

AVIS D'EXPERT

BIG (ET) DATA DANS TOUTE SA COMPLEXITE

Cécile Godé

Professeure agrégée des universités
Sciences de gestion

Aix Marseille Univ, CRET-LOG,
Aix-en-Provence, France

Site Pauliane-MEGA
424, chemin du Viaduc
13080 Aix-en-Provence
06 12 54 78 56
cecile.gode@univ-amu.fr

Amandine Pascal

Professeure agrégée des universités
Sciences de gestion

Aix Marseille Univ, CNRS, LEST,
Aix-en-Provence, France

LEST, UMR 7317 CNRS
35, avenue Jules Ferry
13626 Aix-en-Provence Cedex 01
amandine.pascal@univ-amu.fr

Résumé

Les définitions les plus connues du Big Data font une large place à ses caractéristiques d'ampleur ou aux capacités technologiques de stockage et traitement des données massives. Après être revenu sur ces deux dimensions, cette chronique en développe une troisième, peu traitée par la littérature en management : la complexité. Le Big Data ne fait alors plus forcément référence à des ensembles massifs de données mais à des combinaisons de données hétérogènes qui interagissent et se transforment de façon imprévisible. Cette perspective incite à abandonner toute pensée simplifiante pour appréhender pleinement le Big Data et embrasser les nouveaux défis qu'il soulève.

Introduction

« Mais madame, le Big Data, c'est quoi exactement ? », question récurrente de nos étudiants en sciences de gestion et du management lorsque nous abordons la notion en cours. Contre toute attente, la réponse est moins aisée qu'il n'y paraît au premier abord. Faut-il insister sur l'attribut de volumétrie ? Ou plutôt sur la dimension technologique du Big Data ? Et qu'en est-il de la nature même de ces ensembles de données ? À ce jour, il n'existe pas de vision unique ou de définition univoque du Big Data. Les milieux académiques, de l'entreprise et du conseil, se sont tour à tour essayés à le caractériser sans réel effort intégratif, alimentant souvent la confusion. Ainsi, si le Big Data n'est pas un phénomène récent, les tentatives de synthèse et de définition sont encore nombreuses et d'actualité (voir par exemple, [Fosso-Wamba et al., 2015](#) ; [Lee, 2017](#) ; [Favaretto et al., 2020](#)). Cette chronique s'inscrit dans cette mouvance. A partir d'éléments de la littérature académique et professionnelle, elle revient sur les deux principaux attributs du « Big » (et) « Data » puis apporte un éclairage nouveau et peu traité dans la littérature : la complexité du Big Data et les implications qui en découle.

Le « Big » de Data

Cette première dimension, le « Big » de Data, renvoie à la définition la plus connue aujourd'hui, celle qu'en tant qu'enseignants nous mobilisons fréquemment tant elle est facile à transmettre et illustrer auprès de nos étudiants. Il s'agit de la définition en V(s), qui se concentre sur les caractéristiques d'ampleur et de quantification des ensembles de données.

Les 3 V du Big Data ont été caractérisés dès 2001 par Douglas Laney, alors Vice-Président de Gartner Research. Le premier V fait référence à l'explosion du Volume de données produites, principalement due aux progrès technologiques (par exemple, objets connectés, infrastructures d'échanges de données) et aux modes d'échanges et de communication évoluant vers toujours plus de partage d'informations. Le deuxième V renvoie à la Variété des sources d'information (par exemple, les médias sociaux, les interactions Machine to Machine, les terminaux mobiles) et des types de contenus (texte, audio, vidéo). Les données récoltées ne s'inscrivent plus dans des structures nettes, traditionnelles, mais sont diversifiées, incluant contenus, données géolocalisées, de mesure, mobiles, de connexion, etc. Enfin, le troisième V décrit la Vitesse (ou vélocité) de production, de collecte et de partage des données. Celles-ci sont dynamiques, en mouvement permanent. Elles peuvent de ce fait devenir rapidement obsolètes, leur cycle de génération étant très rapide. Un quatrième V a été ajouté : la Véracité, qui met en avant l'importance de la qualité et de l'intégrité des données récoltées. Ces dernières peuvent être incomplètes, voire inexactes pour certaines : la fiabilité de la source et la qualité du contenu doivent donc être vérifiées. Aujourd'hui, les contributions spécialisées (essentiellement issues de revues professionnelles en management) énoncent jusqu'à dix V pour caractériser le Big Data, insistant notamment sur la Valeur créée à destination des clients et la Vulnérabilité des données produites. Ces contributions sont certes éclairantes en ce qu'elles permettent de comprendre, de façon simple et didactique, le phénomène Big Data. Elles restent cependant limitées dans leur capacité à appréhender l'ensemble de ses dimensions.

Le « Data » de Big

La deuxième dimension du Big Data met en exergue les capacités technologiques de traitement des ensembles massifs de données ainsi que les méthodes qui les sous-tendent. Selon [George et al \(2016\)](#), les technologies et méthodes du Big Data renvoient en effet à la science des données (ou data science), insistant sur les fonctionnalités technologiques des outils à disposition, à savoir leurs capacités de collecte, de stockage, de traitement, d'analyse et de transmission/visualisation des ensembles de données. Cette dimension est essentiellement portée par les grands acteurs du web et communautés qui développent, commercialisent ou diffusent les principales solutions disponibles sur le marché et en open source. Elles représentent une alternative aux solutions traditionnelles qui ne suffisent plus pour collecter, stocker et traiter les masses de données disponibles. Parmi les solutions technologiques les plus connues, on retrouve par exemple MapReduce, un modèle de programmation développé par Google qui permet d'exécuter des calculs parallèles sur un énorme volume de données en utilisant des clusters de serveurs : Hadoop, mis au point par Apache, qui permet de distribuer les grandes quantités de données collectées à travers un cluster de serveurs et de les indexer ; ou encore les bases de données non relationnelles NoSQL dont les grands acteurs du web comme Google ou Amazon ont été parmi les premiers à s'emparer. Ici, l'angle privilégié pour définir le Big Data reste la prise en charge technique des divers V, à partir d'infrastructures techniques et applications logicielles dédiées.

La « Big » complexité des « Data »

Enfin, le caractère complexe du Big Data est une troisième dimension moins connue mais qui nous paraît devoir être mise en avant pour une appréhension complète du phénomène. Comme le remarque [Davenport \(2014\)](#), lorsqu'on évoque le Big Data « *l'accent est souvent mis sur le volume, mais la complexité des grandes données réside en réalité dans le manque de structure* » (p. 1). Les données non structurées n'ont pas de format prédéterminé et ne sont pas organisées ou classées au préalable. Elles évoluent et s'enchevêtrent à d'autres types de données (structurées, issues des bases de données, mais également semi-structurées), renvoyant à un Big Data constitué de « *combinaisons de différents ensembles de données, à différents moment et issus de différentes sources* » ([Van Rijmenam, 2014](#), p. 16). Ces aspects ne sont pas toujours associés à des caractéristiques d'ampleur et de grandeur. Comme le précise par exemple la communauté open source MIKE 2.0 (Method for an Integrated Knowledge Environment), le Big Data ne fait pas forcément référence à des ensembles massifs de données. Certains peuvent être relativement peu volumineux, mais néanmoins constitués de combinaisons de données hétérogènes difficiles à traiter, analyser et interpréter ([Godé et al., 2019](#)).

Le Big Data peut donc aussi être caractérisé par sa complexité, liée à l'hétérogénéité et à la multiplicité des combinaisons de données indépendantes qui interagissent et se transforment de façon imprévisible. Les modèles d'analyse de réseaux (qui permettent d'identifier les rôles des nœuds et de ne pas dissocier la partie du tout), les modèles hiérarchiques bayésiens (qui permettent de traiter les variations du comportement individuel) ou encore les modèles économiques structurels (qui tiennent compte de l'anticipation des acteurs et de leur interconnectivité) sont à privilégier dans la mesure où ils facilitent le traitement et l'analyse du caractère complexe du Big Data ([Glady, 2017](#)). De la même façon, les technologies orientées analyses prédictives et intelligence artificielle permettent d'adresser la nature non structurée des données, d'en comprendre les enchevêtrements et d'en extraire la valeur.

Ici, les principes de la complexité que la pensée d'[Edgar Morin \(1990\)](#) nous enseigne permettent d'appréhender différemment le Big Data, et notamment d'accepter l'idée d'abandonner toute pensée simplifiante. Privilégier la Big complexité des datas plutôt que les Big Vs ou la Data Science pour comprendre le Big Data implique un changement de paradigme et l'abandon du cartésianisme pour embrasser de nouveaux défis, pour « *penser l'un et le multiple ensemble, penser ensemble l'incertitude et le certain, le logique et le contradictoire, et l'inclusion de l'observateur dans l'observation* » ([Morin, 1990](#), p. 92).

Références

- Davenport, T. (2014). *Stratégie Big Data*. Paris, France: Pearson Education.
- Favaretto, M., De Clercq, E., Schneble, C. et Elger, B. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*, 15(2), e0228987. <https://doi.org/10.1371/journal.pone.0228987>
- Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G. et Gnanzou, D. (2015). How 'big data' can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234-246. <https://doi.org/10.1016/j.ijpe.2014.12.031>

- George, G., Osinga, E., Lavie, D. et Scott, B. (2016). Big Data and Data Science methods for management research. *Academy of Management Journal*, 59(5), 1493-1507. Doi: 1493-1507. 10.5465/amj.2016.4005
- Glady, N. (2017). MOOC Lecture 32: Complexité et Big Data, ESSEC TV. Repéré à <https://fr.coursera.org/learn/lavenir-de-la-decision/lecture/mdtw4/complexite-et-big-data-nicolas-glady>
- Godé, C., Lebraty, J.-F. et Vazquez, J. (2019). « Le processus de décision naturaliste en environnement Big Data : le cas des forces de Police au sein d'un Centre d'Information et de Commandement (CIC) ». *Systèmes d'Information et Management*, 24(3), 67-96.
- Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, 60(3), 293-303. <https://doi.org/10.1016/j.bushor.2017.01.004>
- Morin, E. (1990). *Introduction à la pensée complexe*. Paris, France : Editions du Seuil.
- Van Rijmenam, M. (2014). *Think bigger: Developing a successful Big Data strategy for your business*. New York: USA: Amacom.