive, Préventive, Personnalisée et Participative.

Cela permet d'avancer notamment sur des thérapies liées au diabète ou au cancer. Mais cela peut aussi être utilisé pour établir des statistiques sur un très grand nombre de données, comme le projet de la CNAMTS et de l'École polytechnique qui a débuté en 2015 sur les données de tous les Français avec pour objectif la détection de signaux faibles ou anomalies en pharmaco-épidémiologie, l'identification de facteurs utiles pour mieux analyser les parcours de soins, la lutte contre les abus et la fraude.

Par contre, l'échec de Google vient de démontrer la difficulté de comprendre le comportement humain avec l'arrêt de son projet sur la prédiction des épidémies, en particulier de la grippe ; ses prédictions étaient surévaluées de 50% par rapport à la réalité.

Mais que l'on ne se trompe pas : ce ne sont que les premiers essais, et demain les résultats s'amélioreront rapidement. L'avenir nous le dira...

Restent les premières questions posées par l'arrivée de ces technologies quand elles touchent l'humain.

Qui sera propriétaire et/ou détenteur des données issues des analyses de son propre corps et qui pourra en disposer ?

Comment construire les processus de validations qualitatives ou imposer une régulation sur ces applications, intelligence artificielle ou algorithmes dont on ne maîtrise pas le fonctionnement et encore moins la qualité des résultats ?

Presaje a encore bien du travail.

 [Télécharger le PDF de l'article](#)

<< [Retour au sommaire](#)

PRES@JE.COM

Une publication de l'Institut PRES@JE
(**Prospective, Recherche et Etudes Sociétales Appliquées à la Justice et à l'Economie**)

30 rue Claude Lorrain 75016 Paris

Tél. 01 46 51 12 21 - E-mail : contact@presaje.com - www.presaje.com

Directeur de la publication : Michel Rouger

Pour ne plus recevoir d'e-mails de la part de Presaje, [cliquez ici](#) >> [CONSULTER LES PRECEDENTS NUMEROS](#)