

# LE BIG DATA

Mythe ou réalité? Une architecture cognitive permettant de tout évaluer, après avoir tout transformé en chiffres Wilfrid Azan, Serge Bolidum

ESKA | « Revue internationale de psychosociologie et de gestion des comportements organisationnels »

2015/Supplément HS   pages 31 à 45
ISSN 2262-8401
Article disponible en ligne à l'adresse :
https://www.cairn.info/revue-internationale-de-psychosociologie-de-gestion-des- comportements-organisationnels-2015-Supplément-page-31.htm
Pour citer cet article :
Wilfrid Azan, Serge Bolidum« Le Big Data. Mythe ou réalité? Une architecture cognitive permettant de tout évaluer, après avoir tout transformé en chiffres », Revue internationale de psychosociologie et de gestion des comportements organisationnels 2015/Supplément (HS), p. 31-45.
Distribution électronique Cairn.info pour ESKA.

© ESKA. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

# PROLOGUE: LES MYTHES DE LA TOUTE PUISSANCE

# **LE BIG DATA**

Mythe ou réalité? Une architecture cognitive permettant de tout évaluer, après avoir tout transformé en chiffres,

Wilfrid AZAN<sup>6</sup>



Serge BOLIDUM<sup>7</sup>



Cette recherche vise à répondre à la question de recherche suivante : « Dans quelle mesure, les architectures technologiques actuelles permettent-elles de gérer un Big-Data ? ». Après une étude de cas dans le secteur pharmaceutique, nous pouvons répondre globalement que le Big-Data implique de nos jours un saut technologique que les entreprises devront franchir et qui sera de la même ampleur que l'intégration des technologies Internet dans les compagnies.

Maître de Conférences, Université de Strasbourg, Faculté de Droit BETA, UMR CNRS F-67000 Strasbourg, France

Doctorant, Université Paris-Dauphine, DRM- UMR CNRS 7088, F-75016 Paris, France. Membre de l'Équipe de Recherche Management & Organisation (ER M&O). Sa thèse: De l'alignement des cultures numériques au sein des organisations vers la minimisation de la surcharge informationnelle liée à l'usage des technologies de l'information et de la communication. Il est aussi chef d'entreprise. Sa carrière de cadre international au sein du Groupe Sanofi commence en 1998. Il est successivement développeur, consultant interne, chef de projet, directeur de département, puis enfin responsable mondial de la gestion de base de données en produits pharmaceutiques. En 2007, il crée son entreprise et choisit le nom de McErnest

### INTRODUCTION

Le 14 février 2013, lors d'une conférence « TechDays », Bruno Walther, fondateur d'une startup orientée systèmes décisionnels et appelé Captain Dash, déclarait : « Jusque-là, leurs analyses reposaient en grande partie sur les retours d'informations issues des millions de tickets de caisse, mais désormais, il leur faut aussi traiter des milliers d'infos issues des journaux, de la télévision, des blogs. Ce sont des centaines de canaux différents qu'il faut analyser<sup>8</sup>. »

Ce constat est relativement aisée à comprendre, cependant il pose une question délicate, qui sera d'ailleurs notre question de recherche dans cette communication : « Dans quelle mesure, les architectures technologiques actuelles permettent-elle de gérer un Big-Data ? »

Nous définissons les architectures technologiques actuelles comme étant fondées sur des modèles relationnels classiques.

Nous envisagerons notre question de recherche du point de vue du volume, de la vitesse et de l'hétérogénéité des données et nous nous fonderons sur l'étude d'une architecture spécifique, appelée **EAM** (Enterprise Architecture Model), dans une grande compagnie pharmaceutique.

Définissons, tout d'abord ce que nous entendons par Big-Data.

Trois caractéristiques permettent de décrire un Big-Data.

- Premièrement, ce type de système rencontre un approvisionnement continu d'un grand volume de données. Les volumes traités aujourd'hui se mesurent en Peta-Octets (multiple de Tera-Octets).
- Deuxièmement, l'hétérogénéité des données domine. Ces dernières possèdent des formats variables. En effet, il peut y avoir du texte sous différentes formes (txt, HTML, XML, doc. Pdf etc.), mais aussi des images et des vidéos aux formats parfois exotiques.
- Troisièmement, le *Big-Data* est également le fait d'une vitesse accrue lors de la création des données grâce à la combinaison d'algorithmes élaborés et la réplication souvent en temps réel entre systèmes connectés.

Le mode d'utilisation suit la tendance de la mobilité et intègre des objets du quotidien notamment les outils tactiles comme les téléphones intelligents de type smartphones, les tablettes et des possibilités d'utilisation en réunion ou dans le cadre de projets (GFII, 2012).

Cette rigueur des méthodes et la pluralité des données aboutiront à des synthèses et des agrégations toujours plus avancées. Les différents supports seront intégrés, assurant la restitution des données quel que soit le format de départ : multimédia, interface sonore, exploitation et analyse des communications orales et courantes,

http://pro.01net.com/editorial/586777/techdays-le-Big-Data-et-le-machine-learning-au-secours-desmetiers/

8

storytellisation de l'organisation des informations. Toutes ces caractéristiques impliquent des technologies complexes, la mise en œuvre d'algorithmes d'intelligence collective, et une approche prédictive et durable supposant un apprentissage. Les cartographies résultantes s'intégreront dans des tableaux de bord d'analyse et d'outils décisionnels stratégiques associés à une restitution sophistiquée via une interface interactive, de la 3D, des cartes animées, voire une représentation matricielle en fonction de la nature des données traitées.

Le *Big-Data* conduit donc à des traitements effectués en amont et en aval afin de préparer l'analyse des données, les décisions, d'éviter l'info-obésité et la surcharge informationnelle soit le fait de recevoir une trop grande quantité d'informations (Eppler et Mengis 2004). Ce prolongement de l'intelligence artificielle, dans le sens d'analyse sémantique de données complexes, n'est pas possible sans introduire une vision extensive de l'utilisation des capacités cognitives humaines et s'assortit d'une réflexion sur les limites de la rationalité des acteurs.

Pour Loudcher (2011) un des défis est d'extraire et d'analyser (en ligne) la sémantique contenue dans les données complexes. Selon ce même auteur, ce point constitue un véritable verrou scientifique mais qui n'est que partiellement abordé par la communauté scientifique

L'augmentation exponentielle des volumes de données, suivie de l'augmentation des volumes d'informations, l'une des trois dimensions de la surcharge informationnelle identifiées par Isaac et al. (2007), la vitesse de réplication des données entre les systèmes et la maîtrise de l'hétérogénéité, impliquent donc de pouvoir fournir « la bonne information à la bonne personne au bon moment », ceci afin d'accompagner la transformation tant du travail que des structures informationnelles des organisations. L'activation puis le développement accru des *knowledge workers* et des entreprises étendues par la perpétuation des équipes virtuelles (Baskerville et Nandhakumar, 2007), pourraient en être une mise en application.

Dans une première partie, nous présentons un cadre théorique d'une architecture de *Big-Data* que nous testons dans une deuxième partie. Nous présentons la méthode de recherche employée pour la validation de ce cadre. La troisième partie présente et discute les résultats intermédiaires obtenus. Enfin, nous abordons les limites de la recherche et les perspectives possibles.

# ÉLÉMENTS DE DÉFINITION

Nous soulignons ici l'importance de l'architecture dans la gestion du *Big-Data* et dans l'utilisation.

# o Architecture et Big-Data

Les *Big-Data*, un ensemble de données numériques, sont caractérisés par des besoins de capacité et de possibilité de traitements qui excèdent largement celles des logiciels courants. Il n'est pas rare de raisonner à présent en Peta-Octets au lieu de Tera-Octets. Pour supporter un *Big-Data*, une architecture spécifique apparaît nécessaire. Il faut prendre en compte le fait que cette architecture se situe dans une

organisation pour qui la gestion de l'information et des connaissances est essentielle (Reix et al. 2012; Cohendet et al. 2003). Le traitement de données à croissance exponentielle, n'est plus à la portée des systèmes d'information et technologique classiques. A une architecture de données, on fait correspondre une architecture informatique et un système de communication. La théorie de la richesse des média (Daft & Lengel, 1986) suppose que le choix d'un média résulte de l'adéquation entre les caractéristiques du moyen de communication et le contenu du message. Ce choix est lié à la nature du message que l'on veut transmettre, mais également des capacités de chacun à utiliser les différentes possibilités offertes par le média. L'architecture des solutions de *Big-Data* devient un enjeu en fonction de la multitude des connaissances à délivrer. Les systèmes classiques n'ont pas été conçus pour des données à croissance exponentielle. Les organisations « processeurs de connaissances » ne sont pas configurées pour traiter, sans changement de structure, des taux de croissance comparables.

L'architecture définit d'une manière générale la structure d'un système. Dans ce cadre, une architecture de *Big-Data* doit pouvoir faire face aux trois caractéristiques énoncées en introduction. Le tableau 1 illustre les besoins en termes d'architecture pour faire face aux caractéristiques d'un *Big-Data*.

Caractéristique	Architecture requise
Volume	Capacité de stockage en Peta-Octets (matérielle et
	logicielle)
	Capacité d'accès rapide aux données (qualité et
	fiabilité des liens avec le centre de données)
	Capacité de recherche rapide dans les données
	(moteurs de recherche intégrés et analyse
	sémantique)
Vitesse et dynamique	Flexibilité et résilience des liens entre l'extérieur et
d'arrivée des données	le centre de données
Hétérogénéité des données	Variété des modèles de stockage des données
	(Relationnel, NOSQL, taggage automatique,
	Vrac)

Tableau 1 : caractéristiques et contraintes d'architecture

#### Architecture modulaire

Face aux contraintes d'un *Big-data*, examinons la pertinence d'une approche modulaire de l'architecture. L'Analyse Modulaire des Systèmes (Mélèse, 1972) se propose de décomposer l'organisation en modules avec des entrants et sortants permettant de cartographier les interactions entre les services et les contrôles associés. Ce n'est que dans un deuxième temps que les flux de données sont traduits en schémas entité relation (à l'époque MERISE). La modularité est présente dans le « *computer mediated communication*» c'est-à-dire dans les suites logicielles. Elle est associée à l'apprentissage et devient rapidement un enjeu théorique important dans

les années 1990 (Alavi, M, 1994; Alavi, Wheeler & Valacich, J. 1995; Alavi & Gallupe, 2000; Alavi et Leidner, 2001). Dans l'industrie, l'approche modulaire, constitue l'expression d'une volonté d'utiliser au mieux les ressources. La conception de la modularité technique des savoirs (Azan et al., 2012) suppose des formes organisationnelles et informationnelles particulières qui « économisent » de la hiérarchie et de l'administration par les instructeurs et qui sont donc génératrices de baisses de coûts de transaction (principale source de performance d'une architecture modulaire) (Lerch et al., 2005).

Pour faire face aux contraintes d'un *Big-Data*, une architecture modulaire doit être privilégiée.

# Big-Data et décision

Pour Tien (2012), les *Big-Data* sont à l'origine d'une révolution comparable à la révolution industrielle et qui scinde l'architecture informatique autour de modules. Quand on ajoute des capacités d'analyse, les potentialités sont immenses. On parle alors de *Big-Data* analytics. Il s'agit du regroupement de toutes les techniques existantes d'analyse au sein de suites logicielles qui travailleront sur ces grands volumes de données. Certaines sont en différé et d'autres en temps réel (real time BI). Le domaine de la Business Intelligence qui avait été en vogue dans les années 2000 et qui connaissait un certain ralentissement devrait connaître ainsi un renouveau. Cette tendance est dû également à des besoins à moyen terme de compétences d'expertes en analyse de données complexes et plus encore de manager pouvant interpréter ces données (Chen et al., 2012). En adoptant une perspective culturelle des usages des TI dès le début d'un projet Big-Data (Walsh et al., 2010), l'entreprise pourrait anticiper ainsi les stratégies d'usage et d'adoption des TI par ses collaborateurs.

## ETUDE DE CAS : LE SYSTÈME EAM

L'étude de cas est logo-centrée et fondée sur l'expérience d'un chef de projet. Six interviews avec ce chef de projet ont été réalisées entre mars et juin 2012. L'analyse de contenu qui en résulte est soumise à l'interviewé pour relecture.

L'entreprise sujet de notre cas est l'un des leaders mondiaux de l'industrie pharmaceutique, résultat de la fusion en 2004 du numéro un français de l'industrie pharmaceutique AG, 26000 collaborateurs et d'EVE groupe pharmaceutique Germano-Franco-Américain, 75000 collaborateurs. Le groupe représente en 2012 environs 80 usines de productions et 50 centres de distribution. Avec plus de 400 000 références produits à gérer par ses flux logistiques, ayant pour certaine jusqu'à 50 caractéristiques différentes, la direction générale d'AG a du étudier dès 1998 l'implémentation d'une technologie de capture et d'analyse de l'information tout à la fois innovante et efficace. Les bases de données à forte capacité étant largement disponibles à des prix de plus en plus bas, AG a néanmoins dû étudier les avantages

et les inconvénients d'une architecture tout à la fois fonctionnelle et technique, permettant l'accès rapide aux informations essentielles, et cela à tous les niveaux de l'organisation ; ces processus devaient notamment permettre d'orienter à moyen terme les décisions en temps réel ; ce modèle d'architecture d'entreprise (EAM) offre un référentiel mondial des données de base et devait finalement limiter la prolifération des codes produits lors de créations incontrôlées dans les divers systèmes connectés au réseau. Les modèles d'architectures implémentés ont ainsi permis l'émergence de processus techniques et métiers innovants et à forte valeur ajoutée.

Le projet d'implémentation d'un référentiel mondial de gestion des masters data des produits pharmaceutiques a été lancé en 1998 par la direction monde de la « qualité pharmaceutique » et non pas à l'initiative de la direction logistique bien que celle-ci soit le « propriétaire historique » des processus de gestion des flux produits. Étant soumis continuellement à des réglementations strictes et changeantes (SoX, FDA, GxP), les données permettant de décrire précisément les produits pharmaceutiques devaient répondre aux caractéristiques suivantes :

- fiabilité (reliable)
- précision (accurate)
- à jour (up to date)
- complétude (complete)

Il est à noter que la nécessité de codifier dès 1998 plus de 200 000 références de tout type, produit fini, semi fini, etc... avec des caractéristiques bien précises, demanda la création d'une structure organisationnelle appropriée, prenant en compte tant l'architecture fonctionnelle que l'architecture technique.

Les données de base des produits pharmaceutiques étant la « propriété partagée » de plusieurs domaines de l'entreprise, notamment les départements de la qualité pharmaceutique, de la logistique, de la finance et des usines, les responsabilités partagées de la mauvaise qualité de ces données étaient compliquées à trouver. Dès le lancement du projet en 1998, tous les dirigeants appuyèrent le projet d'EAM. Il fut décidé d'abandonner l'organisation précédente afin d'en créer une nouvelle orientée métier sur laquelle est venue se greffer une nouvelle architecture informatique. Ce modèle organisationnel est de type collaboratif en matrice ; un collaborateur décentralisé de chaque entité légale des pays « propriétaires » du contenu des données rend compte à la direction centrale en charge du contenant du référentiel produit, tout en restant sous l'autorité juridique de son responsable local. L'architecture technique se compose de trois serveurs hébergeant l'application de

L'architecture technique se compose de trois serveurs hébergeant l'application de gestion du référentiel des données de base, la fiche article, de SAP R/3 :

- Le premier serveur est dédié au développement de nouvelles fonctionnalités applicatives et aux tests unitaires,
- Le second serveur est celui de la consolidation et de la validation avant mise en production,
- Le troisième serveur est le système de production, système qualifié.

Il est intéressant de noter que l'architecture technique est qualifiée et validée pour l'exploitation dans un environnement réglementaire pharmaceutique et ainsi répondre aux audits demandés par les instances validant les autorisations de mise sur le marché d'un produit.

Du fait de la centralisation de la saisie des données sur un seul système, la dispersion involontaire des informations dans divers systèmes d'information est fortement réduite. La dispersion vers l'ensemble des systèmes clients à partir de l'application centralisée se fait par l'intermédiaire d'interfaces techniques, de type IDOC (SAP) ou fichiers, connectées aux systèmes clients. En outre, toute création de donnée dans un système non centralisé, ne sera pas pris en compte lors de la réplication vers les autres systèmes connectés, notamment les systèmes de consolidation financière et de consolidation des stocks. Le serveur de production hébergeant la solution ERP SAP R/3 est connecté à 36 systèmes tiers. Ces 36 systèmes sont pour la plupart des solutions ERP, SAP ou non, localisées dans les pays et les filiales du groupe, puis des systèmes SAP Netweaver BI, contenant les outils de reporting et de tableaux de bord décisionnel, ainsi que des systèmes de consolidation financière des stocks. L'ensemble des systèmes prend donc en compte comme référentiel unique le système centralisé de gestion des références produit. L'architecture est en étoile, gravitant autour du système central et l'effet « spaghetti » d'une architecture décentralisée est largement réduit : «Dès que AG rachète une entreprise elle est simple à ramener dans notre Big-Data.

L'industrie pharmaceutique est secteur fort un en mouvement depuis l'affermissement des pratiques réglementaires, la réduction des remboursements, ainsi que le passage dans le domaine public de molécules à forte valeur ajoutée. AG s'est positionnée sur le marché des médicaments génériques en rachetant régulièrement des entreprises fabricant ce type de produit. De ce fait, les fusions et les acquisitions d'entreprises concurrentes sont régulières. Il est à noter que le groupe AG a intégré depuis sa création en 1971 plus de 250 entreprises dans le groupe. En cas de fusion et d'acquisition, le temps d'intégration des processus de l'entreprise dans le SI du groupe se doit d'être optimisé. Il en est de même de l'intégration des systèmes d'information et des technologies de ces entreprises. Le système centralisé permet de connecter ces systèmes d'informations à l'aide des bonnes pratiques du modèle d'architecture d'entreprise. Pouvoir distribuer l'ensemble des produits de l'entreprise racheté par la chaîne logistique et financière dans un délai optimal en est l'objectif principal.

Le volume de données est très élevé tant pour la production que pour les tests pharmaceutiques. La base contenant toutes les données est de type ORACLE. La capacité de stockage est de l'ordre de trois téraoctets répartis sur 3 serveurs dont celui de production. Il est à noter que le système de sauvegarde mis en place répond aux réglementations pharmaceutiques, répliquant à 100% dans un emplacement distant toute la configuration de base en cas de destruction inopinée de l'ensemble des systèmes. L'étude du projet d'EAM global a pris de l'ampleur en 2004, au moment de la fusion d'AG et d'EVE. D'une approche standardisée SAP chez EVE, il

a fallu prendre en compte dans les meilleurs délais l'architecture technique non standardisée et non SAP d'AG. Une équipe d'architectes a été créée pour répondre à ce nouveau besoin d'intégration. L'équipe d'architecte est composée de 5 personnes et représente une organisation réduite tant en terme de personnel qu'en terme de coût. En effet, la décentralisation de la saisie des caractéristiques des produits à un réseau mondial d'expert, a permis d'optimiser les processus décisionnel des Systèmes d'Information. L'ensemble des développements est traité par un centre de service dédié au management des requêtes des clients, des incidents et des évolutions.

La vitesse est la synthèse d'information et l'agrégation de données qui s'effectuent dans une autre solution informatique dans ce cas un outil d'aide à la décision. Elle permet la flexibilité et la résilience des liens entre l'extérieur et le centre de données. L'architecture est utilisée au départ comme de la Business Intelligence, par exemple pour la consolidation des stocks (200 000 références). Très rapidement, les mises à jour des informations sur les stocks s'effectuent non plus quotidiennement mais en temps réel. L'information devient alors disponible sur un portail auprès des utilisateurs qui en expriment le besoin et elle devient ainsi plus pertinente.

L'hétérogénéité des données augmente dans le système, puisqu'un même utilisateur peut avoir accès à la fois à des données opérationnelles, (par exemple les nomenclatures produits des produits génériques) et à des données financières (rapport trimestriel). La variété des modèles de stockage des données s'accroît. Les domaines métiers pris en compte sont hétérogènes et l'accès dans l'EAM s'effectue autour de données uniquement structurées. Les données non structurées ne sont pas accessibles et l'hétérogénéité du format est une question en suspens (Tableau 2).

Volume	Vitesse	Hétérogénéité	
Condition	Condition remplie	Condition non remplie	
remplie			

Tableau 2 - Variables, Big-Data et architecture

La réflexion d'une architecture *Big-Data* a démarré dès 1998, bien que celle-ci n'en portait pas le nom. Son architecte, un des chefs de projet de la maîtrise d'ouvrage disposait d'un double doctorat en chimie et en pharmacie que vient compléter une excellente compréhension des SI.

#### DISCUSSION ET CONCLUSION

D'un point de vue global, la première constatation est que les architectures traditionnelles permettent de traiter de gros volumes de données ainsi qu'une dynamique de vitesse relativement élevée. Mais ni l'hétérogénéité ni la vitesse autour de données « non structurées » ne sont prises en compte. En effet, les bases Oracle fondées sur des modèles relationnels classiques ne peuvent pas prendre en compte des données dites non structurées (Tournant et al., 2008).

Ce n'est probablement pas la seule explication au fait que les architectures technologiques et informatiques actuelles ne permettent pas de gérer un *Big-Data*. Les quantités d'informations traitées et leur vitesse ne sont pas ici source d'une pertinence accrue du système d'information. Les savoirs sont modulaires et ne se retrouvent qu'imparfaitement dans l'architecture informatique. Mélèse s'appuie dès 1972 sur les travaux de J.P. Dupuy. Il évoque la condition première de l'autonomie qui est l'auto-référence et qui signifie que les actions de l'individu portent du sens pour lui-même, par rapport à lui-même. Pour Mélèse, cette caractéristique de l'information est à rapprocher de la notion de valeur d'usage. A l'opposé, des actions hétéronomes ne prennent de sens que dans un système plus large qui englobe l'utilisateur (ce qui pour Mélèse correspond à la valeur d'échange de l'information). La mauvaise prise en compte du caractère modulaire et contextuel des savoirs ne permet pas un meilleur passage de l'information à l'interprétation.

De plus, les procédures de nettoyage et de vérification des données entrantes sont trop contraignantes pour pouvoir faire face à un afflux continu de données non structurées. Le coût du passage de cette architecture à une architecture plus adaptée apparaît élevé. En effet, rien que la phase de modélisation des données représentatives de la modularité des métiers et des savoirs s'avère difficile. Il faut tenir compte des modèles existants et les faire cohabiter avec des modèles moins structurées.

D'un point de vue technique, le raisonnement en Tera-Octets doit laisser place à une approche en Peta-Octets. Ce passage de 1 à 1000 implique un renouvellement d'une partie de l'architecture matérielle et notamment de la qualité du réseau. Des « fermes de données » seront peut être nécessaires et il se posera alors la question stratégique de l'externalisation des données. Une telle décision, impliquant une réflexion en termes de contrôle et de sécurité des données, est généralement assez longue à prendre et à mettre en œuvre.

Du point de vue des utilisateurs, l'étude terrain met en évidence que ceux-ci sont d'abord des co-constructeurs de l'architecture. Ils créent des représentations qui deviennent pertinentes collectivement. Il s'agit d'une innovation incrémentale davantage que radicale (pas de révolution en cela) et d'un juste retour des choses dans la mesure où les TI retournent à leur fonction première, la distribution d'informations.

#### Contributions attendues

Cette recherche en cours présentera à terme des contributions tant pour la recherche que pour les managers. Nous espérons pouvoir relier l'analyse des comportements des utilisateurs de BigData à celle de la conception d'artefacts technologique d'analyse de données comme le préconise Baskerville (2011) ou encore pouvoir structurer par l'analyse Modulaire des Systèmes (Mélèse, 1972) un modèle de business process management. D'un point de vue managérial, cette étude pourra donner de précieuses indications aux managers qui cherchent à mettre en place une

architecture distribuée *Big-Data* et cela sans mettre « à feu et à sang » les données existantes de l'entreprise, ni les équipes clientes de ces systèmes. La surcharge quantitative influençant positivement l'anxiété informatique (Elie-Dit-Cosaques et al., 2012), les résultats seront importants en ce qu'ils permettront de mieux comprendre comment limiter ces comportements.

#### Limites de notre recherche et recherches futures

L'Analyse Modulaire des Systèmes de Mélèse (1972) est ancienne et ce simple fait pourrait être identifié comme une limite de notre recherche. D'autres approches de systémique pourraient être mobilisées, notamment celles de Le Moigne (1999) ou plus récemment celle de Meunier (2006), pour la compréhension de l'approche systémique des relations sociales. Notre premier choix a néanmoins été guidé par notre posture épistémologique.

Nous avons analysé une seule étude de cas. Dans ce contexte, nous pouvons faire valoir qu'il pourrait y avoir un biais positif dans l'analyse des données ainsi que dans les premières techniques méthodologiques. Néanmoins, une première étude de cas, dans le cadre d'une approche méthodologique de théorie enracinée (Glaser et Strauss, 1967), modélisation ad hoc d'une situation d'entreprise (Roemelaer et Kalika, 2011), pourra faire émerger de puissants concepts. Cette démarche exploratoire, que nous avons déjà débuté par des entretiens individuels, semi directifs, enregistrés, nous permettra finalement de poser les hypothèses puis d'articuler les situations d'usage constatées sur le terrain. En outre, les études futures pourraient analyser comment certaines zones de l'IS sont influencés par les architectures *Big-Data*, à savoir identifier les relations de cause à effet, ou encore d'appliquer différentes perspectives théoriques à cette problématique pour arriver à une approche globale et une connaissance approfondie des phénomènes induits.

Nous pouvons répondre globalement que le *Big-Data* implique de nos jours un saut technologique que les entreprises devront franchir et qui sera de la même ampleur que l'intégration des technologies Internet dans les compagnies.

En conclusion, le chemin est encore long et les travaux, par exemple, en matière de *data design* ou en matière de *text mining* (Davey et al., 2012), qui visent à offrir un niveau d'analyse supérieur, tout en privilégiant la simplicité et l'intuitivité pour une compréhension par les utilisateurs, sont encore loin de ce qui se passe en entreprise. Ainsi, l'obligation qui est faîte aux entreprises de gérer des *Big-Data* sera, comme nous venons de l'expliciter, à la source de pistes de recherches fructueuses.

### **BIBLIOGRAPHIE**

- Alavi M. (1994), "Computer-mediated collaborative learning: An empirical evaluation", MIS Quarterly, 18(3): 159-174.
- Alavi M., Wheeler, B. & Valacich, J. (1995), "Using IT to reengineer business education: An exploratory investigation or collaborative telelearning". *MIS Quarterly*, 19(2): 293-311.
- Alavi M. & Gallupe B. (2000), "The transformation of business education". New York: Prentice Hall.
- Au N., Ngai E.W.T. & Cheng T.C.E. (2008), "Extending the Understanding of End User Information Systems Satisfaction Formation: An Equitable Needs Fulfillment Model Approach", *MIS Quarterly*, march, Vol. 32 No. 1, pp. 43-66.
- Azan W., & Sutter I. H. (2010). "Knowledge transfer in post-merger integration management: case study of a multinational healthcare company in Greece". *Knowledge Management Research & Practice*, 8(4), 307-321.
- Azan W. & Tournant L. (2008), "Pilotage des connaissances et projets industriels", *Revue Française de Gestion Industrielle*, Vol. 27, n° 2, juin.
- Bandura A. (1986), *Social Foundations of Thought and Action*. Prentice-Hall, Englewood Cliffs, NY.
- Bandura, A., Adams, N. E., & Beyer (1977), "Cognitive Processes Mediating Behavioral Change," *Journal of Personality and Social Psychology* (35:3), pp. 125-139.
- Barclay, D., Higgins, C, & Thompson, R. (1999), "The Partial Least Squares Approach to Causal Modeling; Personal Computer Adoption and *MIS Quarterly* Vol. 23 No. 2/June.
- Baskerville R., & Nandhakumar. J. (2007) "Activating and Perpetuating Virtual Teams: Now That We're Mobile, Where Do We Go?" *Professional Communication, IEEE Transactions On* 50, no. 1:17–34.
- Baskerville, R. "What Design Science Is Not." European Journal of Information Systems 17, no. 5 (2008): 441–443.
- Davey J., Mansmann F., Kohlhammer J., & Keim D. (2012), "Visual Analytics: Towards Intelligent Interactive Internet and Security Solutions", in F. Alvarez & al. (Eds.): FIA 2012, LNCS 7281, pp. 93–104.
- Bughin J., Chui M. & Manika J. (2010), "Clouds, big Data, and smart assets, ten tech-enabled business trends to watch", *McKinsey Ouarterly*.

- Kim B.J. (2012), "A Classifier for Big Data", G. Lee et al. (Eds.), ICHIT 2012, CCIS 310, pp. 505–512.
- Chan Y. (2000), "IT value: the great divide between qualitative and quantitative and individual and organizational measures", *Journal of management information systems*, vol.16, n° 4, pp. 225-261.
- Chen H., R.H.L. Chiang, Storey V.C. (2012), Business Intelligence and analytics: from Big-Data to big impact, *MIS Quarterly*, Vol. 36 N°4, pp. 1165-1188.
- Cohendet P., M. Diani, C. & Lerch (2005), "Stratégie modulaire dans la conception.

  Une interprétation en termes de communautés", *Revue Française de Gestion*, N° 158, septembre-octobre, pp. 121-143.
- Core Techniques and Technologies for advancing Big-Data Sciences & Engineering (BIGDATA), Programm solicitation, NSF 12-499.
- Daft, R.L. & Lengel, R.H. (1986), Organizational information requirements, media richness and structural design. *Management Science* 32(5), pp. 534-571.
- Elie-Dit-Cosaque, C., Pallud J., & Kalika M. (2011). "The Influence of Individual, Contextual, and Social Factors on Perceived Behavioral Control of Information Technology: A Field Theory Approach." *Journal of Management Information Systems* 28, no. 3 (2011): 201–234.
- Eppler, M. J. & Mengis J. "The Concept of Information Overload: A Review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines." *The Information Society* 20, no. 5 (2004): 325–344.
- Glaser, Barney G., and Anselm L. Strauss. *The Discovery of Grounded Theory:*Strategies for Qualitative Research. Transaction
  Publishers, 1967.
- GFII (2012), « Les nouveaux usages de la veille, 5 pratiques en émergence, groupe de travail sur l'intelligence économique et l'économie de la connaissance », livre blanc.
- Hartwick, J., & Barki, H. (1994), "Explaining the role of user participation in information system use", Management Science, 40: 440– 465.
- Heimgärtner R., Kindermann H. & Huang R. (2012), "Revealing Cultural Influences in Human Computer Interaction by Analyzing Big Data in Interactions", AMT 2012, LNCS 7669, pp. 572–583.
- Isaac, H. et al. (2007), "Surcharge Informationnelle, Urgence Et TIC. L'effet Temporel Des Technologies de L'information." Management & Avenir no. 3 : 149–168.

- Lee, A. S. (1989), "A scientific methodology for MIS case studies". *MIS Quarterly*. 13:1, pp. 33-50.
- Le Moigne, Jean Louis (1999). La Modélisation des systèmes complexes. Dunod.
- Loudcher, S (2011), "Vers l'OLAP Sémantique Pour L'analyse En Ligne Des Données Complexes." HDR, Université Lumière Lyon 2.
- Manyika, J., M. Chui, B. Brown, Bughin J., Dobbs R., Roxburgh C. & Byers A.H. (2011), "Big data: the next frontier for innovation, Competition and productivity", McKinsey Global Institut.
- Mate A., Llorens H., & de Gregorio E. (2012), "An Integrated Multidimensional Modeling, Approach to Access Big Data, in Business Intelligence Platforms", S. Castano & al. (Eds.): ER Workshops, Springer-Verlag Berlin Heidelberg, pp. 111–120.
- Melèse J. (1972), "Analyse Modulaire des Systèmes", les Editions d'Organisation, Paris.
- Meunier, J.P. (2003), "Approches systémiques de la communication: systémisme, mimétisme, cognition." De Boeck Supérieur, 2003.
- Orlikowski W.J. (1992), "The duality of technology: Rethinking the concept of technology in organizations", *Organization science*, 3, 3, pp 398-427.
- Reix R., Fallery B., Kalika M., Rowe F. (2012), Systèmes d'information et Management, Editions Vuibert, Paris.
- Romelaer P. et Kalika M., (2011). « Comment réussir sa thèse : la conduite du projet de doctorat ». Seconde édition, 222 p. *DUNOD Paris*.
- Sanchez R., J. Mahoney (1996), "Modularity, Flexibility, and Knowledge Management in Product and Organization Design", *Strategic Management Journal*, vol. 17 (Winter Special Issue), p. 63-76.
- Tapscott D. (2005), "The digital Economy", Mc Graw Hill, New York.
- Tien J.M. (2012), "The next industrial revolution: Integrated services and goods", Journal of System Science and System Engineering, 21(3), pp. 257-296.
- Von Hippel E. (1990)," Task partitioning: an innovation Process Variable", Research Policy, Vol. 19, pp. 407-418.
- Walsh, I., Kefi H., & Baskerville R. (2010) "Managing Culture Creep: Toward a Strategic Model of User IT Culture." *The Journal of Strategic Information Systems* 19, no. 4: 257–280.