



**Research Network on Innovation
Réseau de Recherche sur l'Innovation**

**WORKING PAPERS
DOCUMENTS DE TRAVAIL**

N°48 / 2015

BIG DATA, OPEN DATA ET VALORISATION DES DONNEES

Jean-Louis MONINO

BIG DATA, OPEN DATA ET VALORISATION DES DONNEES

BIG DATA, OPEN DATA AND VALORIZATION OF DATA

Jean-Louis MONINO¹

Résumé : Si l'information est au cœur de l'intelligence économique, les données en sont des éléments indispensables pour construire les savoirs afin de prendre une bonne décision qui dans l'instant peut paraître comme optimale dans les champs des possibles de ses connaissances... C'est l'utilisation des données qui donne le pouvoir. Les entreprises, de plus en plus conscientes de l'importance des données et de l'information, se pressent à réfléchir sur la façon de les « gérer », de les enrichir et d'en tirer profit. L'objectif de cet article est de montrer que les enjeux de l'ère de « la révolution des données » se focalisent autour de l'intégration et de la valorisation des données dans l'entreprise. Il s'inscrit dans une thématique reliée à la montée en puissance de l'économie de l'immatériel mobilisant les connaissances et les savoirs, et met en évidence l'importance des données.

Abstract: If information is at the heart of economic intelligence, data is the essential element to develop knowledge to make a good decision which in the moment may seem to be optimal within the reach of knowledge ... It is the use of data that empowers decision-making. Being increasingly aware of the importance of data and information, companies are pressing to rethink the way to "manage", to enrich and to benefit from them. The objective of this paper is to show that the challenges of the era of "data revolution" focus on the integration and utilization of data in the firm. It is linked the rise of the intangible economy that mobilizes knowledge and highlights the importance of data.

BIG DATA, OPEN DATA ET VALORISATION DES DONNEES

BIG DATA, OPEN DATA AND VALORIZATION OF DATA

Jean-Louis MONINO

TABLE

INTRODUCTION	4
1. LA « DATA »	5
2. LE BIG DATA : ENJEUX ET DÉFIS POUR LE MILIEU DÉCISIONNEL	11
2.1. Définition et dimensions du Big Data	11
2.2. Quelques exemples de sources du « Big Data »	12
2.3. Pour une « éthique des données »	13
2.4. Utilité du « Big Data » pour les entreprises et autres organisations	13
CONCLUSION	15
BIBLIOGRAPHIE	16

INTRODUCTION

Si l'information est au cœur de l'intelligence économique, les données en sont des éléments indispensables pour construire les savoirs afin de prendre une bonne décision qui dans l'instant peut paraître comme optimale dans les champs des possibles de ses connaissances... C'est l'utilisation des données qui donne le pouvoir. Les entreprises, de plus en plus conscientes de l'importance des données et de l'information, se pressent à réfléchir sur la façon de les « gérer », de les enrichir et d'en tirer profit.

Ainsi, l'explosion d'une quantité phénoménale de données et le besoin de les analyser, remet en avant le modèle hiérarchique bien connu : « Données, Information et Connaissance ». Ce modèle est souvent exploité dans la littérature relative à la gestion des informations et des connaissances. Plusieurs études affirment que la première apparition de la hiérarchie des connaissances se trouve dans le poème de T.S. Elliot «The Rock» en 1934. Dans la littérature récente, plusieurs auteurs citent la publication « From data to wisdom » de R.L. Ackoff parue en 1989 comme étant une source de la hiérarchie des connaissances. En effet, ce modèle hiérarchique met en évidence trois mots : « Données », « Information » et « Connaissance ». La relation entre ces trois mots peut être représentée sous la forme ci-dessus où les connaissances occupent la place la plus haute pour souligner le fait que de nombreuses données sont nécessaires à l'acquisition d'une connaissance.

L'arrivée des « Big data² » est en train de bouleverser les pratiques classiques du management et invite l'organisation à adopter de nouvelles visions lui permettant d'avoir « l'information stratégique » susceptible de générer plus de valeur. En d'autres termes, il faut faire passer le message que l'information en elle-même n'est pas le pouvoir: c'est utiliser l'information qui donne le pouvoir, et plus on donne et on échange de l'information, plus on en reçoit (Martinet. B et Marti. Y.M, 2001).

Il faut alors s'adapter et tenter de nouvelles approches, de nouvelles méthodes, de nouveaux savoirs et de nouvelles manières de travailler, ce qui entraîne de nouvelles propriétés et de nouveaux enjeux puisqu'un référencement logique doit être créé et mis en œuvre. La question n'est donc plus d'identifier quelles données stocker, mais, qu'est-ce qu'on peut faire avec ces données ?

Mais, atteindre le plein potentiel des données dépend de la façon dont elles sont présentées, elles doivent être utilisées et réutilisées de multiples façons sans diminuer leur valeur. Cela signifie rendre les données disponibles sous la bonne forme et au bon moment à toute partie intéressée désirent les exploiter et y apporter de la valeur ajoutée. Générer de la valeur aux différents stades de la chaîne de valeur des données sera au centre du croisement de l'ouverture et des quantités volumineuses des données. De plus, l'un des principaux avantages liés aux données ouvertes est qu'elles favorisent le développement d'une culture axée sur le partage des informations et sur la collaboration intersectorielle (GTDO, 2011). L'augmentation des volumes de données, la vitesse de traitement et la diversité de ces dernières ..., posent de nouveaux défis aux entreprises et affectent leurs processus de prise de décisions.

² Le « Big Data » est un concept récent qui fait son apparition en 2012 les entreprises sont confrontées à des volumes de données (data) à traiter de plus en plus considérables et présentant un fort enjeu commercial. Ce terme est traduit par « Données massives »

L'objectif de ce document est de montrer que les enjeux de l'ère de « la révolution des données » se focalisent autour de l'intégration et de la valorisation des données dans l'entreprise. Il s'inscrit dans une thématique reliée à la montée en puissance de l'économie de l'immatériel mobilisant les connaissances et les savoirs, et met en évidence l'importance des données. L'apport en informations permet de générer des connaissances utiles à la prise de décision, dans le cadre des différentes activités propres au développement de l'entreprise. Pour cela, il faut bien clarifier qu'est ce plus précisément qu'une « Donnée » pour pouvoir comprendre et formuler l'expression les enjeux du « Big Data ».

1. LA « DATA »

Les données sont des éléments incontournables de notre société numérique. Ces données pourront être organisées et transformées en informations qui seront ensuite transformées en connaissance par une opération cognitive ou intellectuelle. Cela démontre aujourd'hui que la richesse la plus importante n'est plus matérielle, ni physique. Elle est désormais fondée sur l'immatériel et conduite par le savoir (Pinte, 2006). Le patrimoine immatériel a pris le dessus sur le patrimoine matériel (Monino, 2005, 2015). Nous en produisons à tous les instants et de toutes les formes. Les données lorsqu'elles sont contextualisées et matérialisées permettent de formaliser des échanges entre individus ou entre machines. Ces données peuvent être des documents papiers ou numériques, images, vidéos, bandes sonores, microfilms, discussions informelles d'individus à individus,...

Les dernières technologies comme les ; Smart, 4G, Cloud ..., et l'avènement d'Internet, les logiciels de traitement et d'exploration des données...ont permis l'exploitation des données. Aujourd'hui les données nous proviennent de toute part : des capteurs de géolocalisation, des Smartphone, des réseaux sociaux...où l'on partage fichiers, vidéos, photos,...des transactions d'achats sur Internet des clients, des transactions bancaires par l'intermédiaire des cartes bancaires...

En France, sur 65 millions de personnes, 83% sont internautes. 42% sont inscrits sur Facebook, soit 28 millions de membres. Plus de 72 millions de téléphones sont activés et les français passent en moyenne plus de 4h devant leur ordinateur à naviguer sur Internet chaque jour. Les mobinautes français passent 58 minutes et plus 68% de la population est inscrite sur les réseaux sociaux. Les français passent plus 1h30 chaque jour à la consultation des réseaux sociaux.

Les données ont toujours eu une valeur stratégique. Selon Taylor (1980), la valeur de l'information commence par une donnée qui acquiert de la valeur tout au long de son évolution, pour atteindre l'objectif à exprimer une action de prise de décision. Mais avec l'ampleur des données disponibles aujourd'hui et les capacités à les traiter, elles sont devenues une nouvelle catégorie d'actifs.

L'accélération de l'ère numérique et de l'avancée technologique ont transformé le manque en information en surinformation. Le développement de ces masses des données et leurs accès représente ce qui est appelé « Big Data ». Cette masse de données immatérielles qui arrive en flot continu, son traitement posent des problèmes en particulier dans l'extraction de connaissances. Les défis autour des données sont nombreux : comment les collecter, les stocker, les analyser, les diffuser et les protéger ?

Ainsi, de nouvelles méthodes d'extraction automatique de données sont mises en place comme par exemple le « Data Mining³ » ou le « Text mining⁴ » (Monino, Boya⁵, 2011) qui sous-tendent de profonds bouleversements, qui touchent à la fois l'économie, le marketing, mais également à la recherche. La quantité de données va s'accroître fortement avec l'arrivée sur le marché des objets connectés dont nous allons nous servir. Beaucoup d'objets de la vie courante sont déjà connectés, comme la voiture, la télévision, certains appareils ménager, ... Ils sont ou seront dotés d'une puce chargée de collecter et transmettre des données à leurs utilisateurs par le biais d'un ordinateur, d'une tablette, d'un smartphone... Le plus important réside dans le fait que ces objets pourront également échanger entre eux ! Nous pourrions ainsi, manipuler à distance les équipements de notre domicile, de notre véhicule, ... en nous connectant soit à notre domicile, soit hors de notre domicile en utilisant les Smartphones ou tout autre matériel... C'est ce que l'on appelle « L'Internet des objets ».

En effet, de plus en plus d'objets connectés mesurent des quantités gigantesques de données à des fins de prévision, d'analyse ou de prédiction. Les objets connectés vont en effet récolter beaucoup plus de données sur les individus, le nombre de data utilisables et monétisables va donc augmenter.

« Big Data », « Open Data », « Data Mining », « Data Management », ... tous ces termes ont en commun « la donnée » ou plus tôt les termes liés à la famille « Data ». Elles sont partout, nous en produisons tous les jours dans notre vie quotidienne, et ceux avec l'utilisation croissante des réseaux sociaux, de l'Internet. Ces données sont véhiculées par les smartphones, les tablettes, et sont accessibles de façon plus ou moins simple par les entreprises ou les particuliers. Ces termes renvoient la façon dont les entreprises sont amenées à gérer les données dont elles disposent ou peuvent disposer sur leurs clients, prospects, souscripteurs, fournisseurs, ... Bref sur toutes les parties avec lesquelles elles interagissent.

- L'information : est un ensemble de données contextualisé et matérialisé par un support afin de formaliser des échanges entre individus ou entre machines. Ces supports peuvent être des documents papiers ou numériques, images, vidéos, bandes sonores, microfilms, discussions informelles d'individus à individus, ...

- Le flux Information : est une représentation conceptuelle de l'ensemble des mouvements d'informations existant pour une organisation, quelle que soit leurs formats, natures, valeur, enjeux et leurs origines (internes ou externes). Le flux Information est représenté par 7 familles d'informations : Engageante, porteuse d'éléments de preuve pouvant être utilisée au cours d'un litige - Patrimoine, représente une valeur pour l'organisation qui doit rester accessible dans le temps - Structurante, permet de classer les autres informations, elle se formalise par des référentiels de données - Sensible, porteuse de risques qui peut mettre tout ou partie d'activités en péril - Personnelle, porte des éléments qui permettent d'identifier une personne physique ou de la rendre identifiable - Contributive, porte un certain nombre d'éléments permettant d'alimenter ou de constituer les 5 familles précédentes - Bruyante, n'a d'autre importance ou incidence en soi-même que de permettre de créer de l'information

³ Le « Data Mining » ou « fouille de données » désigne l'ensemble des techniques et méthodes dans les domaines des statistiques, des mathématiques et de l'informatique qui permettent de sortir d'un grand volume de données, des connaissances précises sur des éléments inconnus auparavant.

⁴ Le « Text Mining » ou « Texte Mining » est une technique permettant d'automatiser le traitement de gros volumes de contenus texte pour en extraire les principales tendances et répertorier de manière statistique les différents sujets évoqués

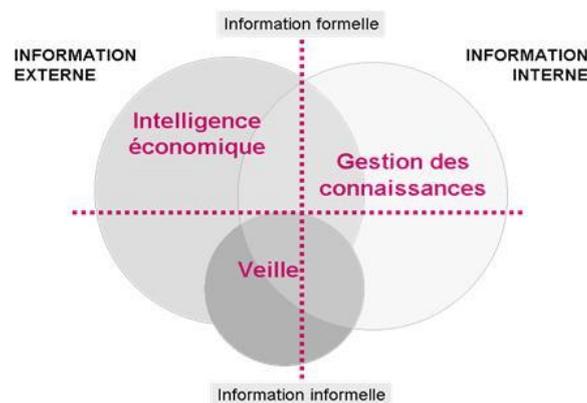
⁵ « La coloration de l'information dans l'efficiences semi forte » *Innovations* - 2011/3 (n°36) Pages : 208

contributive- Les fondamentaux : sont des documents ou ensembles d'informations structurantes pour les usages et comportements d'usages. Les fondamentaux portent des règles, des guides, des indicateurs ainsi que tous éléments structurants aidant à améliorer la maîtrise du flux d'information. Exemple : politiques, directives, référentiels de données, descriptif de processus,...

- La gouvernance de l'Information : est l'ensemble des processus, des fondamentaux, et des structures qui permettent à une organisation de mieux maîtriser et valoriser son flux Information. L'ensemble des actions de gouvernance de l'information entre dans le cadre d'une démarche, qui peut se décliner sur des outils en vue de la réalisation de ses objectifs.

- Les outils spécifiques de gouvernance de l'information : tous les systèmes techniques qui favorisent la création et l'entretien des fondamentaux et qui permettent à l'organisation de maîtriser tout ou partie des familles d'information. Les principaux sont les outils de Records Management, Master Data Management, Gouvernance Risque et Conformité, Systèmes d'Archivage Electronique, Les outils qui sont impactés par cette démarche, et qui diffuse cette gouvernance sont ceux qui permettent la manipulation de l'information : Systèmes collaboratif, Réseaux Sociaux d'Entreprise, GED dites "lourdes" d'affaire, industrielle, Portails, Messagerie, ...

Les deux types d'outils sont indissociables dans le cadre d'une démarche globale de gouvernance de l'information. Elles sont dès lors confrontées à la gestion de ces traitements de données, leur stockage et leur sauvegarde, la sécurité et la protection de ces données. Selon Taylor (1980) la valeur de l'information commence par une donnée, qui acquiert de la valeur tout au long de son évolution, pour atteindre l'objectif à exprimer une action pour prendre une décision. Elle est considérée comme un flux de données interprétée, alors que la connaissance est créée et organisée à partir de ce flux d'information. Cette exploitation donne encore plus de valeur à l'information, qui doit être protégée. Dans ce cadre, nous pouvons proposer une approche de la connaissance comme étant une modélisation comprise de l'information.



Source : J.L Monino, 2005

Ce phénomène intéresse désormais les décideurs opérationnels (managers marketing, financiers...) dès qu'il s'agit d'analyser, en temps réel, l'immense potentiel de données détenues par les entreprises. Pour relever le défi du « Big Data », des mesures devraient être prises incluant tous les outils permettant le traitement des données; de façon plus restrictive, ainsi que tous les acteurs qui analysent ces données. Cela, ne sera possible que par une prise de conscience des gains offerts par « la valorisation des données ». Les données des bases

organisées, réorganisées, traitées par des méthodes statistiques, ou des modélisations économétriques, deviennent des connaissances.

Pour une entreprise il est essentiel de disposer de données, de plus en plus, nombreuses sur l'environnement dans lequel elle opère ou va opérer. On ne travaillera plus sur des classes de comportements, mais sur une analyse individuelle. On peut comprendre aisément que cette révolution entraîne la création d'entreprises dites « start-up » qui auraient pour but de traiter automatiquement cette masse de données qui forment ce que l'on appelle le « Big Data ». On est certainement devant une des composantes de ce que certains appellent la nouvelle révolution industrielle. L'Internet, le numérique, les objets connectés ont ouvert de nouveaux horizons dans une multitude de domaines.

A titre d'exemple, l'accès aux données permet d'enrichir les analyses quantitatives et qualitatives. On peut analyser les contacts clients avec les données recueillies par un « call center », on peut également offrir ce genre de produit en nombre limité comme le fait e-prospects. En effet, Le web offre des possibilités d'informations immenses, les utilisateurs d'Internet demandent une information claire, précise et immédiate. Or, cette réactivité de la demande d'information est un point névralgique de ce type de demande. En effet, Il faut souvent attendre pour avoir un résultat plus que décevant. La start-up e-prospects et le laboratoire TRIS⁶ de l'Université de Montpellier développent un service innovant basé sur une plate-forme communautaire permettant d'acheter et de vendre des contacts et des rendez-vous multi produits et multi pays. La start-up e-prospects développe⁷ des outils de communication permet un cheminement rapide de l'information : de la demande de contact à la réponse par le professionnel.

La start-up e-prospects permet aux entreprises commerciales de trouver sur une même plate-forme les contacts et les rendez-vous alimentés par de nombreux fournisseurs, experts du marketing direct. Les points forts de la société sont :

- De permettre à des marchés, qui ne pouvaient se côtoyer à cause de leur taille ou de leur surface financière, de pouvoir se rencontrer et de développer leur CA ;
- D'informer un acheteur sur la qualité des fournisseurs (chaque acheteur note la prestation du fournisseur sur différents critères) ;
- De mettre en avant les fournisseurs bien notés qui seront recherchés et verront leurs prix de vente progresser grâce à un bonus. A l'inverse, un fournisseur mal noté sortira naturellement du système à cause d'un malus effectué sur son prix de vente. E-prospects met à disposition de ses clients un service novateur sur un besoin marché clairement identifié au prix le plus juste grâce à plusieurs algorithmes, différents systèmes de ventes et à de nombreux moyens de communications. C'est un gain de temps non négligeable que de trouver l'ensemble des acteurs du marché, centralisé dans un même espace.

Ainsi, en explorant le contenu des emails, des appels vocaux, et mixer ces informations avec la navigation sur le site Internet, ou encore, d'étudier les messages échangés sur les réseaux sociaux (facebook, twitter, linkedin, ...) pour identifier les nouveautés intéressantes ou savoir de quels produits parle-t-on le plus?... Encore faut-il, pour que le plein potentiel de valeur que représente les données soit atteint, qu'elles soient facilement accessibles par toutes les parties intéressées, sans obstacles supplémentaires et à des coûts économiquement supportables. Si les données sont ouvertes aux utilisateurs (Monino, 2014) d'autres entreprises de traitement spécifiques des données pourront être créées. Ces entreprises permettront de les traiter selon

⁶ Le laboratoire TRIS : Traitement et Recherche de l'Information et de la Statistique

⁷ Mise en place de R&D par E-prospects avec l'aide des chercheurs du laboratoire TRIS

les besoins des usagers sans qu'ils aient besoin de se livrer eux-mêmes à la construction de modèles et d'équations.

L'ouverture des données, outre sa potentialité en matière économique, et de création d'activités nouvelles, ressortit aussi à un choix philosophique, ou éthique. Elles chiffrent des comportements humains collectifs, et appartiennent donc aussi à ceux dont on a mesuré ces comportements. La culture de ce phénomène se fonde sur la disponibilité des données vers une orientation de communication,

Pour se limiter aux avantages économiques de l'« Open Data », une étude réalisée pour le compte de la Commission Européenne estimait le marché total de l'information du secteur public en 2008 dans l'UE à 28 milliards d'euros. Selon cette étude, les avantages économiques globaux liés à une plus grande ouverture des données du secteur public représenteraient environ 40 milliards d'euros par an pour l'UE. Pour l'ensemble de l'économie de l'UE, le total des gains économiques directs et indirects découlant de l'utilisation des ISP (Internet Service Provider) et des applications fondées sur ces données serait de l'ordre de 140 milliards d'euros par an.

L'ère de l'Internet a fait exploser la recherche d'informations et les entreprises sont submergées par le flot de données qui en découle par une simple navigation sur Internet. Autrement dit, elles sont contraintes à l'acquisition des informations pertinentes pour développer leurs stratégies à haute valeur ajoutée afin de maîtriser le changement incessant de l'environnement. Désormais, la conduite des stratégies industrielles repose largement sur la capacité de l'entreprise à accéder aux informations stratégiques pour mieux agir dans son environnement. Du fait que, cette information peut être à la base d'une connaissance nouvelle (pyramide de connaissance).

Le processus qui permet la collecte, le traitement, et l'interprétation de l'information, représente une pratique qui ne se limite pas à définir des idées, mais surtout à les concrétiser pour assurer une meilleure production de connaissances qui génèrent des innovations. C'est l'intelligence économique qui permet à chaque entreprise d'optimiser son offre, quantitativement, qualitativement, et d'optimiser aussi ses technologies de production.

Par ailleurs, l'action du lobbying⁸⁹ rend un grand service à l'entreprise grâce aux flux de retour d'informations. C'est la connaissance et la compréhension des processus de décision qui permettent aux lobbyistes d'informer les décideurs au moment utile. Soit au niveau opérationnel ou au niveau stratégique. Ce qui permet d'évaluer les résultats obtenus, et par conséquent agir efficacement en mettant en place de nouvelles techniques de livraison, de nouvelles procédures de production, de nouvelles dispositions de surveillance du marché/clients/fournisseurs, etc....

En plus de l'avènement des TIC et la vitesse de production et de diffusion et les capacités de traitement de ces données. Un autre élément devenu important ces dernières années : le temps. Cet élément temporel implique une notion de vitesse de circulation de l'information. Ce qui appelle à repenser la stratégie des entreprises, au-delà des difficultés posées par le traitement

⁸ Un lobby est une structure organisée pour représenter et défendre les intérêts d'un groupe donné en exerçant des pressions ou influences sur des personnes ou institutions détentrices de pouvoir. ...

⁹ Un exemple de lobbying est le travail que l'association « UFC a effectué Que choisir » qui a fourni des documents aux parlementaires pour les sensibiliser aux risques sanitaire que représente l'obésité chez les collégiens et les lycéens. (Voir : <http://www.lobbyiste.org/>)

des grands volumes de données. La valeur d'une donnée augmente au cours du temps et dépend des multiples utilisations qui en sont faites.

Dans ce cadre, les entreprises doivent disposer d'une capacité d'absorption de l'ensemble des données disponibles, leurs permettant d'assimiler et de reproduire les connaissances. Cette capacité suppose l'existence de compétences spécifiques permettant l'utilisation de ces savoirs et donc, la formation des « Data Scientist¹⁰ » et plus que nécessaire pour pouvoir identifier les démarches utiles d'ouverture ou d'exploitation interne des données et quantifier les bénéfices en termes d'innovation et de compétitivité. Par ce que, le « Big Data » n'est qu'un élément d'un nouvel ensemble d'outils et de techniques appelé « Data Science ».

Le « Data Scientiste » a pour mission d'extraire de la connaissance à partir des données de l'entreprise. Il sera appelé à exercer des fonctions stratégiques au sein des entreprises. Pour cela, il doit maîtriser les outils nécessaires. Il doit aussi bien faire preuve de plus de pédagogie et accroître encore sa maîtrise du « Data Mining », car le volume de données exige l'augmentation de panel de techniques à maîtriser.

Face à cette multiplication des données, les entreprises sont amenées à mobiliser des techniques sophistiquées de traitement de données. De fait, la maîtrise des techniques de traitement devient aujourd'hui un réel enjeu stratégique et utile pour la différenciation concurrentielle pour les entreprises (Bughin et al., 2011). Le traitement de ces masses de données joue un rôle primordial dans la société de demain, car il trouve des applications dans des domaines aussi variés que les sciences, le marketing, les services client, le développement durable, les transports, la santé, et même l'éducation.

Le « Big Data » regroupe à la fois le traitement de ces grandes masses de données, leur collecte, leur stockage jusqu'à leur visualisation et leur analyse. Ces données deviennent ainsi le carburant de l'économie numérique. Elles constituent la matière première indispensable à l'activité reine du nouveau siècle : l'intelligence des données « Data Intelligence ». Les enjeux du « Big Data » se focalisent autour de l'intégration et la valorisation des données dans l'entreprise. Il traite la problématique de la valorisation des données dans un contexte de forte concurrence.

Plus précisément, il représente un sujet de recherche qui implique plusieurs champs de recherche (Big data, open data, traitement de données, innovation, intelligence économique...). Certes, cette pluridisciplinarité permet d'apporter un enrichissement considérable aux études et recherches sur la valorisation des entrepôts des données en sa globalité.

Le monde est devenu numérique et les avancées technologiques ont démultiplié les circuits d'accès aux données, à leurs traitements, et à leurs diffusions... Les nouvelles technologies atteignent aujourd'hui une certaine maturité. Les données sont accessibles à tous et partout dans le monde, le nombre d'internautes en 2014 était de 2,9 milliards soit 41% de la population mondiale. Le besoin de connaissance se fait sentir afin d'appréhender ces masses de données. Il faut renseigner, former, informer, en masse. L'essor des technologies connexes, comme l'avènement de l'Internet, des réseaux sociaux, des « Cloud Computing¹¹ », a accru

¹⁰ On peut écrire également : « Data Scientiste » ou en anglais « Data Scientist »

¹¹ Le « Cloud Computing » ou « Informatique en Nuage » ou « Usines Numériques » est une infrastructure dans laquelle la puissance de calcul et le stockage sont gérés par des serveurs distants auxquels les usagers se connectent via une liaison Internet sécurisée.

les volumes disponibles de données (Data). Actuellement, chaque individu crée, consomme, utilise de l'information numérique. Plus de 3,4 millions d'emails envoyés dans le monde chaque seconde, soit 107 000 milliards par an avec 14 600 mails par an et par personne, mais plus de 70% sont des Spam¹². Des milliards de contenus sont partagés sur les réseaux sociaux comme par exemple sur Facebook avec plus de 2,46 millions de contenus partagés toute les minutes,,...Le temps passé sur Internet est de plus 4,8 heures par jour via un ordinateur, et de 2,1 heures via un mobile. Cette nouvelle manne immatérielle qu'est « la Donnée » est produite en temps réel, et arrive en flots continu et provient d'une multitude sources qui sont généralement hétérogènes. Cette l'accumulation de données en tous genres (audio, vidéo, fichiers, photos,...), génère de nouvelles activités qui ont pour objet d'analyser de ces énormes masses de données. Il faut alors s'adapter et tenter d'apporter de nouvelles approches, de nouvelles méthodes, de nouveaux savoirs et de nouvelles manières de travailler ce qui lui conféré de nouvelles propriétés, et de nouveaux enjeux puisqu'elle est référencée de façon disparate. Au niveau des entreprises, cette masse de données est difficile à gérer, et l'interprétation de ce large volume constitue principalement un défi. Cela, impacte les entreprises qui sont amenées à « manipuler »des multitudes de données nécessitant une infrastructure spécifique pour leur création, leur stockage, leur traitement, leur analyse et leur récupération. Le plus gros challenge réside dans « la valorisation des données » disponibles, en quantité, diversité et vitesse d'accès.

2. LE BIG DATA : ENJEUX ET DÉFIS POUR LE MILIEU DÉCISIONNEL

2.1. Définition et dimensions du Big Data

Actuellement dans les journaux scientifiques ou professionnels on trouve le terme de « Big Data », il est généralement relié à plusieurs méthodes, comme la « Data Science », la « Data Visualisation », le « Data Mining », le « Texte Mining », pour ne citer que quelques-unes des méthodes qui permettent d'intégrer des jeux de données de plus en plus volumineux et dont les flux de données nous arrivent très rapidement. L'intégration de la donnée est devenu incontournable dans les processus métier et organisationnels.

Cette masse de données immatérielles qui arrive en flot continu et son traitement posent des problèmes en particulier dans l'extraction de connaissances. Chaque minute, plus de 200 millions de mails sont envoyés dans le monde et plus de 2,5 milliards de contenus (textes, photos, vidéos, images,...) sont mis sur Facebook, Google enregistre 2 millions de requêtes différentes sur son moteur de recherche ... Des chiffres importants qui ne vont pas cesser d'augmenter, sachant également que les robots eux aussi sont producteurs de données et qu'ils représentent plus de 60% de l'activité sur Internet. La quantité de données va s'accroître fortement avec l'arrivée sur le marché des centaines d'objets connectés dont nous allons nous servir. Beaucoup d'objets de la vie courante sont déjà connectés, comme la voiture, la télévision, certains appareils ménagers,...Ils sont ou seront dotés d'une puce chargée de collecter et transmettre des données à leurs utilisateurs par le biais d'un ordinateur, d'une tablette, d'un smartphone...Le plus important réside dans le fait que ces objets pourront également échanger entre eux ! Nous pourrons ainsi, « manipuler » à distance les équipements de notre domicile, de notre véhicule, de notre entreprise,...

¹²Courrier électronique non sollicité envoyé en grand nombre à des boîtes aux lettres électroniques ou à des forums, dans un but publicitaire ou commercial.

Tout le monde est producteur de données, qu'on soit simple consommateur, ou bien entrepreneur, responsable d'entreprise, formateur,... Tout le monde est intéressé par la donnée ! « Les données sont le nouveau pétrole » (Rotella, 2012). Elles constituent la matière première indispensable à l'activité reine du nouveau siècle : « l'intelligence », base de l'information pour ne pas transposer simplement le terme anglo-saxon. (Data Intelligence), sous ses diverses formes, notamment l'intelligence économique.

Le « Big Data » se définit par la croissance des données disponibles, en quantité, diversité et vitesse d'accès en plus de la valeur :

- **Volume** : Un volume exceptionnel de données ;
- **Vitesse** : Une vitesse de traitement en temps quasi-réel ou réel ;
- **Variété** : Les données sont variées, prenant différentes formes. Vidéo, audio, texte, image, lien, tweet, like,... ;
- **Véracité** : Pertinence des données. Fiabilité des données.

Quelques chercheurs comme Mc Nulty (2014) ajoutent 3 autres V à savoir : Variabilité, Véracité et Visualisation. Le premier V se réfère aux données dont la signification est en constante évolution. Le deuxième indique le résultat de l'usage des données, c'est-à-dire que bien qu'il y ait un large consensus sur la valeur potentielle de « Big Data », les données sont pratiquement sans valeur si elles ne sont pas exactes. Cela est particulièrement vrai dans les programmes qui impliquent la prise de décisions automatisées, ou alimentant les données dans un algorithme d'apprentissage machine sans surveillance. Le dernier V, qui est l'un des défis de « Big data », se trouve dans la façon de présenter les résultats de traitement des données (Information) qui rend les conclusions plus claires.

Les données des bases organisées, réorganisées, semi-structurées, traitées par les méthodes statistiques, notamment économétriques, deviennent des connaissances. Encore faut-il, pour le plein potentiel de valeur des données soit atteint, qu'elles soient facilement accessibles par toute partie intéressée, sans obstacles supplémentaires et à des coûts économiquement supportables.

Avant de faire l'historique de cette nouvelle activité humaine, nous allons donner quelques exemples des sources des données du « Big Data » :

- *Perry Rotella (2012), is Data a new Oil ?- <http://www.forbes.com>*
- *Gartner (2011), Big Data - <http://www.gartner.com>*

2.2. Quelques exemples de sources du « Big Data »

Tout d'abord, les données officielles (publiques) constituent un gigantesque vivier de données : statistiques officielles, données économiques, données démographiques et géographiques, données financières, données de consommation énergétique, données sur le trafic routier,... Des organismes comme les banques centrales, les instituts de statistique notamment l'INSEE, France), les départements ministériels,... délivrent une masse croissante de données relatives à leur domaine. De nombreuses institutions produisent des données sociologiques et des résultats de sondages. Ceci est le cas dans tous les pays susceptibles de financer de telles institutions. La rigueur budgétaire général s'applique malheureusement à toutes ces institutions et, certaines sont amenées à rendre payantes des données qui étaient jusque-là gratuites. La question de l'ouverture de ces données ou encore « Open Data » est actuellement sujet de débat.

Ces données ne sont pas complètement brutes, mais déjà traitées simplement. Ainsi l'INSEE fournit les chiffres bruts, mais aussi leurs accroissements relatifs, en valeur et volume, ou les valeurs par tête. A leur tour ces données peuvent être utilisées comme variables explicatives de « modèles » dont se servent divers utilisateurs pour faire des prévisions, ou, plus exactement, des corrélations sur le passé, qu'ils peuvent ensuite projeter sur l'avenir. Encore faut-il que ces utilisateurs bâtissent eux-mêmes ces modèles économétriques, avant de les utiliser en allant chercher les données dans les divers chapitres de l'INSEE (ou l'OCDE ou l'INED...).

Dans le même champ, les millions d'articles stockés sur les entreprises, les hommes politiques, les secteurs d'activité,...représentent des gisements importants de données, sur les mêmes popularités qualitatives, les « images » des uns et des autres, la coloration (Monino, Boya – Revue d'économie et de management de l'innovation), ce qu'on appelait autrefois la sémantique quantitative.

On peut ajouter aux sources précédentes, toutes les données semi-structurées ou non structurées, issues d'informations du web, des sites de médias sociaux, d'images numériques et de vidéos publiées en ligne, d'enregistrements transactionnels d'achats en ligne et de signaux GPS de téléphones mobiles, des tickets de caisse, des traces de navigation sur internet, des réclamations de clients,...

2.3. Pour une « éthique des données »

La « mine » de données est aussi alimentée, et de plus en plus, par une autre catégorie de données bien particulières, qui sont collectées, souvent à notre insu, par de multiples organismes, les données personnelles. Nos civilités, âge, tranches de revenus, goûts, achats plus ou moins fréquents, motifs de ces achats....constituent une part croissante du « Big Data », mais une part majoritairement cachée, dont certains utilisateurs seuls se réservent l'usage pour leur plus grand profit. Si elles sont collectées et stockées très souvent, et même de plus en plus, à notre insu, c'est qu'elles ont une grande valeur commerciale pour ceux qui se les approprient. Ils leur permettent de cibler de plus en plus précisément leur offre sur des particuliers en fonction de ces données.

Le « Big Data » va s'accroître terriblement avec les centaines d'objets connectés (Smartphones, tablettes, ordinateurs, GPS, ...), permettant la manipulation à distance des équipements professionnels ou domestiques. Un exemple qui a fait récemment la « une » de la presse, l'arrivée en France de la firme américaine « Netflix » qui vend de la « vidéo à la demande » sous forme de films et séries, concurrençant ainsi à bas prix, les opérateurs classiques (TV d'Orange). Comment décide-t-il des vidéos proposées, allant jusqu'à faire des offres personnalisées? Par une analyse progressivement enrichie de chacun de nos achats successifs, qui finissent par définir assez précisément nos goûts. Dans ce cas, bien sûr, nous sommes libres de ne pas « acheter » et de ne pas fournir ainsi à Netflix les données dont il va se servir. Mais l'exemple illustre ce qu'on peut réaliser – éventuellement d'ailleurs avec l'aide d'entreprises spécialisées, comme on le verra plus loin- en stockant progressivement les données personnelles. Il faut, donc, à tous prix que se bâtisse et se fixe dans des textes contraignants, une éthique des données » ou « de la donnée », qui sanctionne durement l'usage sans notre autorisation des données personnelles.

2.4. Utilité du « Big Data » pour les entreprises et autres organisations

Mais revenons aux données « produites » par des organismes officiels, et à l'utilité des bases de données collectées légalement. Il existe d'abord une activité d'évidence liée au stock

massif de données très diverses. La puissance informatique nécessaire pour celui qui voudrait stocker par devers lui la totalité des données dont il a besoin et plus encore de celles dont il pourra avoir besoin au vu de l'évolution de ses travaux intellectuels divers, et dont il ignore évidemment la nature et le « poids » est bien trop grande pour qu'elle soit envisageable au sein de son entité, entreprise ou administration. Cette incapacité des utilisateurs a donné naissance à une activité en plein développement, le stockage externe, que les anglais ont joliment baptisé « cloud computing ». Le simple stockage de données est donc en soi une activité nouvelle dans laquelle se sont lancées nombre d'entreprises informatiques. Surtout, le « Big Data » lui-même ouvre des perspectives nouvelles de mode de gestion tant aux entreprises qu'aux entités publiques qu'elle soit de gouvernance ou d'administration. Les données sont en effet la base de l'« intelligence » au sens le plus général de ce mot.

Pour une entreprise il est essentiel de disposer de données de plus en plus nombreuses sur l'environnement dans lequel elle opère et va opérer. C'est l'intelligence économique. Celle-ci, pour chaque entreprise lui permet d'optimiser son offre, quantitativement et qualitativement, et d'optimiser aussi ses technologies de production. Il est donc très important pour elle de transformer les données auxquelles elle a accès en « information » par divers traitement. Nous verrons plus loin que ces traitements eux-mêmes, pratiqués jusqu'à présent par les sociétés de conseil, comme base de leurs « rapports d'études », sont devenus des opportunités de création d'entreprises en ligne remplaçant en grande partie les entreprises de conseil.

De même, une collectivité, un territoire, un Etat, ont un grand besoin d'intelligence économique, mais également de ce qu'on pourrait appeler « Intelligence Politique » ou « Intelligence Sanitaire »... Le « Big Data » apporte à ces entités une masse énorme d'informations qui, cumulées et traitées, doivent évidemment améliorer grandement sa gestion. Répétons ici, par exemple, que la libre exploitation des données de santé, codées pour protéger les données personnelles des patients et, moins légitimement, des médecins, permettrait, en France, des économies massives pour l'assurance maladie certainement de l'ordre de 2% du PIB. Elle faciliterait grandement des comparaisons d'expérimentations sociales, comme le demande Martin Hirsh pour la Sécurité Sociale des Indépendants RSI. Rappelons, pour mémoire, que le simple croisement de données libérées permettrait de traquer au mieux les fraudes.

L'entreprise intervient dans un environnement où les données circulent rapidement et en grande quantité. A titre d'exemple, « TomTom », le leader des produits et services de localisation et de navigation, reçoit quotidiennement des milliards d'enregistrements issus des GPS bidirectionnels de ses clients. Ou encore, la base de données de Wal-Mart (Leader de distribution aux Etats-Unis) est passée de 1.000 téraoctets (soit 1.000.000 gigaoctets de données) en 1999 à plus de 2,5 pétaoctets (2,5 millions de gigaoctets) de données en 2012.

La recherche, le traitement et l'utilisation de ces données ont toujours été au cœur du développement de l'entreprise. L'ère de l'information, qui a donné aux données de nouvelles propriétés, puisqu'elles sont méta taguées de façon disparate, produites en temps réel, arrivent en flots continues et proviennent de plusieurs sources, met l'entreprise devant une double contrainte : elle doit la rechercher (dans la masse disponible) et l'émettre (l'information stratégique). Au niveau des entreprises, cette masse de données est difficile à gérer, et l'interprétation de ce large volume constitue principalement un défi. Du fait que, l'information peut être à la base d'une connaissance nouvelle. Désormais, la conduite des stratégies industrielles repose largement sur la capacité de l'entreprise à accéder aux informations stratégiques pour mieux agir dans son environnement.

Avec la révolution numérique, les données prennent par ailleurs une place centrale dans l'économie. Ouvrir et partager les données, c'est organiser la mise en ligne de données essentielles, qui vont enrichir les analyses de nombreux décideurs, permettre des gains de temps de travail ou permettre, dans les différents secteurs, des prises de décisions mieux informées. C'est créer de grands référentiels partagés par tous les acteurs et encourager le développement de nombreux services à forte valeur ajoutée.

CONCLUSION

La quantité de données générées par des personnes, des appareils connectés à Internet et des entreprises, se développe à un rythme exponentiel. Les organismes financiers, les entreprises, les services de santé ... génèrent de grandes quantités de données au cours de leurs interactions avec les fournisseurs, les patients, les clients et les employés. En dehors même de ces interactions, de nombreuses données sont en cours de création à travers les requêtes de recherche sur Internet, les réseaux sociaux, les systèmes de géolocalisation GPS, les transactions boursières... Le résultat est ce qui constitue la « révolution de données » ou ère du « Big Data ».

Le « Big Data » est le terme général utilisé pour décrire l'augmentation exponentielle des volumes de données, à côté bien sûr de la croissance dans les capacités à transférer, stocker et analyser ces données. L'expression « Big Data » fait ainsi référence aux technologies, processus et techniques permettant à une organisation de créer, manipuler et gérer des données à grande échelle (Hopkins & Evelson, 2011) et d'en extraire de nouvelles connaissances pour en faire une valeur économique. Mais stocker des quantités considérables de données est une chose, les traiter en est une autre.

Il y a deux manières pour transformer les données en un apport important pour une entreprise :

- La transformation des données en information est un véritable stade de valeur de données, qui sont exploitées afin d'avoir des informations utiles, et mener à bien les stratégies des entreprises. Ce qui apporte systématiquement des informations de base de données dans le processus de prise de décision de l'entreprise.
- La transformation des données en produits ou processus apporte de la valeur à une entreprise. Cela se produit lorsque l'analyse des données doit être mise en œuvre dans le mode physique.

Les données massives amènent nécessairement l'entreprise à regarder ses univers, directs ou plus lointains, d'une autre manière, à penser son quotidien et son avenir différemment et, parfois, selon de nouvelles règles. Certes, il y a des secteurs d'activité qui sont plus concernés que d'autres. Citons par exemple les études réalisées par Benkaraache & Benabdelouahed (2015) sur l'impact des « Big Data » sur les activités bancaires, ou encore (Monino et autres, 2015) sur les défis des « Big Data » sur le Supply Chain Management.

Des entreprises de toutes tailles et de tous secteurs, peuvent envisager de nouveaux modèles économiques basés sur leurs données. Ces modèles peuvent reposer tout autant sur des extensions de leurs cœurs de métier que sur la création de nouveaux produits et services.

BIBLIOGRAPHIE

- AGERON BLANDINE, MARIE-LYNEGOURY, A. SPALANZANI. « LE KNOWLEDGE MANAGEMENT APPLIQUE AUX PROBLEMATIQUES DE DEVELOPPEMENT DURABLE DANS LA SUPPLY CHAIN ». ÉTUDE REALISEE PAR LE CENTRE D'ÉTUDES ET DE RECHERCHES APPLIQUEES A LA GESTION. CNRS. CAHIER DE RECHERCHE N° 2010-03 E5. VERSION 1. 2010.
- BENKARAACHE T., BENABDELOUAHED R. (2015) « L'EMERGENCE DU BIG DATA DANS LE SECTEUR BANCAIRE : CAS DES BANQUES MAROCAINES ». ACTES DU CONGRES VSST 2015, GRENADE, ESPAGNE.
- DAVID, A. (2006), « L'IMPACT DES TIC : LOGISTIQUE, TRANSPORT, RELATION DE SERVICE, ORGANISATION, LA DOCUMENTATION FRANÇAISE, PARIS.
- DUPUIS A. (2004)., « GOUVERNANCE, AJUSTEMENT MUTUEL, COMMUNAUTES HUMAINES ET INTELLIGENCE COLLECTIVE », CAHIER DE RECHERCHE DU CERGO.
- GETZ I., ROBINSON A.G. & LEROY D. (2007), VOS IDEES CHANGENT TOUT !, PARIS, ÉD. D'ORGANISATION.
- GROUPE DE TRAVAIL SUR LES DONNEES OUVERTES (GTDO), (2011), RAPPORT SUR L'OUVERTURE DES DONNEES DE LA VILLE DE MONTREAL : UN CAPITAL NUMERIQUE GENERATEUR D'INNOVATION ET DE PARTICIPATION, DISPONIBLE SUR : [HTTP://WWW1.VILLE.MONTREAL.QC.CA/BANQUE311/WEBFM_SEND/1453](http://www1.ville.montreal.qc.ca/banque311/webfm_send/1453)
- HOPKINS, B. ET EVELSON, B. (2011), "EXPAND YOUR DIGITAL HORIZON WITH BIG DATA", FORRESTER RESEARCH, ACCESSIBLE VIA LE LIEN [HTTP://WWW.ASTERDATA.COM/NEWSLETTER-IMAGES/30-04-2012/RESOURCES/FORRESTER_EXPAND_YOUR_DIGITAL_HORIZ.PDF?LF1=97073534D99767367906A1885620](http://www.asterdata.com/newsletter-images/30-04-2012/resources/forrester_expand_your_digital_horiz.pdf?lf1=97073534d99767367906a1885620)
- JACOB R. & TURCOT S. (2000), « LA PME APPRENNANTE, INFORMATION, CONNAISSANCE, INTERACTION, INTELLIGENCE », RAPPORT DE VEILLE PRESENTE A DEVELOPPEMENT ECONOMIQUE CANADA.
- JULIEN P. A., VAGHELY I. & CARRIER C. (2003), « PME ET CONTROLE DE L'INFORMATION : LE ROLE DU TROISIEME HOMME », COLLOQUE D'AGADIR.
- LE GUEN. F, (2013) ; ANALYSE ET VISUALISATION DES DONNEES, PEARSON, FRANCE.
- MARTINET. B ET MARTI. Y-M, (2001) L'INTELLIGENCE ECONOMIQUE : COMMENT DONNER DE LA VALEUR CONCURRENTIELLE A L'INFORMATION, EDITIONS D'ORGANISATION, PARIS.
- MCMAHON, A., LIU Y., CROSSLAND R., BROWN D., LEAL D., ET DEVLUKIA J. (2004), "A BEST PRACTICE ADVICE SYSTEM TO SUPPORT AUTOMOTIVE ENGINEERING ANALYSIS PROCESSES", ENGINEERING WITH COMPUTERS, VOL.19, NO.4.
- McNULTY. E, (2014), BIG DATA: THE SEVEN V's, DISPONIBLE SUR: [HTTP://DATACONOMY.COM/SEVEN-VS-BIG-DATA/](http://dataconomy.com/seven-vs-big-data/)
- MOINET, N., (2009), « MYTHES ET REALITES DE L'INTELLIGENCE ECONOMIQUE. ANALYSE D'UNE DYNAMIQUE A LA RECHERCHE DE SES CONCEPTS OPERATOIRES & PERSPECTIVES DE RECHERCHE EN COMMUNICATION », MEMOIRE D'HABILITATION A DIRIGER DES RECHERCHES.
- MONINO. J.L, LUCATO. G, (2006), MANAGER L'INFORMATION : COMMENT FAIRE AUJOURD'HUI ?, LES MATINS DE LA CITE, LA CCI DE MONTPELLIER.
- MONINO J-L INTRODUCTION GENERALE INTELLIGENCE ECONOMIQUE, ENTREPRISES ET TERRITOIRES L'HARMATTAN MARCHÉ ET ORGANISATIONS 2013/2 (N° 18)
- MONINO J-L L'INFORMATION AU CŒUR DE L'INTELLIGENCE ECONOMIQUE STRATEGIQUE L'HARMATTAN MARCHÉ ET ORGANISATIONS 2013/2 (N° 18)
- MONINO J-L, SEDKAOU.S LES TIC UN OUTIL INDISPENSABLE POUR UNE DEMARCHE D'INTELLIGENCE ECONOMIQUE L'HARMATTAN - MARCHÉ ET ORGANISATIONS 2013/2 (N° 18)

- MONINO J-L BOYA C LA COLORATION DE L'INFORMATION DANS L'EFFICIENCE SEMI FORTE – REVUE INNOVATIONS 2011/3 (N°36)
- MONINO. L.L, SEDKAOUI.S ET MATOUK. J, (2014), " BIG DATA, ETHIQUE DES DONNEES, ET ENTREPRISES ", LES CAHIERS DU CEDIMES, DOSSIER: « ECONOMIE ET GOUVERNANCE », VOL. 8, No2
[HTTP://CEDIMES.COM/IMAGES/DOCUMENTS/CAHIERS_CEDIMES/CAHIER_CEDIMES_VOL_8_N2_2014.PDF](http://CEDIMES.COM/IMAGES/DOCUMENTS/CAHIERS_CEDIMES/CAHIER_CEDIMES_VOL_8_N2_2014.PDF)
- PINTE, J.P (2006), LA VEILLE INFORMATIONNELLE EN EDUCATION POUR REpondre AU DEFI DE LA SOCIETE DE LA CONNAISSANCE AU XXI EME SIECLE : APPLICATION A LA CONCEPTION D'UNE PLATEFORME DE VEILLE ET DE PARTAGE DE CONNAISSANCE EN EDUCATION, [HTTP://HAL.ARCHIVES-OUVERTES.FR/TEL-00143990/](http://HAL.ARCHIVES-OUVERTES.FR/TEL-00143990/)
- ROTELLA. P, (2012), IS DATA THE NEW OIL?, DISPONIBLE SUR: [HTTP://WWW.FORBES.COM/SITES/PERRYROTELLA/2012/04/02/IS-DATA-THE-NEW-OIL/](http://www.forbes.com/sites/perryrotella/2012/04/02/is-data-the-new-oil/)
- SAÏD, A. (2004), « ROLE DE L'INTELLIGENCE ECONOMIQUE DANS LA STRATEGIE DE L'ENTREPRISE », THESE DE DOCTORAT, MENTION SCIENCES DE GESTION.
- TAYLOR R.S. (1980); VALUE-ADDED ASPECTS OF THE INFORMATION PROCESS, COMMUNICATING INFORMATION: PROCEEDINGS OF THE 43RD, ASIS ANNUAL MEETING, VOL: 17, ANAHEIM, CALIFORNIA, OCTOBER 5-10, 1980.
- ELIOT, T.S. « THE ROCK », FABER & FABER. (1934). DISPONIBLE SUR : [HTTP://WWW.WISDOMPORTAL.COM/TECHNOLOGY/TSELIOT-THEROCK.HTML](http://www.wisdomportal.com/TECHNOLOGY/TSELIOT-THEROCK.HTML)
- ZARA. O, (2008) LE MANAGEMENT DE L'INTELLIGENCE COLLECTIVE, VERS UNE NOUVELLE GOUVERNANCE, M21 EDITIONS, 2EME EDITION.