

Jean-Charles COINTOT
Yves EYCHENNE

La Révolution Big Data

Les données au cœur
de la transformation
de l'entreprise

DUNOD

Tout le catalogue sur
www.dunod.com



Le pictogramme qui figure ci-contre mérite une explication. Son objet est d'alerter le lecteur sur la menace que représente pour l'avenir de l'écrit, particulièrement dans le domaine de l'édition technique et universitaire, le développement massif du photocopillage.

Le Code de la propriété intellectuelle du 1^{er} juillet 1992 interdit en effet expressément la photocopie à usage collectif sans autorisation des ayants droit. Or, cette pratique s'est généralisée dans les établissements

d'enseignement supérieur, provoquant une baisse brutale des achats de livres et de revues, au point que la possibilité même pour

les auteurs de créer des œuvres nouvelles et de les faire éditer correctement est aujourd'hui menacée.

Nous rappelons donc que toute reproduction, partielle ou totale, de la présente publication est interdite sans autorisation de l'auteur, de son éditeur ou du Centre français d'exploitation du droit de copie (CFC, 20, rue des Grands-Augustins, 75006 Paris).



© Dunod, 2014

5 rue Laromiguière, 75005 Paris
www.dunod.com

ISBN 978-2-10-071142-0

Le Code de la propriété intellectuelle n'autorisant, aux termes de l'article L. 122-5, 2° et 3° a), d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective » et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite » (art. L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles L. 335-2 et suivants du Code de la propriété intellectuelle.

TABLE DES MATIÈRES

Introduction	1
1 L'ère de l'information et le big data	7
L'émergence des réseaux sociaux	9
Le social business	11
Le consomm'acteur	15
La révolution du mobile	19
Le cloud	27
2 Le big data et les 3V : Volume, Vitesse et Variété	35
Volume	36
Vitesse	42
Variété	44
3 Les outils du big data	51
Les entrepôts de données	52
Hadoop et les outils de découverte des données	54
Les outils de prédiction	59
L'analyse temps réel	64
Les outils de visualisation	65
Les outils de gouvernance	69
4 Vie privée et big data	73
Les données personnelles	74
Les limites de l'utilisation des données personnelles	78
La confiance	83

5	Comment implanter le big data dans l'entreprise	91
	Les modèles de mise en place du big data	91
	Les compétences nécessaires	96
	Les changements organisationnels	99
	Le démarrage	102
	Le déploiement	105
6	Les bénéfices du big data pour chaque métier de l'entreprise	111
	Marketing et vente	113
	Production et logistique	121
	Risques et fraudes	126
	Ressources humaines	135
7	Le big data dans les industries, les services et le secteur public	143
	Le secteur des télécommunications	146
	Le secteur de l'énergie	155
	Le secteur bancaire	166
	Le secteur du commerce	177
	Le secteur public	182
	Le secteur de la santé	193
8	Prospectives	203
	Données ouvertes	203
	Le cognitive computing	212
	Glossaire	221
	Bibliographie	227
	Index	231

INTRODUCTION

Les techniques de l'information se sont sophistiquées au cours des âges, tant au niveau des codes qu'au niveau des supports. Écriture cunéiforme, hiéroglyphes, calligraphies, idéogrammes et alphabets divers et variés ont permis à travers les siècles de représenter l'information. En ce qui concerne les supports de stockage, nous sommes passés de l'argile, la pierre, le bois, le cuir, le métal, ou le parchemin, au silicium aujourd'hui. Mais le but n'a pas changé depuis la nuit des temps. Stocker toujours plus d'information, la reproduire et la diffuser toujours plus vite pour transmettre les idées, l'expérience et le savoir, afin de permettre à chacun, des simples individus aux grands de ce monde, de s'informer pour décider « en connaissance de cause ».

Fin 2012, certains experts annonçaient que l'humanité a créé plus d'information en deux ans qu'au cours de toute son histoire. En réalité, il y a eu entre-temps un changement de paradigme. Les données ne sont plus stockées sous forme scripturale mais digitale. Nos données, quelles qu'elles soient sont stockées sous forme de 0 et de 1 sur des supports magnétiques, ce qui a accéléré leur création, duplication et transmission.

Dans son étude de 2011, McKinsey Global Institute¹ estime que la plupart des entreprises de plus de 1 000 employés stockent en moyenne 200 téraoctets de données soit à peu près

1. Étude de McKinsey Global Institute : « Big data : the next frontier for innovation, competition and productivity », 2011. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

l'équivalent de la bibliothèque du Congrès américain. Selon l'étude menée en 2012 par l'IDC pour le compte d'EMC¹, le volume de données va être multiplié par 50 entre 2005 et 2020 pour atteindre 40 zettaoctets². Difficile de se rendre compte de ce que cela représente... Mais si entre 2010 et 2012, l'humanité a généré autant d'information que tout au long de son histoire passée, ce volume sera amené à doubler à nouveau tous les deux ans tout au long de la décennie.

L'ère numérique est caractérisée par cette croissance exponentielle de la création de données et des technologies qui permettent de les traiter afin d'en retirer de l'information. Cette information, toujours plus riche et plus rapidement disponible, augmente la vitesse et l'ampleur du renouvellement de la connaissance au sein de l'économie et de la société. La révolution industrielle, expression popularisée par Friedrich Engels et par Arnold Toynbee, désigne le processus historique du XIX^e siècle qui fait basculer – de manière plus ou moins rapide selon les pays et les régions – une société à dominante agricole et artisanale vers une société commerciale et industrielle. Elle repose au début sur l'exploitation du charbon et de la vapeur inventée par James Watt en 1769. Elle s'intensifie à la fin du XIX^e siècle et trouve son essor avec l'exploitation de l'électricité et du pétrole. Au début, le pétrole n'est qu'un combustible. 130 ans plus tard, grâce aux progrès de la chimie et de bien d'autres disciplines scientifiques, le pétrole et ses dérivés font partie de notre quotidien. Pas moins de 150 produits allant des antiseptiques aux vêtements, en passant par les produits

1. Étude IDC pour le compte d'EMC « The Digital Universe in 2020 : big data, bigger digital shadows and biggest growth in the Far East » IDC, 2012.

2. Zettaoctet : 10¹⁵ mégaoctets ou 10 000 000 000 000 000 mégaoctets.

cosmétiques et la nourriture, sont répertoriés¹. Les progrès faits par la recherche et l'innovation dans la recherche et l'exploitation des ressources pétrolières ont permis de tirer toujours plus des mêmes champs pétroliers et d'en découvrir de nouveaux.

La révolution digitale est le passage d'une société industrielle à une société de service. Elle est basée sur les données. On dit que ce sont les données qui alimentent l'économie (« *data fuel the economy* »). L'expression anglaise met mieux en évidence ce parallèle entre une économie qui a été alimentée par du pétrole (« *fuel* ») et une économie dont le carburant est devenu « la donnée ». À l'instar de l'évolution industrielle, l'évolution de l'ère digitale repose sur la capacité à toujours mieux exploiter les informations existantes et à exploiter de nouvelles sources d'information. C'est ce que l'on appelle généralement le big data. Nous reviendrons sur des définitions plus précises du big data mais on peut dire, en première approche, que le big data recouvre l'ensemble des technologies, des métiers, des approches conceptuelles permettant d'exploiter l'ensemble des données générées par les hommes de façon consciente ou non et par tous les objets connectés ou non.

Ce nouveau paradigme nous impacte tous, à titre personnel ou professionnel. Pour l'entreprise, il s'agit de canaliser le flot de données incessant généré en interne ou en externe, pour en tirer l'information nécessaire à une meilleure prise de décision. Décisions stratégiques au niveau des métiers : production, marketing, vente, finance, sécurité, mais aussi décisions/actions dans le monde de l'Internet des objets. Ce livre s'adresse aux différents

1. Source : Save and conserve. http://www.saveandconserve.com/2007/05/petroleum_based_products_a_long_list.html : "Petroleum based products : a long list", 2007.

acteurs de l'entreprise pour les aider à comprendre cette révolution et les fondamentaux du big data.

Il n'est toutefois pas réservé aux seuls acteurs de l'entreprise mais peut intéresser toute personne cherchant à exploiter des sources volumineuses de données, riches et variées, en temps réel ou non, car les fondamentaux du big data ne sont pas propres à un type d'exploitation ou un type de métier mais sont les mêmes quel que soit le domaine d'application : domaine public, santé, e-citoyenneté, villes connectées, domaine de la recherche, domaine de l'enseignement.

Le premier chapitre revient à la source du big data, à savoir : l'émergence des réseaux sociaux, la fluidité et l'ubiquité de l'information, résultant du développement de la mobilité et du cloud. Les deux chapitres suivants traitent du big data en tant que tel : comment le définir et quels sont les outils associés. Nous abordons dans le chapitre 4 les problématiques inhérentes à la protection de la vie privée dans ce monde du big data, qui tend à analyser toutes les données que nous générons et qui nous entourent. Nous terminerons cette première partie dédiée au big data par le chapitre 5 qui s'adresse à ceux qui seront amenés à déployer les solutions big data au sein de l'entreprise dans le cadre de leur *transformation digitale*.

La deuxième partie du livre intéressera toute personne qui veut comprendre comment évolue la société. Nous traitons de l'impact du phénomène big data sur les entreprises, les métiers et les secteurs d'activité dans les chapitres 6 et 7. Le big data permettra-t-il de développer une société citoyenne et durable malgré la globalisation ? Va-t-on vers un « *big brother* » ou vers une société assistée par des ordinateurs qui nous comprennent et nous assistent de façon naturelle ? L'ère du big data ne fait que commencer.

Pour terminer ce livre, nous traitons au chapitre 8 de deux sujets prospectifs importants du big data : la donnée ouverte ou l'open data – pour que le big data ne soit pas la propriété de certains – et le cognitive computing – l'ordinateur qui s'instruit et réfléchit grâce aux techniques du big data.

L'ÈRE DE L'INFORMATION ET LE BIG DATA

L'ère numérique est caractérisée par la croissance exponentielle de la création de données digitales. Mais, c'est bien plus qu'« une question de zéro et de un », comme le souligne Peter Sondergaard¹, directeur de la recherche du Gartner. C'est une transformation radicale de nos métiers, de nos usages, de nos façons de vivre, de consommer, de nous comporter en société. En fait, cette transformation radicale est moins le fait des données elles-mêmes que celui des technologies qui permettent de les traiter toujours plus vite et d'en tirer toujours plus d'informations – ce qui étend et renouvelle notre connaissance du monde et de nos sociétés, d'où les expressions « ère de l'information » ou « société de l'information ».

La société de l'information n'est pas qu'une expression ou un concept. Elle a pris corps à Tunis en 2005, lorsque l'Assemblée générale de l'Organisation des nations unies à Tunis a adopté le projet de résolution A/RES/60/252 de l'Union internationale des télécommunications². Depuis, tous les 17 mai de chaque année, se tient la journée mondiale de la société de l'information.

1. Peter Sondergaard, directeur de la recherche du Gartner. <http://www.usine-digitale.fr/article/bienvenue-dans-l-ere-de-la-digital-industrial-economy.N217400>

2. Résolution A/RES/60/252 de l'Union internationale des télécommunications. <http://www.itu.int/wisd/2006/res-60-252-fr.html>

Son but est de promouvoir l'accès, le développement et l'utilisation des technologies de l'information et de la communication dans les domaines économique et social.

Trois termes caractérisent le citoyen de la société de l'information :

- **Social** : l'homme d'aujourd'hui est « social ». Il habite le village « monde » et communique avec la planète à travers les réseaux sociaux. Il a des amis, rencontrés « à l'ancienne », et de nouveaux amis qu'il ne connaît qu'à travers la toile. Est-ce le début d'une conscience collective ? Cela signe-t-il l'émergence d'une noosphère¹ ? Ce qui est certain, c'est que cette communication joue un rôle important dans le comportement au quotidien de l'homme de l'ère de l'information.
- **Mobile** : à l'origine de la croissance exponentielle de la communication au sein de notre société de l'information, se trouve le téléphone mobile. Il est devenu une extension « naturelle » de l'homme. C'est son point d'entrée et son point de contact avec la société.
- **Cloud** : résultante naturelle des deux premiers phénomènes, le cloud est le substrat qui porte le développement de l'intelligence collective générée par l'avènement de l'ère de l'information. Il permet la diffusion, le partage de l'information. Tant que l'information était stockée sur un support autonome, elle pouvait s'échanger mais elle n'était pas omniprésente. Accéder à l'information où que l'on soit, au moment de son choix, avec le terminal dont on dispose, tel est le paradigme de l'ère de l'information.

1. Comme l'imaginait Theilhard de Chardin.

L'ÉMERGENCE DES RÉSEAUX SOCIAUX

Nous parlons ici des « réseaux sociaux » au sens anglo-saxon du terme. Le mot *social networks* a en effet été traduit et adopté un peu rapidement en français sous le terme de « réseaux sociaux ». Il faudrait plutôt parler de « réseaux sociétaux », car ceux-ci sont relatifs à un *fait de société*, et non à un quelconque but humanitaire (même s'ils peuvent relayer de telles actions comme cela a été le cas pour Facebook lors de la catastrophe du réacteur nucléaire de Fukushima¹).

L'origine des réseaux sociaux

Wikipedia voit l'origine de ces réseaux sociaux dans le *computerized bulletin board system*, inventé par deux passionnés d'informatique, Ward Christensen et Randy Suess, lors du grand blizzard de 1978 (une tempête de neige historique qui a bloqué Chicago pendant deux jours en janvier 1978). Le but en était bien de pouvoir échanger des informations entre membres du club CACHE, *Chicago area computer hobbyists' exchange*. Depuis, les sites se sont multipliés sur la toile. Parmi eux : Six Degrees, en 1997, reconnu par beaucoup comme le premier réseau social au sens actuel du terme, LinkedIn et MySpace en 2003, Facebook, Viaduc qui deviendra plus tard Viadeo et Google, sous le nom de Orkut, en 2004, YouTube en 2005, Twitter en 2006. Google Wave qui a fait une courte apparition en 2009 puis Google Buzz en 2010 et finalement Google+ en 2011.

1. A Viable Plan for Emergency Containment at Fukushima,
<https://www.facebook.com/CoverFukushima>

Le poids des réseaux sociaux aujourd'hui

Fin 2013, Facebook, Twitter, Google+ et LinkedIn totalisaient à eux seuls près de 2,5 milliards d'abonnés et YouTube fait état de plus d'un milliard de visiteurs uniques chaque mois. Les réseaux sociaux ne sont plus un phénomène de génération. Les tranches d'âge les plus en croissance en 2013 ont été celle des 55-64 ans pour Twitter et celle des 45-54 ans pour Facebook et Google+. LinkedIn n'est pas en reste puisque la majorité de ses abonnés sont répartis de façon assez équilibrée entre les différentes tranches d'âge de 25 à 64 ans. La tranche au-delà de 65 ans reste significative, et le fait le plus marquant récemment a été l'ouverture de LinkedIn en septembre 2013 aux jeunes de moins de 18 ans¹.

En moins de 10 ans, les réseaux sociaux se sont imposés comme le phénomène sociétal et économique le plus disruptif. Plus de limite, plus de frontière dans la « blogosphère ». Les gens partagent leurs contenus audio, vidéo, photo, créent des communautés virtuelles, mettent en commun des informations, des conseils, des recommandations.

L'utilisation même des réseaux sociaux a évolué pour passer de la sphère strictement privée à la sphère économique et professionnelle. Les marques l'ont compris et affirment leur présence dans les réseaux sociaux.

93 % des responsables marketing les utilisent comme média de communication de marque et 60 % des personnes interrogées disent être plus susceptibles de recommander une marque si elle est présente sur leur réseau social². Les 10 marques les

1. Digital Insights, <http://blog.digitalinsights.in/social-media-facts-and-statistics-2013/0560387.html>

2. Source : <http://www.socialbakers.com/blog/2183-socialbakers-quarterly-facebook-advertising-report>

plus populaires sur Facebook totalisent près de 400 millions de « fans »¹. On comprend l'intérêt de l'utilisation de ces médias pour le marketing.

Les réseaux sociaux sont aussi devenus des lieux d'échange entre employeurs et employés. LinkedIn compte plus de 3 millions de pages d'entreprises et 1,5 million de groupes d'intérêts professionnels. Un nouvel abonné rejoint LinkedIn toutes les deux secondes. 42 % des inscrits mettent à jour leur profil régulièrement et cherchent à se faire « recommander » car ils savent que la plupart des sociétés de recrutement consultent LinkedIn pour leurs recherches.

Cette communication perpétuelle entre individus, sociétés, partenaires, employeurs, employés est à la base du « social business ».

LE SOCIAL BUSINESS

De même que les réseaux sociaux n'ont aucune vocation sociale, le « social business » n'est en rien une activité à but non lucratif. Nous sommes loin du concept de l'entreprise sociale développée par Muhammad Yunus, Prix Nobel de la paix en 2006.

Social business et entreprise sociale : parenté et différences

Le social business dont il est question ici est basé sur un modèle tout à fait classique : la mission de l'entreprise est de créer de la valeur, de dégager des bénéfices et de distribuer des dividendes à ses actionnaires. Ce qui change fondamentalement par

1. Statistiques Facebook : <http://www.socialbakers.com/facebook-statistics/>

rapport à l'entreprise « traditionnelle », c'est la façon d'opérer pour atteindre ce but. Le social business est un nouveau business model, adapté à la « société de l'information ».

Mais si l'entreprise sociale de Yunus et le social business sont très différents, il est intéressant de remarquer certaines similitudes entre ces approches, notamment dans les moyens qu'ils emploient. On peut penser au crowdsourcing par exemple, que ce soit au niveau du financement ou au niveau de l'innovation et de la conception des produits et services.

Même si le financement participatif existe depuis très longtemps, par le biais de proches, d'amis ou de business angels, la parenté du crowdfunding et du micro-financement pratiqué aujourd'hui (pour financer des projets de start-ups ou des projets artistiques) est souvent attribuée au concept de micro-crédit de Yunus. Le *jugaad*, c'est-à-dire la débrouillardise, la capacité à l'innovation participative, est souvent associé à l'entreprise sociale. Puisque le développement passe par l'innovation, l'innovation ne doit pas être réservée aux seuls laboratoires.

Aujourd'hui, cette approche « sociale » de l'innovation est utilisée dans nombre de projets de start-ups mais aussi de grandes sociétés comme la SNCF ou Renault. Pour les start-ups l'innovation participative peut être stimulée par les incubateurs avec espaces participatifs ou des laboratoires participatifs. Pour les grandes entreprises, l'innovation participative consiste à intégrer les employés, les partenaires ou les clients dans la définition de nouveaux produits, de nouveaux services, de nouveaux processus, etc.

L'écosystème du social business

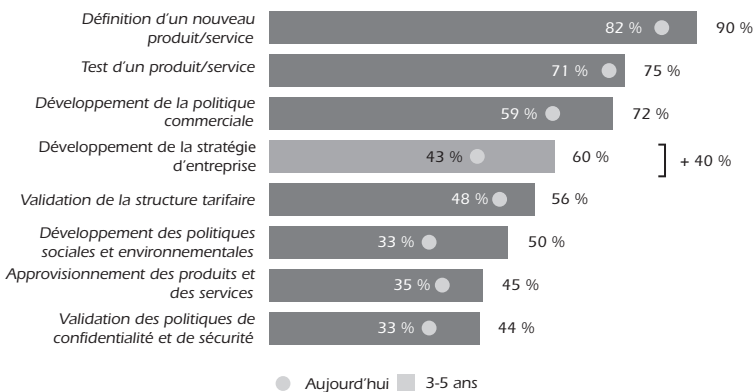
Ceci amène à une autre caractéristique commune aux deux modèles, l'importance de l'écosystème. Si ce n'est pas nouveau dans la sphère du social, c'est en revanche tout à fait nouveau dans le monde de l'entreprise. Le partenaire, le client, l'employé ont toujours existé, mais sous un mode structuré. Le partenaire fournit ou consomme un produit ou service, le client consomme un produit ou un service, l'employé exécute la mission qui lui a été fournie. Il s'agit d'un mode monodirectionnel entre acteurs. Dans l'écosystème du « social business », l'échange est multidirectionnel. Chacun des acteurs participe de façon active. Le partenaire ne se contente plus de fournir ou consommer un produit ou un service, il participe à son amélioration ou à sa création. Il développe des synergies, voire de nouveaux modèles de rétribution entre acteurs.

De même, dans le modèle « social business », l'employé n'est plus un exécutant mais un acteur de la marche de l'entreprise. C'est depuis longtemps un moteur pour les start-ups, mais c'est encore très nouveau pour les grandes entreprises. À cela, plusieurs raisons : la structure hiérarchisée, avec de nombreux niveaux, tendant à déresponsabiliser, mais aussi le manque d'outils. Comment faire participer des milliers d'employés ? Comment structurer la discussion ? Comment repérer les bons interlocuteurs, les bons experts sur des domaines qui ne sont pas strictement liés à leurs missions ? Les outils collaboratifs de dialogues, tels que les blogs, les wikis, les forums rendent cela possible. IBM met à disposition de l'ensemble de ses salariés dans le monde, soit environ 400 000 personnes, un environnement collaboratif, IBM Connections, qui permet à chacun de créer blog, forum ou wiki sur les thèmes qu'ils souhaitent allant des produits aux valeurs de la société. De même, tous les ans, IBM organise une « Jam session » au cours de laquelle chaque employé peut s'exprimer. Les « Jam sessions »

sont limitées dans le temps : 3 jours, 24 heures sur 24, et impliquent des modérateurs, qui permettent de canaliser et regrouper les thèmes similaires. Ils font émerger des centaines de milliers de contributions des points importants pour l'entreprise, que ce soit au niveau des produits ou des valeurs communes aux employés d'IBM, qui forgent l'esprit de la société. Au début réservées aux employés, ces « Jam sessions » sont désormais ouvertes aux partenaires et aux clients.

Le client participe à part entière à l'écosystème du « social business ». Il est actif parce qu'on le sollicite à travers les réseaux sociaux (pages Facebook, forums, blogs et autres outils collaboratifs de marque) et qu'on lui demande son avis (*scoring* et autres « vous avez aimé »). Mais il peut et sait aussi s'exprimer directement sur la toile, avec ces mêmes outils pour donner son avis de façon volontaire et spontanée. C'est un des points fondamentaux du changement de paradigme du « social business ». L'entreprise doit savoir travailler dans une relation bilatérale avec son client et non plus lui pousser ses offres.

Activités pour lesquelles les PDG veulent prendre en compte les attentes clients



Les PDG veulent que les clients participent davantage... jusqu'à contribuer au développement de leur stratégie d'entreprise

L'étude « The Customer Activated Enterprise » menée par l'Industry of Business Value d'IBM en 2013 montre que 82 % des responsables d'entreprise tiennent compte aujourd'hui des attentes de leurs clients dans la définition de leurs produits. Ils seront 90 % dans 3 à 5 ans. Plus étonnant, 60 % d'entre eux pensent qu'ils devront tenir compte de l'avis de leurs clients dans la définition de leur stratégie d'ici 3 ans. C'est l'avènement du consomm'acteur.

LE CONSOM'ACTEUR

On parle beaucoup aujourd'hui du consomm'acteur¹. Par opposition au consommateur passif, cible du marketing traditionnel, ce nouveau consommateur, connecté, informé entend bien rester maître de sa consommation. Il n'est plus la victime d'un discours publicitaire monodirectionnel de la marque vers le client, mais un consommateur responsable qui réfléchit sa consommation. Il s'émancipe des circuits commerciaux qui ont été pensés pour lui.

Buycott et boycott

À l'origine, la consom'action était très liée au développement durable. À l'instar du citoyen qui infléchit les politiques par son bulletin de vote, le consomm'acteur utilise son pouvoir d'achat pour infléchir les politiques commerciales de certaines marques. Il « vote » avec son caddie, en choisissant à qui il donne son argent. En achetant bio, en favorisant le commerce équitable et le commerce de proximité, le consomm'acteur exprime ses convictions sociétales. Cela peut aller jusqu'au

1. *À nouveaux consommateurs, nouveau marketing*, Philippe Jourdan, François Laurent et Jean-Claude Pacitto, Dunod, 2011.