

# ANTHROPEN

Le dictionnaire francophone d'anthropologie ancré dans le contemporain

## **BIG DATA**

Duclos, Vincent

Université du Québec à Montréal, Canada

Date de publication : 2023-09-12

DOI : 10.47854/anthropen.v1i1.52071

[Voir d'autres entrées dans le dictionnaire](#)

Le terme *big data* désigne des ensembles extrêmement volumineux et complexes de données qui ne peuvent pas être facilement analysées à l'aide des techniques traditionnelles de traitement des données. Le *big data* est parfois défini par les « 3V » : volume, vitesse et variété. Le volume fait référence à la quantité massive de données générées et collectées. La vitesse fait référence à la vitesse à laquelle les données sont générées, transmises et traitées. La variété fait référence à la diversité des types de données et des sources. Le *big data* est souvent utilisé pour des applications telles que l'apprentissage automatique, la modélisation prédictive, l'exploration de données et le traitement du langage naturel. Elles sont utilisées dans des domaines divers incluant la finance, les soins de santé, le marketing et la recherche scientifique. Les données des *big data* proviennent de sources multiples. Celles-ci incluent les traces laissées par les usages d'Internet et des réseaux sociaux (Facebook, Instagram et Twitter au premier plan), les appareils intelligents et connectés (téléphones, tablettes, montres, etc.), les plateformes de commerce en ligne, les plateformes de diffusion en continu, les marchés financiers, la recherche scientifique, les données de secteurs publics (par exemple agences gouvernementales, systèmes de transport, programmes d'aide sociale, etc.), entre autres. La collecte de données est ainsi contingente de processus de production et de consommation d'informations par des individus qui vivent leur vie numérique ordinaire. C'est-à-dire que l'explosion du *big data* doit être située dans le contexte d'un capitalisme numérique faisant de la capture, l'extraction et la récupération des données son fonds de commerce.

Depuis une quinzaine d'années, le *big data* a reçu une importante attention médiatique et scientifique. Pour les plus optimistes, avec les quantités massives de données générées aujourd'hui, le *big data* permet de détecter des phénomènes, modèles ou relations qui étaient auparavant imperceptibles ou trop complexes à identifier. Le *big data* offre par le fait même des réponses à des questions fondamentales de la vie individuelle et collective. Cette promesse fut par exemple évoquée par Chris Anderson, le rédacteur en chef de la revue *Wired*, qui a publié en 2008 un article précurseur intitulé « The End of Theory », annonçant une

ISSN : 2561-5807, Anthropen, Université Laval, 2021. Ceci est un texte en libre accès diffusé sous la licence CC-BY-NC-ND, <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Citer cette entrée : Duclos, Vincent, 2023, « Big Data », *Anthropen*.  
<https://doi.org/10.47854/anthropen.v1i1.52071>.

transformation radicale de notre rapport à la connaissance. Anderson suggérait qu'il y avait eu un renversement dans la structure de la connaissance : avec le *big data*, le monde exprime en données sa vérité, indépendamment des modèles, principes ou théories cherchant à trouver des causes plus profondes ; il suffit d'extraire des structures de significations déjà existantes. Nul besoin de théorie ou de causalité, cette dernière étant supplantée par la corrélation – la corrélation suffit. Quelques années plus tard, dans un ouvrage qui deviendra un succès de librairie, Mayer-Schönberger et Cukier (2013) insisteront sur ce point : une vision du monde que nous pensions être composée de causes est remise en question par une prépondérance de corrélations. Par le fait même, c'est notre compréhension la plus élémentaire de la manière de prendre des décisions et d'appréhender la réalité qui est bouleversée.

On le voit bien, comprendre l'impact du *big data* sur le rapport à la connaissance implique d'insister sur la distinction entre causalité et corrélation. La causalité, telle qu'on l'entend habituellement, fait référence à une relation dans laquelle une variable (la cause) produit un effet sur une autre variable. Pour établir la causalité, il est nécessaire de démontrer qu'il existe un mécanisme plausible reliant la cause et l'effet. La corrélation, par contraste, fait référence à une relation statistique entre deux ou plusieurs variables. Lorsque deux variables sont corrélées, cela signifie qu'elles ont tendance à varier ensemble de manière prévisible. En suivant une variable, on peut donc également suivre l'autre. Prenons un exemple fréquemment cité de corrélation n'impliquant pas de causalité : lorsque les ventes de crèmes glacées augmentent, les taux de criminalité ont tendance à augmenter également. Bien qu'il soit peu probable que la consommation de crèmes glacées incite les gens à commettre des délits, il est possible qu'une troisième variable, par exemple le temps plus chaud, soit à l'origine de l'augmentation des ventes de crèmes glacées *et* des taux de criminalité. Ainsi de Bobby Henderson qui, dans une lettre ouverte (vite devenue virale) au conseil scolaire du Kansas, en réponse à la décision d'enseigner la théorie du dessein intelligent dans les écoles, s'amusait à noter une forte corrélation négative entre la température moyenne de la planète et le nombre de pirates qui ont navigué au cours des deux derniers siècles. Une seule suite logique semblait alors s'imposer : nous avons besoin de davantage de pirates. Bien qu'illustrant des corrélations particulièrement fallacieuses, ces exemples ont le mérite de soulever la question : si presque n'importe quelle corrélation peut être découverte, comment peut-on distinguer le vrai du faux ? Plus largement, quel rôle les sciences sociales, et l'anthropologie en particulier, sont-elles appelées à jouer face au déploiement rapide du *big data* ?

Une première grande approche anthropologique du *big data* consiste à l'étudier en tant que processus sociaux. S'inspirant de la riche tradition des études sociales de la science (STS) cherchant à ouvrir la « boîte noire » des artefacts technologiques (Akrich 1987), des travaux ont éclairé les valeurs, attachements et relations qui composent le *big data*. Ils ont montré comment les données ne sont jamais brutes et objectives – elles sont toujours créées, déduites et interprétées. Les données sont façonnées par les terrains desquels elles découlent : toutes les données sont locales, dans le sens qu'elles ont des liens complexes avec le lieu dont elles proviennent (Loukissas 2019). Plusieurs travaux récents ont montré comment les corrélations propres au *big data* et aux techniques associées – incluant les algorithmes et l'apprentissage automatique – sont susceptibles de reproduire des *patterns* du passé, incluant des inégalités sociales de toutes sortes. Le rôle joué par les algorithmes et le *big data* dans la propagation de préjugés sexistes et raciaux ou sociaux est par

exemple bien documenté (Benjamin 2019 ; Noble 2018). Une littérature foisonnante a montré comment le *big data* constitue des outils de reproduction sociale qui structurent les chances de vie de manière profonde, mais aussi très inégale. Loin d'être « neutre » ou « objectif », le *big data* peut amplifier et automatiser les inégalités sociales existantes (Dionne et Paquette 2019). Dans d'autres contextes, il peut rendre certaines personnes invisibles, notamment les populations « pauvres en données », qui utilisent moins les technologies numériques et génèrent donc moins de données.

Le *big data* permet d'entraîner des modèles à reconnaître certaines situations ou tendances. Mais par-delà son travail technique, il génère de nouvelles façons de se projeter dans le temps (Caduff 2014). Derrière la popularité croissante du *big data*, on retrouve une promesse de prévoir l'imprévisible et de traduire l'avenir, qu'il soit climatique, commercial, financier ou sanitaire, entre autres, en une série de calculs. Mais le *big data* ne fait pas que prédire : il participe à la mise en scène de l'avenir que ces calculs anticipent (Duclos 2019). Il transforme l'avenir en une série de futurs calculables. Dans *Weapons of Math Destruction*, par exemple, O'Neil suggère qu'en l'absence de données pertinentes pour les comportements et phénomènes qu'ils cherchent à prédire, les modèles utilisant le *big data* les remplacent souvent par d'autres données agissant comme des « proxy ». Des corrélations statistiques sont établies, par exemple, entre le code postal ou la langue d'une personne et sa capacité à rembourser un prêt ou occuper un emploi. De telles corrélations, note O'Neil, sont souvent discriminatoires. Dans *Automating Inequality*, Eubanks examine également l'impact de systèmes automatisés de prise en charge de services sociaux tels que l'aide sociale et l'assistance publique. L'un des principaux arguments d'Eubanks est que ces systèmes reflètent et renforcent les biais et les préjugés qui existent au sein de la société. Elle souligne que ces systèmes sont souvent formés à partir de données historiques qui contiennent des préjugés intégrés, tels que les pratiques discriminatoires en matière de logement, d'éducation et de justice pénale. À un niveau moral et politique, Eubanks suggère qu'une telle automatisation génère une distance morale qui justifie des décisions inhumaines pour ce qui est de la distribution de ressources telles que la nourriture, les soins et le logement (Eubanks 2017).

L'examen du *big data* évoque évidemment des objets familiers à l'anthropologie. Les anthropologues se sont depuis longtemps attachés à montrer comment les structures du pouvoir opèrent par le biais de la ressemblance, de la similitude et de l'imitation. Ils ont également étudié les multiples manières par lesquelles les personnes sont représentées, catégorisées, rassemblées, divisées et marquées. Il n'est donc pas très surprenant de voir le *big data* devenir l'objet d'une étude critique. Mais le *big data* interroge également la spécificité de la pratique anthropologique, bousculant ses repères épistémologiques, l'entraînant peut-être à admettre que, même si elle n'est pas causalité, la corrélation peut être intéressante en soi (Boullier 2015). Au cours des dernières années, la question a ainsi commencé à se poser avec de plus en plus d'insistance : comment faire une anthropologie non seulement *du*, mais bien *avec* et *par* le *big data* ? Animés par une telle réflexion, certains cherchent à saisir comment l'anthropologie pourrait dialoguer avec le *big data*, voire contribuer à ses pratiques (Moats et Seaver 2019). Ceci implique entre autres la remise en question d'une certaine sensibilité ethnographique qui tend à faire sienne la dichotomie vieillissante entre le qualitatif et le quantitatif, les récits et les chiffres. Voilà qui exige de refuser les portraits un peu trop simplistes du *big data* pour mettre « la main à la pâte » et expérimenter, voire surmonter, la séparation entre sciences

sociales et *big data*. D'autres poussent la chose plus loin et en appellent à l'émergence d'une anthropologie « machinique » (*machinic anthropology*) qui élargirait son champ d'enquête par l'adoption de la pensée quantitative et les méthodes computationnelles (Pedersen 2023). À tout le moins, la montée en puissance du *big data* semble exiger de l'anthropologie qu'elle aborde cette nouvelle « matière première » avec les qualités qui lui sont propres, en s'intéressant à la science « en train de se faire », dans un souci d'apprentissage réciproque, avec toutes les possibilités, les risques et les incertitudes que cela comporte.

## Références

Akrich, M., 1987, « Comment décrire les objets techniques ? », *Techniques et culture*, 9 : 49-64. <https://doi.org/10.4000/tc.863>.

Anderson, C., 2008, « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », *Wired*, 23 juin. <https://www.wired.com/2008/06/pb-theory/>.

Benjamin, R., 2019, *Race after Technology: Abolitionist Tools for The New Jim Code*. Cambridge (R.-U.), Polity Press.

Boullier, D., 2015, « Les sciences sociales face aux traces du *big data*. Société, opinion ou vibrations ? », *Revue française de science politique*, 65 (5) : 805-828. <http://doi:10.3917/rfsp.655.0805>.

Dionne, E. et J. Paquette, 2019, « Introduction. Données massives, médiation algorithmique et savoirs à l'ère du numérique – Le rêve politique de la grippe : réflexions introductives en temps de pandémie et variations sur le thème d'un “réel” en faisance », *Global Media Journal*, 11 (2) : 1-17. <http://gmj-canadianedition.ca/wp-content/uploads/2020/06/02-Editorial-Volume-11-issue-2-Final.pdf>.

Duclos, V., 2019, « Algorithmic Futures: The Life and Death of Google Flu Trends », *Medicine Anthropology Theory*, 6 (3) : 54-76. <https://doi.org/10.17157/mat.6.3.660>

Eubanks, V., 2017, *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York, St. Martin's Press.

Loukissas, Y.A., 2019, *All Data are Local: Thinking Critically in a Data-Driven Society*. Cambridge (MA), MIT Press.

Mayer-Schönberger, V. et K. Cukier, 2013, *Big Data: A Revolution that Will Transform How we Live, Work, and Think*. Boston et New York, Houghton Mifflin Harcourt.

Moats, D. et N. Seaver, 2019, « “You social scientists love mind games”: Experimenting in the “divide” between data science and critical algorithm studies », *Big Data & Society*, 6 (1). <https://doi.org/10.1177/2053951719833404>.

Noble, S., 2018, *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York, NYU Press.

Pedersen, M.A., 2023, « Editorial Introduction: Towards a Machinic Anthropology », *Big Data & Society*, 10 (1) : 1-9. <https://doi.org/10.1177/20539517231153803>.