Revue COSSI

N° 4 | 2018
Méthodes et stratégies de gestion de l'information par les organisations : des big data aux thick data

Introduction : Méthodes et stratégies de gestion de l'information par les organisations : des "big data" aux "thick data"

Christian MARCON *Professeur des universités*
IAE
CEREGE
University of Poitiers
Sylvie GROSJEAN *Professeur des universités*
Communication
Laboratoire d'innovations Com&Tech
University of Ottawa
Monica MALLOWAN *Professeur agrégé*
Observatoire Provis
Université de Moncton

Édition électronique :
URL :
https://revue-cossi.numerev.com/articles/revue-4/67-introduction-methodes-et-strategies-de-gestion-de-l-information-par-les-organisations-des-big-data-aux-thick-data
DOI : numerev_2009
Date de publication : 04/09/2018

Cette publication est sous licence **CC BY-NC-ND** (Attribution - No commercial - No derivatives).

Pour **citer cette publication** : MARCON, Ch., GROSJEAN, S., MALLOWAN, M. (2018) Introduction : Méthodes et stratégies de gestion de l'information par les organisations : des "big data" aux "thick data". *Revue COSSI*, (4). https://doi.org/10.34745/numerev_2009

La problématique de la gestion des données par les organisations n’est pas récente, mais l’accès à des données massives produites par le monde digital (e-commerce, requête internet, capteurs e-santé, objets connectés, etc.) conduit indéniablement les organisation à gérer, traiter, utiliser et réutiliser leurs données différemment voire à exploiter celles d’autres organisations. Confrontées à la pression concurrentielle, les organisations comptent sur la performance des technologies de l’information pour soutenir leurs processus organisationnels et pour les aider à maîtriser la masse d’information en circulation dans leur environnement interne et externe. Face à l’accumulation de données massives (big data) en milieu organisationnel (Bollier, 2010; Rudder, 2014), l’approche privilégiée pour en tirer un sens est celle de l’analyse quantitative menant à des démarches d’intelligence d’affaires (business intelligence), en vue de s’en servir pour la prise de décision et le passage à l’action (Cohen, 2013; Fernandez, 2013). Ceci conduit à l’idée dangereuse que des données statistiques seraient plus utiles et objectives et contribueraient à rendre les organisations plus efficaces et rentables (Bollier, 2010).

Mots-clés :

La problématique de la gestion des données par les organisations n’est pas récente, mais l’accès à des données massives produites par le monde digital (e-commerce, requête internet, capteurs e-santé, objets connectés, etc.) conduit indéniablement les organisation à gérer, traiter, utiliser et réutiliser leurs données différemment voire à exploiter celles d’autres organisations. Confrontées à la pression concurrentielle, les organisations comptent sur la performance des technologies de l’information pour soutenir leurs processus organisationnels et pour les aider à maîtriser la masse d’information en circulation dans leur environnement interne et externe. Face à l’accumulation de données massives (big data) en milieu organisationnel (Bollier, 2010; Rudder, 2014), l’approche privilégiée pour en tirer un sens est celle de l’analyse quantitative menant à des démarches d’intelligence d’affaires (business intelligence), en vue de s’en servir pour la prise de décision et le passage à l’action (Cohen, 2013; Fernandez, 2013). Ceci conduit à l’idée dangereuse que des données statistiques seraient plus utiles et objectives et contribueraient à rendre les organisations plus efficaces et rentables (Bollier, 2010).

Toutefois, cette accumulation de données chiffrées analysées statistiquement est

limitée par l'absence d'un contexte significatif riche (Alles et Vasarhelyi, 2014), qui pour sa part est généré par une accumulation d'éléments non chiffrés, non quantifiables, non structurés et donc difficilement mesurables, de type socio-culturel (thick data), qui doivent être étudiés par une analyse qualitative, et qui mènent à des démarches d'intelligence stratégique (competitive and strategic intelligence). Cette rencontre entre les big data et les thick data peut générer l'information actionnable et compléter la perspective qu'une organisation doit avoir de son environnement et donc l'aider à mieux orienter ses voies d'action (Hoppe, 2015). Il est donc nécessaire de s'interroger sur la tendance qu'ont les organisations à prendre des décisions stratégiques en se fondant uniquement sur des algorithmes. « Les « datapartisans », représentants d'une pensée pro-positiviste, soutiennent une exploitation logico-mathématique des données par les machines et tentent par ce biais d'analyser des objets de connaissance, allant jusqu'à croire que la machine pourrait remplacer l'esprit humain dans le processus d'analyse. Ils font ainsi fi des recherches menées sur la logique des significations de Piaget, sur la logique déontique de Simon ou encore sur la dialogique de Morin, lesquelles doivent être mobilisées pour concevoir et construire des projets de connaissance » (Verlaet, 2015, p.253). La gestion des données massives n'est pas uniquement une question d'outils mais de stratégies et cela nécessite qu'une réflexion soit menée sur les stratégies et méthodes mises en œuvre par les organisations pour traiter, visualiser et « faire sens » de leurs données.

Thème 1 - Les organisations face aux big data : des méthodes d'analyse à la réflexion critique en passant par les transformations organisationnelles

Le développement rapide des outils et méthodes visant à exploiter les données, informations massivement accessibles par Internet (big data) génère un discours enthousiaste et techniciste. Les big data nous sont annoncées comme la promesse d'un avenir radieux : meilleure approche de notre santé, de notre vie quotidienne, de nos besoins domestiques, etc. Qualifié de nouvelle révolution industrielle comparable à celle de la vapeur, le big data annoncerait une transformation profonde de la société et de nos organisations. Les organisations de santé portent attention aux big data afin d'aider les cliniciens dans leur diagnostic (Sybord, 2016), en santé publique des campagnes de promotion sont conçues en s'appuyant sur des analyses menées à partir de bases de données en épidémiologie (Dumez, Minvielle et Marrauld, 2015; Colloc et Hénocque, 2016; Summons et Regan, 2016), dans le domaine des relations publiques, les big data transforment les pratiques (Mercanti-Guérin, 2013), sans oublier les organisations publiques qui cherchent des méthodes, stratégies afin de « faire parler » ces données massives et ainsi orienter les politiques publiques (Hamel et Marguerit, 2013).

Il est indéniable que ces données fascinent du fait des potentialités qu'elles laissent entrevoir en terme de performance organisationnelle, de prise de décision, de stratégies (Vayre, 2014; Brasseur, 2013). Mais, le big data soulève des enjeux techniques, méthodologiques, organisationnels, communicationnels et éthiques auxquels ils faut réfléchir (Cardon, 2012; Béranger, 2016; Boyd et Crawford, 2012; Parks, 2014; Crawford, Miltner et Gray, 2014). Par exemple, nos organisations vont-elles s'en remettre uniquement à des algorithmes pour les aider dans leur prise de décisions

stratégiques ? En quoi l'introduction de nouvelles méthodes de traitement et de gestion de données massives transforment-elles la culture organisationnelle, la collaboration inter-professionnelle, ou l'activité même des travailleurs ? Par ailleurs, confronté à des discours optimistes face aux big data, le chercheur ne peut que questionner cet engouement et plusieurs questions méritent d'être soulevées :

- Quelle éthique impliquent ces procédés et outils de traitement des données massives ? Quel droit régit le big data et garantit le respect des libertés personnelles tandis que fleurit un discours quasi-propagandiste qui soutient qu'un peu moins de vie privée permettrait beaucoup plus de bien-être collectif ?
- Quels enjeux en terme de communication pour nos organisations ? Comment de telles données transforment-elles les relations avec les clients, les partie-prenantes, et les employés? En quoi et comment de telles données peuvent-elles soutenir la communication des organisations ?
- Quelles questions cognitives pose le big data? Pour Rouvroy (2014), par exemple, les risques « ne tiennent pas tant à une plus grande visibilité, ou à une perte relative d'anonymat ou d'intimité des individus qu'à […] un court-circuitage des capacités d'entendement, de volonté et d'énonciation des individus. » De surcroît, confier l'analyse des données à des algorithmes ne revient-il pas à exposer ses connaissances aux risques liés aux schémas cognitifs des concepteurs de ces outils ?
- Quelles questions le big data pose-t-il aux managers ? Pour McAfee et Brynjolfsson (2012) par exemple, le management avec les big data est un enjeu managérial qui commence par la responsabilité de l'équipe de cadres supérieurs. Ne risque-t-on pas simplement de voir les managers se porter vers le big data parce qu'il est à la mode et qu'il en va de leur image de suivre le courant porteur (Zerbib, 2013) ? Ou parce qu'ils sont à la recherche d'une béquille (walking stick) (Hafsi et Thomas, 2005) ?

Autant de questions et bien d'autres que soulève l'essor des datas dans les organisations que ce soit en terme communicationnel, méthodologique, éthique, etc.

Thème 2 – Les organisations de recherche face à l'essor des datas

De plus en plus, les structures comme les centres de documentation, les bibliothèques, le réseau de lecture publique, les bibliothèques spécialisées, les centres d'archives, doivent intégrer dans leurs modalités de gestion et de management des ensembles de données, baptisés « jeux de données » les obligeant non seulement à compléter leur approche, et les intégrer dans leur schéma de politique documentaire, tout en considérant des scénarii d'usages et de besoins pour valoriser ces jeux et les mettre à disposition des usagers. Parallèlement, ces jeux de données, pour la plupart ouverts et gratuits, obligent les professionnels de l'information et du document à penser des modes de valorisation, d'information et de médiation documentaire (Labelle et Le Corf, 2012) pour les rendre visibles, compréhensibles et accessibles aux publics. C'est ainsi

toute la chaîne d'activités documentaires qui s'en trouve modifiée ; plus que jamais dans l'histoire de la gestion des ressources, la gestion des datas oblige à considérer les autres et les écosystèmes d'information environnants (Millerand, 2012).

De plus, les organisations de recherche et gouvernementales sont amenées à gérer des données dites ouvertes. Ces données ouvertes désignent des chiffres, relevés, mesures, réponses à des enquêtes, statistiques, comptages, et autres données quantitatives collectées par les organismes publics et mises à disposition en format numérique sur des plateformes gouvernementales, régionales ou locales permettant leur accès et leur réutilisation par les citoyens et les entreprises. Ces données sont collectées par des chercheurs, des institutions et peuvent permettre d'élaborer des hypothèses, de les infirmer ou de les valider (Bégault, 2011). Les données de recherche, surtout lorsqu'elles sont issues de projets subventionnés par des organismes de recherche, sont considérées comme devant être accessibles à la communauté des chercheurs et du public. Elles sont perçues comme étant un facteur essentiel de soutien à l'innovation pour les chercheurs, les décideurs et l'industrie, puisqu'elles constituent le point de départ de nouvelles découvertes et recherches (Données de recherche Canada, 2011; Australian National Data Service, 2015). Ces données de recherche, après avoir servi aux fins des projets initiaux, se voient le plus souvent perdues parce que non communiquées ou non traitées (Interagency Working Group on Digital Data, 2009; Fondation canadienne pour l'innovation, 2015). Le défi consiste à transformer des données éparses en collections gérées, repérables et réutilisables (Australian National Data Service, 2015).

Millerand (2012) évoque le travail des professionnels de l'information sur les données de recherche. Ici entre en jeu une séquence d'activités documentaires, notamment la collecte, le traitement, le stockage, l'analyse, l'accès, le partage, la conservation et la réutilisation des données de recherche. Les méthodes et outils documentaires doivent s'adapter aux types de données quantitatives ou qualitatives, structurées ou non, recueillies dans le cadre des projets de recherche, considérant que ces données peuvent être accompagnées non seulement de documents textuels, mais aussi de documents iconographiques, cartographiques, sonores, d'images en mouvement, de dessins d'architecture et de dessins techniques. Diverses méthodes propres aux sciences de l'information peuvent être sollicitées pour gérer et mettre à disposition les données de recherche (Salaün et Arsenault, 2009). Ainsi, les méthodes de traitement incluent l'évaluation de l'information en vue de son élimination ou de sa conservation permanente, la classification (par exemple, hiérarchique ou à facettes), la description et l'indexation de l'information, de même que la conservation, l'accès et la diffusion de l'information. La fouille de textes, pour sa part, permet plutôt « la découverte, à l'aide d'outils informatiques, de nouvelles informations en extrayant différentes données provenant de plusieurs documents textuels » (Hearst, 2003, traduction libre). L'accès à l'information doit quant à lui tenir compte des comportements informationnels des utilisateurs, tout en s'adaptant à la nature des documents (numériques natifs ou numérisés, structurés ou non structurés) et à la structure ouverte du Web pour favoriser le libre accès.

Au cœur du traitement, de l'accès et de la réutilisation des données de recherche, se voit posée la question de leur durabilité qui devrait s'appuyer sur une politique et des modalités de gestion concertées entre chercheurs et professionnels de l'information. Les chercheurs doivent rajuster leurs pratiques en les articulant au travail des professionnels de l'information. Ces derniers, qu'ils soient bibliothécaires, documentalistes ou archivistes, doivent à leur tour coordonner leurs actions afin d'établir un cadre de gouvernance cohérent au bénéfice des chercheurs. Il importe de faciliter, tout en les orientant, les pratiques informationnelles sous-jacentes à une culture d'accès et de partage des données de recherche et, au-delà de celles-ci, de la science en train de se faire.

Références bibliographiques :

Alles, Michael et Miklos A. Vasarhelyi. (2014). Thick data: adding context to big data to enhance auditability. *International Journal of Auditing Technology*, 2(2), p. 95- 108.

Australian National Data Service. (2015). *Better data for Australian research*. [Brochure]. 8 p. Disponible à : http://ands.org.au/__data/assets/pdf_file/0006/387843/better-data-for-australian-research.pdf.

Bégault, Béatrice. (2011). *Enjeux de la diffusion électronique des recherches : pratiques informationnelles et validation des connaissances en sciences de l'ingénieur*. Sarrebruck : Éditions universitaires européennes. 292 p.

Béranger, Jérôme. (2016). La valeur éthique des big data en santé. *Les Cahiers du numérique*, 12, p.109-132.

Bollier, David. (2010). *The promise and peril of big data*. Technical report. Washington, D.C. : The Aspen Institute. 56 p.

Boyd, Dana et Kate Crawford. (2012). Critical questions for big data. *Information, Communication and Society*, 15(5), p.662-679.

Brasseur, Christophe. (2013). *Enjeux et usages du big data : technologies, méthodes et mise en œuvre*. Editions Hermeès. 203 p.

Cardon, Dominique. (2012). *Regarder les données. Multitudes*, 49, p. 138-142.

Cohen, Corine. (2013). *Business intelligence : the effectiveness of strategic intelligence and its impact on the performance of organizations*. Hoboken, NJ : Wiley-ISTE. 304 p.

Colloc, Joël et Bruno Hénocque. (2016). Enjeux du big data et identifications des données médicales, *Les Cahiers du numérique*, 2, 216 p.

Crawford, Kate, Kate Miltner et Mary Gray. (2014). Critiquing big data: politics, ethics, epistemology. *International Journal of Communication*, 8, p. 1663-1672

Données de recherche Canada. (2011). Portrait de la situation des données. *Rapport sur le Sommet 2011 sur les données de recherche canadiennes*. 49 p. Disponible à : http://www.rdc-drc.ca/wp-content/uploads/f-Report-of-the-Canadian-Research-Data-Summit.pdf.

Dumez, Hervé, Étienne Minvielle et Laurie Marrauld. (2015). *Le Big Data en santé*. Note complémentaire N.1 du rapport « État des lieux de l'innovation en santé numérique », 29 mai 2015.

Fernandez, Alain. (2013). *Les nouveaux tableaux de bord des managers : le projet Business Intelligence clés en main*. Paris : Eyrolles. 468 p.

Fondation canadienne pour l'innovation. (2015). *Élaborer la stratégie canadienne sur l'infrastructure de recherche numérique : point de vue de la FCI*. 22 p. Disponible à : http://www.innovation.ca/sites/default/files/Funds/cyber/elaborer-strategie-irn-pour-le-canada-nov6.pdf.

Hafsi, Taïeb et Howard Thomas. (2005). The field of strategy : in search of a walking stick. *European Management Journal*, 23(5), p. 507–519.

Hamel, Marieâ□□Pierre et David Marguerit. (2013). Quelles possibilités offertes par l'analyse des big data pour améliorer les téléservices publics ? *Revue française d'administration publique*, 146(2), p. 437-448

Hearst, Marti. (2013). *What is text mining ?* Unpublished essay, 17 octobre 2003. Disponible à : http://people.ischool.berkeley.edu/~hearst/text-mining.html.

Hoppe, Magnus. (2015). Intelligence as a discipline, not just as a practice. *Journal of Intelligence Studies in Business*, 5(3), p. 47-56.

Interagency Working Group on Digital Data. (2009). *Harnessing the power of digital data for science and society*. Report of the Interagency Working Group on Digital Data to the Committee on Science of the National Science and Technology Council, January 2009. 24 p. Disponible à : https://www.nitrd.gov/about/harnessing_power_web.pdf.

Labelle, Sarah et Jean-Baptiste Le Corf. (2012). Modalités de diffusion et processus documentaires, conditions du « détachement » des informations publiques. Analyse des discours législatifs et des portails open data territoriaux. *Les enjeux de l'information et de la communication*, 3/2, p. 59-71.

McAfee, Andrew et Erik Brynjolfsson. (2012). Big Data : the management revolution. *Harvard Business Review*, 90(10), p. 61-68.

Mercanti-Guérin, Maria. (2013). L'amélioration du reciblage par les Big Data: une aide à

la décision qui menace l'image de marque ? *Revue Internationale d'intelligence Économique*, 5, p. 153-165.

Millerand, Florence. (2012). La science en réseau. *Revue d'anthropologie des connaissances*, 6(1), p. 163-190.

Parks, Malcom. R. (2014). Big data in communication research: its contents and discontents. *Journal of Communication*, 64, p. 355-360

Rouvroy, Antoinette. (2014). Des données sans personne : le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data. Dans France, Conseil d'État. *Étude annuelle 2014 du Conseil d'État – Le numérique et les droits fondamentaux*. Paris : La Documentation française, p. 407-421. Disponible à : http://www.ladocumentationfrancaise.fr/var/storage/rapports-publics/144000541.pdf.

Rudder, Christian. (2014). *Dataclysm: who we are (when we think no one's looking)*. New York, NY: Crown Publishing. 304 p.

Salaün, Jean-Michel et Clément Arsenault (dir.). (2009). *Introduction aux sciences de l'information*. Montréal : Presses de l'Université de Montréal. 235 p.

Summons, Peter et Brian Regan. (2016). Social impact of big data in Australian healthcare. *Les Cahiers du numérique*, 12, p. 13-30.

Sybord, Christine. (2016). Big data et conception d'un système d'aide à la décision clinique. Vers une gestion sociocognicielle de la responsabilité médicale ? *Les Cahiers du numérique*, 12, p. 73-108.

Vayre, Jean-Sébastien. (2014). Manipuler les données. Documenter le marché. Les implications organisationnelles du mouvement big data. *Les Cahiers du Numérique*, 10(1), p. 95-125.

Verlaet, Lise. (2015). La deuxième révolution des systèmes d'information : vers le constructivisme numérique. *Hermès La Revue*, 71(2), p. 249-254.

Zerbib, Romain. (2013). Les modes en gestion : une arme de diffusion massive. *Revue Internationale d'intelligence économique*, 5(2), p. 197-216.