

## Le « Big Data » : opportunité et prudence

Anaïs Thibault-Landry<sup>1</sup>, Lara Manganelli<sup>2</sup> et Kathleen Bentein<sup>3</sup>

<sup>1</sup> Université du Québec à Montréal, [anais.thibaultlandry@gmail.com](mailto:anais.thibaultlandry@gmail.com)

<sup>2</sup> Université du Québec à Montréal

<sup>3</sup> Université du Québec à Montréal

**L'essor de la technologie dans les organisations rend désormais possible la collecte et l'accès à des quantités massives d'informations de diverses natures et provenant de différentes sources (c.-à-d. des mégadonnées ou « Big Data »), permettant ainsi de mesurer et de comprendre sous un nouvel angle la complexité du vécu des employés en milieu de travail. Le premier objectif de cet article est de présenter sommairement la façon dont l'accessibilité croissante des mégadonnées a le potentiel de révolutionner la recherche et la pratique dans le monde du travail. Par ailleurs, le second objectif est d'exposer certaines limites et certaines vigilances à avoir en tant que chercheur, à l'égard des enjeux éthiques liés à l'accès, à l'utilisation, et à l'interprétation de ces mégadonnées. Quelques pistes de réflexion concernant l'élaboration de devis, le consentement lié à la participation et la multiplicité des acteurs impliqués dans ce nouveau type de recherche organisationnelle sont proposées.**

*Mots clés* : mégadonnées (ou « Big Data »), technologie, devis de recherche, enjeux éthiques, consentement.

### Introduction

Qu'il s'agisse d'étudier la satisfaction en emploi ou, à l'opposé, l'épuisement professionnel, mesurer de près le vécu des employés représente depuis longtemps un défi de taille. Dans cette optique, la venue des mégadonnées permet désormais d'élargir les perspectives des chercheurs et des praticiens s'intéressant aux phénomènes organisationnels complexes à étudier en raison de la pluralité des facteurs impliqués et de l'évolution du vécu subjectif des employés à travers le temps, comme le bien-être et l'engagement au travail (Langley, Smallman, Tsoukas, & Van de Ven, 2013; Luciano, Mathieu, Park, & Tannenbaum, 2017). Bien des études en milieu organisationnel reposent jusqu'à présent sur des données issues de sondages réalisés auprès d'échantillons ciblés d'employés, à un moment précis, et portent sur un nombre limité de facteurs, reflétant les postulats de nos questions de recherche présentées ci-dessous (Kozlowski, 2015). Ceci dit, au fur et à mesure que les organisations se tournent vers de nouvelles solutions technologiques, telles que les logiciels de gestion intégrée (p. ex. : ERP, SAP, Oracle) et de relations-clients (p. ex. : CRM, Salesforce) ; les outils d'automatisation de processus (p. ex. : Slack, Teamwork), de travail collaboratif et de gestion de documents communs (p. ex. : Sharepoint, Google Docs) ; les outils virtuels de conférence (p. ex. : Skype, Cisco) et de communication (p. ex. : WhatsApp, Messenger Outlook) ; les plateformes

sociales internes (p. ex. : Yammer, Wiki interne, Jive) ; et les intranets, les mégadonnées, également désignées sous l'appellation anglophone de « Big Data », deviennent de plus en plus accessibles et peuvent désormais nous aider à mieux comprendre l'univers complexe du travail (Davenport, 2014 ; George, Haas, & Pentland, 2014 ; McAbee, Landis, & Burke, 2017 ; Parks, 2014). Ces nouveaux développements requièrent également un certain niveau de prudence et une vigilance en ce qui concerne la manière dont nous concevons nos études, notamment en ce qui a trait aux processus d'élaboration conceptuelle et à la collecte de données, ainsi qu'aux enjeux éthiques liés à la participation à la recherche.

### L'ère de la technologie : La nature et la collecte de mégadonnées

De nouvelles technologies, en termes d'ordinateurs portables, de tablettes, de téléphones, de montres, d'applications mobiles et de logiciels, permettent désormais aux organisations d'amasser des quantités massives de données, offrant de ce fait l'occasion de révolutionner la façon dont les entreprises sont étudiées. On parle généralement de « mégadonnées » ou « Big Data » lorsqu'un ensemble de données possède trois principales caractéristiques : le volume, la vitesse et la variété (Laney, 2001 ; McAbee et al., 2017 ; Tonidandel, King, & Cortina, 2016). Le volume fait référence à la taille ; ainsi, les mégadonnées constituent des ensembles de données d'une taille extrêmement

large, dépassant généralement la capacité d'un ordinateur moyen de l'utilisateur typique et la capacité des logiciels statistiques classiques (p. ex. : SPSS, SAS) pour stocker et analyser les données (Bollier, 2010 ; Fayyad, Piatetsky-Shapiro, & Smyth, 1996 ; McAbee et al., 2017 ; Tonidandel et al., 2016). Ainsi, il est important de souligner que le volume fait autant référence à un échantillon de grande taille (et/ou de l'inclusion de plusieurs variables) qu'à des exigences de stockage de données (p. ex. : téraoctets, pétaoctets, etc.).

Une deuxième grande caractéristique des mégadonnées, soit la vitesse, fait référence à la rapidité avec laquelle les données sont générées, ainsi qu'à la rapidité avec laquelle elles peuvent être traitées pour être analysées et interprétées (McAbee et al., 2017 ; Power, 2014). Par exemple, déjà en 2012, Walmart collectait plus de 2,5 pétaoctets, soit l'équivalent de 1015 bits de données, de transactions par heure (McAfee, Brynjolfsson, Davenport, Patil, & Barton, 2012). Enfin, la troisième caractéristique, la variété, fait référence à la multiplicité de sources et de formes que ces données peuvent prendre, incluant des sources intra- et inter-organisationnelles privées, comme les courriels et les messages instantanés, incluant les données de transactions de clients, et des sources publiques, incluant le contenu des réseaux sociaux et les données financières du marché économique (Chen et al, 2013 ; Guzzo, Fink, King, Tonidandel, & Landis, 2015 ; Laney, 2001 ; McAbee et al., 2017 ; Power, 2014). Récemment, certains chercheurs ont ajouté deux autres caractéristiques aux mégadonnées, à savoir la valeur de celles-ci, en d'autres termes leur utilité (pour la recherche et pour le retour d'investissement dans les organisations), et leur véracité, définissant leur niveau de précision et d'exactitude, afin de mieux les distinguer des autres types de données (De Mauro, Greco, & Grimaldi, 2015).

Nous verrons comment ces mégadonnées sont collectées et peuvent nous permettre, à titre de chercheurs, d'étudier l'expérience des employés dans leurs milieux de travail sous un angle différent et nouveau, pour ainsi mieux comprendre certains phénomènes organisationnels, comme les dynamiques de groupe et d'équipes de travail (George et al., 2014 ; Luciano et al., 2017), la satisfaction au travail (Waber, Olguin, Kim, & Pentland, 2008) ou le stress vécu par les employés (Adams et al., 2014). À titre d'exemple, les nouvelles

technologies rendent à présent simples et rapides la collecte et l'accès à des mégadonnées comme les données comportementales, les données physiologiques et les données textuelles (Luciano et al., 2017). Les données comportementales font référence à des données observables, quantifiables et objectives, représentant la géolocalisation des employés, incluant leurs emplacements et leurs déplacements. L'émergence d'appareils portables comme les téléphones, les montres, les bracelets et autres accessoires intelligents, également appelés « *wearables* », peuvent aussi indiquer la proximité, les mouvements et la posture, ainsi que toute autre information liée à la géolocalisation des employés dans le temps et l'espace (Pentland, 2007 ; Luciano et al., 2017). De récentes études ont démontré que certains comportements non verbaux, incluant les mouvements corporels tels que la gestuelle, les déplacements et le temps passé dans certaines positions, sont associés à la cohésion d'équipe (Hung & Gatica-Perez, 2010) et peuvent servir à prédire la perception des réunions par les employés comme étant collaboratives ou conflictuelles (Bousmalis, Mehu, & Pantic, 2013 ; Gatica-Perez, 2009).

Par exemple, Waber et al. (2008) ont utilisé un appareil portable sociométrique pour collecter divers types de données sur une période d'un mois auprès d'une vingtaine d'employés d'une banque allemande. Cet appareil a collecté une panoplie d'informations sur les activités quotidiennes des employés liées à leur géolocalisation incluant leurs déplacements, leurs mouvements et leurs positions, ainsi que leurs interactions sociales (proximité et fréquence) dans le milieu de travail. Les auteurs ont également collecté et analysé les échanges de courriels des employés. Au total, les auteurs ont collecté 880 courriels, soit 40 courriels par employé, et 2 200 heures de données, soit 100 heures par employé. En s'appuyant sur cette technologie, les auteurs ont pu tirer des conclusions spécifiques sur les comportements adoptés par les individus en milieu de travail. Plus spécifiquement, les résultats de cette étude suggèrent que la quantité totale de communication face à face et virtuelle (via courriels) par employé est négativement associée à sa satisfaction par rapport aux interactions de groupe, ainsi qu'au travail en général. Ce genre de recherche illustre donc la façon dont les mégadonnées peuvent être utiles pour étudier sous un nouvel angle la vie en milieu de travail, incluant l'expérience des employés avec

leurs collègues dans le cadre de la communication entre les membres d'une équipe.

Dans la même optique, les appareils portables sont conçus pour collecter facilement une panoplie de données physiologiques, telles que le rythme cardiaque, la pression artérielle et la température corporelle (Imani et al., 2016). Combinées à la diversité croissante d'applications mobiles disponibles pour ces derniers, ils deviennent des outils fort utiles pour les chercheurs désireux de mieux comprendre l'expérience des employés au travail, comme leurs réactions physiologiques lors de situations de diverses natures, par exemple lors de conflits, de présentations, de travail d'équipe, de réunions, de pauses-café, ou de dîners.

D'autres études menées en laboratoire vont même jusqu'à utiliser des appareils portables pour mesurer le niveau d'activité cérébrale lors de situations de travail d'équipe pour prédire l'émergence d'un leader, ainsi que le niveau d'engagement des autres membres de l'équipe (Waldman et al., 2013 ; Luciano et al., 2017). Plus précisément, Waldman et al. (2013) ont évalué l'engagement individuel, l'émergence de leadership et l'engagement d'équipe auprès de 150 étudiants au MBA, répartis en une trentaine d'équipes, dans une mise en situation fictive portant sur la responsabilité sociale d'une organisation. Lors de cette étude, les chercheurs ont demandé aux membres de chaque équipe de porter des appareils sans fils, légers et capables d'enregistrer le niveau d'activité neurologique, et de remplir un questionnaire d'engagement individuel. Ce faisant, les chercheurs ont pu utiliser cette technologie pour évaluer le niveau d'activité neurologique et le jumeler au niveau subjectif d'engagement et ainsi observer la façon dont ces deux mesures fluctuaient pour chacun des membres des équipes tout au long de la simulation de travail d'équipe. Les évaluations des membres de l'équipe ont par ailleurs permis l'identification d'un leader pour chaque équipe. En ce sens, les résultats de l'étude suggèrent que, lorsqu'un membre est très engagé à la fois au niveau subjectif (de manière autorapportée) et neurologique (tel qu'enregistré par l'appareil portable), ses coéquipiers ont tendance à le considérer comme un leader. De plus, les membres identifiés par leurs coéquipiers comme ayant le plus de leadership démontraient le plus haut niveau d'activité neurologique lorsqu'ils prenaient la parole dans leur groupe.

Néanmoins, comme l'illustre l'étude de Waldman et al. (2013), la plupart des études reposant sur des données physiologiques restent limitées au contexte de laboratoire. Certains chercheurs commencent à utiliser des technologies portables semblables, quoique simplifiées, pour collecter de telles données à l'extérieur des laboratoires, dans les environnements de travail. Par exemple, Adams et al. (2014) ont utilisé une application pour téléphones intelligents (SESAME - Échantillonnage et mesure de l'expérience de stress) pour collecter des données reflétant les fluctuations de niveau de stress des employés. Dans leur étude, ils ont collecté des données non verbales (intonation indiquant un ton stressé ou non) via des capteurs sur les téléphones mobiles, ainsi que des données physiologiques (activité électroépidermique indiquant le stress) grâce à des appareils portés au poignet, qu'ils ont combinées à des mesures d'autoévaluation de stress et d'émotions générales. Pendant dix jours, sept participants ont été invités à utiliser l'application de 8h00 à 23h00 et à s'autoévaluer en réponse à des notifications sur leurs appareils intelligents. Les résultats de l'étude offrent une démonstration préliminaire de l'utilité et de l'efficacité d'une telle application pour évaluer les fluctuations quotidiennes des niveaux de stress des individus. Plus précisément, leurs résultats ont révélé que les capteurs de stress non verbal et les données électroépidermiques produisaient des mesures de stress corrélant significativement avec les autoévaluations des participants. Ainsi, de tels capteurs pourraient constituer des médiums plus efficaces et moins invasifs pour évaluer avec précision le niveau de stress en milieu de travail. En somme, comme l'illustre l'exemple ci-haut, de plus en plus d'études pourraient être menées pour mieux évaluer l'utilisation et l'efficacité de ce type d'appareils pour aider les employés à surveiller en temps réel l'évolution de leur niveau de stress au travail.

Enfin, en ce qui concerne les données textuelles, les appareils intelligents, tels que les téléphones, les tablettes et les ordinateurs de toute taille, en jonction avec l'émergence de panoplies de logiciels et d'applications de communication et de gestion de documents, permettent d'amasser une quantité impressionnante d'informations textuelles provenant d'une multitude de sources, comme les services de messageries instantanées et de courriels, ainsi que les plateformes sociales internes

et publiques. Les chercheurs commencent également à utiliser la vaste quantité de données textuelles disponibles sur les plateformes sociales publiques comme Twitter pour mieux comprendre le bien-être psychologique lié au travail. Par exemple, Wang, Hernandez, Newman, Hem et Bian (2016) ont analysé plus de 2 millions de Tweets afin de comprendre l'influence des journées de congé (par exemple, les jours de fin de semaine) sur le niveau de stress et de récupération des individus après une semaine de travail. Les auteurs ont procédé à l'analyse linguistique du contenu des messages textuels envoyés tous les jours pendant 18 mois. Leurs analyses ont révélé deux dimensions significatives et importantes pour la récupération des travailleurs : (a) une dimension liée aux émotions négatives, incluant le stress et la somatisation, et (b) une dimension liée aux émotions positives, incluant les repas, la famille et les loisirs. Les résultats ont révélé que les émotions positives diminuaient de façon significative en milieu de semaine, de mardi à jeudi, puis augmentaient pendant les fins de semaine, de vendredi à samedi, et s'accompagnaient d'une baisse significative d'émotions négatives et de stress au travail les vendredis.

### **Comment les mégadonnées peuvent-elles aider les chercheurs à mieux appréhender l'expérience des employés ?**

Les mégadonnées provenant des organisations offrent un certain nombre d'avantages en comparaison aux données collectées via des médiums alternatifs comme les questionnaires, les sondages et les enquêtes d'opinions. Avant les développements technologiques de la dernière décennie, une grande majorité de la recherche conduite en organisation était limitée par les méthodes de collecte de données existantes. En raison de leurs caractéristiques, les mégadonnées permettent de saisir avec précision d'autres dimensions de la complexité inhérente au monde du travail.

Pour illustrer ceci, prenons la notion du temps. Jusqu'à tout récemment, bien des questions de recherche devaient omettre l'aspect temporel et être étudiées de façon statique en s'appuyant sur des données de sondages et de questionnaires recueillies à un moment circonscrit dans le temps (Kozlowski, 2015; McGrath & Tschan, 2007; Luciano et al., 2017). Grâce à leur flux continu, les

mégadonnées peuvent être utilisées pour mener des études intensives comportant de nombreux points de données et des mesures répétées (Kozlowski et al., 2013; Luciano et al., 2017). Nos questions de recherche n'ont plus besoin de se limiter à quelques semaines et peuvent dorénavant prendre en compte plusieurs mois, voire des années. Ainsi, l'utilisation de mégadonnées permet d'élargir le champ temporel de nos recherches. En contrepartie, ceci requiert une plus grande considération conceptuelle entourant la notion de temps, et donc de fréquence de mesure. En ce sens, les considérations liées à la temporalité, incluant le nombre, la fréquence et l'espacement entre les différents moments de collecte de données, ne devraient pas être uniquement basées sur les possibilités d'ordre technique, mais plutôt sur la compréhension et la conceptualisation théorique du phénomène en question. Dans une même veine, il faut à présent consciemment choisir la période durant laquelle circonscrire la collecte des mégadonnées. L'accès et l'utilisation de mégadonnées impliquent donc que nous soyons dorénavant plus attentifs à l'aspect temporel, en prenant en compte l'importance d'une perspective longitudinale et de mesures répétées, dans nos questions de recherche et que nous en tenions compte dans l'élaboration de nos devis de recherche et dans nos choix méthodologiques.

De plus, jusqu'à présent, les phénomènes organisationnels, tels que le climat de travail (Kozlowski et al., 2013), ont également largement été étudiés de façon quantitative, en agrégeant les scores individuels des employés ayant répondu aux questionnaires, aux sondages et aux enquêtes d'opinion. L'avenue des mégadonnées encourage l'adoption d'une méthodologie qui permet d'incorporer des sources de données non numériques, afin de les transformer plus aisément en données quantitatives par la suite (Tonidandel et al., 2016). De plus, l'accès et l'utilisation de mégadonnées permettent de dépasser le cadre d'une seule source de données (p. ex. : données autorapportées) pour faciliter l'inclusion de différentes sources (p.ex. : données objectives, ou rapportées par les collègues ou les superviseurs). Plus précisément, l'évolution de nos méthodes de recherche pour étudier le monde du travail implique la possibilité de combiner différentes sources d'information, telle que dans le cas de communications intra- et inter-organisationnelles, les données provenant de divers acteurs (superviseurs, collègues, employés, et/ou clients,

fournisseurs, usagers) communiquant via différents médiums (courriels, messagerie instantanée, appels, téléconférences).

Illustrant ceci, une étude de Pentland (2012) a collecté des données de géolocalisation et de communication entre les membres d'une équipe grâce à un appareil portable sociométrique permettant de mesurer la fréquence des interactions et des communications verbales face à face sur plusieurs jours consécutifs de travail. Les auteurs ont également mesuré plusieurs variables, incluant l'intelligence, la personnalité et les habiletés des employés, ainsi que le contenu des discussions entre les membres des équipes. En analysant ces différents types de mégadonnées, cette étude a permis d'identifier des facteurs clés menant à un plus haut niveau de rendement d'équipe. Elle a ainsi mis en évidence l'importance des communications informelles entre collègues pour la satisfaction et le rendement des membres des équipes au travail. Comme on le voit avec l'étude de Pentland (2012), l'utilisation des mégadonnées dans le cadre d'un devis misant sur différents types de mégadonnées peut contribuer à l'élaboration de conclusions plus riches et complètes du vécu des employés en organisations.

Dans cette optique, on voit également que les mégadonnées peuvent fournir un niveau additionnel de précision et de granularité de mesure, puisqu'elles permettent d'avoir accès à des éléments du vécu expérientiel des employés qui sont plus difficilement observables et mesurables. Les mégadonnées peuvent donc aider à contrer certains biais de présentation associés à des méthodologies de recherche alternatives reposant sur des évaluations subjectives ou rapportées directement par les employés. En ce sens, contrevalider les constats issus d'une recherche basée sur des mesures autorapportées, par exemple provenant d'un sondage sur la satisfaction des employés vis-à-vis de leur climat de travail, à l'aide de mégadonnées obtenues de façon plus impartiale en codant les données textuelles de courriels internes, pourrait permettre d'observer le degré de convergence et de divergence entre la façon dont les employés évaluent qualitativement le climat de leur milieu de travail lorsqu'on leur présente des dimensions spécifiques à l'aide d'échelles validées et la façon dont ils en parlent naturellement entre collègues. Dans cette optique, des développements récents de logiciels spécialisés dans l'analyse de texte assistée par ordinateur comme AutoMap

(Carley, Columbus, & Landwehr, 2013 ; IBM, 2015) ont permis de faire des avancements importants dans le traitement des communications échangées verbalement et textuellement entre les employés. Ces logiciels permettent désormais d'identifier des thèmes généraux de manière systématique, en plus de pouvoir fournir des informations sur le degré de personnalisation, la ponctuation, et même la valence, le ton et les émotions associés au discours (Krippendorff, 2004 ; Luciano et al., 2017). Ainsi, en plus de pouvoir déceler la présence d'écarts éventuels entre les évaluations formelles et le discours spontané, ceci pourrait même permettre d'explorer et d'analyser la nature de ces écarts.

Une étude illustrant ceci est celle de Jehn et Bezrukova (2004) dans laquelle ils ont analysé les rapports écrits par des superviseurs pour identifier l'effet de facteurs comme la stratégie adoptée par l'organisation et l'accès des employés à de la formation sur la diversité culturelle, dans la relation entre la diversité culturelle des équipes de travail et leur niveau de performance. Plus précisément, ils ont analysé des rapports concernant plus de 10 000 employés dans plus de 1 500 équipes de travail d'une entreprise nommée dans « Fortune 500 ». Les compétences évaluées dans les rapports des superviseurs ont été utilisées comme mesure indirecte du contexte du groupe reflétant la culture, les stratégies et les pratiques en matière de ressources humaines. Les résultats ont révélé, entre autres, que les primes monétaires les plus élevées étaient accordées aux membres d'équipes valorisant davantage une culture collectiviste (orientée vers la collaboration entre les membres), tandis que les primes monétaires les plus basses étaient accordées dans les équipes où l'on survalorisait des stratégies organisationnelles axées sur la stabilité financière.

Enfin, nous pouvons prendre avantage du fait que les mégadonnées permettent à présent d'intégrer des données allant d'un contexte limité au cadre professionnel, par exemple le niveau de ressources et de requêtes au travail, à des contextes plus larges, tels que le temps et la distance parcourue en transport en commun pour se rendre au travail, pour comprendre le vécu des employés et même pour mieux prédire — pour pouvoir prévenir et intervenir — le risque de stress accru, d'épuisement professionnel, de désengagement ou de démission des employés. Cette intégration d'information permettrait d'avoir une compréhension plus précise quant à l'impact de divers facteurs sur la

santé psychologique et le vécu des employés dans leurs milieux de travail (p. ex. : l'influence de la vie personnelle sur la performance au travail). Ceci dit, bien que l'accès et l'utilisation d'information issue des mégadonnées offrent des occasions prometteuses, ils soulèvent également des considérations importantes quant aux enjeux éthiques et aux droits de la personne, tels que le respect de la vie privée, le consentement libre et éclairé, ainsi que la propriété des données.

### **Les limites de l'accès et de l'utilisation du « Big Data » : enjeux éthiques en recherche**

L'un des principaux rôles des chercheurs, psychologues de formation, étant de veiller au bien-être des individus, cette nouvelle ère d'essor technologique accentue l'importance de nous porter garants de normes de consentement, de confidentialité, de respect de la dignité et/ou d'anonymat dans le contexte de la recherche en organisation. Plus précisément, l'émergence de la technologie avec toutes les opportunités qu'elle offre en matière d'accès à des données de toutes sortes rend cruciale une prise de conscience entourant les enjeux et les considérations éthiques associés à l'accès, l'utilisation et l'interprétation des mégadonnées. Par leur formation et leur code déontologique, les psychologues organisationnels auront donc un rôle clé à jouer pour aider à identifier les meilleures pratiques en termes de protection des droits des participants et de consentement à la recherche. Ceci inclut de déterminer les acteurs auprès desquels obtenir le consentement pour avoir accès aux données, par exemple le fournisseur de la solution technologique, l'organisation mandataire et l'individu à l'emploi de l'organisation. Similairement, comme il deviendra possible, en croisant les informations collectées dans plusieurs bases de données, d'identifier les individus, il faudra faire attention tout particulièrement à mettre en place des solutions pour préserver la confidentialité des données des participants (McAbee et al., 2017). Il faudra également trouver une façon d'informer tous les usagers des différents produits et solutions technologiques pour respecter la définition d'un consentement éclairé, puisque les mégadonnées sont souvent collectées sans consentement préalable, sans objectif spécifique, et sans même que les individus soient au courant que leurs données sont utilisées dans le cadre de recherches.

En outre, à titre de chercheurs, nous serons donc responsables non seulement de questionner nos pratiques en faisant preuve de rigueur, d'ouverture d'esprit, de prudence, et de bienveillance, et d'acquiescer une expertise propre quant à ces enjeux, mais aussi d'établir des partenariats avec des experts légaux afin de mieux comprendre les lois relatives à l'accès aux données, au partage et à la divulgation d'informations.

De plus, malgré l'enthousiasme que le « *Big Data* » peut générer, on ne peut mettre de côté l'importance pour les chercheurs organisationnels de rester conscients des biais méthodologiques pouvant découler de l'utilisation de mégadonnées. Ainsi nous devons également nous porter garants de protéger les postulats de recherche, par exemple en trouvant des explications alternatives potentielles et en faisant attention à ne pas tirer de conclusions causales (McAbee et al., 2017). En ce sens, il sera important de ne pas perdre de vue la possibilité de biais introduits par les méthodes d'échantillonnage de participants et les caractéristiques de collecte des mégadonnées, qui peuvent avoir des implications critiques pour la qualité des observations et des conclusions émergeant de ces données (Antonakis, Bendahan, Jacquart, & Lalive, 2010; Tonidandel et al., 2016). Ceci est d'autant plus vrai quant aux conclusions pouvant découler de la recherche où les mégadonnées sont collectées dans un contexte autre que la question de recherche initiale.

Par exemple, il pourrait y avoir un biais de sélection, ou d'échantillonnage, si le devis de recherche repose sur l'utilisation de données collectées à partir de montres intelligentes, car les employés présentant certaines caractéristiques sociodémographiques pourraient être plus enclins à posséder et à porter ce type de montres et ainsi être surreprésentés par rapport à ceux n'en portant pas (par exemple, si les hommes avaient plus tendance à posséder une montre intelligente ou à la porter durant leur journée de travail). Cette différence pourrait alors introduire un biais méthodologique dans une question de recherche portant sur le lien entre la proximité physique et la cohésion des membres d'une équipe de travail. Dans un même ordre d'idées, certaines technologies sont plus facilement accessibles et plus couramment utilisées dans certains pays ou à certains niveaux socioéconomiques (Antonakis et al., 2010).

Ces considérations limitent donc l'interprétation et la généralisation des résultats de recherche et requièrent impérativement d'être prises en compte dans les analyses statistiques sous forme de variables modératrices reflétant les caractéristiques des échantillons de participants, des sources de collecte de données et dans l'interprétation des constats obtenus. Négliger de prendre en compte des tels facteurs pourrait indéniablement biaiser et influencer la qualité des conclusions tirées (Antonakis et al., 2010; Tonidandel et al., 2016).

Les mégadonnées encouragent également une révision de notre façon d'élaborer et de tester nos théories afin d'intégrer une approche à la fois inductive et déductive (Kitchin, 2014; Tonidandel et al., 2016). Plus précisément, cette révision implique l'élaboration d'hypothèses sur une base déductive en s'appuyant sur des connaissances et des théories existantes. Tester empiriquement ces hypothèses pourrait ensuite être effectué en collectant des données et en les couplant à diverses sources de mégadonnées (approche déductive). Les techniques d'analyse statistique des mégadonnées pourraient aussi être utilisées pour valider les observations et explorer des facteurs émergeant des données, qui autrement n'auraient pas été découverts ni pris en compte (approche inductive). En utilisant de façon complémentaire ces deux approches déductive et inductive, les chercheurs seront plus habiles à réviser leurs hypothèses et à les ajuster en fonction de ce qu'ils observent dans les données à travers différents modèles conceptuels et statistiques. Cette approche intégrant des processus déductifs et des processus inductifs pourrait ainsi améliorer nos pratiques de recherche et nos théories en nous permettant d'incorporer de nouvelles variables, de tester des modèles alternatifs, de reproduire les résultats et d'adapter les modèles avec le temps et le cumul d'évidence, pour ultimement parvenir à une meilleure compréhension des phénomènes organisationnels sous investigation (Tonidandel et al., 2016).

D'un point de vue pratique, les organisations ont également un rôle important à jouer dans la collecte et l'utilisation des mégadonnées. Selon Laviolette (2016), jusqu'à très récemment, les mégadonnées n'étaient accessibles qu'à de grandes organisations comme Google et Amazon, mais les organisations de toute taille ont maintenant de plus en plus d'occasions de collecter ce type de données. Il est

donc à leur avantage de créer des partenariats avec des chercheurs et/ou des experts en analyse de mégadonnées afin de réfléchir aux divers types, sources et natures des mégadonnées avant qu'elles ne soient collectées pour ainsi pleinement profiter de la technologie. À cet effet, les organisations devraient s'attarder davantage à cerner avec précision la façon dont divers types de mégadonnées pourraient leur être utiles pour analyser l'atteinte de leurs objectifs organisationnels (par exemple : la satisfaction des besoins et des attentes de la clientèle; la mobilisation et le bien-être du personnel; la gestion efficace des processus organisationnels, etc.) et, ce faisant, éviter d'investir du temps, de l'argent, et plus généralement des ressources dans la collecte de données peu utiles.

## Conclusion

En conclusion, avec l'essor de la technologie dans les milieux organisationnels, les chercheurs organisationnels ont désormais accès au « *Big Data* », soit à des quantités massives d'informations appelées « mégadonnées » qui peuvent être utilisées pour mieux comprendre le vécu complexe des employés en milieu de travail. La nature de ces mégadonnées divergeant considérablement de ce que nos méthodes de recherche nous offraient comme données jusqu'à présent, il semble venu le temps d'adapter et d'élargir notre façon de conceptualiser et de conduire la recherche en milieu organisationnel. Dans ce bref article, nous invitons les chercheurs et les praticiens à réfléchir à l'évolution des méthodes et de la conduite en recherche dans une ère où le « *Big Data* » est sur toutes les lèvres.

## Références

- Adams, P., Rabbi, M., Rahman, T., Matthews, M., Volda, A., Gay, G., ... Volda, S. (2014). *Towards Personal Stress Informatics: Comparing Minimally Invasive Techniques for Measuring Daily Stress in the Wild*. Proceedings of the 8<sup>th</sup> International Conference on Pervasive Computing Technologies for Healthcare. doi:10.4108/icst.pervasivehealth.2014.254959
- Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2010). On making causal claims: A review and

- recommendations. *Leadership Quarterly*, 21, 1086-1120. doi:10.1016/j.leaqua.2010.10.010
- Bollier, D. (2010). *The promise and peril of big data*. Washington, DC: The Aspen Institute.
- Bousmalis, K., Mehu, M., & Pantic, M. (2013). Towards the automatic detection of spontaneous agreement and disagreement based on nonverbal behaviour: A survey of related cues, databases, and tools. *Image and Vision Computing*, 31(2), 203-221. doi:10.1016/j.imavis.2012.07.003
- Carley, K. M., Columbus, D., & Landwehr, P. (2013). *Automap user's guide*. Pittsburgh, PA: Carnegie Mellon University, School of Computer Science, Institute for Software Research.
- Chen, J., Chen, Y., Du, X., Li, C., Lu, J., Zhao, S., & Zhou, X. (2013). Big data challenge: a data management perspective. *Frontiers of Computer Science*, 7(2), 157-164. doi:10.1007/s11704-013-3903-7
- Davenport, T. H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. USA: Harvard Business School.
- De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. In *AIP conference proceedings* (Vol. 1644, No. 1, pp. 97-104). AIP.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17, 37-54.
- Gatica-Perez, D. (2009). Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12), 1775-1787. doi:10.1016/j.imavis.2009.01.004
- George, G., Haas, M. R., & Pentland, A. (2014). Big Data and Management. *Academy of Management Journal*, 57(2), 321-326. doi:10.5465/amj.2014.4002
- Guzzo, R. A., Fink, A. A., King, E., Tonidandel, S., & Landis, R. S. (2015). Big Data Recommendations for Industrial-Organizational Psychology. *Industrial and Organizational Psychology*, 8(04), 491-508. doi:10.1017/iop.2015.40
- Hung, H., & Gatica-Perez, D. (2010). Estimating Cohesion in Small Groups Using Audio-Visual Nonverbal Behavior. *IEEE Transactions on Multimedia*, 12(6), 563-575. doi:10.1109/tmm.2010.2055233
- IBM. (2015). *IBM SPSS Modeler 17.0 user's guide*. Armonk, NY: Author.
- Imani, S., Bandodkar, A. J., Mohan, A. M. V., Kumar, R., Yu, S., Wang, J., & Mercier, P. P. (2016). A wearable chemical-electrophysiological hybrid biosensing system for real-time health and fitness monitoring. *Nature Communications*, 7(1). doi:10.1038/ncomms11650
- Jehn, K. A., & Bezrukova, K. (2004). A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior*, 25(6), 703-729. doi:10.1002/job.257
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1, 1-12. doi:10.1177/2053951714528481
- Kozlowski, S. W. J. (2015). Advancing research on team process dynamics. *Organizational Psychology Review*, 5(4), 270-299. doi:10.1177/2041386614533586
- Kozlowski, S. W. J., Chao, G. T., Grand, J. A., Braun, M. T., & Kuljanin, G. (2013). Advancing Multilevel Research Design. *Organizational Research Methods*, 16(4), 581-615. doi:10.1177/1094428113493119
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology (2nd ed.)*. Thousand Oaks, CA: Sage.
- Laney, D. (2001). *3D data management: Controlling data volume, velocity and variety*. Repéré à : <https://blogs.gartner.com/douglaney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Langley, A., Smallman, C., Tsoukas, H., & Van de Ven, A. H. (2013). Process studies of change in organization and management: Unveiling temporality, activity, and flow. *Academy of Management Journal*, 56(1), 1-13. doi:10.5465/amj.2013.4001



- Lavolette, F. (2016, 18 février). *La recherche et le poids des données massives*. Découvrir magazine, Dossier: Chiffres, Acfas.
- Luciano, M. M., Mathieu, J. E., Park, S., & Tannenbaum, S. I. (2017). A Fitting Approach to Construct and Measurement Alignment. *Organizational Research Methods*, *21*(3), 592–632. doi:10.1177/1094428117728372
- McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, *27*(2), 277–290. doi:10.1016/j.hrmmr.2016.08.005
- McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard business review*, *90*(10), 60–68.
- McGrath, J. E., & Tschan, F. (2007). Temporal matters in the study of work groups in organizations. *The Psychologist-Manager Journal*, *10*(1), 3–12. doi:10.1080/10887150709336609
- Parks, M. R. (2014). Big Data in Communication Research: Its Contents and Discontents. *Journal of Communication*, *64*(2), 355–360. doi:10.1111/jcom.12090
- Pentland, A. (Sandy). (2007). Automatic mapping and modeling of human networks. *Physica A: Statistical Mechanics and Its Applications*, *378*(1), 59–67. doi:10.1016/j.physa.2006.11.046
- Pentland, A. (2012). The new science of building great teams. *Harvard Business Review*, *90*(4), 60–69.
- Power, D. J. (2014). Using “Big Data” for analytics and decision support. *Journal of Decision Systems*, *23*(2), 222–228. doi:10.1080/12460125.2014.888848
- Tonidandel, S., King, E. B., & Cortina, J. M. (2016). Big Data Methods. *Organizational Research Methods*, *21*(3), 525–547. doi:10.1177/1094428116677299
- Waber, B. N., Olguin Olguin, D., Kim, T., & Pentland, A. (2008). Understanding Organizational Behavior with Wearable Sensing Technology. *SSRN Electronic Journal*. doi:10.2139/ssrn.1263992
- Waldman, D. A., Wang, D., Stikic, M., Berka, C., Balthazard, P. A., Richardson, T., ... Maak, T. (2013). Emergent Leadership and Team Engagement: An Application of Neuroscience Technology and Methods. *Academy of Management Proceedings*, *2013*(1), 12966. doi:10.5465/ambpp.2013.63
- Wang, W., Hernandez, I., Newman, D. A., He, J., & Bian, J. (2016). Twitter Analysis: Studying US Weekly Trends in Work Stress and Emotion. *Applied Psychology*, *65*(2), 355–378. doi:10.1111/apps.12065