

Les données peuvent-elles faire le « bien » ?

Peut-on appliquer directement les méthodes du *big data* aux questions sociales et solidaires ? Une analyse des promesses et des dangers de l'utilisation des datas sur le terrain de la solidarité, en particulier sous le regard de la mesure de l'impact social des actions.

Habités à faire parler les données dans l'intérêt économique d'une entreprise, les *data scientists*, ces ingénieurs mathématiciens au cœur de la mécanique algorithmique, sauraient-ils répondre à des besoins forts d'intérêt général ? C'est le pari de Datakind et de son fondateur Jake Porway. Il réunit, en effet, des *data scientists* bénévoles, lors d'événements ponctuels autour de la problématique d'une structure jugée d'utilité sociale. En vingt-quatre ou quarante-huit heures, ils analysent et combinent différentes sources de données, conçoivent des modèles et des modes de visualisation : cartographies, graphes, etc. C'est ainsi qu'a été conçu un modèle de prédiction des collisions entre voitures et piétons dans les rues de New York. Global Witness, un lobby citoyen, a pu identifier grâce à Datakind une liste d'entreprises potentiellement coupables de fraude fiscale. LONG Give Directly, qui organise des dons directs aux personnes pauvres en Afrique, est à présent équipée d'un algorithme de reconnaissance de villages pauvres qui s'appuie sur une analyse automatique des images satellites de Google Earth. Enfin, Simpa Networks, une entreprise

sociale indienne qui loue des panneaux solaires au temps d'utilisation à des foyers sans accès à l'électricité et les donne après un montant de location atteint, a obtenu un modèle prédictif permettant d'identifier, parmi ses nouveaux clients, les plus susceptibles d'aller au bout du processus de location.

Mais ces résultats ont-ils du sens d'un point de vue solidaire ? N'y a-t-il pas quelque risque que leur usage *a posteriori*, par exemple pour Simpa Networks, ne soit pas aussi socialement « juste » qu'anticipé ? Enfin, utiliser ainsi les données pour mieux répondre aux missions que se sont fixées ces acteurs, provoque-t-il, par ricochet, une transformation de l'ensemble du secteur social et solidaire ? Et dans quel sens ?

LES TRÉPIDATIONS DE LA MACHINE « DATA »

Pour être pertinent, le monde des *data scientists* doit agréger un nombre important de données ; il incite alors à l'ouverture et au partage de celles-ci, comme l'illustre l'application Entourage⁷. En

effet, en agrégeant les données des associations et des citoyens qui vont à la rencontre des SDF (parcours, matériels distribués, personnes rencontrées, besoins exprimés, etc.), cette application montre qu'il serait possible d'identifier des zones blanches, des doublons, des manques, etc. Une telle cartographie en temps réel de l'aide aux plus démunis pourrait permettre d'ajuster en continu les différentes actions sur un territoire, afin de mieux couvrir les besoins des sans-abri. En outre, les outils numériques permettent de faire ce que l'on appelle du *crowdsourcing* solidaire en collectant des données directement auprès d'individus en difficultés. Ainsi, des SDF expriment directement leurs besoins *via* l'application Entourage, ou des personnes en détresse se manifestent après un tremblement de terre, comme ce fut le cas au Népal en 2015 avec Quakemap. Les données ainsi collectées s'affichent sur une *datavizualisation* (cartographie interactive, graphes) qui facilite l'organisation des services d'aide et d'urgence.

Appliquer les recettes des *data scientists* et de leur *big data* aux enjeux sociaux risque de faire obstacle à un usage éthique et solidaire de l'analyse des données.

Quakemap, Entourage ou encore Mutum⁸ sont des plateformes. Cette forme d'organisation est au monde des datas ce que la structure associative est

au monde de l'ESS (économie sociale et solidaire). Les plateformes se fondent sur l'utilisation des données pour optimiser à grande échelle la mise en relation entre une demande et une offre, entre un besoin et une capacité d'aide. Ce modèle est fortement porté par l'entrepreneuriat social, dont les acteurs veulent «faire le bien en se faisant du bien», rapporte Arnaud Mourot, directeur général d'Ashoka France. Ils ont constitué de façon plus ou moins informelle un réseau mondial d'entrepreneurs sociaux qui croient en un capitalisme d'intérêt général. Il existe cependant un mouvement alternatif et, il est vrai, plus confidentiel, le «coopérativisme de plateforme» : il considère que les pratiques sociales et solidaires se marient mieux à des plateformes dont les acteurs et simples usagers sont tous les propriétaires, plutôt que d'appartenir à quelques-uns, actionnaires ou même fondateurs.

La logique économique et entrepreneuriale semble pour le moment l'emporter. Si l'on regarde, par exemple, l'ensemble des projets réalisés par Datakind, tous semblent empreints d'un même objectif : maximiser l'impact social d'une structure, même si cela signifie exclure certains profils jugés moins «performants» ou pour lesquels l'aide apportée serait moins pertinente du point de vue de l'aidant. Plusieurs projets traitent, par exemple, de l'opportunité d'ouvrir un nouvel établissement. Le croisement de différentes sources de données permet de déterminer l'emplacement où il serait le plus utile, c'est-à-dire où il y aurait le plus de personnes en demande d'aide selon les critères de la structure concernée. L'application de ces principes de maximisation d'impact *via* les données aux services publics de transports ou de santé ne risque-t-il pas d'aggraver les inégalités entre territoires ?

Jusqu'à où peut-on, dès lors, appliquer la logique du nombre et la politique du chiffre dans le monde social et solidaire ? Jusqu'à où peut-on ajuster les

actions solidaires en fonction de leur potentiel bénéfique ? Il est certes possible de «maximiser» un impact social sans accepter aveuglément la course à la productivité et à la performance, souvent perçue dans le monde de l'ESS comme l'une des origines du «mal» de nos sociétés contemporaines. «Data-kind», cette *gentille* initiative de Jake Porway, qui nous laisserait penser en poussant le sens des mots qu'habituellement les *data scientists* travailleraient pour des «méchants», ne risque-t-elle pas de pervertir les *good guys* en leur «apprenant» des méthodes de *bad guys* ? Bien sûr, la réalité du terrain est loin d'être aussi manichéenne que cela. Mais les données sont-elles vraiment dignes de confiance ?

DES DATAS OUI, MAIS AVEC DES PINCETTES

«Les données ne peuvent pas produire de la certitude, elles ne peuvent qu'estimer l'incertitude», avance Heather Krause, fondatrice de Datassist, une agence canadienne qui accompagne les structures non lucratives dans l'utilisation des données. Pour bien comprendre cette phrase, il faut «ouvrir» les données, les regarder pour ce qu'elles sont, en faire la biographie, nous dit-elle... D'où viennent-elles ? Qui les a produites ? Qu'est-ce qu'elles mesurent ? Comment et pourquoi ont-elles été collectées ? Heather Krause prend l'exemple des données de l'ONU sur l'évolution de la violence conjugale dans différents pays. Intuitivement, on a tendance à faire confiance à cette source. Pourtant, en y regardant de plus près, on découvre de multiples biais qui génèrent des erreurs de lecture : là, un changement de loi qui rend possible de porter plainte pour violence conjugale ; ici, une modification des catégories d'âge ; là encore, des études qui mesurent la violence physique, ou bien d'autres qui y ajoutent la violence psychique, etc. Les données agrégées se révèlent, en vérité, souvent impossibles à comparer. Leur «nettoyage» est donc un

préalable obligatoire qui, dans la pratique, n'est quasiment jamais respecté. Mieux vaut donc suivre le conseil d'Heather Krause et considérer les données comme une information et non comme une connaissance, et encore moins comme une vérité.

« Les données ne peuvent pas produire de la certitude, elles ne peuvent qu'estimer l'incertitude. »

HEATHER KRAUSE

Ensuite, il importe de ne jamais perdre de vue que les datas sont le reflet d'un monde partiel : celui que l'on peut enregistrer, calculer, modéliser. On pourrait imaginer les données comme des reliefs, faisant apparaître en creux des vallées qui dessinent en réalité l'impensé du *big data*. Ce «hors-champ» contient ce qui échappe à la mesure, soit en raison de biais au moment de la collecte ou de l'analyse, soit parce qu'il existe probablement, à l'image du silence qui disparaît dès qu'on le nomme, des réalités que nous ne pourrions jamais représenter à l'aide de données.

Avoir conscience des biais, du cadre et du hors-champ des données ne remet pas en cause la contribution qu'elles peuvent apporter aux questions sociales et solidaires. En revanche, cela incite à plus de retenue quant aux résultats obtenus. Les données ne sauraient, sans danger, être le seul indicateur de la réussite d'une action.

C'est pourtant le risque que nous prenons à mesure que l'intelligence artifi-

L'usage du *big data*...



... permettrait de sauver chaque année 13 000 vies *via* l'anticipation des tremblements de terre.¹



... permet d'établir dans 88 % des cas l'orientation sexuelle d'un individu *via* ses *likes* sur Facebook.²

Sources : 1. « Big Data: Saving 13,000 Lives a Year by Predicting Earthquakes? », Bernard Marr, forbes.com, avril 2015.

2. « Private Traits and Attributes Are Predictable from Digital Records of Human Behavior », Michal Kosinski, David Stillwell et Thore Graepel, « Proceedings of the National Academy of Sciences of the United States of America », 2013.

cielle (IA) progresse. En effet, l'IA, nourrie et éduquée par les masses de données collectées, automatisera probablement un nombre croissant de processus de décisions : de l'obtention d'un crédit à celle d'un logement ou d'un travail *via* des plateformes qui présélectionnent automatiquement des profils en fonction des attentes. Perspective pour beaucoup inquiétante : le financement des structures sociales et solidaires s'ajustera-t-il un jour automatiquement en fonction de la mesure de leur impact, elle-même automatisée ? Sous ce regard, pour le monde de l'ESS, les données représentent bien plus qu'un outil de travail. Il s'agit de comprendre le monde des datas, de s'en saisir afin de se donner les moyens de peser sur les débats qui pourraient décider, demain, des modalités de calcul et de rétribution de l'action sociale et solidaire.

LE CHOC DU FINANCEMENT SELON L'IMPACT SOCIAL

L'une des tendances fortes, dans la mesure de l'impact social, est de juger de l'efficacité d'une action au nombre d'euros qu'elle ferait économiser à la collectivité - estimé grâce au recueil de ces fameuses datas. Soit l'économie du coût de X mois de chômage pour telle structure d'insertion, de X prises en charge hospitalières pour telle association de prévention, ou encore de X reconduites à la frontière pour tel incubateur à destination des réfugiés, etc. Cette mesure monétaire de l'impact social s'accompagne du modèle économique suivant : l'auteur de l'action se

verrait reverser une partie des économies qu'il aurait contribué à réaliser. Ainsi, demain, une action solidaire envers les SDF pourrait être financée en fonction du nombre de SDF « sortis » de la rue ou même du nombre de jours, de semaines ou de mois en moins que le SDF y aurait passé. Pourtant, comment arriverons-nous à individualiser de la sorte les impacts sociaux, alors que l'action solidaire se mène la plupart du temps en réseau et sur un temps long ? Qui pourra dire « c'est nous qui avons sorti telle personne de la rue » ?

L'usage de datas pour mesurer l'impact social d'une action peut s'avérer précieux, mais peut provoquer de grosses erreurs d'appréciation.

En outre, le secteur public ne pourra pas financer à fonds perdus les actions d'utilité sociale. Il s'agit donc, dans cette démarche de monétisation de l'action sociale et solidaire, d'attirer les investisseurs désireux - et aussi de plus en plus contraints par la réglementation -

de placer leur argent dans des projets à impact social ou environnemental positif. Se profile alors, en contrepartie d'un appel d'air financier sans précédent, un risque de captation de la valeur sociale par des acteurs loin des enjeux de solidarité, à travers les jeux d'actionnariat de l'entrepreneuriat social.

Enfin, sachant que la mesure de l'impact social serait, à terme, issue de calculs algorithmiques, le jeu des corrélations pourrait inclure, sans que l'on en ait conscience, des bénéfiques indirects et hors de propos, comme la baisse évitée ou la hausse du prix au mètre carré de l'immobilier grâce à l'absence ou au départ de SDF. Comme cela a déjà été démontré sur les questions d'obtention de crédits ou de sélection d'étudiants, certains algorithmes utilisent des critères discriminants pour effectuer leurs calculs : appartenance communautaire, genre, etc. Il paraît dès lors essentiel que les processus de calcul des algorithmes soient rendus lisibles par et pour des humains.

Pour toutes ces raisons, et bien d'autres encore, le monde de l'ESS aurait sur ce terrain des datas un rôle déterminant à jouer : celui de défendre une mesure d'impact social qui préserverait la dignité des personnes et qui protégerait de la spéculation financière les liens de solidarité. Que les acteurs sociaux et solidaires s'emparent des données et s'approprient les tenants et les aboutissants du monde numérique, ce n'est pas qu'une opportunité pour eux, c'est une nécessité pour nous tous.

Christèle Bazin

1. Fred Turner, *Aux Sources de l'utopie numérique - De la contre-culture à la cyberculture*, Stewart Brand, un homme d'influence, C&F éditions, 2012.

2. Le logiciel libre, imaginé par Richard Stallman, a pour principe le libre accès aux codes sources, ainsi que sa libre utilisation et modification. Firefox est un logiciel libre, par exemple.

3. La neutralité du Net est un principe de non-discrimination des données

circulant sur le réseau Internet. Ainsi, il ne peut y avoir de traitement différent en fonction de la source, du type ou de la destination d'un contenu.

4. Le *bitcoin* est une monnaie virtuelle produite par un algorithme. Elle n'est donc sous l'autorité d'aucun État. C'est une monnaie spéculative en raison de son design original basé sur une masse monétaire maximale et un ralentissement de la production de

bitcoins à mesure que l'on s'approche de la limite, rendant toujours plus coûteuse la production de nouvelles unités.

5. La *blockchain* est une base de données distribuée, c'est-à-dire qu'elle n'est pas contrôlée par un serveur central. Elle enregistre, gère et certifie des transactions - notamment en *bitcoin*. Son principe est réputé infalsifiable grâce à son procédé de validation en pair-à-pair.

6. Evgeny Morozov, *Pour tout résoudre cliquez ici - L'aberration du solutionnisme technologique*, FYP éditions, 2014.

7. Lire l'article « Un entourage bienveillant des personnes sans abri », page 40.

8. Lire l'article « Mutum : les paradoxes d'une économie du partage », page 70.



SOLIDARUM

Base de connaissances pour
l'invention sociale et solidaire

Cet article en format PDF est directement tiré de ***Visions solidaires pour demain***, revue papier annuelle dont l'objet est de réfléchir à ce qu'est, et ce que pourrait être dans le futur, la solidarité sociale. Ce fichier PDF est accessible au sein de la base de connaissances ***Solidarum***, plateforme en ligne, gratuite et évolutive, qui propose à la consultation et au téléchargement des médias vidéo, texte, son et image : des visions et reportages créés spécifiquement pour elle, en *Creative Commons*.

Solidarum et ***Visions solidaires pour demain*** sont édités par la Fondation Cognacq-Jay et réalisés par une rédaction autonome dédiée, avec l'appui d'un comité éditorial composé en majorité de personnalités extérieures à la Fondation.

www.solidarum.org