

"L'arrivée des Big Data marque une nouvelle révolution technologique"

Interview de Avner Bar-Hen, professeur du Cnam

Publié le 27 septembre 2017 – Mis à jour le 12 octobre 2018

Big Data ! Depuis quelques temps, tout le monde a ce mot sur les lèvres. Il impressionne autant qu'il semble dissimuler des enjeux importants. Pourtant, il n'est pas toujours facile de savoir ce qu'il recouvre précisément et pour quelles raisons on lui accorde tant d'importance. Plongée dans l'univers des mégadonnées avec Avner Bar-Hen, professeur du Cnam, chaire Statistique et données massives.



Depuis quelques années, les mentions aux Big Data fleurissent un peu partout... Mais, concrètement, de quoi parle-t-on ?

Le mot *Big Data* est assez récent. À l'origine, dans les années 1980, il n'était utilisé que par des chercheurs qui étaient submergés par les données, en astrophysique par exemple. Puis, il y a eu comme un effet de mode, et tout le monde s'est emparé de ce mot qui entre dans Le Petit Robert en 2016, avec une définition assez vague : « l'ensemble des données générées par les nouvelles technologies, caractérisées par leur volume colossal ». On peut d'ailleurs noter que ce dictionnaire préconise l'utilisation de « mégadonnées », terme choisi par la Commission générale de terminologie et de néologie en août 2014 pour les « données structurées ou non dont le très grand volume requiert des outils d'analyse adaptés ». Mais, dans un cas comme dans l'autre, la définition reste extrêmement vague, basée essentiellement sur le volume de données.

Tout ne serait donc qu'une question de volume et de masses de données ?

Pas seulement car ces définitions sont trompeuses ! Dès 2001, les *Big Data* ont en effet été caractérisées par les 3V (le volume, la vélocité, la variété) auxquels se sont ensuite ajoutés de nouveaux éléments comme la véracité ou la valeur de la donnée. Aujourd'hui, on peut ainsi recenser une quarantaine de « V » possibles, de la volatilité à la viscosité en passant par la viralité... Autrement dit, ce qu'il faut retenir ce n'est pas le *Big*, qui n'a pour moi que peu de sens et serait plutôt à considérer comme une approche publicitaire, mais bien la *Data*. L'important réside en effet dans cette nouvelle représentation du monde sous forme de données car elle bouleverse notre manière de concevoir les choses et donc les questions que l'on se pose.



“ Le Big Data est un bouleversement aussi important que celui de l'invention de la perspective à la Renaissance. ”

Avner Bar-Hen, professeur du Cnam, chaire Statistique et données massives

Derrière l'effet de mode, il y a donc quelque chose de plus profond...

Oui, l'arrivée des *Big Data* marque une nouvelle révolution technologique qui est née en deux temps. D'une part, lorsqu'il a été possible de stocker de l'information. Et, d'autre part lorsque nous nous sommes aperçus qu'il y avait de la valeur dans le stockage de ces traces numériques, ces empreintes que nous laissons autour de nous et que l'on arrive à récupérer sous des formes diverses et variées. Nous sommes ainsi rentrés dans une époque où nous représentons le monde à travers des données chiffrées et numérisées. C'est un bouleversement aussi important que celui de l'invention de la perspective à la Renaissance qui a permis de représenter de manière réaliste l'espace qui nous entoure.

Qui est touché par cette révolution ?

Cette révolution est d'autant plus importante qu'elle touche tous les domaines – le marketing, l'assurance, la génomique, les sciences humaines... – puisqu'elle change notre manière de voir, de réfléchir, d'aborder les questions. C'est ce qui explique que tout le monde se sent propriétaire des *Big Data*. Pour un commercial, elles ont bouleversé le marketing ; pour une informaticienne, l'informatique ; idem pour une biologiste ou un documentaliste... Et, ils ont tous un peu raison ! Que ce soit dans les domaines de la santé publique ou de la génomique, du stockage ou de la sécurité informatique, du calcul haute performance ou des statistiques... les mégadonnées ont révolutionné notre domaine d'expertise, développé de nouveaux métiers, nécessité des formations spécifiques...

Cette révolution ne fait que débiter. Nous sommes en effet loin d'être en fin de cycle, et le monde de demain sera extrêmement différent de celui d'aujourd'hui. Parallèlement au développement soutenu des embauches pour les fonctions les plus qualifiées de *Data Architect* ou de *Data Scientist* comme les moins qualifiées, notamment pour le stockage ou le nettoyage des données, on assiste déjà à une transformation permanente des métiers. À terme, ils seront certainement tous touchés par l'exploitation des données numériques, même si certains le seront nécessairement plus que d'autres. Ce qui pose d'ailleurs un réel problème en matière de formation : les métiers se transformant au fur et à mesure, avec des cycles beaucoup plus rapides que ceux liés à la formation initiale, il est indispensable que la formation continue puisse prendre le relais.

Les mégadonnées ont-elles aussi révolutionné notre manière de faire de la science ?

Depuis Karl Popper, la recherche s'inscrit principalement dans une démarche hypothético-déductive : le chercheur formule une hypothèse qu'il expérimente afin de l'infirmier ou de la confirmer. À partir de ce résultat, il pose de nouvelles hypothèses et ainsi de suite. Au contraire, avec les mégadonnées, nous sommes plutôt dans une démarche empirique. C'est une façon totalement différente de faire de la science qui a pour corollaire négatif que nous ne savons pas toujours ce que nous faisons.

Mais il y a aussi des aspects très positifs puisque les *Big Data* permettent d'avancer sur des questions que nous ne pouvions pas aborder avant. Si l'on prend l'exemple de la génomique, fouiller à la volée dans des données pour évaluer les risques de développer un cancer, est peut être suffisant dans un premier temps. En tout cas avant que l'on ne puisse expliquer pourquoi certaines personnes sont résistantes à telles pathologies et d'autres non.

Les *Big Data* peuvent aussi nous permettre de gagner du temps dans le recueil des données pour se concentrer sur l'analyse et la réflexion. Pour un collègue historien qui travaillait sur la grippe espagnole dans la presse française, quelques heures ont été suffisantes pour extraire les données des journaux contre plusieurs mois s'il avait dû les lire. Je ne prétends pas que le résultat est meilleur, mais il est évident que l'opération est beaucoup plus rapide et que cela permet de se concentrer sur l'analyse des données et non plus sur leur recueil.

Et dans votre propre approche des statistiques ?

Personnellement, l'arrivée des *Big Data* a modifié deux aspects de mon travail. J'ai commencé à faire des statistiques à une époque où la recherche française était surtout théorique, avec peu de ressources logicielles. Au contraire, aujourd'hui, les recherches en statistiques sont fortement dirigées vers des questions applicatives. Ce changement de paradigme a généré énormément de nouvelles questions structurelles et sociétales si bien que désormais, lorsqu'on développe des méthodes statistiques ou qu'on en étudie les propriétés mathématiques, la dimension applicative est très marquée.

Les mégadonnées seraient donc le nouvel eldorado de l'innovation ?

Oui et non. Structurellement, de nombreuses entreprises se dotent aujourd'hui d'une direction « innovation et *Big Data* » car pour beaucoup d'entre elles les mégadonnées représentent un important vecteur d'innovation. Depuis des années, elles ont en effet recueilli une multitude de données, qui ne sont pas nécessairement structurées puisqu'on y trouve aussi bien des variables numériques que des images ou des sons. Stockées dans des *Data Laks*, sous forme de donnée brute, elles permettent de travailler sur de très gros jeux de données, souvent en temps réel.

Mais, si cela représente effectivement un vecteur d'innovation, il n'est pas exclusif. La fin annoncée du pétrole, par exemple, entraînera elle aussi de nombreux bouleversements qui transformeront radicalement notre société. C'est déjà le cas, et cela touche directement notre quotidien. Depuis quelques mois, les « pochons » en plastique ont ainsi été remplacés par des sacs biodégradables en amidon de maïs ou de pomme de terre, matière première renouvelable qui offre une solution durable à la raréfaction des combustibles fossiles. De même, demain, nous arrêterons probablement d'importer des tee-shirts de Chine car le coût du transport sera devenu prohibitif. Il sera alors nécessaire d'innover pour combler ce vide.

Propos recueillis par Yvan Boude,
rédacteur en chef du *Cnam mag'*

Innovation Numérique

Le dernier Cnam mag'

LE CNAM MAG' #9

Société numérique, société inclusive ?

1 mai 2018

[+ Retrouvez tous les numéros](#)

Les cours d'Avner Bar-Hen au Cnam

Master Statistique Sciences des données

Master Statistique du risque pour la finance et l'assurance

Unité d'enseignement Modélisation et prévision des séries chronologiques

L'auteur

Avner Bar-Hen
Professeur Cnam
[+ tous ses articles](#)



