

# Création de valeur par les données massives

Sylvain WILLART

PhD, Risk analyst, Advanzia Bank ; Maître de Conférences  
Université de Lille

Dominique CRIÉ

Professeur des Universités, Université de Lille,  
Institut d'administration des entreprises,  
Lille Economie Management (UMR CNRS 9221)



L'information est source de pouvoir. La décennie qui suivra sera celle de l'émergence ou du renforcement de géants informationnels marquant le Web de leur empreinte et de leur pouvoir. Ces entreprises ont ouvert un chemin que beaucoup d'autres se pressent aujourd'hui d'emprunter, celui de l'exploitation des données massives, ou Big Data. Cependant, la mise en place d'un projet orienté données n'est pas nécessairement synonyme de retour sur investissement. Il peut être intéressant de s'interroger sur les leviers de création de valeur permis par les Big Data afin de maximiser la potentialité d'un retour sur investissement (ROI) positif.

## Un renouvellement de la production et des usages des données

La production des Big Data est largement assurée par la « multitude » et ce de façon plus ou moins consciente. Les données de clics (*clickstream*) sont générées par les internautes presque à leur corps défendant, celles des réseaux sociaux relèvent d'une démarche plus pro-active mais sans doute non conscientisée ; certaines enfin représentent un réel investissement collaboratif dans une mission précise (OpenStreetMap). Mais la production des Big Data a également une dimension automatisée du fait de la multiplication des capteurs (GPS, RFID...) et des objets connectés (*Internet of Things*). Ces deux éléments se conjuguent avec le phénomène de « *datafication* » qui renvoie à cette propension toujours plus grande dans toutes les sphères de la société de mesurer tout ce qui peut l'être et de tenter de conserver une trace chiffrée de tous les événements. Les données sont ainsi de plus en plus souvent récoltées en continu et concernent tout type d'activité.

La question de l'usage de ces données n'intervient souvent qu'après la récolte ! Et c'est là le second aspect important: la plupart de la valeur extraite des données vient d'usages secondaires, non envisagés au moment de leur collecte. Ces usages ex-post sont également contingents du croisement de plusieurs sources de données. Dans cette optique, le développement de l'open-data a une importance capitale : en mettant à disposition de nombreuses données, les possibilités de croisements et donc de découverte d'informations augmentent exponentiellement. Mais ces croisements et usages secondaires ne sont pas sans problème, notamment lorsqu'il s'agit de données à caractère personnel, qui ne peuvent normalement être recueillies sans un consentement relatif à un usage précisé et donc prédéfini.

## Des données à la valeur ou de la valeur des données

Si la récolte de données ne nécessite pas de but prédéfini, il faut souligner toutefois que la valorisation requiert en revanche un travail important de raffinage et d'extraction. L'objectif est alors de transformer la donnée en information, laquelle pourra produire de la valeur si délivrée au bon moment dans un contexte adapté. La valorisation des données peut s'entendre de façon duale : valeur d'usage et valeur d'échange. La première apparaît lorsqu'un service est construit à partir des données, la valeur de ce service pour l'utilisateur peut alors être corrélée à la quantité des données et à la qualité des informations qu'il intègre. De tels services peuvent être destinés à des consommateurs (suivi de consommation, points de fidélité récoltés, recherche d'une offre optimale...) ou à des entreprises (aide à la décision, tableau de bord, tendances, prédictions...). Quant à la valeur d'échange, le traitement préalable des données est moins crucial. En revanche, l'échange nécessite la structuration d'un marché autour de vendeurs et d'acheteurs. Il est aujourd'hui largement oligopolistique.

## Des contraintes techniques et méthodologiques nouvelles

Le traitement des Big Data dans l'optique de l'extraction d'une valeur d'usage suppose un changement radical des pratiques en termes d'équipement, de logiciels, et de méthodes d'analyse.

Au plan de l'équipement, « l'informatique dans les nuages » (Cloud computing) s'impose : la capacité de stockage et la puissance de calcul n'ont plus vocation à être hébergés au sein de l'entreprise mais consommés auprès d'un prestataire, variabilisant les coûts fixes de l'informatique.

Le stockage des données dans le cloud requiert l'utilisation d'outils spécifiques, capables de les traiter à partir de différents serveurs et d'effectuer des calculs en utilisant les nombreux processeurs de ces serveurs. Ce « parallélisme<sup>1</sup> » informatique nécessite des solutions logicielles capables de traiter le stockage distribué (e.g. Hadoop), d'effectuer de façon performante des tris sur les données (e.g. MapReduce) et de réaliser sur ces données des analyses statistiques (clustering, classification, moteur de recommandation e.g. Mahout).

Une dernière contrainte tient à la nature des données intégrant de plus en plus souvent des dimensions spatiales, temporelles, et de réseau. Les liens qui en découlent entre les observations (plus ou moins proches dans le temps, dans l'espace, et dans leurs relations sociales) complexifient les analyses notamment via la forme particulière de la matrice de variance-covariance : la mise en œuvre pratique est encore un défi en termes de puissance de calcul et de rapidité d'obtention des résultats.

Le processus de production, les nouveaux usages potentiels, les défis de traitement des données ou l'extraction des informations qu'elles contiennent peuvent modifier en profondeur la façon dont les entreprises créent une valeur pour elles-mêmes et leurs clients. De nouveaux rapports de pouvoir s'instaurent sur un marché qui se structure pour le moment autour de quelques majors. En conséquence, il est crucial de s'interroger sur la, ou les, valeur(s) que les données peuvent apporter à l'entreprise, ainsi que sur les conditions sous lesquelles cette valeur peut être capturée.

---

1. Utilisation de plusieurs ordinateurs/serveurs en même temps, en « parallèle », de façon « distribuée ».

## Une approche économique et comptable de la valeur des données

En démultipliant les pistes d'exploitation des données, la révolution des Big Data positionne l'information comme une ressource stratégique. Mais elle soulève également la question de la valorisation de cette ressource.

Dans une certaine mesure, les données peuvent être comptabilisées comme des actifs immatériels, ce qui pose la question de leur valorisation dans le bilan de l'entreprise. De plus, les données peuvent être considérées comme un bien public, ou plus précisément collectif (non-rival et exclusif), dont la valeur ne s'épuise pas avec l'usage. En fait, la valeur des données ne s'érode qu'avec le temps, parfois très rapidement, notamment pour les moteurs de recommandation ou la publicité ciblée. La valorisation comptable se heurte donc à un double problème : actif immatériel et à amortir dans un temps court mais non dépendant des usages. Les partisans de l'orientation Big Data soulignent que la valeur des données réside dans l'optimisation des décisions de l'entreprise. Sur un échantillon de 179 entreprises cotées, Brynjolfsson et al. (2011) montrent que les entreprises engagées dans une démarche Big Data augmentent leurs ratios de productivité de 5 à 6% en moyenne.

La valorisation boursière de certaines entreprises valide cette thèse : Facebook a été introduit en bourse le 18 mai 2012 pour une capitalisation de 104 milliards de dollars alors que son total bilan de 2011 s'élevait à 6,3 milliards. Nonobstant les cash-flows espérés, ces chiffres valorisent les données de l'entreprise à 97,7 milliards, soit 102,3 dollars pour chacun de ses 955 millions de comptes actifs à la date de son introduction.

Les données ont, depuis longtemps, également une valeur économique d'échange. La révolution des Big Data tend à structurer et massifier ce marché en y incluant de nouvelles sources de données et en multipliant le nombre d'acteurs. Les prix constatés y sont extrêmement variables en fonction de la précision des données (tableau 1).

**Tableau 1 : exemples de prix pour différents types de données sur le marché**

Entreprise	Type de donnée	Valorisation
<b>Axiom</b>	Adresses email (accompagnées éventuellement d'un profil)	<b>2 à 5 cents par contact</b>
<b>Plateformes RTB et AdExchange</b>	Profils de navigation (sans identification)	<b>0 à 1 dollar pour 1000 affichages en Europe</b>
<b>Facebook</b>	Valorisation boursière d'un profil	<b>102,3 dollars</b>
<b>Federico Zannier</b>	Données de navigation, localisation du pointeur de souris, GPS, webcam, fichiers log <sup>2</sup>	<b>2 dollars pour un jour, 5 pour une semaine</b>

2. « A bit(e) of Me » : <<http://www.kickstarter.com/projects/1461902402/a-bit-e-of-me>>, accédé le 05 Mai 2016

Entreprise	Type de donnée	Valorisation
<b>Datasift</b>	Tweets et analyses (tarif fonction de la charge de calcul; prix du Data Processing Unit : 20 cents par heure)	<b>10 cents pour 1000 tweets plus un coût de traitement exprimé en DPU</b>
<b>Datacoup</b>	Données de profil agrégées (navigation, réseaux sociaux, transactions de carte bancaire) type panel	<b>8 dollars par mois offert à chaque membre du panel</b>
<b>Appannie.com Intelligence</b>	Données de téléchargement d'applications mobiles et ebooks par plateformes et par pays	<b>Gratuit pour l'éditeur d'une application. 15000 dollars par an pour un accès à l'ensemble des données.</b>
<b>Datamarket.com</b>	Données mises en vente par les entreprises qui les détiennent ou les créent	<b>Prix fixé par le vendeur</b>

La valorisation des données par le marché souffre aujourd'hui d'un certain manque de recul. Les exemples sont trop peu nombreux ou biaisés par le fait que les vendeurs proposent des produits trop spécifiques et peu comparables. Il est donc difficile d'extrapoler une tendance, un prix, ou même le contour exact du produit vendu. Ces données sont également rarement vendues par les personnes auxquelles elles se rapportent (les exemples de Facebook, mais également de DataSift, Appannie, ou du Real-Time Bidding concernent en effet des acteurs intermédiaires qui récoltent des données de tiers, les analysent, les agrègent, et les revendent sous forme d'information – profils de navigation par exemple).

En cas de vente de données personnelles par les personnes concernées, la valeur accordée à ces données diffère entre le déclaratif et le comportemental et elles sont souvent échangées contre la gratuité de certains services. On pourrait en conclure que, du point de vue du consommateur, ses données ont moins de valeur que le service qu'il reçoit en retour (dans le cas contraire, il refuserait le service). Mais il existe un biais cognitif réduisant l'importance perçue des conséquences de long terme (sollicitations commerciales) au profit des bénéfices de court terme (obtention immédiate du service). Enfin pour ne pas dévoiler ses informations, l'internaute peut également recourir à différents stratagèmes dont le mensonge. Une nouvelle option émerge pour la monétisation des données personnelles par l'intermédiaire de plateformes dédiées (coffres-forts numériques), les internautes pourront stocker et louer ou vendre leurs informations personnelles (e.g. enliken.com).

## La valeur d'usage des données

Si les résultats de Brynjolfsson et al. (2011) valident l'importance de l'orientation données, ils précisent mal comment et à quels niveaux de l'entreprise les données peuvent délivrer leur valeur d'usage. Pourtant, l'ensemble des fonctions de l'entreprise semble devoir être impacté par les Big Data. Une analyse en termes de chaîne de valeur peut clarifier comment cette révolution se décline dans l'ensemble de l'entreprise, pour ensuite soulever la question de l'impact sur la prise de décision stratégique de l'usage des données et des informations qu'elles peuvent receler.

Certaines activités sont déjà largement impactées par les données. La logistique par exemple

a intégré le suivi des livraisons en temps réel par l'application de puces RFID sur les colis. Le marketing en ligne a profité de l'analyse des données de clic issues des fichiers LOG des sites internet pour développer des mesures d'audience ou les techniques de recommandation, de retargeting (publicité comportementale), et de *Real Time Bidding* (enchères en temps réel d'espaces publicitaires).

Pour la logistique comme pour le marketing, la datafication a permis l'amélioration des services existants (meilleur suivi des livraisons, optimisation du design des sites, personnalisation des contenus publicitaires), mais également une accélération du cycle de décision. La fonction logistique en intégrant le suivi des ventes à celui des stocks gagne en réactivité et automatise les réapprovisionnements. Certaines entreprises spécialisées utilisent leurs masses de données pour calculer avec précision le coût de chaque livraison et ajuster automatiquement leurs prix.

Au-delà de l'amélioration des solutions existantes, les produits vendus peuvent être assortis de divers capteurs permettant la récolte des données d'usage des consommateurs. Ces données peuvent ensuite être utilisées pour améliorer le produit, ou proposer un service après-vente personnalisé. L'objectif ici étant d'automatiser la récolte des retours d'expérience clients en prenant en compte l'ensemble de ces derniers et non seulement ceux qui prennent le temps de répondre aux sollicitations des entreprises ou de contacter le service après-vente.

Le stockage, la sécurisation, et l'analyse de ces données liées aux objets et/ou aux individus peut également représenter un marché important qu'investissent déjà les fournisseurs de « coffres-forts numériques » ([adminium.fr](http://adminium.fr), [personal.com...](http://personal.com...)).

## Limites et risques de l'orientation données

La stratégie orientée données comporte également des risques pour les entreprises notamment ceux liés à l'utilisation de mauvaises données (en termes de qualité ou de nature) et aux mauvaises utilisations des données.

Le contrôle de la nature des données revient quant à lui à s'assurer que le type de données utilisé est le mieux adapté pour répondre à la question posée. Sur ce point, on peut citer notamment le point de vue hétérodoxe de la société CtrlShift (2011) qui soulève la question de savoir si les Big Data sont effectivement les mieux à même de servir au pilotage de la relation client. L'idée ici est que ces données, trop massives, génèrent des segmentations trop floues qui, du point de vue du consommateur final, sont toujours mal adaptées à ses intentions d'achats et provoquent un sentiment d'intrusion. A l'inverse, l'utilisation de « *small data* », émises directement par les consommateurs via leur coffre-fort numérique, pourrait permettre de mieux répondre à cette attente. Au plan statistique, il s'agit notamment d'une amélioration du rapport signal/bruit : plutôt que d'essayer de récolter un grand nombre d'informations, on se concentre sur les sources les plus pertinentes i.e. les consommateurs eux-mêmes. Cela étant, on peut souligner que ces SmallData, si elles ne sont plus aussi volumineuses, conservent les caractéristiques de variété et de vélocité de leurs grandes sœurs.

## Pour conclure....

D'une part les données vont continuer de croître en volume, en diversité de sources, en rapidité de captation et de traitement, et en utilisations. Les entreprises mal préparées à ce déluge de données risquent de rencontrer des difficultés d'adaptation. D'autre part, et paradoxalement, les données disponibles risquent de ne jamais être complètes, la complexité des phénomènes humains dépassant, certainement encore pour longtemps, la capacité de mesure et d'analyse. Si la quantité d'information et la capacité de calcul peuvent repousser les frontières de la

rationalité limitée, elles ne peuvent la transformer en rationalité parfaite.

L'adaptation des entreprises est nécessaire afin de profiter des leviers de création de valeur que permettent les stratégies orientées données. De l'adaptation des produits et services existants à l'investissement de nouveaux marchés, de nombreuses opportunités restent ouvertes.

## Référence

- Brynjolfsson, E., Hitt, L. et Kim, H., "Strength in Numbers: How does data-driven decision-making affect firm performance?", (6 décembre 2011), International Conference on Information Systems

