

Mégadonnées, grandes idées – Les connaissances sont-elles toujours la clé du pouvoir à l'ère numérique ?

par Thomas Gehling et Bernhard Wolters, Gen Re, Cologne

À la toute première ligne du roman de Charles Dickens intitulé *Les Temps difficiles*, Thomas Gradgrind déclare « Or, ce que je veux, ce sont des faits ». Remplacez le mot « faits » par « données » et les inconditionnels des mégadonnées seront tout à fait d'accord. Certains assurent même qu'il n'y aura à l'avenir plus besoin de théories, puisque les modèles de données nous racontent d'ores et déjà toute l'histoire. Ces affirmations sont-elles fondées ? Après tout, les scientifiques travaillent avec des données depuis des siècles. Les données empiriques constituent la pierre angulaire des sciences sociales et naturelles, depuis que ces dernières ont émergé pour remplacer ce qui était de plus en plus perçu comme des chicaneries scolastiques non empiriques.

Cet article examine la validité de l'affirmation selon laquelle la théorie a perdu son but concret, dans un monde où l'abondance et la grande disponibilité des données sont prétendument suffisantes pour guider nos actions. Dans un futur article de Risk Insights, nous nous pencherons sur les conséquences potentielles de cette évolution sur l'industrie de l'assurance vie, en étudiant plus particulièrement le rôle des connaissances spécialisées de l'Homme à l'avenir.

La fin de la théorie

La dernière personne à avoir proclamé la fin d'un principe capital a été le politologue américain Francis Fukuyama. En 1989, il a affirmé que l'histoire était sur le point de culminer en une étape finale, lors de laquelle le monde entier serait dirigé par des démocraties libérales.¹ Hélas, l'histoire ne s'est pas arrêtée, loin de là, et lui a donné tort.

Mais les agrégateurs de données tels que Google seront-ils les fossoyeurs de la théorie ? Il est « temps de se demander : Qu'est-ce que la science peut apprendre de Google ? »² Ou bien, les inconditionnels des mégadonnées seront-ils entraînés par son incontestable dynamique, de la même manière que Fukuyama a été submergé par les changements politiques marquants du début des années 1990 ?

Sommaire

La fin de la théorie	1
Avons-nous besoin des connaissances ?	2
Avantages de l'acquisition de connaissances	4
Conclusion	4

La lettre d'information en bref

Risk Insights est une revue technique publiée par Gen Re pour les cadres de l'assurance vie et santé du monde entier. Les articles portent sur des thèmes ayant trait à la gestion des risques, la médecine, les sinistres, la souscription et les questions actuarielles. Les produits étudiés sont les assurances maladie grave, soins longue durée, pensions d'invalidité, santé et vie.

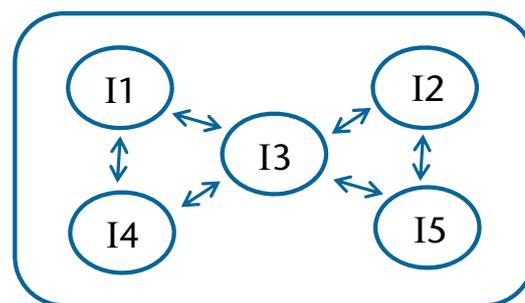
L'idée selon laquelle les mégadonnées donneront un accès direct et empirique au monde, sans aucune théorie, a été régulièrement évoquée ces dernières années.³ En réalité, il s'agit seulement d'une nouvelle manifestation d'une dispute séculaire dans l'histoire de la science et de la philosophie. De grands philosophes rationalistes, parmi lesquels Descartes et Leibniz, affirmaient qu'il n'est possible d'acquérir un certain niveau de connaissance qu'en usant de la raison et de la logique. Inversement, l'école empiriste de Bacon, Locke et Hume estimait que l'expérience constituait le fondement ultime de la connaissance. Bacon est même allé jusqu'à affirmer que les faits empiriques parlent d'eux-mêmes.⁴

Le fondement de la connaissance a de nouveau été contesté durant la première moitié du XXe siècle, lorsque les partisans du « falsificationnisme » se sont opposés aux défenseurs du « positivisme logique ». La connaissance scientifique était cette fois en jeu, ou plus précisément les critères auxquels une connaissance revendiquée doit répondre pour pouvoir être qualifiée de scientifique.

D'après le philosophe Karl Popper, les connaissances scientifiques revendiquées (théories) doivent en principe être réfutables (falsifiables) sur la base de l'observation empirique.⁵ Par exemple, la théorie qui affirme « Il va pleuvoir demain » est scientifique car elle sera réfutée s'il ne pleut pas demain. En revanche, la théorie selon laquelle « Il va ou ne va pas pleuvoir demain » n'est pas scientifique, puisqu'aucun énoncé observationnel concevable ne peut réfuter cette théorie.

Les positivistes logiques soutenaient quant à eux que l'essence de la connaissance scientifique réside dans les énoncés observationnels empiriques de base et que la méthode inductive est un élément structurant. Sur cette base, la revendication (théorie) selon laquelle il pleuvra demain ne peut être considérée comme scientifique que s'il pleut déjà depuis plusieurs jours. Dans le cas contraire, cette affirmation doit être considérée comme une supposition non scientifique.

Figure 1 – La connaissance, un réseau d'informations liées entre elles⁹



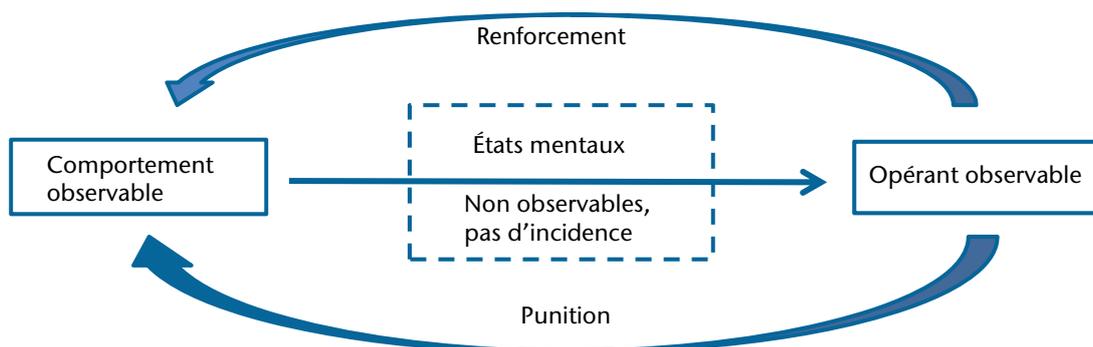
Avons-nous besoin des connaissances ?

Pour réussir dans un monde numérique, les partisans des mégadonnées affirment que les connaissances ne sont plus nécessaires. Selon les termes qu'ils emploient, la « corrélation [série de schémas] remplace la causalité [comprendre : les connaissances] et la science peut progresser même en l'absence de modèles différents, de théories unifiées ou même d'explications mécanistes. »⁶

Encadré 1 – Comportementalisme

Contrairement à l'approche psychodynamique freudienne, qui est axée sur les processus mentaux subconscients (non observables), le comportementalisme était considéré par ses partisans comme « une branche purement objective des sciences naturelles ». ¹² Son objectif théorique consistait en « la prédiction et le contrôle des comportements ». ¹³ La théorie comportementale a été entièrement élaborée par B.F. Skinner, qui a ajouté à la focalisation existante sur le conditionnement classique, à savoir le fait qu'un stimulus provoque une réponse, ce qu'il a appelé un conditionnement opérant, à savoir un comportement accidentel renforcé ou puni par un opérant et ainsi le renforcement ou la diminution de la probabilité d'un tel comportement. Selon la théorie comportementale, il s'agissait là de l'essence de l'apprentissage. Skinner ne niait pas l'existence d'états mentaux accompagnant le conditionnement opérant mais il affirmait qu'ils n'avaient aucune incidence (voir Figure 2).

Figure 2 – Comportementalisme



Ce point de vue fait écho au scepticisme inhérent à l'école de pensée empiriste en ce qui concerne la certitude intemporelle de la connaissance. Examinons la Figure 1, qui montre un réseau d'informations (I) liées entre elles et dans lequel chaque information explique les autres (flèches bleues). Ce réseau représente des connaissances. En l'absence d'explications et de rapports pour relier une information à une autre, il ne reste que des fragments d'information, qui « ne peuvent pas aider à donner un sens à la réalité qu'ils cherchent à décrire. »⁷

Les informations peuvent être décrites comme la somme de données (D) et d'un sens (S), soit $I = D+S$.⁸ Les défenseurs des mégadonnées affirment que les connaissances sont totalement superflues puisque le « sens » peut être déduit de cette équation. Des données bien formulées, compilées selon certaines règles (syntaxe) suffisent – comme dans certaines traductions ou publicités – et il n'y a nul besoin d'informations concernant les langues, la culture ou les conventions.

Pour déterminer si les connaissances deviennent réellement inutiles, nous devons examiner la distinction heuristique entre connaissances spécifiques et connaissances de fond.¹⁰ Les connaissances spécifiques incluent les explications qui rendent les corrélations statistiques plausibles, à la lumière de nos connaissances de fond. En plus d'étayer les connaissances spécifiques, les connaissances de fond nous permettent de recueillir des données de façon ciblée et de passer au crible les modèles de données intéressants.¹¹ En outre, elles nous aident à exploiter les corrélations qui ont été identifiées et, éventuellement, expliquées. Les connaissances de fond rendent nos prédictions plus fiables, tandis

que les connaissances spécifiques les rendent plus rentables, et pas seulement dans un sens commercial.

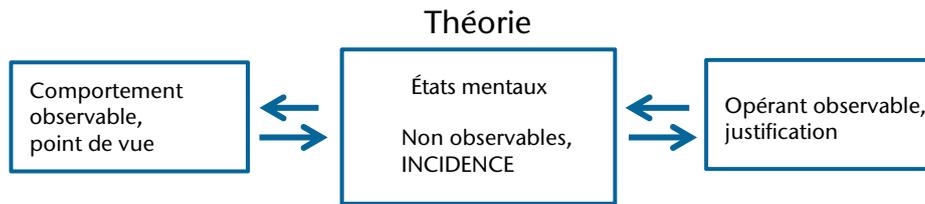
Pour illustrer les avantages des connaissances spécifiques, nous pouvons reprendre un exemple issu de l'histoire de la psychologie. Au cours de la première moitié du siècle dernier, une approche théorique appelée « comportementalisme » a acquis une certaine influence (Encadré 1).

Par son refus d'expliquer la corrélation entre opérants observables et comportements observables, le comportementalisme devançait le rejet de la connaissance par les mégadonnées. La validité du comportementalisme a par la suite été remise en cause par une autre approche théorique, appelée psychologie cognitive.

Encadré 2 – Dissonance cognitive

Selon la théorie de la dissonance cognitive, les conflits qui émergent dans l'esprit humain entre des points de vue incompatibles ou entre des points de vue et des comportements incitent les sujets à changer de point de vue ou bien à adapter leur comportement, afin de rétablir leur tranquillité d'esprit. Ce processus dépend aussi de la pertinence de la justification d'un nouveau point de vue ou comportement contradictoire (la quantité d'argent reçue, dans l'expérience décrite). Une justification insuffisante (une récompense monétaire moindre, par ex.) cause un trouble psychologique (dissonance cognitive) plus important et, par conséquent, une plus forte propension à changer de comportement ou de point de vue.

Figure 3 – Psychologie cognitive



La psychologie cognitive affirme que l'état mental et la manière dont l'esprit fonctionne ont une réelle importance. Dans le cadre d'une expérience, il a été demandé aux participants de donner leur véritable point de vue sur certains sujets puis de contredire ces points de vue dans de brèves rédactions, en contrepartie d'une somme d'argent variable.¹⁴ Lors d'une étape finale, il leur a été demandé de remplir des questionnaires supposément anonymes pour recueillir à nouveau leur véritable point de vue.

Les comportementalistes s'attendaient à ce que, parmi les sujets de l'essai recevant les sommes les plus importantes pour aller à l'encontre de leur point de vue initial (renforcement), une plus forte proportion changerait d'opinion dans le questionnaire final, ce qui signifie qu'ils auraient assimilé un nouveau point de vue. En réalité, c'est tout le contraire qui s'est produit : plus la récompense était faible, plus on a observé un changement de point de vue induit par l'expérience.¹⁵

Ce résultat était correctement prédit par la psychologie cognitive, ou plus précisément par la théorie de la dissonance cognitive (Encadré 2). Les étudiants recevant une faible somme d'argent ont été davantage affectés par leur comportement contradictoire et, par conséquent, ont davantage ressenti le besoin de changer de point de vue pour le rendre cohérent avec leur comportement. La Figure 3 illustre le fait que la théorie peut s'avérer pertinente pour expliquer une corrélation (observable).

Avantages de l'acquisition de connaissances

Les avantages de l'acquisition de connaissances de fond apparaissent clairement lorsqu'on étudie l'exemple suivant.¹⁶ Des analystes de données employés par la société de distribution américaine

Target ont découvert que les habitudes d'achat de 25 produits leur permettaient d'attribuer des scores de prévisibilité de grossesse ainsi que des estimations de dates de naissance. S'étant renseignée dans le domaine neurologique et psychologique sur l'acquisition d'habitudes (notamment la boucle déclencheur-routine-récompense), Target savait que les gens sont réceptifs aux déclencheurs de changement d'habitude lorsque se produisent de grands événements de la vie, comme une grossesse.

La société Target a alors lancé des publicités sur mesure pour des produits de puériculture, afin d'inciter les femmes enceintes et leurs proches à se rendre dans ses magasins. Là, ils seraient poussés à acheter d'autres produits, mus par leur envie de modifier leurs habitudes d'achat. Mais il planait là une menace de désastre imminent du point de vue des relations publiques. Target a su la déceler à temps, grâce à ses connaissances de fond culturelles et psychologiques.

La société a réalisé que les femmes enceintes pourraient s'inquiéter de publicités créées sur mesure pour des produits associés à la grossesse. Ces femmes auraient pu, à juste titre, se demander comment Target avait appris leur grossesse ou même si elles étaient épiées. Target a donc ajouté des publicités pour des produits que des femmes enceintes n'achèteraient jamais, comme des tondeuses ou des verres à vin.

Conclusion

Personne ne doute réellement que les connaissances de fond resteront importantes même dans un monde numérique régi par les mégadonnées. Nous aurons toujours besoin de poser les bonnes questions, ce qui est impossible en l'absence de connaissances de fond. Les partisans des mégadonnées le concéderont certainement. Leur scepticisme est plutôt

dirigé vers les connaissances spécifiques, qui se présentent sous la forme de théories expliquant des observations statistiques, et notamment des corrélations. Selon eux, ces connaissances étaient nécessaires dans le passé pour améliorer notre degré de confiance envers les prédictions, qui étaient basées sur un nombre limité d'observations (données). La théorie n'était qu'un dispositif auxiliaire et ne pouvait pas prétendre à une plus haute dignité. Avec l'avènement des mégadonnées, l'échantillonnage, l'inférence inductive et les hypothèses sont remplacées par l'observation en temps réel et le suivi de populations entières.¹⁷ La théorie ne semble plus être nécessaire pour sauter le pas inductif entre un nombre limité d'observations et la population cible en question. Pour les applications commerciales, ce point de vue paraît valable, en particulier lorsqu'on sait avec quelle rapidité il est possible de corriger des erreurs, à peu de frais. Cependant, des pièges subsistent, comme le montre l'exemple précédent, issu d'une expérience de psychologie comportementale.

Un examen rapide des définitions possibles pour le terme « mégadonnées » reflète la plupart des points controversés que nous avons évoqués. Selon une interprétation donnée, ce terme désigne un ensemble de données qui est devenu trop important pour pouvoir être géré avec les techniques traditionnelles des bases de données.¹⁸ Le volume joue un rôle supplémentaire, dans le sens où les applications de mégadonnées n'ont pas besoin de réaliser d'échantillonnages et peuvent se contenter d'observer et suivre les événements.¹⁹ Selon un autre point de vue, les mégadonnées ne se définissent pas tant par le manque de puissance de calcul que par la détection de nouveaux modèles, de corrélations qui peuvent apporter de la valeur ajoutée.²⁰ En plus du volume et de la diversité, la vitesse est une autre caractéristique importante des mégadonnées, étant donné que les calculs sont généralement réalisés en temps réel.²¹

Dans un futur article de Risk Insights, nous présenterons les caractéristiques des mégadonnées qui peuvent réellement être considérées comme nouvelles et transformatrices, et nous examinerons le nouvel élan donné par l'accélération des changements sociaux. Dans cet article, l'accent

sera davantage mis sur les questions (éthiques) pratiques qui entourent l'utilisation des mégadonnées. Pour conclure, nous analyserons comment les mégadonnées influencent déjà et influenceront à l'avenir la manière dont les compagnies d'assurance vie abordent leurs enjeux majeurs, à savoir les besoins des clients, le risque de changement, l'anti-sélection et le risque moral.

À propos des auteurs

Thomas Gehling, *actuaire principal, est spécialisé dans la conception de produits dans le domaine de l'invalidité et de la mortalité pour le marché allemand. Il est membre de l'association actuarielle allemande (DAV) et peut être contacté par téléphone au +49 221 9738 746 ou par e-mail à l'adresse thomas.gehling@genre.com.*



Bernhard Wolters *est vice-président du département d'assurance vie et santé de Gen Re et a été responsable pendant plusieurs années des marchés néerlandais et scandinaves. Il peut être contacté par téléphone au +49 221 9738 848 ou par e-mail à l'adresse wolters@genre.com.*



Notes

- 1 Francis Fukuyama, *The End of History?*, *The National Interest*, 1989, p. 1.
- 2 Chris Anderson, *The End Of Theory: The Data Deluge Makes The Scientific Method Obsolete*, <http://www.wired.com/2008/06/pb-theory/>, consulté le 21 juin 2016.
- 3 Voir par exemple Luciano Floridi, *The Fourth Revolution*, 2014, pp. 129-131.
- 4 Bertrand Russell, *History of Western Philosophy*, 2000 (première publication en 1946), p. 529. « Il espérait qu'un simple arrangement ordonné des données ferait apparaître la bonne hypothèse, mais ce n'est que rarement le cas. »
- 5 Karl Popper, *The Logic of Scientific Discovery*, 2002 (première édition en anglais en 1959; la version originale en allemand « *Die Logik der Forschung* » a été publiée en 1935).
- 6 Chris Anderson, *ibid* note 2. Le philosophe David Hume rejetait déjà la notion de causalité au XVIII^e siècle, en affirmant que « les objets n'ont pas de connexion identifiable entre eux ; ils ne sont pas non plus liés par un quelconque autre principe mais fonctionnent d'une manière spécifique répondant à la notion imaginaire selon laquelle on peut déduire de l'apparition de l'un qu'un autre existe. » David Hume, *A Treatise Of Human Nature* (initialement publié en 1739), 2003 ed., p. 73.
- 7 Luciano Floridi, *Information: A Very Short Introduction*, 2010, p. 51.
- 8 Luciano Floridi, *ibid* note 7, p. 20.
- 9 Source des graphiques : les auteurs, sauf indication contraire.
- 10 Cette distinction ne prétend pas à une plus haute dignité. Elle est seulement établie aux fins de notre argumentation.
- 11 Luciano Floridi, *ibid* note 3, p. 130.
- 12 John B. Watson, *Psychology as the Behaviourist Views it*, *Psychological Review* 1913, <http://psychclassics.yorku.ca/Watson/views.htm>, consulté le 26 juin 2016.
- 13 *Ibid*.
- 14 Martin Carrier, Jürgen Mittelstrass, *Mind, Brain, Behaviour. The Mind-Body Problem and the Philosophy of Psychology*, 1991, p. 136.
- 15 Il convient de noter que le comportement des sujets (réponses à un questionnaire au début et à la fin, afin de recueillir leur véritable point de vue) a également été enregistré sur la base de leurs rédactions qui, comme il le leur avait été demandé, contredisaient leur point de vue d'origine. Les réponses données aux trois étapes ne pouvaient pas être modifiées.
- 16 Charles Duhigg, *How Companies Learn Your Secrets*, *The New York Times Magazine*, 16 février 2012.
- 17 Les termes « hypothèse » et « théorie » sont utilisés de manière interchangeable.
- 18 Louis Rossouw, *Big Data – Big Opportunities*, *Gen Re's The Bulletin*, mai 2013.
- 19 https://en.wikipedia.org/wiki/Big_data, consulté le 13 juillet 2016.
- 20 Luciano Floridi, *ibid* note 3, p. 16.
- 21 https://en.wikipedia.org/wiki/Big_data, consulté le 13 juillet 2016.



The difference is...the quality of the promise.

genre.com | genre.com/perspective | Twitter: @Gen_Re

General Reinsurance AG—Succursale Paris

21, rue Balzac
75008 Paris
Tel. +33 1 5367 7676
Fax +33 1 5367 4646

Editors:

Ulrich Pasdika, ulrich.pasdika@genre.com (responsible)
Ross Campbell, ross_campbell@genre.com

Photos: © Thinkstock – FactoryTh, NicoElNino, Vimvertigo

© General Reinsurance AG 2017

This information was compiled by Gen Re and is intended to provide background information to our clients, as well as to our professional staff. The information is time sensitive and may need to be revised and updated periodically. It is not intended to be legal or medical advice. You should consult with your own appropriate professional advisors before relying on it.