



HAL
open science

Mesurer la justice socio-spatiale : de l'ancien au nouveau monde, promesses et menaces du “ Big Data ”

Elisabeth Tovar

► To cite this version:

Elisabeth Tovar. Mesurer la justice socio-spatiale : de l'ancien au nouveau monde, promesses et menaces du “ Big Data ”. Justice spatiale = Spatial justice, 2016, Liberté, Egalité, Computer, 10, <http://www.jssj.org/article/mesurer-la-justice-socio-spatiale-de-lancien-au-nouveau-monde-promesses-et-menaces-du-big-data/>. <halshs-01507163>

HAL Id: halshs-01507163

<https://shs.hal.science/halshs-01507163v1>

Submitted on 12 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Mesurer la justice socio-spatiale : de l'ancien au nouveau monde, promesses et menaces du « Big Data »¹

Elisabeth Tovar

Résumé (180 mots)

Qu'ils le célèbrent ou le maudissent, certains auteurs présentent le Big Data comme l'avènement d'un nouveau monde statistique où la causalité et la méthode scientifique seraient remplacées par de puissants algorithmes statistiques capables d'inférer des modèles prédictifs à partir des corrélations entre les masses de nos empreintes numériques quotidiennes. Dans cette réflexion, nous proposons de nous intéresser à ce phénomène en mobilisant les outils de la science économique et en confrontant, dans le cadre de la mesure de la justice socio-spatiale, les avantages et inconvénients normatifs de 'l'Ancien monde' de la statistique traditionnelle, hiérarchisé et déductif et ceux du 'Nouveau monde' du Big Data, décentralisé et inductif. Finalement, ni magie noire ni magie blanche, le Big data ne semble être qu'un outil statistique de plus, un construit social qui peut et doit être soumis à la critique des Sciences sociales. Selon le point de vue conséquentialiste des économistes, pour qui dire le juste implique la meilleure mesure du bien-être possible, le monde du Big Data apparaît comme une réelle amélioration par rapport à la pénurie de données de l'Ancien monde statistique... dont il reste à déterminer, d'un point de vue procédural, la régulation.

Mots-clefs : Big Data, science économique, justice conséquentialiste, justice procédurale, justice socio-spatiale

¹ Une première version de cette réflexion a été rédigée pour la Journée d'études « Liberté, égalité, computer. Gouvernamentalité algorithmique et justice spatiale » organisée par la revue Justice Spatiale, Justice Sociale le 28 novembre 2014 à l'Université Paris Ouest. L'auteur remercie les organisateurs de cette journée ainsi que intervenants et les participants pour le débat, contradictoire mais fécond, qui a nourri la rédaction de cette réflexion. Elle remercie également deux rapporteurs anonymes dont les commentaires ont été féconds pour l'amélioration du manuscrit.

[Introduction] 'Big Data' et justice socio-spatiale : questions d'économiste

Le Big Data, parfois présenté comme l'un des éléments de la 4^{ème} révolution industrielle (ANDERSON, 2012), est le plus souvent défini comme l'avènement d'un monde statistique tridimensionnel, caractérisé par les célèbres « 3V » : volume, vitesse et variété croissants des données échangées et analysées (*cf.* rapport du groupe META-Gartner rédigé par LANEY, 2001). Au-delà, on met aujourd'hui en avant un changement de paradigme dans la *nature* et l'utilisation des données : alors que l'analyse informatique traditionnelle des données (*data mining* compris) repose sur des raisonnements déductifs, le Big Data marquerait le passage à une analyse de nature inductive. Dans ce « nouveau monde » statistique, l'analyse inférentielle permettrait, à partir d'un très grand nombre de données décentralisées à faible densité d'information, d'inférer des modèles dotés d'une capacité *prédictive*, étant donnée une certaine incertitude (DELORT, 2015).

« Le Big Data est fondamentalement différent de la fouille de donnée, ou *data mining*. Cette différence ne porte pas sur le volume de données. Elle est de nature conceptuelle. Un *datawarehouse*, dans lequel s'effectue la fouille de données, s'appuie sur un modèle. À l'inverse, (...) le Big Data consiste à créer *en exploratoire et par induction sur des masses de données à faible densité en information des modèles à capacité prédictive*. (...) Nous passons des faits à des règles et les mathématiques permettent de mesurer l'incertitude pesant sur ces règles, dépendant notamment des faits sur lesquels ces règles sont basées » (DELORT, 2015).

L'émergence de ce nouveau monde statistique du Big Data engendre un riche et vif débat dans lequel s'inscrit ce numéro spécial de la revue *Justice Spatiale, Spatial Justice*.

Ainsi, dans *Politique Étrangère*, MAYER-SCHÖNBERGER se place du point de vue du 'néo-positivisme numérique' (MOSCO, 2014, cité par OUELLET *et al.*, 2014) et salue dans « *La Révolution Big Data* » une évolution d'une magnitude comparable au remplacement de la notion newtonienne d'absolu par la relativité d'Einstein. Accroissant la rationalité de la prise de décision, le Big Data déconnecterait enfin notre perception du monde de nos fragiles postulats préconçus et, surtout, de notre illusoire besoin de causalité. Par ailleurs, il accorderait une valeur économique à la

chronique de nos existences : juxtaposées, elles feraient sens, le Big Data révélant des façons de faire société jusque-là insoupçonnées. Il serait donc impératif d'en mesurer les retombées économiques et d'en organiser la bonne la gouvernance politique (MAYER-SCHÖNBERGER, 2014).

Par contraste, d'autres auteurs adoptent le point de vue des *surveillance studies* (OUELLET *et al.*, 2014) : ainsi, dans *Les Cahiers du Numérique*, CARMES et NOYER (2015) soulignent dans « *L'irrésistible montée de l'algorithmique* », les dangers de l'externalisation du traitement de données trop abondantes (« *Too Big to Know* », WEINBERGER, 2012) hors des processus de la pensée humaine. Dans la lignée de la critique de la gouvernementalité statistique de DESROSIÈRES (2008a et 2008b), ces arguments font écho à la critique de la « *raison computationnelle* » évoquée par BACHIMONT (2008). À l'extrême, ANDERSON (2008) jouait la provocation en annonçant l'obsolescence de la méthode scientifique, déductive et contenue dans les limites de la pensée humaine, comme outil de lecture et d'ordonnement du monde.

Ainsi, le Big Data pose un défi de fond aux Sciences Sociales : l'avènement de ce « nouveau monde statistique » peut-il réellement bouleverser notre façon de comprendre la société – mais aussi de travailler à cette compréhension ?

Dans cette réflexion, nous proposons de contribuer à ce débat en l'indexant à un champ d'études particulier, celui de la mesure de la justice socio-spatiale, et en mobilisant les outils d'une science sociale particulière, la science économique. Cette dernière se révèle en effet singulièrement avide de données quantifiables permettant de dresser une cartographie fidèle du monde pour en corriger les dysfonctionnements, qu'il s'agisse d'inégalités ou d'inefficacités. Cette démarche positiviste est celle d'un grand nombre d'économistes empiristes « standard » dont l'effort de production scientifique est tourné vers la création d'outils de mesure pertinents au regard des problématiques sociales étudiées, et pour qui la théorie en découle et doit être amendée à l'épreuve des faits – positionnement méthodologique parfois critiqué au regard d'autres paradigmes des sciences économiques (*cf.* dans LABROUSSE, 2010, la critique adressée à Esther DUFLOT sur son positivisme).

Productrice proluxe d'indicateurs statistiques, la science économique « standard » semble donc constituer un débouché tout trouvé pour le Big Data, corde supplémentaire à un arc économétrique déjà bien étayé... Inversement, parce qu'elle se pose la question de la mesure, cette économie positiviste sait bien qu'aucune donnée n'est neutre, et que tout indicateur statistique n'est valable qu'au regard de ses propriétés axiomatiques rapportées à l'objectif de sa mesure. Cela la conduit à poser un regard critique sur la promesse – quelque peu fantasmée – d'un monde de chiffres rendu intelligible, comme par magie, par l'algorithmique inductive déshumanisée du Big Data.

Plus généralement, la science économique peut éclairer le débat à la lumière des méthodes d'analyse qui lui sont propres ; dans cette réflexion nous proposons de mobiliser pour ce faire deux outils propres à la science économique contemporaine : le raisonnement à la marge et le recours à l'économie normative. Raisonner à la marge conduit à s'intéresser au *coût d'opportunité* des changements étudiés, et confronter les avantages et les inconvénients des états du monde avant et après la révolution du Big Data. Par ailleurs, les outils de l'économie normative permettent d'apprécier l'opportunité des changements mis à jour : l'état du nouveau monde est-il plus juste, donc désirable, que l'état du monde ancien ? Quelles évolutions supplémentaires semblent nécessaires pour le rapprocher d'une situation réellement juste ?

Sur ces bases méthodologiques, nous tentons dans cet article de réflexion de contribuer au débat à partir de ces « tics de pensée » d'économiste, en confrontant, dans le cadre de la mesure de la justice socio-spatiale, les avantages et inconvénients normatifs de 'l'Ancien monde' de la statistique traditionnelle, hiérarchisé et déductif et ceux du 'Nouveau monde' du Big Data, décentralisé et inductif.

'L'Ancien Monde' hiérarchisé et déductif : quantifier pour dire le juste, sous contrôle procédural

Quantifier pour dire le juste

Pour un économiste, il est impensable de remettre en cause l'impérieuse nécessité de quantifier le réel pour dire le juste.

Ainsi, Jérémie BENTHAM, fondant l'utilitarisme, paradigme normatif dominant en économie, adopte une posture explicitement *conséquentialiste* où le juste n'est autre que le bien, celui-ci étant appréhendé à l'aune, hédoniste, d'une utilité nourrie par 14 plaisirs et 12 peines², qu'un *felicific calculus* permet à chacun de compiler dans l'intimité de son être. Sur cette base, le juste défini à l'échelle de la société n'est rien d'autre que la plus grande utilité (c'est-à-dire le plus grand bonheur) pour le plus grand nombre, avec, dans les modèles d'économie, la fameuse maximisation de la somme des utilités comme programme du planificateur bienveillant.

De cette fondation il découle qu'en économie énoncer le juste implique nécessairement de mesurer le bien-être ; réciproquement, il ne peut y avoir de discours économique intelligible sur le juste sans quantification. La quantification de la justice sociale ne peut être un mal, mais s'impose comme un bien nécessaire à la réflexion sur le bien commun. Sur cette base, il est donc impératif de se doter des outils de mesure les mieux fondés normativement, ce qui explique les 'guerres d'indicateurs' parfois dénoncée par les détracteurs de la science économique. Proposer de meilleures mesures du bien-être et de meilleurs indicateurs de sa

² La liste des plaisirs de BENTHAM est la suivante : « 1. Les plaisirs des sens. 2. Les plaisirs de richesse. 3. Les plaisirs de compétence. 4. Les plaisirs de la bonne entente. 5. Les plaisirs de renommée. 6. Les plaisirs du pouvoir. 7. Les plaisirs de piété. 8. Les plaisirs de bienveillance. 9. Les plaisirs de malveillance. 10. Les plaisirs de mémoire. 11. Les plaisirs d'imagination. 12. Les plaisirs d'attente. 13. Les plaisirs dépendant de l'association. 14. Les plaisirs de soulagement. ». La liste des peines est la suivante : « 1. Les douleurs de privation. 2. Les douleurs des sens. 3. Les douleurs de maladresse. 4. Les douleurs de discorde. 5. Les douleurs de mauvaise réputation. 6. Les douleurs de piété. 7. Les douleurs de bienveillance. 8. Les douleurs de malveillance. 9. Les douleurs de mémoire. 10. Les douleurs d'imagination. 11. Les douleurs d'attente. 12. Les douleurs dépendant de l'association. » (BENTHAM, 1789, chap. v, § 3).

distribution parmi les membres de la société, c'est, in fine, progresser dans l'appréciation de la justice sociale.

Réfléchir sur la justice socio-spatiale c'est donc avant tout, pour un économiste, construire des mesures géolocalisées du bien-être les plus pertinentes possibles, ainsi que des indicateurs spatialisés de la distribution de ce bien-être reposant sur les bases axiomatiques les plus satisfaisantes possibles.

Pauvreté des données, cécité normative

Dans ce contexte, la pauvreté des données géolocalisées disponibles dans l'Ancien monde statistique constituait un obstacle problématique à la mesure du juste, et donc à la possibilité de penser une action publique susceptible de rapprocher l'état réel de la société de la société idéale.

Une première source de données typique de l'Ancien monde est la mise à disposition de bases données administratives. Renvoyant à leur nature étymologique, ces « *statistiques* »³ (revenus fiscaux communaux de la Direction Générale des Impôts, déclarations administratives de données sociales des entreprises collectées par les centres de transfert des données sociales, nombre et caractéristiques des demandeurs d'emploi collectés par Pôle Emploi...) sont construites par et pour la bonne administration publique. D'autres données, moins nombreuses, proviennent de la réalisation d'enquêtes nationales pensées par l'appareil statistique pour informer les décideurs publics de la réalité des territoires qu'ils dirigent (recensement général de la population, enquêtes sur la mobilité des ménages, connaissance locale de l'appareil productif...).

Une seconde source de données géolocalisées est issue du travail « bottom-up » des chercheurs en sciences sociales autoproduisant des données (enquêtes, études de terrain...) nécessaires à la poursuite de leurs programmes de recherche.

Qu'elle soit administrative ou scientifique, cette production de données est symptomatique de la nature déductive de 'l'Ancien monde', où les statistiques

³ Le terme français 'statistique' est un emprunt récent au mot *Statistik*, forgé par l'économiste allemand ACHENWALL (1719-1772), qui l'a dérivé de l'italien *statista* « Homme d'État », la statistique représentant pour lui l'ensemble des connaissances que doit posséder un homme d'État (TLFI, 2015).

étaient issues d'une volonté humaine consciente et réfléchie – de bien gouverner dans un cas, de bien comprendre dans l'autre.

Ces dispositifs de production de données prêtent le flanc à de nombreuses critiques. Ils sont, tout d'abord, très coûteux, à tel point qu'un rapport de l'Assemblée Nationale indiquait récemment que « *Le coût de l'ancienne forme de recensement a été l'un des principaux motifs qui ont conduit à élaborer une nouvelle méthode. En effet, le dernier recensement général, qui a eu lieu en 1999, était initialement prévu pour 1997, mais a été repoussé pour des raisons budgétaires, puisque le surcoût lié à sa réalisation aurait pu notamment entraîner le non-respect des critères de Maastricht. Le coût budgétaire du recensement était en effet concentré sur une seule année et demandait donc un effort ponctuel important. Ainsi, le recensement de 1999 a coûté environ 1,2 milliard de francs, soit environ 180 millions d'euros.* » (GOSSELIN, 2008).

Dans le cas particulier des données géolocalisées s'ajoute la marge d'erreur statistique qui limite **leur exploitation statistique** pour les plus petites échelles ou les unités spatiales les moins peuplées.

Par ailleurs, les données géolocalisées sont aussi **difficilement accessibles** du fait, notamment, de la nécessaire protection de la vie privée : bien que l'INSEE travaille actuellement à la mise à disposition de données synthétiques carroyées selon une maille très fine, il n'est pas aisé, pour qui n'est pas affilié à un centre de recherche institutionnel et doté d'un solide projet de recherche, d'obtenir des données individuelles géolocalisées. Dans le même ordre d'idées, les données sur la localisation des individus (adresse postale, lieu de naissance, adresse IP, géolocalisation...) font partie des données sensibles dont la collecte et le traitement informatique par les chercheurs en sciences sociales nécessitent une demande d'autorisation préalable à la Commission Nationale Informatique et Libertés.

D'un point de vue normatif, ces restrictions, gênantes d'un point de vue conséquentialiste car limitant, in fine, le pouvoir de la société de se connaître elle-même, peuvent conduire à un optimum de justice sociale si l'on adopte le point de vue rival de la justice procédurale : leur existence permet le respect des droits des individus au sein de la société.

Au-delà, le principal inconvénient des données de l'ancien monde est qu'elles sont, le plus souvent, marquées du sceau des impératifs de l'administration publique : elles ne permettent de voir, de la société, que ce qui fait sens pour les nécessités de l'action publique. Cette subordination de la création de données géolocalisées aux contingences pratiques et politiques de la marche de l'État légitime toutes les critiques qui font écho à la « *gouvernementalité statistique* » de DESROSIERES (2008a et 2008b).

Par exemple, les données issues du Recensement Général de la Population permettent d'apprécier, avec une grande précision, l'étendue des privations des ménages défavorisés : il est ainsi possible de savoir dans quels quartiers la proportion de logements avec sanitaires est la plus faible, ou dans quelles communes enclavées le taux de ménages non motorisés est le plus élevé. Symétriquement, il est en revanche impossible de mesurer l'ampleur de la richesse des ménages favorisés, car les menus du questionnaire du recensement sont systématiquement tronqués vers le haut : au-delà d'un certain confort, les caractéristiques précises des logements 'sortent du radar' de la statistique publique. On peut savoir qu'un logement possède plus d'une salle d'eau ou plus de six pièces, mais rien ne permet de faire la différence entre un grand appartement, un hôtel particulier ou un château avec dépendances... Si la statistique publique permet d'étudier les terres de relégation sous toutes leurs coutures, les terres d'abondance échappent ainsi au regard des acteurs publics et des chercheurs : s'il est aisé d'identifier les zones de relégation marquées par la pauvreté et dont le destin socio-économique diverge de celui du reste du territoire (*cf.* par exemple PRÉTECEILLE, 2007, 2012 ou TOVAR, 2014), il est impossible d'identifier proprement, à partir des données de la statistique publique, les zones d'entre soi où les riches organisent leur sécession du reste du territoire. Cependant, du point de vue qui préside à la construction de ces données, celui de l'action publique, il n'est pas nécessaire d'en savoir plus : les données collectées suffisent pour guider les politiques publiques de désenclavement et de lutte contre la concentration géographique de la pauvreté.

À travers cet exemple, on conçoit comment l'origine et la nature des données de « l'Ancien monde » peuvent biaiser le regard que nous portons sur la société.

Coûteuses, leur portée est conditionnée par les finalités qui ont présidé à leur construction ; leur utilisation rigoureuse ne saurait s'affranchir d'une connaissance fine de leur processus de production – sous peine de tomber dans les travers dénoncés par ceux qui dénoncent l'implacable gouvernement des nombres. Dans l'univers 'déductif' et finaliste de l'Ancien monde, le danger est que les œillères statistiques deviennent des œillères intellectuelles⁴.

Or, dans le domaine de l'étude de la justice sociale et, plus précisément, de la dimension territoriale de la justice sociale, cet état de choses induit une relative pauvreté des indicateurs statistiques disponibles, alors que les théories de la justice contemporaines proposent des définitions théoriques riches et nuancées du bien-être, base informationnelle de la mesure de la justice sociale.

Prenons un autre exemple, celui du bien-être 'capabiliste' proposé par Amartya SEN, Nobel d'économie en 1998, en concurrence de l'utilité maniée par les économistes utilitaristes standards – et bien souvent résumée par le simple revenu disponible des ménages. L'étalon de mesure du bien-être défendu par SEN est, par essence, de nature multidimensionnelle (SEN, 1993, 2010). S'il se compose des réalisations effectives des individus mesurées selon des fonctionnements pluriels (revenu, santé, éducation, reconnaissance sociale, logement...) cet étalon de mesure intègre également deux dimensions de la liberté individuelle : la matrice des capacités renvoie à l'ensemble des réalisations effectives qu'une personne pourrait potentiellement mener (i.e. sa liberté *d'opportunité*), alors que la liberté procédurale (i.e. liberté de *choix*) reflète le degré de maîtrise des individus sur leur propre destinée. Sa finesse théorique rend l'approche par les capacités séduisante pour étudier le bien-être des personnes, mais sa mise en œuvre pratique est loin d'être aisée. Où trouver les indicateurs de mesure qui rendent justice à la richesse et à diversité des réalisations effectives des personnes ? Plus ardu encore, il est

⁴ On pense à cette blague bien connue sur les économistes : une nuit, un policier trouve un économiste cherchant quelque chose par terre sous un lampadaire. Il lui demande s'il a perdu quelque chose. L'économiste répond « J'ai perdu mes clefs dans la ruelle sombre de l'autre côté de la rue ». Le policier lui demande alors pourquoi il cherche ses clefs sous le lampadaire, et non pas dans la ruelle. L'économiste répond : « Parce qu'on y voit mieux ici pour chercher ».

impossible, à l'heure actuelle, de trouver des indicateurs statistiques (a fortiori géolocalisés) qui permettent de mesurer la liberté d'opportunités et la liberté procédurale de façon satisfaisante.

Au total, si l'on adopte un point de vue normatif conséquentialiste – celui de l'analyse économique où le caractère juste ou non d'une organisation sociale dépend de ses effets sur les êtres humains qui le composent – alors dans l'Ancien monde la production des données géolocalisées nécessaires à l'appréciation de la justice socio-spatiale ne peut être considérée comme satisfaisante.

Garanties procédurales

Cependant, ce jugement sans appel est plus que nuancé dès lors que l'on apprécie le juste d'un point de vue procédural, celui de la justice des règles qui président à l'ordonnement du monde. À cet égard, il semble intéressant de relater l'argument proposé par SEN lui-même au sujet des difficultés de mise en œuvre pratique de son approche. Défendant la nécessité, sous peine de paternalisme débridé, d'une indétermination théorique des composantes statistiques nécessaires à la mesure du bien-être capabiliste, il fait remarquer que le fait que les données statistiques soient si coûteuses à collecter contraint l'évaluateur à se contenter, *in fine*, des informations sur les éléments du bien-être disponibles... c'est-à-dire celles qui sont nécessairement celles les plus valorisés par la société.

D'une certaine façon, de par sa nature délibérée et hiérarchisée, le processus de production de statistiques de l'Ancien monde garantit la 'traçabilité' des données produites, *a fortiori* dans un système politique démocratique où l'action publique se soumet à la délibération publique – sous l'œil toujours critique des chercheurs.

Malgré cette absolue procédure qui permet de s'accommoder de la situation actuelle, on peut déplorer de n'avoir accès, en tant que chercheurs, qu'à des données légitimement valorisées par la société, via le filtre de l'action publique et de la discussion démocratique. On pourrait faire valoir, par contraste, l'intérêt d'avoir des données moins « orthodoxes » pour explorer des dimensions novatrices, ou marginales, dans la mesure du bien-être, détachées de toute finalité posée a priori.

C'est précisément la promesse que propose l'ère du 'Nouveau monde statistique', celui du 'Big Data'.

Un 'nouveau monde' décentralisé et inductif : promesses conséquentialistes, questionnement procédural

L'abondance statistique au service de la connaissance

L'émergence du Big Data ouvre des perspectives vertigineuses pour la mesure géolocalisée du bien-être. Pour un économiste, c'est un progrès inestimable pour la meilleure mesure des inégalités socio-spatiales et, d'un point de vue conséquentialiste, pour la mise en place de mesures de politiques publiques visant à corriger ces inégalités.

Tout d'abord, la quantité mais surtout la *nature* des données géolocalisées désormais accessibles ont subi une profonde mutation : l'immense océan des données qui constituent le Big Data est le fruit d'un processus de production inédit, décentralisé et marqué par une intentionnalité « de basse intensité ». Par contraste avec la collection réfléchie de données publiques, sont désormais à portée de clic la collection des innombrables traces statistiques laissées à jamais par nos existences numériques : contenu des recherches effectuées sur les navigateurs, courriels, profils et activité sur nos réseaux sociaux en ligne familiaux, professionnels, amicaux ou sentimentaux, achats, comptes bancaires en ligne, déplacements, données biologiques collectées par nos appareils connectés... Parce que notre vie se déroule désormais partiellement en ligne, la mémoire infinie du Web possède la comptabilité exacte de nos goûts et de nos opinions politiques, des multiples enchevêtrements de nos sociabilités, de notre activité productive, de notre rôle d'homo oeconomicus échangeant sur les marchés... mais aussi de l'empreinte de notre existence physique et peut-être, demain, de notre intimité physiologique. Cette accumulation de données s'opère le plus souvent indépendamment de notre volonté (même si les outils permettant de masquer nos traces numériques existent) ; elle est également parfois le résultat de la construction consciente de notre identité numérique : profils

de réseaux sociaux et professionnels, blogs, souscription à des services de stockage en ligne...

Il est donc désormais possible de percevoir, mesurer et quantifier comme jamais la réalité – notamment géographique – de nos existences : parce que les communications numériques médiatisent une part croissante de la réalité de nos vies, elles produisent des amas anarchiques de données non contrôlées, décentralisées, 'spontanées', d'une richesse et d'une pertinence sans commune mesure avec celles de l'Ancien monde.

Un exemple de ce « rêve d'économiste », de ce « brave new world » (SHEARMUR, 2015) dans le domaine de la mesure de la ségrégation résidentielle : deux chercheurs estoniens ont tout récemment confronté la mesure des interactions sociales des habitants de Tallinn issue des données de recensement (ségrégation nocturne) à celle qui résulte de la mesure de leur géolocalisation collectée par le biais de leurs téléphones portables (ségrégation diurne) (SILM et AHAS, 2014). Ils ont montré que, bien que les lieux de résidence soient nettement ségrégués, les différents groupes ethniques faisaient un usage partagé de la ville pendant la journée, avec une probabilité élevée de contacts inter-ethniques.

Cette étude, greffant des connaissances issues du 'nouveau monde' statistique à l'état des connaissances issues de « l'Ancien monde », ouvre de nouveaux questionnements prometteurs pour une meilleure compréhension de la ségrégation. De nombreux auteurs (comme le sociologue PRETECEILLE, 2007, 2014) expliquent que la ségrégation résidentielle est largement le fruit de la recherche de l'entre soi par les catégories économiquement dominantes, au demeurant fortement ségréguées. Par contraste, les catégories populaires, statistiquement moins ségréguées, vivraient dans des quartiers pauvres car délaissés par les plus riches. Si, comme le montre cette étude mobilisant des données issues du 'nouveau monde' statistique, la ségrégation diurne est plus faible que la ségrégation nocturne, cela signifie-t-il que les groupes dominants sont moins performants dans leurs stratégies d'évitement lors de leur usage quotidien de l'espace urbain ? Qu'ils perçoivent différemment leur projection diurne au travers du territoire de la ville et leur repli statique nocturne dans 'leur' quartier résidentiel ? Que les groupes dominés sont intégrés au

fonctionnement économique de la ville mais relégués dès lors qu'il s'agit de la sociabilité ? Certes, à l'aide d'outils éprouvés comme les monographies et les entretiens, les Science sociales pourraient jeter un faisceau lumineux ciblé sur des territoires, des groupes, des individus soigneusement sélectionnés au regard de ces questionnements. Au-delà, l'ère du Big Data laisse espérer des réponses plus globales grâce à la mise en lumière de régularités statistiques appréciées à l'échelle de l'ensemble de la société.

Il est un autre aspect des évolutions numériques qu'il faut souligner : la simplification de l'accès à l'ensemble des bases de données de l'Ancien monde : statistique publique, bases de données créées par les entreprises publiques et privées ou par des chercheurs isolés... Récemment organisée par la puissance publique elle-même (cf. www.data.gouv.fr) ces données sont plus visibles, plus facilement exploitables, en particulier par des non spécialistes.

Décentralisation et démocratisation du discours statistique

Nous progressons ainsi vers une plus grande démocratisation de la connaissance de la société. En particulier, l'ère numérique implique la diffusion des technologies de traitement des données, tout particulièrement géolocalisées, comme par exemple le logiciel libre R (www.cran.r-project.org) et ses modules sans cesse plus perfectionnés de cartographie et de d'analyse statistique des données géolocalisées. En prolongement de cette production d'outils collaboratifs, la pratique de l'analyse statistique est facilitée par l'émergence de communautés d'utilisateurs qui vulgarisent et diffusent outils, méthodes et bonnes pratiques économétriques. La production de discours statistiques n'est plus l'apanage des « sachants », spécialistes des administrations publiques ou chercheurs académiques accrédités.

Côté pile, le 'Big data' et la révolution numérique laissent espérer l'affranchissement de la production de données (notamment géolocalisées) vis-à-vis de la tutelle publique, et l'avènement d'une ère *d'abondance statistique* décentralisée et démocratique : le conséquentialiste saluera ainsi le dynamitage des verrous techniques posés à la connaissance des inégalités socio-spatiales. En allant jusqu'au bout de ce raisonnement, le Big Data pourrait constituer un outil de défense des

citoyens s'ils se servent de ces données pour dénoncer les abus éventuels de l'appareil d'Etat.

Menaces procédurales : entre fantasme et réalité

Côté face, cela pose d'épineuses questions procédurales : consentement à la diffusion de données personnelles, marchandisation de cette nouvelle ressource économique, contrôle de l'accès et de l'utilisation de ces données, protection des libertés individuelles, maîtrise des dispositifs d'analyse économétriques propres au Big Data...

On peut commencer par souligner l'invisibilité numérique des existences vécues hors de vue du 'tout social' virtuel. C'est le danger d'une nouvelle forme de relégation, cette fois entre habitants et exclus (par choix ou par défaut de communication ou de consommation) de la virtualité partagée. Si je n'ai pas de smartphone, comment puis-je être inclus dans la mesure de la ségrégation proposée par les chercheurs estoniens ? Ici, c'est entre espace et cyberspace que se nouent les enjeux de la justice spatiale. Plus généralement, SHEARMUR (2015) explique que le Big Data ne peut traiter que de l'information codifiable et quantifiable, mais ne peut conduire, sans médiatisation humaine, à la compréhension de l'Humanité.

Symétriquement se pose la question du défaut d'opacité et de profondeurs temporelles et spatiales du cyberspace. Dans l'espace physique, réel, il est possible de faire (même temporairement) sécession du social et de protéger son intimité dans des lieux privés, soustraits à la vue publique et de mieux en mieux protégés⁵. L'inégale répartition de cette possibilité de « vivre cachés pour vivre heureux » est d'ailleurs, comme nous l'avons vu plus haut, une clef de lecture de la justice spatiale, le privilège de l'invisibilité étant celui des catégories dominantes. Dans le cyberspace, les barrières qui garantissent notre intimité semblent bien minces : ordinateurs peu sécurisés, traçage et écoute systématiques de nos courriels et déplacement sur la Toile, appareils connectés indiscrets... On peut peut-être s'amuser

⁵ A tel point que l'une des raisons avancées par la statistique publique allemande pour renoncer à un recensement exhaustif est la multiplication des barrières (portes, concierges, digicodes...) empêchant le contact entre les citoyens et les agents recenseurs, œil statistique de l'État.

de la mise à niveau des puissants et des misérables face aux grandes oreilles d'Échelon et de la NSA, François Hollande et Angela Merkel leur étant tout aussi accessibles que le plus modeste citoyen européen. La conséquence est, cependant, de mener nos vies sous un 'œil de Sauron' enregistrant tout ce qui devrait pouvoir être tu : actions, opinions, faiblesses...

On pourrait réinterpréter l'impératif de justice spatiale comme la nécessaire étanchéité, procédurale, entre l'espace intime et un espace public fossilisé par le virtuel.

Une troisième inquiétude procédurale porte sur le contrôle de ces données nouvelles. L'abondance informationnelle du Big Data fournissent les moyens de la connaissance orwellienne de l'ensemble des espaces de notre existence : celui du monde réel, celui du monde virtuel et, bientôt, celui du monde intime de nos corps et de nos esprits. D'un point de vue procédural, la question est ouverte de dire si un tel niveau de transparence est une mauvaise chose en soi ; d'un point de vue conséquentialiste, tout dépend de ce qui en est fait.

Un élément de débat porte sur la marchandisation de ce que MAYER-SCHÖNBERGER (2014), qualifie de « nouvelle ressource économique ». Les données du Big Data *« peuvent être utilisées indéfiniment à des fins multiples et nouvelles, une valeur supplémentaire étant produite à chaque opération. (...) [Leur réutilisation] ouvre la voie à la création de nouveaux produits et services, et donc à de nouveaux flux de revenus pour les entreprises – ce qui pourrait conduire à l'évolution de leur business model. »* Ainsi, le modèle économique des grands acteurs du numérique qui proposent des services « gratuits » en ligne (moteurs de recherche, sites marchands, réseaux sociaux, applications...) repose de plus en plus sur la collecte et la revente de ces précieuses données.

Cette marchandisation de nos empreintes virtuelles est mal connue des citoyens et échappe de facto aux autorités dotées de la légitimité procédurale garantie par le suffrage universel. Dans ce contexte, on pourrait proposer un nouvel idéal, procédural, de justice spatiale, avec la possibilité, pour des administrations représentant la légitimité démocratique sur un territoire, d'imposer des procédures

de contrôle justes et transparentes à des entités qui apparaissent, à tort, comme pouvant exister « hors sol », déconnectées du monde réel.

Il ne s'agit cependant pas d'une question de fond : bien que lents à se mettre en place, on peut espérer, à terme, une régulation démocratique prochaine de ce nouveau monde numérique. En revanche, la nature de l'analyse statistique du Big Data soulève une question autrement épineuse.

Avec le Big Data, comme nous l'avons vu, il ne serait plus nécessaire de chercher pour trouver, l'algorithmique informatique permettant, mieux que l'intentionnalité humaine, de faire émerger des liens statistiques sans besoin d'avoir recours à un quelconque modèle causal préalable, la masse des données disponibles permettrait d'utiliser des modèles probabilistes capables de *prédire* des événements.

Un exemple souvent cité concerne un enjeu important en termes de justice socio-spatiale, celui du ciblage géographique et du profilage ethnique de l'action policière. De plus en plus utilisés par les forces de police aux Etats-Unis, les logiciels de « *Predictive Policing* » révèlent, au pâti de maisons près, la localisation géographique de la criminalité avant même que celle-ci ne soit enregistrée (PEARSALL 2010). Le recours à ces techniques de Big Data se fait bien évidemment au nom d'un principe tout conséquentialiste de meilleure efficacité de l'action policière, mais des considérations procédurales entrent aussi en jeu. Il s'agit en effet aussi de garantir un meilleur respect du 4^e amendement de la Constitution américaine protégeant les citoyens contre des perquisitions ne reposant pas sur des 'présomptions sérieuses'⁶. En effet, le *Predictive Policing* guide « objectivement » l'action policière, ce qui constituerait un progrès par rapport aux intuitions de policiers subjectifs, partiels et arbitraires – ou tout simplement racistes (GUTHRIE FERGUSON 2012 ; KOSS 2015).

Cela soulève immédiatement de nombreux problèmes de protection des libertés civiques (GUTHRIE FERGUSON 2012 ; KOSS 2015 ; SPRAGUE 2015) : le Big Data

⁶ Cet amendement établit le droit des citoyens « d'être garantis dans leurs personne, domicile, papiers et effets, contre les perquisitions et saisies non motivées ne sera pas violé, et aucun mandat ne sera délivré, si ce n'est sur présomption sérieuse, corroborée par serment ou affirmation, ni sans qu'il décrive particulièrement le lieu à fouiller et les personnes ou les choses à saisir ».

donnerait ainsi une légitimation statistique du harcèlement par la police des populations (essentiellement Noires et Latines) qui vivent dans les quartiers les plus défavorisés – et donc criminogènes (CRAWFORD et SCHULTZ 2014 ; SPRAGUE 2015 ; BAROCAS et SELBST 2016).

Cependant, d'un strict point de vue statistique, le Big Data n'infère aucune causalité et ne fait que proposer des inférences probabilistes. Le logiciel ne dit en aucun cas qu'une personne possédant tel ensemble de caractéristiques, se déplaçant à pied dans telle banlieue américaine, est nécessairement un criminel ; il indique seulement qu'à une heure donnée et à un endroit donné, la probabilité qu'une telle personne commette tel type de crime est élevée. La « *dictature des données* » ne découlerait donc pas du Big Data lui-même, mais des imperfections des humains qui le manipulent : incapables de penser de façon probabiliste, irrémédiablement contaminés par un besoin viscéral de causalité, les utilisateurs humains du Big Data l'utiliseraient à tort pour confirmer leurs préjugés (MAYER-SCHÖNBERGER, 2014).

« *There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot. (...) The opportunity is great: (...) Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?* » (ANDERSON, 2008)

Cet extrait de la célèbre tribune d'ANDERSON (2008) dans *Wired* illustre à la fois le dogme des pro-Big Data et les fantasmes de ses détracteurs. Il laisse entrevoir un monde il faudrait en finir avec la Science, cette obsession humaine pour la modélisation et la recherche de causalité, et laisser penser les machines à notre place. Cette image vertigineuse d'une société régulée par une algorithmique infaillible pose problème. En effet, tant le dogme du Big Data que sa crainte reposent sur deux croyances fragiles : le 'naturalisme' des données collectées par le Big Data et la supériorité des puissantes corrélations inductives du Big Data sur la causalité, graal de la méthode scientifique.

Or, comme le rappellent CRAWFORD, MILTNER et GRAY (2014), les chercheurs en Sciences sociales, habitués de l'Ancien monde statistique, ont depuis longtemps

établi la nature éminemment sociale de la production et de l'utilisation des données statistiques : « *Raw Data is an Oxymoron* », comme le résume le titre de l'ouvrage collectif dirigé par Lisa GITELMAN (2013), et les données collationnées par le Big Data ne font pas exception.

Par ailleurs, si le Big Data permet de déceler des corrélations inédites dissimulées dans la masse des données numériques, il n'y a rien de bien révolutionnaire au regard de ce qu'est la Statistique, définie ainsi par COURNOT en 1843 : « *ensemble de techniques d'interprétation mathématique appliquées à des phénomènes pour lesquels une étude exhaustive de tous les facteurs est impossible* ». Ensuite, il peut proposer une vision probabiliste du monde. Pour autant, au-delà des formules incantatoires, il semble bien hâtif d'en déduire une quelconque supériorité de la corrélation comme outil de compréhension du monde.

Comme en témoignent les corrélations improbables moquées par le site [Spurious Correlations](#) (VIGEN, 2015), il semble peu sérieux, comme tentent de le faire croire et les zélotes du Big Data et ses opposants catastrophistes, de prétendre pouvoir se passer de la rationalité humaine pour saisir la portée des résultats proposés par le Big Data.

« *The big data team simply uncovered better, more meaningful correlations. (...)” Big data analysis can be about correlations OR causation—it all depends, as it has always been, on what question we are asking, what problem we are solving, and what goal we are trying to achieve. I don't think big data will do anything to—and has little to do with—our obsession with causation. But as Big Data successfully demonstrates, this is one technology-driven phenomenon that can improve our lives and require all of us to pay attention and start engaging in a meaningful conversation of what to do about its potential risks.* » (PRESS, 2013)

Les algorithmes ne sont donc pas une nouvelle divinité numérique : ce sont des objets sociaux, construits et utilisés par des humains au regard de motivations particulières. Ainsi, loin de souscrire au spectre dystopique du roman de science-fiction *Minority Report* de Philip K. DICK (1956), on peut souligner la neutralité intentionnelle de ce qui n'est, in fine, qu'un simple outil d'analyse statistique. D'un point de vue conséquentialiste, le Big Data, arme potentielle des oppresseurs, peut tout aussi bien être un outil de libération aux mains des opprimés. Confrontée aux

dérives du '*Predictive Policing*', la société civile américaine a ainsi développé des applications de contrôle de la police comme Cop Watch et utilise des méthodes de Big Data pour prédire... l'usage irraisonné de la force par les départements de la police municipale.

[Conclusion] Que faire ? Programme de travail pour économiste

Le Big Data ne relève donc ni de la magie noire ni de la magie blanche : ce n'est qu'un nouvel outil statistique et, en tant que tel, un construit social dont la portée est limitée et qui peut et doit être soumis à la critique des Sciences sociales :

« *This points to the next frontier: how to address these weaknesses in big data science. In the near term, data scientists should take a page from social scientists, who have a long history of asking where the data they're working with comes from, what methods were used to gather and analyze it, and what cognitive biases they might bring to its interpretation (...). Longer term, we must ask how we can bring together big data approaches with small data studies - computational social science with traditional qualitative methods. (...) This goes beyond merely conducting focus groups to confirm what you already want to see in a big data set. (...) Social science methodologies may make the challenge of understanding big data more complex, but they also bring context-awareness to our research to address serious signal problems. Then we can move from the focus on merely "big" data towards something more three-dimensional: data with depth* ». (CRAWFORD, 2013)

Comment un économiste peut-il contribuer à l'effort collectif d'intelligibilité appelé par CRAWFORD (2013) ?

Du point de vue conséquentialiste qui caractérise la science économique, la participation du plus grand nombre à la grande trace laissée par l'enchevêtrement de nos existences virtuelles est nécessaire. Cela donnerait en effet une connaissance fidèle de la réalité de nos sociétés, et permettrait à tous de peser dans la mesure de la justice sociale, « *chacun comptant pour un, et aucun pour plus d'un* » selon la vénérable maxime utilitariste. Cela implique de soutenir la lutte contre la fracture numérique en incluant les relégués dans la société du tout connecté, et d'organiser l'enregistrement égal de l'existence de chacun.

Dans le même temps, le respect des libertés publiques et la nature procédurale de la justice sociale ne doivent pas être oubliées, avec l'édiction de règles justes encadrant

la participation à ce nouvel espace public (l'économie normative souligne l'importance des critères⁷ de réversibilité, de publicité et de participation).

Par ailleurs, dans le prolongement du travail axiomatique mené sur les indicateurs statistiques de 'l'Ancien monde', les économistes pourraient contribuer à la formulation des propriétés normatives que devraient respecter les algorithmes du Big Data, déconstruisant leur image de boîtes noires divines hermétiques à la pensée humaine.

Enfin, si l'on assimile cette compilation de notre existence virtuelle à un bien commun à protéger et à partager, il faut éviter que le marché ne décide, seul, de sa collecte, de son stockage, de son échange et de sa valorisation. Les données de l'ère numérique sont non rivales : elles ne sont pas consommées et détruites par leur utilisation. Elles sont par ailleurs faiblement *excludables* : il est difficile d'empêcher leur utilisation. Ce sont donc, a minima, des « biens de club » ; a maxima, des « biens publics ». Pour un économiste, la puissance publique doit nécessairement être chargée de leur régulation, afin de garantir leur efficace production et leur juste répartition.

Au terme de cette réflexion, le Big Data et le monde statistique nouveau qu'il symbolise apparaissent comme des améliorations par rapport à la pénurie de données de l'Ancien monde statistique. Cependant – et c'est finalement assez satisfaisant pour un économiste – il s'agit plus d'une amélioration marginale dont il convient de ne pas surestimer l'ampleur... mais dont il faut contrôler l'exploitation.

A propos de l'auteur : Tovar Elisabeth, Maître de Conférences, Université Paris Ouest et EconomiX (UMR 7235)

Pour citer cet article : « Mesurer la justice socio-spatiale : de l'ancien au nouveau monde, promesses et menaces du « Big Data », *justice spatiale | spatial justice*, n°10, Juillet 2016, <http://www.jssj.org>

7 Parmi un grand nombre d'autres : conformité, réversibilité, possibilité de recours, publicité, contradiction, participation, motivation, preuve, indépendance, impartialité, compétence, légalité...

Bibliographie

- ANDERSON Chris**, « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », *Wired*, 2008, [URL: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory].
- ANDERSON Chris**, *Makers: The new industrial revolution*, New York, Crown Business, 2012.
- BACHIMONT Bruno**, « Formal Signs and Numerical Computation: Between Intuitionism and Formalism. Critique of Computational Reason ». In : H. Schramm, L. Schwartz et J. Lazardzig (éds.), *Theatrum Scientiarum: Instruments in Art and Science, on the Architectonics of Cultural Boundaries in the 17th Century*, 362-382. Berlin: Walter de Gruyter Verlag, 2008.
- BAROCAS Solon et SELBST Andrew D.**, « *Big Data's Disparate Impact* », *California Law Review*, n°104, 2016 [URL : http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899].
- BENTHAM Jeremy** (1789) *Introduction to the Principles of Morals and Legislation*, 1789, [URL : <http://oll.libertyfund.org/titles/278>].
- CARMES Maryse, NOYER Jean-Max**, « L'irrésistible montée de l'algorithmique. Méthodes et concepts en SHS », *Les Cahiers du numérique*, Vol. 10, n°4, 63-102, 2014.
- COURNOT Antoine-Augustin**, *Exposition de la théorie des chances et des probabilités*, Paris, Hachette 1843. [URL : <http://gallica.bnf.fr/ark:/12148/bpt6k285042>]
- CRAWFORD Kate, MILTNER Kate et GRAY Mary L.**, « Critiquing Big Data: Politics, Ethics, Epistemology », *Introduction au numéro spécial, International Journal of Communication* 8, 1663-1672, 2014.
- CRAWFORD Kate, SCHULTZ Jason**, « Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms », *Boston College Law Review*, vol 93, 93-128, 2014.
- CRAWFORD Kate**, « The hidden biases in big data », *Harvard Business Review*, 1, 2013 [URL : <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data>].
- DELORT Pierre** (2015) *Le Big Data*. Paris, PUF, Collection Que sais-je ? n°4021, 128 p.
- DESROSIERES Alain**, *Gouverner par les nombres. L'Argument statistique II*, Paris, Presses de l'École des Mines de Paris, 2008b.
- DESROSIERES Alain**, *Pour une sociologie historique de la quantification. L'Argument statistique I*, Paris, Presses de l'École des Mines de Paris, 2008a.
- DICK Philip K.**, « Minority Report », *Fantastic Universe*, 1956.
- DIMINESCU Dana et WIEVIORKA Michel**, « Le défi numérique pour les sciences sociales », *Socio. La nouvelle revue des sciences sociales*, n°4, 2015. [URL : <https://socio.revues.org/1254>]
- GITELMAN Lisa** (éd.) *Raw data is an oxymoron*. Cambridge, MIT Press, 2013.
- GOSSELIN Philippe**, « Rapport d'information sur la nouvelle méthode de recensement de la population », *Rapport d'information de l'Assemblée Nationale*, n°1246, Assemblée Nationale, 2008.
- GUTHRIE FERGUSON Andrew**, « Predictive policing and reasonable suspicion », *Emory Law Journal*, 1, 2012.

- HU Han, WEN Yonggang, CHUA Tat-Seng, LI Xuelong** « *Towards scalable systems for big data analytics: a technology tutorial* » IEEE Access Vol 2, 652–687, 2014.
- KOSS Kelly K.**, « Leveraging Predictive Policing Algorithms To Restore Fourth Amendment Protections In High-Crime Areas In A Post-Wardlow World », *Chicago-Kent Law Review*, vol 90, n°1, 301-334, 2015.
- LABROUSSE A.**, « Nouvelle économie du développement et essais cliniques randomisés : une mise en perspective d'un outil de preuve et de gouvernement », *Revue de la régulation*, n°7, 2010, [URL : <http://regulation.revues.org/index7818.html>].
- LANEY Douglas**, 3D Data Management: Controlling Data Volume, Velocity, and Variety, META Group, 2001 [URL : <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>].
- MAYER-SCHÖNBERGER Viktor**, « La révolution Big Data », *Politique étrangère*, n°4, 69-81, 2014
- MAYER-SCHÖNBERGER Viktor, CUKIER Kenneth**, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, Eamon Dolan/Houghton Mifflin Harcourt, 2013.
- MOSCO Vincent**, *To the Cloud: Big Data in a Turbulent World*, Paradigm Publishers, 2014.
- OUELLET Maxime, MONDOUX André, MÉNARD Marc, BONENFANT Maude, RICHERT Fabien**, « *Big Data*, gouvernance et surveillance », *Cahiers du CRICIS*, n°2014/1, 2014, [URL : http://www.archipel.uqam.ca/6469/1/CRICIS_CAHIERS_2014-1.pdf].
- PEARSALL Beth**, 'Predictive Policing: The Future of Law Enforcement?', *National Institute of Justice Journal*, n°266, 2010 [URL : <http://www.nij.gov/journals/266/Pages/predictive.aspx>]
- PRÉTECEILLE Edmond**, "Segregation, social mix and public policies in Paris", In T. Maloutas et K. Fujita (éds.), *Residential Segregation Around the World. Making sense of contextual diversity*, Ashgate, 153-176, 2012.
- PRESS Gil**, « Big Data News Roundup: Correlation vs. Causation », *Forbes Tech*, 19 avril, 2013 [URL : <http://www.forbes.com/sites/gilpress/2013/04/19/big-data-news-roundup-correlation-vs-causation/>]
- PRÉTECEILLE Edmond**, « Is gentrification a useful paradigm to analyse social changes in the Paris metropolis? » *Environment and Planning A*, Vol 39, n°1, 10-31, 2007.
- SEN Amartya**, *Éthique et économie*, Paris, Presses universitaires de France, 1993.
- SEN Amartya**, *L'idée de justice*, Paris, Le Seuil, 2010.
- SHEARMUR Richard**, « Dazzled by data: Big Data, the census and urban geography », *Urban Geography*, vol. 36, No. 7, 965–968, 2015 [URL <http://dx.doi.org/10.1080/02723638.2015.1050922>]
- SILM Sirii, AHAS Rein**, « The temporal variation of ethnic segregation in a city: Evidence from a mobile phone use dataset », *Social Science Research*, Vol. 47, 30–43, 2014.
- SPRAGUE Robert**, « Welcome to the Machine: Privacy and Workplace Implications of Predictive Analytics », *Richmont Journal of Law and Technology*, 21, 2015 [URL : <http://jolt.richmond.edu/v21i4/article13.pdf>]

TOVAR Élisabeth, « Mesurer la pauvreté : l'apport de l'approche par les capacités. L'exemple de l'aire urbaine parisienne en 2010 », *Informations Sociales*, n°82 (Mars-Avril), 2014.

UNIVERSITÉ DE LORRAINE et UMR ATILF (Analyse et Traitement Informatique de la Langue Française), « Statistique », Dictionnaire Trésor de la Langue Française Informatisé (TLFI), [URL : <http://www.cnrtl.fr/etymologie/statistique>]

VIGEN Tyler, *Spurious Correlations*, Hachette Books, 2015.

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, Basis Books, 2012.