

REGARDER LES DONNÉES

Dominique Cardon

Association Multitudes | « [Multitudes](#) »

2012/2 n° 49 | pages 138 à 142

ISSN 0292-0107

ISBN 9782916940830

Article disponible en ligne à l'adresse :

<https://www.cairn.info/revue-multitudes-2012-2-page-138.htm>

Distribution électronique Cairn.info pour Association Multitudes.

© Association Multitudes. Tous droits réservés pour tous pays.

La reproduction ou représentation de cet article, notamment par photocopie, n'est autorisée que dans les limites des conditions générales d'utilisation du site ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Toute autre reproduction ou représentation, en tout ou partie, sous quelque forme et de quelque manière que ce soit, est interdite sauf accord préalable et écrit de l'éditeur, en dehors des cas prévus par la législation en vigueur en France. Il est précisé que son stockage dans une base de données est également interdit.

Regarder les données

Dominique Cardon

À quelle mise en forme du social devrions-nous nous fier? Comment représenter la société lorsque les outils numériques nous donnent des capacités accrues de représenter les êtres et les collectifs dans des espaces digitaux aux paramètres si flexibles que tout leur semble permis?

L'intérêt d'une confrontation des débats sociologiques avec les figurations que projettent l'art contemporain, dans l'exposition *Multiversités créatives*, est de mettre en tension de nouvelles façons d'architecturer le monde et les manières de penser le social à l'heure des *big data*. Les machines numériques sculptent désormais des objets en référant leurs formes aux flux de données qui jaillissent de l'incessante activité des bases de données du web. L'accélération des processus de calcul et le déluge de nouvelles données que favorise la digitalisation progressive de toutes les traces de la vie quotidienne se déploient dans tous les domaines: prolifération des sources d'information, digitalisation du dossier médical et des informations personnelles, modèles probabilistes de l'assurance, développement des outils de *data mining* dans la relation client, capture et suivi des traces

de mobilité, de communication ou de navigation. Stockées, agrégées, calculées, les données entrent de plus en plus souvent dans des dispositifs de marché, de surveillance, d'évaluation ou de recommandation. D'une utilisation *ex post* réservée à des statisticiens professionnels, les mesures deviennent des indicateurs *ex ante*, s'inscrivent à même les interfaces et s'introduisent dans les usages les plus quotidiens.

Pour l'heure, la mise en place d'un *monde des données* est souvent appréciée, avec enthousiasme ou frayeur, comme une nouvelle puissance susceptible, pour les uns, de réinventer les marchés et l'organisation, de rendre la démocratie plus transparente, de faciliter les interactions avec les choses et l'environnement ou d'élaborer des connaissances prédictives, alors que d'autres s'alarment des usages commerciaux des fichiers, des menaces sur la vie privée, de la dictature de l'hyper-visibilité ou de la colonisation du monde vécu par un esprit de calcul et d'évaluation. Les données sont moins interrogées dans leur matérialité, pour ce qu'elles sont et pour ce qu'elles font, que pour mettre en scène les espoirs et les craintes qu'elles suscitent.

Faire parler des mondes de données

Or, pendant que s'affole le débat sur les risques et les opportunités des *big data*, d'autres s'affairent à régler les algorithmes destinées à faire « parler » les données. Ils entretiennent un rapport beaucoup plus « réaliste » à la façon dont la multiplicité des traces numériques n'engage aucune signification *a priori* et s'échinent à construire de nouveaux artefacts computationnels destinés à mettre en signification les *mondes de données*. Ils composent avec l'hétérogénéité grandissante des traces, testent des modèles, mobilisent de nouveaux savoirs statistiques et conçoivent de nouvelles formes de visualisation. Car des données, il faut faire sens et forme. Les œuvres présentées dans *Multiversités créatives* explorent les nouveaux paradigmes visuels que peuvent composer les mondes de données lorsqu'ils rencontrent des algorithmes.

Dans le domaine du marketing, du design et du journalisme, la question de la visualisation des données est depuis quelques années en animation constante. Elle se nourrit à la fois de l'augmentation des possibilités offertes par les outils informatiques et graphiques et surtout de l'accroissement des sources et des flux de données. Les gigantesques bases de données que les nouveaux vecteurs numériques sont en train de constituer sont pourtant « idiots ». Alors que la construction de la statistique traditionnelle s'est appuyée sur un ensemble de techniques destinées à produire une représentation catégorielle du monde, le monde des *big data* ne cherche pas à représenter le social mais à l'aspirer. À l'échantillonnage catégoriel, il préfère la totalité réticulée et aspire dans ses serveurs gloutons un ensemble proliférant, désordonné et hétérogène de traces et d'informations dont il est difficile de rendre sens.

Désormais plus que le tableau ou la liste classée, c'est le graphe qui est devenu la principale signature visuelle des nouvelles données numériques. Visualisées avec Pajek, Guess ou Gephi, les proliférantes données digitales sont représentées sous formes de nœuds et d'arêtes. Elles se nouent et se dénouent dans des clusters, forment des agrégats et des chemins, se laissent zoomer et dézoomer. Le graphe est devenue l'outil d'exploration des traces numériques et, de façon significative, les algorithmes qui sont nécessaires à cette mise en forme ne sont plus destinés à produire des agrégats catégoriels produisant des indices et des taux de corrélation, mais à visualiser le placement des points sur les nouvelles cartes digitales, comme en témoigne le travail mené par Linkfluence pour représenter les liens entre les sites et blogs politiques du web.

Le calcul ne mesure plus des causes, des déterminations ou des corrélations, il visualise l'espace des relations entre des données hétérogènes. Les interprétations ne sont plus disponibles sous forme de théories et d'hypothèses déjà constituées, mais l'interprète doit tracer son chemin en explorant la carte des données sans boussole ni théorie préalable. Ce changement de paradigme dans les techniques de traitement des données qui déplace les techniques de visualisation du tableau de chiffres vers le graphe nœuds-lien est aussi le témoin d'une transformation des manières de construire et de représenter la société dans le travail des sciences humaines.

Le monde des *big data*

Ce qui anime la croissance continue du web est l'extension vorace du périmètre des données qu'il prélève sur le monde, processus dont l'ambition est la « digitalisation de la vie elle-

même¹ ». Le web s'étend en transformant en objets numériques des choses, des activités, des flux ou des états qui restaient précédemment dans l'ombre. Silencieusement ancrées à leur contexte, ces *traces* n'avaient pas de raisons particulières de circuler dans un espace d'information plus large. Idiosyncrasiques, non ou peu intentionnelles, souvent à peine perçues et isolées en tant que telles, elles ne font pas sens hors de leur contexte et lui sont souvent si profondément attachées qu'elles ne se distinguent pas du flux d'activité des individus. En leur donnant une existence numérique, le web rend perceptibles, mobiles et dénombrables des états du monde jusqu'alors inaperçus. Cet insatiable mouvement d'expansion, dont le point de fuite est le développement d'un Internet des objets, procède d'un processus de rationalisation qui vise à capturer ces états « à bas bruit » du monde afin d'en faire des informations digitales. Il s'agit de transformer des empreintes en indices, de faire signe avec des traces².

Pris individuellement, ces enregistrements du quotidien des individus semblent de peu de signification, mais rendus communs, ils peuvent constituer des espaces documentaires inattendus et initier les coordinations opportunistes qui sont au principe de la formation des communautés sur le web³. Un même processus s'observe pour de nombreuses empreintes de la vie quotidienne. Nous n'avions pas idée de compter notre nombre d'amis, Facebook nous a appris à le faire. Nous ne mesurons pas le

nombre de kilomètres parcourus dans la journée, Fitbit le fait pour nous. Nous ne dénombrions pas le nombre d'actions militantes (réunions, tracts distribués, porte-à-porte ou appels téléphoniques), MyBarackObama.com en a fait un badge de fierté à afficher devant ses proches. Alors que les trajets quotidiens, les efforts sportifs, la consommation électrique ne laissent pas de traces exploitables, étaient oubliés à l'instant et ne devenaient support d'aucune connaissance, ou d'une connaissance simplement locale et à usage strictement personnel, ils sont aujourd'hui devenus des prises exploitables qui peuvent, en certaines circonstances, servir à produire de l'information, du réseau, des métriques ou des conversations. La numérisation des traces participe donc à un mouvement d'emprise sur le monde qui augmente les possibilités de faire sens d'empreintes multiples lorsque celles-ci sont rendues publiques et partageables. La composition des représentations de la société ne cessent de s'élargir à de nouveaux types d'informations qui n'appartiennent pas aux catégories traditionnelles (l'âge, le lieu, la catégorisation socioprofessionnelle, le niveau de revenu...) de figuration du social.

Un renversement temporel

Cependant la signification de ces nouvelles traces numériques est loin d'être acquise d'emblée. Il y a même lieu de se demander à quoi ces empreintes bruyantes, étranges et apparemment sans objet sont destinées. Leur conférer le statut d'« information » paraît surestimer très largement leur capacité indiciaire. Souvent, le web capture de nouvelles empreintes sans finalité explicite. Il laisse le libre jeu de la coopération entre internautes décider de ce que ces données, une fois rendues publiques, pourront leur signifier. Les promoteurs des données

1 Rouvroy (Antoinette), « Détecter et prévenir: de la digitalisation des corps et de la docilité des normes », in Lebeer (Guy), Moriau (Jacques), dir., *(Se) gouverner. Entre souci de soi et action publique*, P.I.E. Peter Lang, 2009.

2 Merzeau (Louise), « Du signe à la trace: l'information sur mesure », *Hermès*, n°53, 2009, p. 23-31.

3 Aguiton (Christophe), Cardon (Dominique), « The Strength of Weak Cooperation: An attempt to Understand the Meaning of Web2.0 », *Communications & Strategies*, n°65, 1^{er} quarter 2007, p. 51-65.

ouvertes disent aux institutions que si elles ne peuvent pas leur dire préalablement pour quels usages les données doivent être rendues publiques, il se trouvera toujours des internautes qui, *a posteriori*, sauront les faire parler⁴.

Le renversement temporel de l'*ex post* vers l'*ex ante* est une caractéristique centrale de l'économie informationnelle de l'Internet. Les données sont publiques par défaut, alors que nous avons l'habitude de les considérer privées par défaut. Le web publie puis filtre, alors que les médias traditionnels filtraient avant de publier. La visualisation précède l'interprétation.

Cette inversion s'applique aussi au sens du travail statistique lorsqu'il passe d'une situation de rareté à un contexte d'abondance des données. Alors que la statistique a longtemps conçu la production de données comme un moyen de répondre à des hypothèses formées préalablement, aujourd'hui, les hypothèses apparaissent comme des conséquences émergentes du travail de fouille effectuée au sein de gigantesques masses de données « idiotes ». Les nouvelles techniques de traitement statistique du *data mining* basée sur la fouille des données et l'apprentissage ne font aucune hypothèse préalable sur les données, aussi hétérogènes soient-elles, qu'elles calculent. Et il n'est pas rare qu'elles prétendent rendre pertinentes des corrélations sans cause⁵. Pour certains, le travail d'interprétation des sciences humaines n'aurait plus de raison d'être dans l'ère des *big data* et pourrait être abandonné au récent mariage du behaviorisme et des algorithmes.

4 Noveck (Beth Simone), *Wiki Government. How technology can make government better, democracy stronger, and citizens more powerful*, Washington, Brookings Institution Press, 2009.

5 Rouvroy (Antoinette), Berns (Thomas), « Le nouveau pouvoir statistique », *Multitudes* 40 (hiver 2010); Weinberger (David), *Too Big to Know. Rethinking Knowledge now that the Facts Aren't the Facts, Experts are Everywhere, and the Smartest Person in the Room is the Room*, New York, Basic Books, 2011.

Dans « L'âge des Petabits », Chris Anderson, rédacteur en chef de *Wired*, écrit : « C'est un monde dans lequel des quantités massives de données et les mathématiques appliquées remplacent tous les autres outils qui pourraient être utilisés. Exit toutes les théories sur les comportements humains, de la linguistique à la sociologie. Oubliez la taxinomie, l'ontologie, et la psychologie. Qui peut savoir pourquoi les gens font ce qu'ils font ? Le fait est qu'ils le font, et que nous pouvons le tracer et mesurer avec une fidélité sans précédent. Si l'on a assez de données, les chiffres parlent d'eux-mêmes⁶ ».

Le moment de la visualisation

C'est dans ce contexte que la visualisation des données numériques sous formes de graphes nœuds-liens est apparue comme une nouvelle étape dans le travail d'analyse des données. Entre la trace et l'interprétation se glisse désormais le moment de la visualisation. Celui-ci est apparu comme un moyen de court-circuiter la connexion immédiate et invisible entre les données et les algorithmes statistiques. Face aux torrents des *big data*, un détour s'impose : regarder les données, circuler dans le flux décousu des traces, tester la solidité des connexions entre entités, préserver une sorte d'état élémentaire des traces. La visualisation est venue interrompre l'interprétation automatique de la statistique.

6 Anderson (Chris), « The End of Theory. Will the Data Deluge Makes the Scientific Method Obsolete? », *Edge*, 2008, cité dans Boyd (Danah), Crawford (Kate), « Six Provocations for Big Data », *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, September 2011. Disponible sur SSRN: <http://ssrn.com/abstract=1926431> (traduction française sur Internet. actu: www.Internetactu.net/2011/09/23/big-data-la-necessite-d%E2%80%99un-debat/)

Dans le domaine des *web sciences*, où se retrouvent informaticiens et sociologues, la créativité des chercheurs s'est considérablement épanouie ces dix dernières années, ainsi qu'en témoigne le site *visualcomplexity* sur lequel Manuel Lima rassemble les multiples expériences de représentations des données sous forme de graphes⁷. Outre le fait que la sélection de ces figurations se fait moins en fonction de leur apport scientifique que de leur esthétique visuelle et de leur capacité à proposer de nouvelles façons de naviguer dans les données, ces représentations introduisent du jeu dans les catégories habituelles de description du monde social.

Sur les outils de Jeffrey Heer, il est possible de voir les réseaux d'« amis » Facebook, Friendster ou Orkut s'organiser en blocs colorés, mais les liens d'amitiés sont-ils vraiment des témoins de ce que l'on entend habituellement par amitié ? Sur Linkfluence, il est possible de naviguer entre les blogs regroupés par territoires thématiques, mais cette territorialisation peut-elle être comparée à une géographie urbaine avec ses routes nationales, ses artères et ses chemins vicinaux ? Sur HistoryFlows, il est possible de naviguer temporellement dans la rédaction coopérative d'un article de Wikipédia, mais cette trajectoire est-elle le reflet de ce que l'on avait habituellement l'habitude d'entendre comme échange d'arguments ? Sur les graphes de Nicholas Christakis et James Fowler qui représentent le poids des personnes et leurs liens sociaux, il semble que l'obésité soit contagieuse au sein d'un réseau amical⁸.

Regarder avant d'interpréter

Ces nouvelles approches entrent en résonance avec les transformations des paradigmes interprétatifs dans les sciences sociales. Elles semblent entériner l'idée d'une dilution des formes sociales instituées, des catégories stables et des interprétations robustes. De façon parfaitement ambivalente, elles peuvent alors à la fois servir la disparition de l'effort interprétatif des sciences sociales au profit d'une simple symptomatologie d'un monde soudainement devenu « liquide⁹ », comme elle peut renforcer le travail d'inventivité interprétative de la sociologie s'attachant à identifier de nouvelles façons d'associer les entités qui composent le monde social. Ce que la visualisation des nouvelles traces digitales fait alors voir, c'est que les catégories d'interprétation traditionnelles ne tiennent plus d'elles-mêmes et qu'il faut commencer par regarder avant d'interpréter¹⁰.

⁷ www.visualcomplexity.com/vc et Lima (Manuel), *Visual Complexity. Mapping Patterns of Information*, New York, Princeton Architectural Press, 2011.

⁸ Christakis (Nicholas A.), Fowler (James), *Connected. The Surprising Power of our Social Networks and How They Shape Our Lives*, New York, Little, Brown & Company, 2009.

⁹ Bauman (Zygmunt), *La vie liquide*, Le Rouergue/Chambon, 2006.

¹⁰ Latour (Bruno), Jensen (Pablo), Venturini (Tommaso), Grauwin (Sébastien), Boullier (Dominique), « The whole is always smaller than its parts. A digital test of Gabriel Tarde's Monads », *British Journal of Sociology*, à paraître, 2012.