

## DES RÉDACTEURS EN CHEF

### NOUVELLES FAÇONS DE VOIR LES BIG DATA

Peu de sujets ont reçu une attention aussi récente des chercheurs de toutes disciplines, des praticiens, des décideurs politiques et des médias " **Big Data**. " **Pourtant, à partir de nos expériences sur le Journal de l'Academy of Management** L'équipe éditoriale, nous pensons que beaucoup d'ambiguïté et même de confusion règne toujours autour de questions clés telles que: qu'est-ce que le Big Data englobe? La grande datamean signifie-t-elle la fin de la théorie? En quoi la recherche sur les mégadonnées diffère-t-elle des méthodes scientifiques conventionnelles de recherche en gestion? Que faut-il pour étudier les mégadonnées topublisha dans les revues de gestion?

Par conséquent, notre objectif dans cet éditorial est d'offrir un aperçu des **mégadonnées aligné avec notre équipe éditoriale " s se concentrer sur " de nouvelles façons de voir. " Nous reconnaissons facilement que les** mégadonnées peuvent étendre notre portée théorique et élargir le répertoire d'approches méthodologiques pour étudier les phénomènes de gestion de nouvelles façons. En tant que phénomène commercial omniprésent mais émergent, les mégadonnées offrent aux chercheurs en gestion des opportunités fructueuses non seulement de remettre en question, de modifier et d'étendre les théories existantes, mais aussi d'informer la pratique des mégadonnées par le biais d'enquêtes systématiques. Le Big Data fournit également des outils d'analyse et de visualisation précieux pour compléter, turbocompresser et même transformer certains domaines de la recherche en **gestion - telles que l'utilisation de données non structurées, le traitement des données en temps réel et la reconnaissance des formes**. Dans le même temps, les mégadonnées obligent également à revoir certaines hypothèses, pratiques, processus et outils de recherche développés dans le contexte de considérations de données limitées.

Nous pensons que le domaine sera dans une position plus forte pour **tirer parti des opportunités du Big Data - et pour éviter les pièges - lorsque nous transférons non seulement des** connaissances d'autres disciplines, mais nous nous engageons également dans la coproduction de connaissances sur les mégadonnées. À cette fin, nous commençons par clarifier la logique des mégadonnées dans une perspective de recherche, en faisant valoir que les chercheurs en gestion peuvent enrichir la perspective de plusieurs façons importantes. Nous décrivons ensuite les opportunités de recherche qui peuvent tirer parti des forces du big data et de la bourse de gestion à l'avantage mutuel. Nous concluons ensuite l'éditorial par une multitude de suggestions pour surmonter les obstacles de base

**à publier ces opportunités de recherche dans le domaine " s revues. Ensemble, les trois thèmes - ( 1) enrichir la perspective, (2) tirer parti des atouts de l'avantage mutuel et (3) surmonter les obstacles à la publication - nous permettent d'offrir de nouvelles perspectives sur la manière et la manière dont les bourses de gestion peuvent façonner le contenu et la trajectoire évolutive des connaissances sur les mégadonnées. Ils permettent également une discussion intégrée sur les problèmes fondamentaux des mégadonnées -**

le paradigmatique et le méthodologique, ainsi que le conceptuel et le phénoménologique.

Notre message est un optimisme tempéré de réalisme. Nous pensons **que des approches de recherche innovantes simultanément exploiter** la puissance des mégadonnées et la pluralité d'approches théoriques et empiriques peuvent se compléter pour faire progresser à la fois la recherche en gestion et les pratiques de mégadonnées. Mais, même si le besoin et les opportunités pour de telles innovations sont multiples, ils restent des activités complexes, difficiles et peut-être risquées pour les chercheurs individuels. Nous espérons que cet éditorial servira de tremplin pour ceux qui souhaitent faire passer la conversation d'un accent unidirectionnel sur les implications des mégadonnées à un dialogue bidirectionnel pour faire avancer à la fois les mégadonnées et l'érudition en gestion.

#### ENRICHIR LA PERSPECTIVE

Nous suggérons que la réflexion sur les données volumineuses en tant qu'approche de recherche devrait faire partie de tout dialogue sur les mégadonnées sur le terrain en raison des implications qu'elles ont pour les questions de recherche, la construction de modèles, la conception de la recherche, la collecte de données et l'analyse et la visualisation des données. La perspective a été caractérisée de plusieurs manières, comme suit: (a) de la théorie ou des données à petit échantillon devant être interprétées par les humains au traitement de grandes quantités de données pour atteindre des découvertes basées sur les données (Eragal & Klischewski,

2017); (b) de la causalité aux schémas et corrélations dans les données (Mayer-Sch" onberger & Cukier, 2013); (c) de tester une théorie aux idées nées des données (Kitchin, 2014); et d) l'importance et le statut acquis par les données en tant que produit et produit reconnu (Leonelli, 2014). Toutes ces descriptions semblent raisonnables, mais, à notre avis,

**ancrer vraiment l'approche est la " loi des grands nombres "- l'idée que,** avec suffisamment de données et d'échantillons, les erreurs (incertitude) sont vouées à la certitude (Succi et Coveney, 2018). Comme l'a déclaré **Cohen (2013: 1921), " Big Data " Les prétentions au privilège** épistémologique découlent de sa fidélité affirmée à la réalité à un niveau de détail très élevé. "

**Le point de vue comme une sorte de " empiricismon stériles " qui** implique de recueillir et de parcourir des données pour trouver des modèles et faire des prédictions sur les dépendances et les causes (Frick

e, 2015; Sætra, 2018). nous

observent également que, jusqu'à présent, les applications de Big Data **semblent avoir la question de ce qui se passe actuellement et se produira probablement ensuite. Par exemple, un objectif commun n'est pas de savoir** pourquoi une seule variable peut expliquer une variable de résultat, mais comment le résultat varie avec de nombreux prédictors potentiels -

avec ou sans théorie quant aux prédictors pertinents (Einav & Levin, 2014).

Nous nous opposerions à l'affirmation selon laquelle la perspective diminue l'importance de l'adéquation causale et de la profondeur de la recherche. Parce que **les données sont un moyen pour une fin, les mégadonnées " s l'informativité** pour parvenir à des conclusions justifiables est plus importante que son volume, sa vitesse ou sa variété (Bowman, 2018). Une constatation corrélative, par exemple, peut ne pas se transformer en une constatation causale en augmentant simplement le volume, la variété et la vitesse des données sous-jacentes. Le vrai problème, cependant, n'est pas les données en soi, mais la perspective qui sous-tend la manière dont les données sont considérées, collectées, conservées et examinées (Coveney, Dougherty et Highfield, 2016). Plus précisément, nous sommes d'accord avec d'autres pour dire que l'affirmation selon laquelle les chercheurs n'ont pas besoin de commencer par la théorie mais pourraient plutôt acquérir des perspectives et des explications plus objectives à partir de modèles et d'analyses de mégadonnées est ténue et peu convaincante (Chan et Moses, 2016; Sætra, 2018).

Au contraire, étant donné la complexité et les ressources nécessaires pour accéder et traiter les ensembles de données volumineuses, il nous semble que poser les bonnes questions est crucial. Sans théorie guidant **les questions, une explication de ce qui se passe et pourquoi, peuvent ne pas être traités de manière adéquate.** De plus, une capacité accrue à détecter les corrélations et les grappes dans les données ne peut guère se substituer à la théorie pour fournir une base plus solide permettant d'éviter les erreurs et de tirer des déductions appropriées de ces corrélations. Sans théorie, par conséquent, les approches purement basées sur le Big Data dans le domaine de la gestion pourraient systématiquement échouer à fournir des comptes conceptuels des **phénomènes et processus de gestion auxquels ils sont appliqués. - comme** cela a été observé avec d'autres

ainsi que la biologie et la médecine (Coveney et al., 2016).

En effet, l'idéal de l'empirisme pur ou de l'induction pure fonctionne rarement en soi, et aucune théorie ne peut être assez bonne pour supplanter le besoin de données et de tests (Calude et Longo, 2017). Ainsi, peut-être que l'interprétation et l'utilisation des mégadonnées **comme perspective devraient ressembler à ce que " enlèvement " ( la** combinaison de logiques déductives et inductives pour dériver des inférences causales). Si oui, quelles sont les implications pour notre modèle prédominant de production et d'utilisation des connaissances? La recherche abductive implique une logique de découverte et de doute (Locke, Golden-Biddle et Feldman, 2008), et de telles dispositions et capacités méritent une plus grande attention avec l'utilisation des mégadonnées. Par exemple, comment les modèles de mégadonnées pourraient-ils servir de source pour le développement d'une nouvelle théorie, qui est ensuite élaborée et testée de manière déductive? Et, plus largement, comment pourrait-on rendre la perspective des mégadonnées davantage axée sur la théorie pour étudier les phénomènes de gestion?

La question de savoir quand et dans quelles conditions une approche Big **Data pourrait produire des informations exploitables sur le plan " plus petite " des** données de haute qualité, et vice versa, sont également intéressantes à considérer. Nous pensons que, à mesure que la complexité et l'ambiguïté de la situation augmentent dans une décision, un processus ou un système organisationnel, l'avantage comparatif des mégadonnées peut diminuer, en particulier lorsque la qualité des données est mitigée et qu'il existe des biais systématiques (inconnus). La collecte de données auprès de millions d'individus peut apporter peu d'avantages dans l'amélioration de la précision prédictive, par exemple, si seul un sous-ensemble entraîne la plus grande variance des données. Plus largement, bigdatamight ne fonctionne pas bien si la qualité des données ne permet pas une véritable répliquabilité des modèles et une riche compréhension des sources spécifiques d'instabilité dans les modèles (Oswald & Putka, 2016). Comme Succi et Coveney (2018: 11) l'ont fait remarquer:

En fin de compte, la plupart des [mégadonnées] se résument à des formes plus ou moins sophistiquées d'ajustement de courbe basées sur la minimisation des erreurs. De telles procédures de minimisation fonctionnent bien si le paysage des erreurs est lisse, mais elles présentent une fragilité vis-à-vis des ondulations dans d'autres situations, qui sont la règle dans les systèmes complexes. Enfin, la perspective exige intrinsèquement que le processus d'exploration des données soit contextuellement informé, mais le contexte plus **large est souvent entièrement ignoré. Que pourrait faire la recherche en gestion - en particulier, les chercheurs qualitatifs - dire sur le comment et le pourquoi du contexte dans la perspective " enrichissement? Sans ces** enrichissements, les mégadonnées

les modèles de recherche en gestion pourraient connaître des progrès **lents et un taux d'échec plus élevé, tout en gênant le chercheur " s capacité à comprendre les échecs " causes profondes.**

#### TIRER PARTI DES FORCES POUR AVANTAGE MUTUEL

Au-delà d'une perspective plus riche, nous encourageons également **l'attention à la recherche sur les mégadonnées alignée avec le domaine " s** les points forts et les priorités des chercheurs. C'est là que le dialogue à double sens pourrait conduire à des avancées spécifiques et permettre aux chercheurs en gestion d'envisager les différentes voies pour une synthèse durable de la gestion et de l'érudition du big data. Pour faciliter une approche organisée, nous discutons ensuite d'un cadre de mégadonnées en tant que concept, méthodologie et phénomène.

#### Concept

**Compte tenu des mégadonnées " s des utilisations diverses dans** différents contextes, disciplines et applications, le concept risque de **devenir " tout et rien. " La définition populaire en termes de propriétés de** données telles que le volume et la variété a créé une ambiguïté sur ce qui pourrait compter comme big data. Par exemple, on ne sait pas exactement ce qui détermine le seuil pour qualifier les données comme

**" gros " à travers différents paramètres et applications. Les chercheurs en** gestion, qui mettent fortement l'accent sur des définitions et des concepts plus clairs, pourraient contribuer à faire avancer l'ambiguïté définitionnelle actuelle en orientant la conversation vers une compréhension plus globale du domaine, des limites et de la précision des concepts et concepts du Big Data. Notre propre définition de travail consiste à considérer les mégadonnées comme une étiquette qui fait référence à la génération, l'organisation, le stockage, la récupération, l'analyse et la visualisation d'ensembles de données impliquant de gros volumes et une variété de données, impliquant de nouveaux types de méthodes méthodologiques, épistémologiques et politico questions et questions éthiques.

De même, même si les chercheurs se sont penchés sur les dimensions des mégadonnées, un consensus n'a pas encore émergé. Trois sont répandus: le volume (l'ampleur des données), la variété (hétérogénéité structurelle dans un ensemble de données) et la vitesse (la vitesse à laquelle les données sont générées et la vitesse à laquelle elles sont analysées et utilisées) (Tonidandel, King et Cortina, 2018). Mais, les chercheurs ont également avancé d'autres dimensions **(curieusement, beaucoup commencent par la lettre " v " ), telles que la** véracité, la vision, la visibilité et la valeur, entre autres. Chaque dimension pose des défis distincts et des moyens de les surmonter pour les chercheurs et les

accéder, stocker et utiliser les mégadonnées. Par exemple, la dimension de la vitesse est associée à des problèmes tels que la vitesse de transfert, l'évolutivité du stockage et le calendrier, tandis que la véracité s'accompagne de problèmes tels que l'incertitude, l'authenticité, la fiabilité et la responsabilité. Un examen des questions de recherche de fond, du niveau d'analyse et des lentilles théoriques utilisées pour construire des hypothèses et des propositions appelle à la clarté concernant ces manifestations dimensionnelles. Un tel classement, classification ou autre agrégation de problèmes et de caractéristiques de niveau inférieur à travers les dimensions du Big Data pourrait également servir de base pour le développement de structures de Big Data plus claires pour les tests.

#### Méthodologie

Les études sur les données volumineuses commencent généralement par un chercheur ayant accès à une source de données ou à un ensemble de données sur un phénomène, plutôt qu'avec une théorie (Johnson, Gray et Sarker, 2019). Par la suite, le processus d'analyse implique des problèmes spécifiques dans l'accès aux données et le nettoyage, la recherche et le traitement qui sont différents des approches conventionnelles. L'exécution des phases peut nécessiter des compétences de calcul et de programmation distinctes (par exemple, R et Python). Données pour

**" plus petite " les recherches sont normalement produites de manière** structurée et saisies à certains points. Un défi clé pour la méthodologie du Big Data est de savoir comment intégrer et stocker des données structurées et non structurées de manière à rendre les analyses et la visualisation ultérieures efficaces et sécurisées. Un autre défi est que les ensembles de mégadonnées ne sont souvent pas créés pour examiner des questions et des constructions spécifiques. Ainsi, le chercheur doit traiter de diverses questions relatives à la construction et à la qualité des données.

C'est à travers ces phases méthodologiques difficiles du Big Data que nous encourageons les chercheurs en gestion à mieux comprendre les **avantages et les inconvénients de " en commençant par la théorie " dans les** études de Big Data. Précisément de quelles manières (et quand) la théorie pourrait-elle aider à guider les différentes décisions relatives au nettoyage, à la construction, à l'agrégation et au stockage des ensembles de données volumineuses? Par exemple, une théorie à niveaux multiples ou méso pourrait éclairer la décision de savoir si un ensemble de données **volumineuses doit être construit " horizontalement profond "( beaucoup de variables mais moins d'observations) plutôt que " verticalement**

**Profond "( moins de variables mais beaucoup d'observations).** Nous encourageons également les chercheurs à approfondir leur compréhension de chaque facette du processus méthodologique. Il pourrait, par exemple, être une pratique productive pour le terrain si les études sur les mégadonnées devaient systématiquement contenir un résumé des étapes méthodologiques entreprises, y compris les défis rencontrés et les solutions mises en œuvre.

Même s'il est possible d'examiner certains grands ensembles de données en utilisant des techniques statistiques et informatiques traditionnelles, beaucoup ne sont pas adaptés à des ensembles de données divers et non structurés. Les statistiques se concentrent massivement sur les inférences à partir des données, tandis que les architectures et algorithmes de calcul qui peuvent extraire et discerner des connaissances précieuses à partir d'ensembles de données complexes sont parmi les considérations clés dans les approches Big Data. Ces architectures et algorithmes sont utilisés pour analyser des ensembles de données volumineuses à des fins spécifiques, telles que le regroupement, l'identification de modèle et la prévision. Certaines des techniques incluaient l'extraction de données, l'apprentissage automatique, les réseaux de neurones et l'apprentissage profond (convolution, croyance profonde et réseaux récurrents). Nous observons également qu'une application directe de certaines de ces techniques, en particulier l'apprentissage automatique non supervisé, dans la recherche en gestion pourrait entraîner plusieurs défis. Par exemple, l'un des points forts des techniques d'apprentissage en profondeur est de rechercher puis d'extraire des modèles à partir d'ensembles de données non structurés. De cette façon, des questions pourraient être posées concernant, par exemple, comment construire un modèle explicatif autour d'un modèle, et comment communiquer les limites et les contraintes du modèle final.

Ce qui est également évident pour nous, après avoir examiné les recherches pertinentes, est que ces techniques ont tendance à être hautement spécialisées dans différentes tâches de recherche et à évoluer de manière dynamique, ce qui rend difficile pour les chercheurs individuels d'exploiter le potentiel de ces techniques. Par exemple, les techniques de bigdatavisualisation nécessitent des connaissances informatiques, statistiques et informationnelles. Un autre ensemble de défis pourrait provenir de la puissance de calcul et de l'infrastructure nécessaires, qui pourraient ne pas être facilement accessibles aux chercheurs individuels. Ensemble, nous suggérons donc que les chercheurs en gestion doivent développer une compréhension plus systématique des avantages et des inconvénients des techniques analytiques de Big Data disponibles dans le contexte des études **et des phénomènes de gestion. - par exemple, comment le champ " Les approches empiriques doivent-elles être combinées avec des techniques de mégadonnées telles que des données ou des résultats expérimentaux et des applications ultérieures de techniques d'apprentissage automatique pour obtenir des informations plus généralisables? Ou,**

comment le terrain pourrait-il tirer parti des techniques d'étalonnage des **covariables ou de la multicollinéarité pour " plus petite " des études de données** dans lesquelles la variable de résultat est complexe et distale, comme la performance organisationnelle? De même, le domaine pourrait bénéficier d'une meilleure compréhension des défis et des opportunités d'investigation des résultats de l'apprentissage automatique et d'autres techniques prédictives. Par exemple, de quelle manière les techniques d'apprentissage automatique non supervisées peuvent-elles être combinées avec la recherche qualitative, comme l'utilisation des grappes et des modèles pour éclairer la sélection des concepts, les agrégations et les thèmes initiaux?

Il convient également de noter que différentes technologies et plates-formes peuvent convenir à diverses fins dans les tâches méthodologiques, telles que le stockage, l'accès et le traitement des données. Bien que certaines plates-formes existent et offrent des capacités impressionnantes de traitement des ensembles de données volumineuses, les chercheurs en gestion devront acquérir une compréhension plus complète de la façon de choisir parmi le menu en constante expansion des technologies de données volumineuses. Ainsi, nous encourageons les spécialistes de la gestion à consacrer plus d'attention aux nouveaux modèles de recherche et aux complémentarités analytiques au carrefour des méthodologies conventionnelles et du Big Data. Nous encourageons également plus d'attention à la compréhension des avantages et des inconvénients des techniques et technologies du big data dans le contexte des phénomènes de gestion, individuellement et comparativement vis-à-vis

vis-à-vis des techniques conventionnelles telles que les statistiques multivariées.

## Phénomène

Des pratiques et applications de Big Data se sont produites dans divers contextes, industries et économies. Nous suggérons donc que les spécialistes de la gestion peuvent et devraient accorder une attention particulière aux mégadonnées en tant que phénomène dans les organisations, les institutions et les sociétés. Il y a un grand besoin de comprendre où et dans quels contextes organisationnel et industriel les applications de Big Data pourraient être plus importantes pour les organisations et les gestionnaires. Plus largement, nous pensons que le domaine devrait conduire au développement de nouvelles théories, approches et cadres qui pourraient aider les dirigeants et leurs entreprises à mieux utiliser et extraire de la valeur des mégadonnées. Par exemple, comment les technologies et les outils de mégadonnées pourraient-ils être utilisés pour soutenir la stratégie d'entreprise, par exemple en intégrant diverses technologies de données dans les unités commerciales? Une **question connexe concerne l'entreprise " s choix stratégique concernant " où et comment jouer " dans l'espace Big Data. Du point de vue de la prise de décision,**

des approches d'apprentissage automatique qui identifient automatiquement des modèles exploitables pourraient aider à alléger une partie du fardeau cognitif des gestionnaires. Ce potentiel soulève plusieurs questions intrigantes. Quelles sont la nature et les conséquences du compromis entre la cognition exécutive limitée et les exigences cognitives des mégadonnées? Comment une capacité d'analyser et de visualiser rapidement des modèles cachés dans les mégadonnées peut-elle façonner la qualité et la rapidité de la prise de décision? Quels types de gestionnaires sont plus susceptibles d'adopter (ou d'éviter) les mégadonnées pour prendre des décisions? Répondre à ces questions pourrait conduire à de nouvelles théories ou à des améliorations des théories existantes telles que la vue basée sur les ressources, l'apprentissage organisationnel, les échelons supérieurs, entre autres, ainsi que d'aider les gestionnaires à améliorer l'utilisation des mégadonnées dans leur propre prise de décision.

Certains chercheurs ont fait valoir que la notion de mégadonnées en tant qu'objectifs et fondés sur des faits est un mythe (Gitelman, 2013). Étant donné la possibilité d'une interprétation subjective, les micro-fondations individuelles pourraient être cruciales pour comprendre les processus et les utilisations des mégadonnées dans les organisations. Plusieurs questions générales appellent l'attention: comment les individus et les groupes choisissent-ils et interprètent-ils les mégadonnées? Quels sont certains des obstacles **psychologiques aux individus " adoption du big data? Ces questions pourraient être étudiées en s'appuyant sur une gamme de lentilles théoriques distinctes.** Les perspectives basées sur l'attention, le jugement et les théories heuristiques, et la pensée contrefactuelle pourraient être particulièrement pertinentes pour comprendre comment les individus pourraient utiliser et interpréter les mégadonnées.

Les mégadonnées pourraient également créer des opportunités de recherche autour de ses propres écosystèmes. Par exemple, alors que beaucoup a été dit sur la façon dont les mégadonnées révolutionnent les processus de gestion et sur les avantages que les équipes décisionnelles peuvent tirer de son utilisation, peu a été dit sur les défis et les processus des équipes de mégadonnées qui génèrent et gèrent les mégadonnées dans les organisations ( Saltz, 2015). Comprendre les nouveaux défis interpersonnels auxquels sont confrontées les équipes de Big Data est une direction importante qui pourrait également apporter une valeur prescriptive considérable, comme dans le contexte du développement de nouveaux produits. Plus largement, la création d'une infrastructure de Big Data nécessite que les cadres supérieurs mettent en place des structures et des capacités appropriées qui prennent en charge l'intégration et l'unification des nombreuses îles de données et de capacités analytiques qui pourraient exister dans toute l'organisation.

les implications des informations analytiques associées. Galbraith (2014: 3) a observé que, lorsque les organisations adoptent les mégadonnées, il **existe " un changement de pouvoir des décideurs expérimentés et avisés vers les décideurs numériques. " Comment les organisations structurent-elles ce changement de pouvoir? L'équipe de direction typique comprend-elle un directeur numérique distinct ou le directeur de l'information porte-t-il deux casquettes: informatique et big data? Une autre question concerne la manière dont les organisations peuvent créer des normes et des valeurs concernant le partage d'informations, la transparence et la confiance.**

Il serait négligent de ne pas aborder les problèmes d'éthique et de confidentialité liés aux mégadonnées. Il est maintenant devenu clair que la génération et le stockage d'ensembles de données volumineuses impliquent plus de défis que prévu, comme le montrent par exemple les récents scandales comme CambridgeAnalytica. Pour commencer, la disponibilité des données n'est pas une garantie que leur utilisation serait éthique ou même légale. En outre, il existe des problèmes et des **exigences contradictoires de transparence et de protection des individus "**

identités et connaissances personnelles (p. ex. Acquisti, Brandimarte et Loewenstein, 2015). Bien que les réglementations légales et les codes d'organisations puissent servir de repères utiles, les différences individuelles sont essentielles pour comprendre la propension des individus à aller au-delà de la conformité minimale, ou, alternativement, la tendance des individus à se livrer à des méfaits éthiques en ce qui concerne l'acquisition et l'utilisation des mégadonnées. Les normes et pratiques concernant les droits des données individuelles, l'éthique et la vie privée sont en cours de développement et de débat à l'échelle mondiale. Ces complexités éthiques dans les mégadonnées fournissent des opportunités d'enrichir les théories dans les domaines de l'éthique et des valeurs, telles que le leadership éthique, les valeurs morales et l'identité.

De plus, bien que les idéaux du big data parlent d'ouverture et d'accès à tous, ce n'est pas tout à fait le cas. Les mégadonnées deviennent une activité de plus en plus importante dans laquelle divers acteurs contrôlent non seulement les bases de données, mais réglementent également le marketing, les ventes et l'utilisation de ces données et capacités d'analyse (Cohen,

2013). Est-ce que cela va conduire à un accès asymétrique et à une nouvelle fracture du Big Data entre les chercheurs et les praticiens, et au sein et entre les sociétés et les nations plus largement? Plusieurs indications suggèrent que **les mégadonnées peuvent conduire à une " Effet Matthew, " par lequel nous** voulons simplement dire, pour paraphraser Merton (1973), que les riches en données et en capacité d'analyse pourraient s'enrichir et les pauvres en données et en capacité d'analyse pourraient devenir plus pauvres. De même, les problèmes de transparence et de réplication de la recherche pourraient devenir problématiques

les ensembles de données et les analyses qui les sous-tendent devaient être gardés secrets pour diverses raisons, telles que l'avantage concurrentiel (Cohen, 2013).

#### SURMONTER LES OBSTACLES DE BASE À LA PUBLICATION

Notre discussion sur les mégadonnées en tant que perspective de recherche et les priorités de recherche associées dans les sections précédentes montrent clairement que les mégadonnées posent des problèmes **distincts à chaque étape du processus et de la conception de la recherche**. - du démarrage et / ou de la construction de la théorie, de l'accès et de l'intégration des données à l'analyse, au rapport et à la visualisation. Il nous semble également que la recherche sur les mégadonnées se développe d'une manière **qui pourrait dépasser le cadre d'un seul chercheur**. " **s les capacités et les ressources**, en raison de l'accès et de la gestion des données, de la puissance de calcul requise et de la connaissance nécessaire des outils et techniques analytiques. Nous pensons que les chercheurs devront également prendre en compte et surmonter certains obstacles plutôt fondamentaux à la publication **d'études sur le Big Data sur le terrain**. " **s revues**.

Premièrement, les mégadonnées ne peuvent se substituer à des plans de recherche minutieux et crédibles et à la prise en compte appropriée des questions de recherche. En l'absence de question claire et théoriquement pertinente guidant leur création et leur préparation, les ensembles de données volumineuses peuvent **apparaître comme un grand échantillon de commodité ou un " mode**. " **Une question** clé pour les chercheurs est donc la suivante: pourquoi les mégadonnées sont-elles les plus appropriées pour étudier la question de recherche qui nous intéresse? Les chercheurs devront donc peut-être fournir une justification supplémentaire de la manière et des types de données et de variables collectées, construites et agrégées. Nous encourageons particulièrement les chercheurs à intégrer (explicitement ou implicitement) la logique de l'accès aux données et de leur collecte, de leur intégration et de leur agrégation, de leur analyse, de leurs rapports et de leur visualisation afin d'élaborer et de communiquer la conception de la recherche de leurs études sur les mégadonnées.

Deuxièmement, il peut être difficile, voire impossible, aux examinateurs et autres auteurs de reproduire et d'étendre les études s'il y a peu de transparence sur la façon dont les données sont créées, manipulées et / ou analysées. Les entreprises privées possèdent et stockent souvent des ensembles de données volumineux. Sans certains contrôles et contrôles de qualité intégrés, la fiabilité et la validité des variables pourraient être systématiquement compromises. Les erreurs systématiques ne peuvent pas être résolues en collectant davantage des mêmes données. Les examinateurs et les lecteurs ont l'habitude de voir des études empiriques qui utilisent généralement de petits échantillons dans lesquels les variables sont opérationnalisées d'une manière spécifique. Alors que certaines variables peuvent avoir une validité apparente et

nécessitent moins de justification, les chercheurs devront peut-être trouver des solutions concernant l'opérationnalisation des constructions latentes et de profil intégrées dans des ensembles de données volumineuses. Une solution consiste à utiliser les petits contextes d'échantillonnage pour établir la validité de ces mesures avant de les utiliser dans les contextes de mégadonnées. Une autre solution pourrait être de combiner l'analyse des mégadonnées avec d'autres **méthodes - soit quantitative soit qualitative -**

pour établir la validité et / ou éclairer les processus ou mécanismes clés en jeu. Encore une autre option consiste à travailler en étroite collaboration avec des praticiens experts pour assurer une forte validité apparente des hypothèses et des approches utilisées par les chercheurs.

Troisièmement, la sélection de constructions ou de variables dans les approches empiriques actuelles se fait généralement sur la base de la théorie sous-jacente. Avec les mégadonnées, le processus de conversion des données en constructions d'intérêts peut manquer de clarté et de transparence car certaines techniques associées, telles que l'apprentissage automatique, peuvent à peine être guidées par une théorie explicite. Ici, on pourrait distinguer entre les techniques d'apprentissage supervisé et non supervisé. Dans les techniques supervisées, le chercheur pourrait spécifier les variables à incorporer dans le modèle et, ainsi, l'approche est similaire à la recherche conventionnelle sur petit échantillon. Mais, dans les algorithmes non supervisés, l'algorithme sélectionnera les variables parmi les variables disponibles à inclure dans le modèle. Les examinateurs et les lecteurs ne sont pas habitués à voir **des articles qui sélectionnent des variables " Aléatoire "( du point de vue du** paradigme de recherche sur les petits échantillons). De même, étant donné que les paradigmes existants contrôlent les variables dans les régressions pour contrôler les influences alternatives corrélées aux variables explicatives, les chercheurs doivent prêter attention à la façon dont ils peuvent convaincre les examinateurs que les modèles d'associations trouvés à partir des données sont raisonnables et ne sont **pas simplement des associations " par chance**. "

Quatrièmement, quelle que soit la ou les techniques analytiques utilisées, mais en particulier pour la machine et l'apprentissage profond, nous encourageons les chercheurs à décrire le contenu et le processus de variables et d'associations spécifiques examinées, plutôt que de les **masquer " boîte noire de calcul**. " **Les techniques d'apprentissage** automatique non supervisées et approfondies, en automatisant les tests d'hypothèses multiples avec des modificateurs et des biais opaques, **pourraient en fait compliquer le sens des constructions et des prédictions - avec** le risque supplémentaire de cracher des corrélations fallacieuses à une échelle sans précédent. Une préoccupation connexe est que, parce que les technologies et les techniques des mégadonnées évoluent rapidement, les chercheurs pourraient relâcher les techniques et la modélisation obsolètes. Les outils et

**la technique devra être justifiée par rapport à l'étude " s question et tests nécessaires pour une réponse crédible.**

Cinquièmement, la plupart des approches Big Data utilisent des techniques prédictives plutôt que des approches d'inférence statistique. Ainsi, les chercheurs utilisant des approches de Big Data doivent convaincre les examinateurs et les lecteurs que les approches sont également bonnes, sinon meilleures, pour tester les théories, par rapport aux approches d'inférence statistique. De plus, lorsque des associations simultanées entre plusieurs variables doivent être présentées dans un seul modèle, les chercheurs doivent les présenter dans un format compréhensible pour les examinateurs formés à différents paradigmes. Enfin, parce que la signification statistique n'est pas pertinente avec des tailles d'échantillon massives, les chercheurs devraient travailler pour **justifier et démontrer l'importance des résultats - par exemple, avec des tailles d'effet et une signification contextualisée.** Une représentation visuelle des résultats serait également probablement une approche nécessaire.

Malgré ces pièges, obstacles et défis, nous soutenons que les bourses de gestion seront en position de force dans la mesure où elles transfèrent non seulement les connaissances pertinentes sur les mégadonnées sur le terrain, mais façonnent également activement le contenu et la trajectoire évolutive de ces connaissances. Nous avons discuté ici de plusieurs directions pour synthétiser les avantages du big data et de la bourse de gestion à l'avantage mutuel. Les divers enjeux paradigmatiques, conceptuels, méthodologiques et phénoménologiques entourant les mégadonnées signifient également pour nous que les chercheurs individuels devront évaluer les avantages et les risques et procéder avec prudence lorsqu'ils poursuivent leurs recherches sur les mégadonnées.

Zeki Simsek

Université de Clemson

Eero Vaara

École de commerce de l'Université Aalto

Srikanth Paruchuri

Université d'État de Pennsylvanie

Sucheta Nadkarni

Université de Cambridge

Jason D. Shaw

Université technologique de Nanyang

## RÉFÉRENCES

- Acquisti, A., Brandimarte, L. et Loewenstein, G. 2015. Confidentialité et comportement humain à l'ère de l'information. *Science*, **347**: 509 - 514.
- Bowman, AW 2018. Grandes questions, données informatives, ex-science cellent. *Statistiques et lettres de probabilité*, **136**: 34 - 36.
- Calude, CS et Longo, G. 2017. Le déluge de faux corrélations dans les mégadonnées. *Fondements de la science*, **22**: 595 - 612.
- Chan, J. et Moses, BL 2016. Le Big Data est-il difficile criminologie? *Criminologie théorique*, **20**: 21 - 39.
- Cohen, J. 2013. À quoi sert la vie privée. *Harvard LawReview*, **126**: 1904 - 1933.
- Coveney, PV, Dougherty, ER et Highfield, RR 2016. Les mégadonnées ont aussi besoin de grandes théories. *Transactions philosophiques de la société royale A: sciences mathématiques, physiques et du génie*, **374**: 20160153. Extrait de <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2016.0153>.
- Einav, L. et Levin, J. 2014. La révolution des données et l'éco-analyse nomique. *Politique d'innovation et économie*, **14**: 1 - 24.
- Elragal, A., & Klischewski, R. 2017. Théorique ou prévision basée sur les processus? Défis épistémologiques de l'analyse des mégadonnées. *Journal of BigData*, **4**: 19. Récupéré à partir de <http://link.springer.com/article/10.1186/s40537-017-0079-2>. Frick' e, M. 2015. Big data et son épistémologie. *Journal de l'Association for Information Science and Technology*, **66**: 651 - 661.
- Galbraith, JR 2014. Défis de conception d'organisation résultant du big data. *Journal of Organization Design*, **3**: 2 - 13.
- Gitelman, L. (Ed.) 2013. " Données brutes " est un oxymore Cambridge, MA: MIT Press. Johnson, P., Gray, P., et Sarker, S. 2019. Revisiter IS re les pratiques de recherche à l'ère du big data. *Information et organisation*, **29**: 41 - 56.
- Kitchin, R. 2014. Big Data, nouvelles épistémologies et paradigm change. *Big Data & Société*, **1**: 1 - 12.
- Leonelli, S. 2014. Quelle différence la quantité fait-elle? Sur l'épistémologie du Big Data en biologie. *Big Data & Société*, **1**: 1 - 11.
- Locke, K., Golden-Biddle, K. et Feldman, M. 2008. Fabrication générateur de doute: repenser le rôle du doute dans le processus de recherche. *Sciences de l'organisation*, **19**: 907 - 918.
- Mayer-Schönberger, V., & Cukier, K. 2013. Big data: une révolution qui transformera notre façon de vivre, de travailler et de penser. Boston, MA: Houghton Mifflin Harcourt. Merton, RK 1973. La sociologie des sciences: théorique et études empiriques. Chicago, IL: University of Chicago Press.
- Oswald, FL et Putka, DJ 2016. Méthodes statistiques pour big data: une visite panoramique. Dans S. Tonidandel, EB King, &

JM Cortina (éd.), *Bigdataatwork: Révolution scientifique des données et psychologie organisationnelle*: 43 - 63. New York, NY: Routledge.

Sætra, HK 2018. La science comme vocation à l'ère de

**Big Data: la philosophie de la science derrière le Big Data et l'humanité " s a continué de participer à la science. Sciences psychologiques et comportementales intégratives, 4: 508 - 522.**

Saltz, JS 2015. La nécessité de nouveaux processus, méthodolo-

et outils pour soutenir les équipes de Big Data et améliorer l'efficacité des projets Big Data. Dans IEEE Computer Society (Ed.), Conférence internationale 2015 de l'IEEE sur

**Big Data: 2066 - 2071. Los Alamitos, Californie: IEEE Computer Society.**

Succi, S., & Coveney, PV 2018. Big data: la fin de

**la méthode scientifique? Transactions philosophiques de la Royal Society A: sciences mathématiques, physiques et du génie, 377: 20180145. Extrait de <https://royalsocietypublishing.org/doi/full/10.1098/rsta.2018.0145>.**

Tonidandel, S., King, EB et Cortina, JM 2018. Big data

méthodes: Tirer parti des techniques analytiques de données modernes pour **construire la science organisationnelle. Méthodes de recherche organisationnelle, 21: 525 - 547.**

Le droit d'auteur de Academy of Management Journal est la propriété d'Academy of Management et son contenu ne peut être copié ou envoyé par courrier électronique à plusieurs sites ou publié sur une liste de diffusion sans l'autorisation écrite expresse du titulaire du droit d'auteur. Cependant, les utilisateurs peuvent imprimer, télécharger ou envoyer par courrier électronique des articles pour un usage individuel.