

L'involution géographique : des données géosociales aux algorithmes

Geographic Involution: from Geosocial Data to Algorithms

Teriitutea Quesnot

**Éditeur**

Netcom Association

Édition électroniqueURL : <http://netcom.revues.org/2545>

ISSN : 2431-210X

Édition imprimée

Date de publication : 16 décembre 2016

Pagination : 281-304

ISSN : 0987-6014

Référence électronique

Teriitutea Quesnot, « L'involution géographique : des données géosociales aux algorithmes », *Netcom* [En ligne], 30-3/4 | 2016, mis en ligne le 22 mars 2017, consulté le 29 mars 2017. URL : <http://netcom.revues.org/2545>



Netcom – Réseaux, communication et territoires est mis à disposition selon les termes de la licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Networks and Communication Studies,
NETCOM, vol. 30 (2016), n° 3-4
pp. 281-304

L'INVOLUTION GEOGRAPHIQUE : DES DONNEES GEOSOCIALES AUX ALGORITHMES

GEOGRAPHIC INVOLUTION: FROM GEOSOCIAL DATA TO ALGORITHMS

QUESNOT TERIITUTEA¹

Résumé - *Les données géolocalisées émanant du web social suscitent actuellement un vif intérêt chez les géographes. Certains voient en elles un potentiel indéniable pour la recherche académique. D'autres, en revanche, se contentent d'émettre des réserves quant à leur réelle utilité. Je soutiens pour ma part que les bénéfices que nous pouvons tirer de leur analyse sont très minces, voire inexistantes ; pour peu que l'on considère l'espace non pas comme une simple étendue, mais comme un véritable contenu de la relation sociale. Trois arguments que j'explique ici me poussent à formuler une telle affirmation : les « données » « géosociales » sont (1) fabriquées de toute pièce par des sociétés privées dont l'objectif premier est d'engranger des capitaux, (2) dépourvues de contexte et de ce fait (3) uniquement interprétables en termes de position.*

Mots-clés - *Big data, Espace social, Géographie critique, Lieu, Spatialité humaine.*

Abstract - *Geolocated data shared on the social web arouse geographers' interest. While some of them advocate their potential for academic research, others remain less enthusiastic. I argue here that the benefits from the analyses of such data are poor, nay absent; if Space is considered not just as a simple extent, but as a meaningful content of social interactions. In particular, I put forward three arguments to support this position: « geosocial » « data » are (1) manufactured by profit organizations, (2) contextless and consequently (3) exclusively interpretable in terms of location.*

Keywords - *Big data, Critical geography, Human spatiality, Place, Social space.*

¹ Docteur en sciences géomatiques - Université Laval, 1055 Avenue du Séminaire, Pavillon Louis-Jacques Casault, Québec (QC) G1V 0A6, Canada - teriitutea.quesnot.1@ulaval.ca.

1. MISE EN CONTEXTE ET CONSIDERATIONS PRELIMINAIRES

Je souhaite défendre à travers ces quelques lignes un point de vue qui va à l'encontre d'une tendance que l'on peut observer chez une partie des géographes : celle de *croire* que l'information géolocalisée partagée sur les réseaux sociaux numériques – ou « données géosociales » – présente un intérêt indéniable pour la recherche académique en géographie. Grâce aux interfaces de programmation (APIs), cette masse de « données » générée en continu est accessible au monde de la recherche.

Pourquoi donc s'en priver ?

En réalité, ce n'est pas l'exploitation des données géosociales qui est ici pointée du doigt, mais plutôt leur prétendue portée géographique et sociale. J'aimerais clarifier ce point d'emblée : à moins de mettre un trait d'égalité entre les termes « géographie » et « localisation », les données géosociales n'ont de « géographique » et de « social » que le qualificatif qu'on veut bien leur prêter.

Si l'on envisage la géographie comme une science sociale dont l'espace, qui en est l'objet, constitue la dimension de toute vie sociale (Lévy, 2003a), l'adjectif « géosocial(e) » relève alors du pléonasme puisque ce qui est « géographique » demeure « social » par essence. Dans cet esprit, je soutiens que les données géosociales ne sont pas non plus des « données géographiques » à proprement parler, mais des « métadonnées géographiques ».

À l'heure actuelle, le phénomène Big data est sur toutes les lèvres sans être clairement défini pour autant. Les réseaux sociaux ont par ailleurs connu de profondes mutations depuis l'apparition de l'Internet mobile et l'essor de la géolocalisation de personnes et de contenus numériques. Ces récents bouleversements ont eu un impact sur notre façon d'appréhender l'espace, de le penser, de communiquer à travers lui et à propos de lui, mais aussi de l'étudier. Une mise en contexte s'impose donc avant de présenter les arguments en faveur de la position que je défends dans cet article.

1.1. Rupture cartographique et néogéographies

La démocratisation des plateformes cartographiques qui s'opère depuis la sortie de Google Maps en 2005 modifie profondément notre rapport à la carte et donc notre manière d'appréhender l'espace. L'échelle figée des cartes papier a perdu de sa signification avec le caractère dynamique des cartes numériques. La possibilité pour l'utilisateur de superposer des orthophotographies annihile l'aspect symbolique de la carte. Grâce aux rendus photographiques en 3D popularisés par le service *Street View* de Google, la perspective spatiale égocentrée cohabite désormais avec l'allocentrisme typique des cartes traditionnelles. Mais encore, le fait que chaque internaute puisse « librement » contribuer à l'enrichissement de contenus cartographiques sur de telles plateformes (p. ex. OpenStreetMap) marque une véritable rupture avec l'art de la cartographie réservé autrefois aux initiés.

Ces individus qui apportent leur pierre à l'édifice participent à un mouvement qualifié de « néogéographique ». C'est précisément dans ce contexte que Goodchild (2007) a introduit le fameux concept de « *Volunteered Geographic Information* » (VGI). Selon lui, le phénomène VGI a été rendu possible grâce à un ensemble de technologies parmi lesquelles figurent le Web 2.0, les méthodes de géoréférencement et de géotagging (c.-à-d. créer des balises de géolocalisation), les systèmes dotés de puces GPS, les ordinateurs dont le gain de puissance est en constante évolution et bien évidemment, l'accès généralisé à l'Internet. Le concept de VGI a rapidement pris de l'ampleur dans le milieu académique ; en particulier chez les géographes et les chercheurs en *Geographic Information Science* ou GIScience (cf. Egenhofer et al., 2016 pour plus d'informations sur cette discipline). Certains collègues ont néanmoins questionné le caractère « volontaire » des VGI en mettant en avant le fait que nous laissons – souvent à notre insu – des traces de nos déplacements ici et là. Ainsi, Fischer (2012) est allé jusqu'à préconiser l'emploi du terme *iVGI* (la lettre *i* signifiant « *Involunteered* ») alors que Harvey (2013) suggère plutôt de dissocier l'information géographique « volontaire » de l'information géographique « contributive ».

À mon sens, ce débat n'a pas lieu d'être, car le terme « *Volunteered* » a été interprété selon l'anglais américain (c.-à-d. accomplir une tâche sciemment) alors que Goodchild a tendance à employer un vocabulaire propre à l'anglais britannique. Il préfère par exemple l'adjectif « *Geographical* » à « *Geographic* » lorsqu'il introduit la GIScience (cf. Goodchild, 1992). « *Volunteered* » doit donc être compris au sens britannique ; soit « Contributif ». En outre, Fischer (2012) et Harvey (2013) justifient leur point de vue respectif en soulignant que la capture de l'information géographique se fait de plus en plus sans notre consentement *via* les dispositifs mobiles « intelligents ». Or, Goodchild (2007) n'évoque à aucun moment ces technologies. Et pour cause, le premier véritable smartphone équipé d'une puce GPS (l'iPhone 3GS) n'est sorti que l'année suivante, en 2008. Toute cette confusion est en grande partie liée à l'essor des Big data et aux rapides avancées technologiques que nous connaissons depuis près d'une décennie.

1.2. Les enjeux des Big data pour les géographes

La recherche autour des Big data s'intensifie, mais force est de constater qu'il n'existe aucun réel consensus sur leur définition (Goodchild, 2013). La formule des « 3V » proposée par Laney en 2001 a longtemps été reprise pour caractériser les Big data : *V*olume (en référence à leur masse imposante) ; *V*itesse (elles sont créées sans interruption) ; *V*ariété (elles se présentent sous plusieurs formes). Kitchin (2014) remarque que quatre critères ressortent dans la récente littérature scientifique : les Big data sont considérées comme (1) exhaustives ($n=all$) ; (2) hyperdétaillées ; (3) relationnelles (possibilité de croisements avec d'autres données) ; (4) flexibles (c.-à-d. aisément modifiables). De façon plus générale, le phénomène Big data s'apparente à un ensemble de discours, de pratiques et de *croyances* centrés sur la donnée, l'Homme et son comportement.

Les approches classiques de stockage et de traitement des données sont devenues obsolètes. Les traditionnels SGBDR (systèmes de gestion de bases de données relationnelles) et leur langage de requête SQL ne suffisent plus. Pour gérer ces données massives, nous devons faire appel à une palette d'outils alternatifs à l'instar des moteurs NoSQL (p. ex. MongoDB) et des *frameworks* assurant des traitements distribués (p. ex. Hadoop). Par conséquent, les chercheurs consacrant leurs travaux aux Big data doivent se former un minimum à ces technologies dites « non conventionnelles ». Les « *Data Scientists* » et autres « *Data Analysts* », ceux dont la tâche est de « faire parler les données », constituent des profils très recherchés dans l'industrie.

Les géographes n'échappent pas à ces contraintes, mais les défis qui viennent d'être exposés ne sont pas si récents que cela pour celles et ceux qui exploitent les savoir-faire techniques de la géographie (SIG, télédétection, etc.). Comme l'évoque Goodchild (2013), ces derniers sont déjà confrontés depuis la fin des années 1980 aux enjeux imposés par les 3V. Les géomaticiens travaillent effectivement sur des données géographiques dont le poids peut atteindre des niveaux considérables (p.ex. les fichiers LIDAR à haute résolution). En particulier, les « SIGistes » jonglent en permanence avec différents types de donnée (fichiers vectoriels, orthophotographies, bases de données spatiales, tables, etc.) qui nécessitent quasi systématiquement un nettoyage en amont. La vitesse me semble donc être le paramètre avec lequel les adeptes des savoir-faire techniques de la géographie sont le moins familiers.

Hormis les 3V, je pense que le changement le plus radical auquel nous – géographes – ayons à faire face est celui de travailler sur des données « géographiques » distribuées par des sociétés privées et, comme je le montre plus loin, produites *en apparence* par les citoyens. Je pense notamment à l'information géolocalisée qui est quotidiennement partagée sur les réseaux sociaux. Celle-ci est d'autant plus intrigante qu'elle se présente aux chercheurs sous plusieurs formes. Pour reprendre la classification de Kitchin (2013), une seule et même donnée géosociale peut être en effet considérée comme « automatisée », « contributive », voire les deux à la fois.

1.3. La place des réseaux sociaux dans la sphère des Big data

Les réseaux sociaux occupent une place importante dans la sphère des Big data. D'ailleurs, le web social est souvent confondu avec ces plateformes. Pourtant, il s'agit bel et bien de deux réalités distinctes. Précisément, le web social est avant tout une composante du web 2.0 (Proulx et Millerand, 2010). Il hérite donc des mêmes traits de caractère que celui-ci : créer, partager et mettre à jour rapidement des contenus numériques constitue le dénominateur commun de tous les outils du web social ; dont font partie les réseaux sociaux. Le web social englobe donc ces derniers, mais également d'autres outils comme les blogues, les plateformes d'échange et de partage, ou encore les sites web collaboratifs.

Claire et concise, la définition des réseaux sociaux numériques proposée par Coutant et Stenger mérite d'être citée (2009 ; p. 5) :

Ils constituent des services web qui (1) permettent aux individus de construire un profil public ou semi-public au sein d'un système, (2) de gérer une liste des utilisateurs avec lesquels ils partagent un lien, (3) de voir et naviguer sur leur liste de liens et sur ceux établis par les autres au sein du système ; et (4) fondent leur attractivité essentiellement sur les trois premiers points et non sur une activité particulière.

Toujours selon Proulx et Millerand, on peut distinguer trois types de réseaux sociaux : (1) amical (p. ex. Facebook), (2) professionnel (p. ex. LinkedIn), et (3) de rencontre (p. ex. Meetic). Dans les faits, les réseaux sociaux amicaux sont ceux qui rassemblent le plus d'individus et qui, par conséquent, génèrent le plus d'activités sur la toile. Il n'est donc pas étonnant que ces plateformes aient suscité l'intérêt des chercheurs dès leur création. Néanmoins, les géographes et autres chercheurs s'intéressant à l'espace ont dû attendre la démocratisation de la géolocalisation pour voir apparaître les premières formes de données géosociales. D'un point de vue historique, une course à la géolocalisation sociale s'est disputée entre 2009 et 2013. Peu médiatisée, celle-ci a principalement concerné Facebook, Google et les réseaux sociaux basés sur la localisation ou LBSNs (*Location-based Social Networks*).

Les LBSNs proposent des contenus aux utilisateurs en fonction de leur localisation. Le point d'entrée des LBSNs est le *lieu* dont les attributs sont enrichis par des notes, des commentaires et d'autres contenus de même nature. Les deux premiers LBSNs, Brightkite et Gowalla, sont sortis en 2007 et ont été respectivement rachetés par Limbo (2009) et Facebook (2011). Google et Facebook ont cherché à détrôner Foursquare, le LBSN en vogue de l'époque, en proposant une déclinaison géolocalisée à part entière de leur service. Google Latitude est lancé en 2009 et Facebook emboîte le pas avec Places en 2010. Le développement de Places prendra fin en 2011. Latitude fermera ses portes en 2013. Facebook conservera néanmoins la fonctionnalité de « *check-in*² ». En 2014, Foursquare scinde son réseau social : Swarm gère désormais le volet « géolocalisation sociale » (check-in, attribution de titres de maire, classement hebdomadaire, etc.) tandis que la plateforme d'origine se concentre sur la notation des lieux (essentiellement les commerces) à l'image de Yelp.

En outre, la majorité des travaux reposant sur des données géosociales privilégie trois « fournisseurs » : Twitter, Foursquare et Facebook. Le réseau de microblogging domine incontestablement le paysage de la recherche académique (Steiger et al., 2015). L'API de Twitter, tout comme celle de Facebook, est sortie il y a 10 ans déjà. Celle-ci est particulièrement appréciée pour sa stabilité. Foursquare a quant à lui lancé son API l'année de sa création (2009). Il serait inutile de présenter dans le détail chacune des spécificités et des restrictions de ces APIs. Le néophyte doit juste garder à l'esprit qu'une API est une bibliothèque de fonctions déclinée en plusieurs langages de programmation. Celle-ci permet aux développeurs (et aux chercheurs)

² En publiant un « check-in » depuis un lieu *P*, l'utilisateur déclare à son réseau de contacts qu'il se trouve à *P* à cet instant.

d'accéder aux bases de données des réseaux sociaux *via* seulement quelques lignes de code. Techniquement, les données géosociales sont collectées grâce à deux types de requêtes spatiales. La première revient à spécifier les coordonnées géographiques d'un point de référence à partir duquel sera établi un rayon de recherche. L'alternative consiste à définir un cadre liminaire communément appelé « *bounding box* ».

2. UNE PRISE DE POSITION NECESSAIRE

La nature même des Big data laisse à penser que toute la complexité de la spatialité humaine peut être appréhendée à travers l'information géolocalisée émanant des réseaux sociaux. Les données géosociales, tout comme les Big data, sont synonymes d'objectivité pour le chercheur qui les exploite : absence de lien avec les sujets étudiés, aucun biais associé aux choix des indicateurs ni aucune montée en généralités, juste une observation silencieuse de « faits » qui se déroulent en temps réel sous ses yeux. Face à ces arguments convaincants, le géographe ne peut qu'être tenté de faire des réseaux sociaux et de leurs APIs un instrument de collecte de données privilégié pour sa recherche. J'ai d'ailleurs moi-même été séduit par ce mouvement « pro-data » (cf. Quesnot et Roche, 2015abc). Et c'est la raison pour laquelle je me permets d'affirmer aujourd'hui, après avoir manipulé « en long et en large » ces données géosociales, que la dimension *géographique* de l'information émanant des réseaux sociaux est en réalité très pauvre ; sinon inexistante.

Encore faut-il s'entendre sur la définition du qualificatif « géographique » ...

Dans une acception très restreinte, l'espace est considéré comme un objet qui sert de *support* aux lieux et aux différents rapports sociaux. Une vision plus englobante envisage l'espace non pas comme une simple étendue, mais comme un véritable *contenu de la relation sociale* ; une ressource avec laquelle nous interagissons au quotidien grâce à nos compétences spatiales (Lussault, 2009). La spatialité humaine est ici considérée comme l'ensemble des pratiques spatiales – aussi bien matérielles qu'idéelles – qui concourent à la production de l'espace (social). La première définition se rapproche plus d'une géographie quantitative, celle de l'analyse spatiale, de la géomatique et de la GIScience. La seconde s'aligne plutôt avec les théories et concepts de la géographie humaine et sociale, voire humaniste. Dans le premier cas de figure, les géographes ont pour la plupart parfaitement conscience de la « réduction spatiale » que je viens d'évoquer. À titre d'exemple et si l'on évacue toute considération éthique, travailler sur des mobilités à partir de données géolocalisées transmises par des smartphones ne pose pas vraiment de problème d'un point de vue conceptuel. En revanche, se persuader que la subtilité de toutes nos pratiques spatiales est « encapsulée » à l'intérieur même des données géosociales est un point de vue réfutable ; une *croyance* qui, selon moi, nécessite d'être évacuée. Face au charisme des Big data, il est tout à fait compréhensible que *certain*s géographes soient tentés d'emprunter ce raccourci. Mais aussi séduisant soit-il, le caractère « géographique » et « social » des données « géo » – « sociales » n'est ni plus

ni moins qu'un vaste écran de fumée qui masque une (vulgaire) paire de coordonnées géographiques.

Au premier abord, cette prise de position semble contribuer à l'accentuation du clivage traditionnel opposant les tenants de la géographie quantitative d'un côté, et les adeptes de la géographie qualitative de l'autre. S'il est vrai que les propos tenus dans cet article ne favorisent en rien la convergence de ces deux visions, ils relèvent néanmoins d'une dynamique visant à maintenir un certain équilibre clairement fragilisé par le poids des Big data. Car comme le soulignent à juste titre Graham et Shelton (2013), leur entrée en scène n'impose pas qu'une adaptation et une modification des techniques d'analyse quantitative conventionnelles. Selon eux, tout porte à croire que les données massives étoufferont à long terme la recherche *qualitative* en géographie (p. 257) :

Although big data presents challenges to conventional notions and practices of 'hard' science, including to the field of geographic information science (Goodchild, 2013; Gorman, 2013), our worry is that the prevalence of the big data meme might lead to a new scientific, positivist, and quantitative turn in the social sciences by reducing the space for critical, qualitative, and postpositivist research.

Les sciences sociales se durcissent ; c'est un fait indéniable. La « science sociale computationnelle » (cf. Lazer et al., 2009) a le vent en poupe. Et dans sa quête permanente d'objectivité, la géographie n'échappe pas à la règle. Déjà « computationnelle » depuis les années 1990 avec l'essor de la GIScience, la géographie *quantitative* subit à l'heure actuelle un véritable processus de « datafication ». Ce mouvement est particulièrement virulent en Amérique du Nord. Le meeting annuel de l'association des géographes américains (AAG) de 2015 a d'ailleurs consacré une session spéciale aux recherches émergentes en « *data-driven geography* ». Exit les théories de la géographie, place aux données géographiques (cf. Miller et Goodchild, 2015). Alors que la GIScience vise en toute humilité à décrire du mieux que possible le Monde qui nous entoure, la *data-driven geography* se donne pour objectif d'expliquer son fonctionnement à travers l'analyse de données géographiques (prétendues) exhaustives.

Dans un papier récemment publié, Kwan (2016) défend une position intéressante que je partage entièrement. Selon elle, la géographie sous l'ère des Big data n'est pas plus axée sur les données qu'auparavant. Parler de « *data-driven geography* » n'est donc pas pertinent. En vérité, la nouvelle géographie qui se dessine actuellement est algorithmique, car outre le traitement des « données géographiques » recueillies, leur capture, leur stockage ainsi que leur récupération par le biais d'APIs requiert systématiquement l'usage d'algorithmes. Les algorithmes n'étant ni plus ni moins que des intentions déguisées (Sadin, 2015), le chercheur qui dresse ses conclusions à partir de ces données se doit de garder à l'esprit cette « filiation algorithmique ». Ainsi, les données géosociales ne sont pas le fait d'un individu totalement désintéressé, mais bien celui d'une société privée dont la motivation première est d'enranger des capitaux.

En définitive et pour schématiser, les données géosociales s'organisent en trois couches que je propose de décortiquer en partant de la plus épaisse (l'écran de fumée) à la plus subtile (l'intention déguisée en algorithmes) (cf. figure 1). Il s'agit là d'un mouvement involutif à double titre. « Involution » car nous passons effectivement du *multiple* (les données géosociales fabriquées en masse) à l'*un* (une société privée = une intention majeure). Involution « géographique » car c'est bien ce qui guette la géographie de demain si le phénomène Big data arrive à prendre le pas sur la géographie qualitative. Réduire l'espace à une étendue, c'est *de facto* assimiler le concept de lieu à une position voire tout au plus une localité ; soit « une association entre un phénomène donné et un point de repère sur une étendue vide [...] qui possède un nom » (Lévy, 2003b ; p. 560-561). Confondre la spatialité humaine et les données géosociales, c'est affirmer que toute la richesse de notre vécu spatial est *quantifiable* ; ce dont je doute fort. Avant de présenter les arguments qui soutiennent ma position, j'aimerais tout naturellement exposer les principaux avantages mis en avant par les chercheurs qui fondent leurs travaux sur les « émanations géographiques » issues des réseaux sociaux.

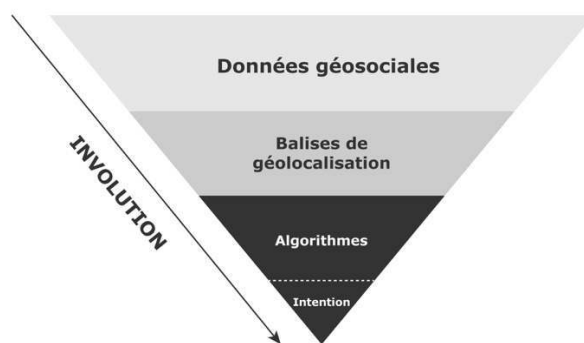


Figure 1 : Décortiquer les données géosociales.

3. LES RESEAUX SOCIAUX AU SERVICE DE LA GEOGRAPHIE (?)

Trois arguments sont systématiquement avancés lorsqu'il est question de données géosociales : elles sont (1) *exhaustives* car disponibles en surabondance, (2) *produites* par les citoyens eux-mêmes et de ce fait, (3) elles permettent d'appréhender la spatialité humaine, le sens des lieux urbains et donc l'*espace social* lui-même.

3.1. Un déluge de données « géographiques »

Parmi les 3V du Big data, le *V*olume est celui qui prend sans nul doute le plus de place sur le devant de la scène géosociale. Le nombre global d'utilisateurs actifs de médias sociaux est imposant en soi : 2,307 milliards pour l'année 2016 selon le dernier rapport de l'agence We Are Social³. À l'heure actuelle, Facebook ne compte pas moins

³ Consulté pour la dernière fois le 24 mars 2016 à l'adresse suivante : <http://wearesocial.com/special-reports/digital-in-2016>

d'un milliard d'utilisateurs actifs tandis que Twitter en fédère 320 millions. De son côté, le LBSN Foursquare/Swarm dépasse tout juste la barre des 50 millions d'utilisateurs. De toute évidence, ce vivier d'utilisateurs génère quotidiennement un volume important de données. Le chiffre de 10 millions est souvent annoncé en ce qui concerne les tweets géolocalisés. Celui-ci descend à 7 millions pour les check-ins publiés sur la plateforme Swarm qui, au passage, est connue pour avoir la plus grande base de données de lieux du web social (65 millions de lieux recensés au total).

Les données géolocalisées de Twitter récupérées via l'API gratuite « *Streaming* » sont largement exploitées dans les productions scientifiques intégrant des données géosociales. Selon Steiger et ses collaborateurs (2015), la majorité de ces travaux porte sur l'étude d'événements se produisant sur la surface de la Terre (p. ex. les mobilités) ; les méthodes d'analyse spatio-temporelle et sémantique étant dans ce cas fortement privilégiées. Les données de Foursquare ne sont pas en reste pour autant, car la plateforme propose à ses utilisateurs d'étendre la visibilité de leurs check-ins sur leurs comptes Facebook et Twitter. C'est la raison pour laquelle une partie des tweets géolocalisés provient directement de Foursquare. Toutes ces passerelles permettent aux utilisateurs de réseaux sociaux de divulguer leur position à un réseau de contacts toujours plus étendu.

Le nombre de recherches faisant usage de *millions* de check-ins, de tweets géotagués et de mentions « J'aime » géolocalisées est impressionnant. Dresser une liste exhaustive de tous ces travaux serait tout bonnement impossible. Les données géolocalisées issues des réseaux sociaux peuvent donc être considérées comme des Big data à part entière. Récolter un volume de données équivalent en aussi peu de temps et sur autant de personnes est difficilement concevable avec une approche scientifique conventionnelle. C'est en particulier cet aspect qui est mis en avant par les partisans des données géosociales et plus généralement des Big data. Mais encore, leur surabondance pousse certains à affirmer qu'elles sont synonymes d'exhaustivité.

À titre d'exemple, la recherche de Noulas et ses collaborateurs (2011) figure parmi les premières à avoir exploité des masses colossales de données « géographiques » provenant des réseaux sociaux. L'API de Foursquare ayant été très restrictive à ses débuts, ces derniers sont passés par l'API *Streaming* de Twitter pour récolter un total de 12 millions de check-ins. Le recueil s'est fait sur un peu plus de trois mois seulement et concerne 679 000 utilisateurs. En analysant la répartition des check-ins dans le temps, Noulas et al. ont pu identifier trois pics d'activités en semaine : (1) le matin, lorsque les individus se rendent sur leur lieu de travail ; (2) à l'heure du déjeuner ; et enfin (3) entre 18h et 20h, lorsque ces derniers quittent leur lieu de travail. Cette répartition change en fin de semaine : elle ne présente aucun pic d'activité, mais plutôt une courbe décrivant une évolution constante tout au long de la journée pour ensuite chuter progressivement à partir de 22h. Noulas et ses collègues concluent ainsi que les données de Foursquare sont parfaitement représentatives des déplacements quotidiens des utilisateurs. Jurdak et al. (2015) sont récemment arrivés à la même conclusion en analysant plus de six millions de messages provenant exclusivement de la Twittosphère.

3.2. Appréhender la spatialité humaine – et donc par extension l'espace social – à partir des données géosociales

Dans un article provocateur paru dans le magazine *Wired* (« *The end of theory: The data deluge makes the scientific method obsolete* »), Anderson (2008) écrivait :

This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves. The big target here isn't advertising, though. It's science. The scientific method is built around testable hypotheses. [...] But faced with massive data, this approach to science — hypothesize, model, test — is becoming obsolete. [...] 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.

Les propos tenus par Anderson ont été tantôt critiqués (p. ex. Boyd et Crawford, 2012), tantôt soutenus par une partie de la communauté scientifique (p. ex. Prensky, 2009). Pour Hey et ses collaborateurs (2009), le phénomène Big data constitue le quatrième paradigme de la science qui succède à l'ère computationnelle. En évacuant les hypothèses formulées *ex ante*, le mouvement Big data accorde une importance capitale à la « sérendipité » ou l'art de faire des découvertes fortuites. Mais au-delà de la sérendipité, nous avons affaire ici à un courant de pensée que je qualifie de « data-enactif » par analogie avec la théorie de l'*énaction* proposée par Varela et ses collaborateurs (1993). Selon ce principe, la connaissance émerge du couplage entre les données et l'*Homo Numericus*. Spécifiquement, l'Homme connecté produit (volontairement ou non) des données sur lui-même. Celles-ci sont ensuite collectées puis analysées afin de mieux saisir toute la profondeur de son expérience en tant qu'*être* humain. Ce cercle n'étant réellement vertueux que si l'on croise des données couvrant différentes thématiques.

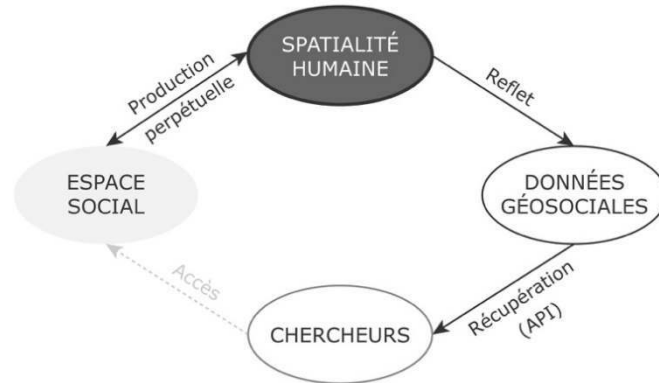


Figure 2 : Les données géosociales : une fenêtre sur l'espace social.

Les fondements du principe « data-enactif » sont bien évidemment applicables aux données géosociales (cf. figure 2). Dans ce cas de figure, les données géosociales reflètent la spatialité humaine et leur analyse permet *de facto* d'accéder à l'espace social lui-même ; celui que nous perpétons par nos pratiques spatiales quotidiennes, nos usages, ainsi que nos représentations (Cailly, 2003). Stefanidis et ses collaborateurs (2013) emploient l'expression « Information Géographique Ambiante » (AGI) pour caractériser ce brouillard persistant formé par les données géosociales. Les AGI sont pour eux une forme dérivée des VGI : elles ne sont pas créées de manière ponctuelle, mais sont présentes sous forme « éthérique » car renouvelées en permanence par les utilisateurs du web social. Toujours selon eux, l'intérêt de l'AGI issue des réseaux sociaux est double : elle est localisable et porte en elle la dimension sociale de l'espace géographique. Dans cet esprit, l'activité que les utilisateurs génèrent en publiant des commentaires, des check-ins, des contenus visuels (principalement des photos et des vidéos), des mentions « J'aime » ou encore des notations, correspond à une traduction numérique des pratiques (spatiales) qui fondent le « système social humain ». En d'autres termes, cartographier les AGI reviendrait *de facto* à visualiser l'espace social lui-même (p. 333-334 ; 336) :

We no longer map just buildings and infrastructure, but we can now map abstract concepts like the flow of information in a society, contextual information to place and linking both quantitative and qualitative analysis in human geography. In a sense one could consider AGI to be addressing the fact that the human social system is a constantly evolving complex organism where people's roles and activities are adapting to changing conditions, and affect events in space and time. By moving beyond simple mashups of social media feeds to actual analysis of their content we gain valuable insight into this complex system. [...] As the popularity of social media is growing exponentially we are presented with unique opportunities to identify and understand information dissemination mechanisms and patterns of activity in both the geographical and social dimensions [...]

Par conséquent, ces derniers ont proposé une architecture informatique nommée « *Social media ingestor* » dont la fonction principale est d'« ingérer » en temps réel les AGI provenant des différentes plateformes composant le web social (p. ex. Twitter, Facebook, Flickr ou encore YouTube). Une fois la « digestion » accomplie, les données en question se retrouvent stockées sur une seule et même base de données ; prêtes à l'emploi pour de futures analyses spatiales et productions cartographiques.

3.3. Extraire le sens des lieux urbains

Les adeptes de la géographie humaine et sociale envisagent le lieu non pas comme un point abstrait référencé dans un système de coordonnées, mais comme une « condition de l'expérience humaine » (Entrikin, 2003). D'après Lussault (2003a), un véritable lieu « n'existe pleinement qu'en tant qu'il possède une portée sociale, en termes de pratiques comme de représentations, qu'il s'inscrit comme un objet identifiable, et éventuellement identificatoire, dans un fonctionnement collectif, qu'il est chargé de valeurs communes dans lesquelles peuvent potentiellement – donc pas systématiquement – se reconnaître les individus. » (p. 562).

Ainsi, un lieu doit être *identifiable* pour exister. Il doit posséder une identité spatiale qui lui est propre pour se différencier des autres lieux. Cette première acceptation de l'identité spatiale s'aligne avec le concept d'« *identity of place* » de Relph (1976). Dans ce contexte, deux lieux qui se distinguent (essentiellement par la distance) peuvent partager des attributs à la fois matériels et substantiels (p. ex. un style architectural et une langue). Une deuxième définition de l'identité spatiale renvoie à l'identification d'un opérateur à un espace (Lussault, 2003b). Cette seconde acceptation fait référence à l'« *identity with place* » de Relph (1976). Le processus d'identification qui entre en jeu ici est étroitement lié à la notion de valeur spatiale : un individu choisit de s'identifier à un lieu qui a une certaine valeur à ses yeux ; cette dernière pouvant s'accorder ou non avec les discours, les images et les représentations collectives qui lui sont associés.

En outre, comme l'explique Stock (2006), la dimension spatiale de l'identité relève dans ce cas de figure de deux mécaniques fondamentales. D'une part, l'identité renvoie à « un lien arbitraire entre lieu et individus » (p.4). D'autre part, elle « n'est pas seulement issue des représentations, mais aussi issue de la pratique des lieux ». Le terme « pratique » doit être ici compris comme une action *contextualisée* qui permet à un opérateur d'entrer en relation avec un espace (Ruby, 2003). Ainsi, tout comme l'espace géographique reste indissociable de sa composante sociale, les représentations relatives à un lieu ne peuvent être pleinement déchiffrées sans la prise en compte des pratiques spatiales auxquelles elles sont associées. Partant de cette conception du lieu, certains géographes et chercheurs en GIScience prêtent aux données géosociales un potentiel sans précédent : celles-ci seraient, pour ainsi dire, la manifestation numérisée de la spatialité humaine. De ce fait, leur analyse permettrait de quantifier ce qui demeure qualitatif par essence : le(s) sens que nous accordons aux lieux par nos pratiques spatiales quotidiennes.

Dans un article publié cette année, Jenkins et ses collaborateurs (2016) soutiennent que la signification (*meaning*) des lieux est extractible à partir de l'analyse de tweets géotaggués. Ils expliquent très clairement qu'avant l'apparition des données géosociales, le concept de lieu n'était qu'un vague construit théorique difficilement appréhendable de manière concrète et objective (p. 1) :

Place can be generally defined as a location that has been assigned meaning through human experience, and as such it is of multidisciplinary scientific interest. Up to this point place has been studied primarily within the context of social sciences as a theoretical construct. The availability of large amounts of user-generated content, e.g. in the form of social media feeds or Wikipedia contributions, allows us for the first time to computationally analyze and quantify the shared meaning of place. By aggregating references to human activities within urban spaces we can observe the emergence of unique themes that characterize different locations, thus identifying places through their discernible sociocultural signatures.

La position défendue par Jenkins et al. (2016) est loin d'être récente. On a pu observer une réappropriation du concept de lieu depuis quelques années déjà. En 2009, le journal interdisciplinaire « *Spatial Cognition and Computation* » sortait un numéro spécial consacré à la modélisation computationnelle des lieux. Le titre de l'article d'introduction rédigé par Winter et al. (2009) est d'ailleurs très explicite : « *Does Place have a place in geographic information science?* ». Avant de prêter attention à la composante spatiale, beaucoup de travaux intégrant des données issues des réseaux sociaux ont porté sur l'analyse des sentiments des internautes ; notamment à partir de Twitter. La fonctionnalité de check-in ne s'étant véritablement démocratisée que depuis la sortie des LBSNs, il aura fallu attendre les années 2010 pour voir apparaître les premiers travaux mettant en avant les potentialités des données géosociales. À travers le projet *Livehoods*, Cranshaw et ses collaborateurs (2012) prétendent par exemple accéder aux « cartes mentales collectives ». Ils soutiennent que les 18 millions de tweets et de check-ins recueillis dans le cadre de ce projet sont une manifestation des pratiques spatiales à l'échelle urbaine. Leur algorithme détermine des aires géographiques dont les limites s'accordent, selon eux, avec la structure interne des personnes qui vivent ces espaces urbains au quotidien. Dans le même registre, Kelley (2013) affirme que les données géographiques provenant des réseaux sociaux nous donnent un accès privilégié aux représentations et à l'imaginaire collectif urbain. Enfin, Roche et moi-même avons proposé d'exploiter l'information géolocalisée partagée sur Facebook et Foursquare pour identifier des points de repère populaires en zone urbaine (Quesnot et Roche, 2015a). Nous avons justifié ce choix en avançant le même argument que nos pairs ; à savoir que les données géosociales reflètent la dimension sociale des lieux urbains.

4. LA FACE CACHEE DES DONNEES GEOSOCIALES

Le potentiel des données géosociales se résume donc à trois arguments spécifiques. Premièrement, leur abondance sous-entend une certaine exhaustivité qui implique que le « réel numérisé » se confonde avec la « réalité » elle-même (cf. Rouvroy, 2014). Deuxièmement, le paradigme des Big data part du principe que l'*Homo Numericus* produit des données brutes sur lui-même auxquelles nous – chercheurs – avons directement accès. Dans cet esprit, les données géosociales sont censées refléter l'essence même de nos pratiques spatiales, de nos représentations ainsi que des stratégies que nous déployons quotidiennement pour entrer en contact avec les lieux urbains. Le troisième et dernier argument découle de ce principe : puisque les données géosociales sont le reflet de nos spatialités quotidiennes, elles nous permettent incontestablement d'approcher *par les nombres* l'expérience spatiale humaine et en particulier le *sens* des lieux urbains. Une fois condensés, ces trois arguments forment la première couche constitutive des données géosociales. La présente section se donne pour objectif de la décortiquer, mais également de se projeter au-delà de ses frontières.

4.1. Des données a-contextuelles

Avant de se rendre sur son lieu de travail, Clovis – un fervent adepte des « technologies intelligentes » – s'arrête au Starbucks Coffee du coin pour apprécier sa boisson favorite. Il y publie un check-in sur Swarm avant de prendre son bus. Durant le trajet, il poste un tweet *automatiquement* géotagué à propos de la victoire des Golden State Warriors de la veille sur les Houston Rockets. Clovis est un fan de NBA. Arrivé sur son lieu de travail, il indique à son réseau de contacts Facebook qu'une dure journée de labeur l'attend *via* un check-in commenté. À la pause de midi, Clovis et son collègue de travail, Charles, ont pour habitude de se rejoindre au bistro B ; une enseigne très peu connue, mais qui présente l'intérêt d'être située non loin de leur lieu de travail. Une fois attablés, les deux amis s'empressent de publier chacun de leur côté un check-in sur Swarm. La compétition est féroce : Clovis et Charles se disputent le titre de « Maire » du bistro B. Ce statut permet notamment d'obtenir un rabais de 5% sur chaque addition. Pendant sa dernière pause de la journée, Clovis, téméraire, se rend une nouvelle fois au bistro B pour prendre un café. C'est l'occasion pour lui – ou plutôt l'excuse – de publier un nouveau check-in et d'augmenter ainsi son rang dans le classement hebdomadaire des prétendants au titre de maire. La fin justifiant les moyens, Charles adopte une stratégie radicalement différente : il a recours à un programme spécial – un *bot* – qui publie *à sa place* des check-ins à intervalles réguliers. À eux seuls, Clovis et Charles génèrent un nombre impressionnant de check-ins qui dépasse largement celui du restaurant R situé à proximité du bistro B. R est très populaire, mais sa clientèle ne compte en revanche qu'une minorité de « Geeks ». 18h, l'heure est venue pour Clovis de faire un tour chez son dentiste afin de soigner cette carie qu'il traîne depuis quelques semaines déjà. Soucieux de l'image qu'il renvoie sur les réseaux sociaux, Clovis se garde bien de publier un check-in avant de rejoindre le groupe d'amis avec lequel il passera le reste de la soirée. Photographe dans l'âme, Clovis a pris plusieurs clichés depuis la place de la Concorde ; dont une photographie *automatiquement* géolocalisée de la Tour Eiffel qu'il a postée sur Instagram. En bon utilisateur de Swarm, Clovis publie un ultime

check-in avant de se coucher. Pour sa propre sécurité, le LBSN brouille automatiquement la localisation des check-ins associés à des lieux privés comme la résidence de Clovis.

Bien que totalement fictif, le résumé qui relate le quotidien du personnage Clovis est applicable à n'importe quel acteur du web social. Il s'agit là d'un récit chargé de significations. Au même titre que les actions qu'il entreprend physiquement (p. ex. se rendre au bistro B ou prendre des photographies), toute l'activité que génère Clovis sur les réseaux sociaux (et en particulier la publication de contenus géolocalisés en temps réel) peut être assimilée à une véritable succession de pratiques spatiales ; c'est-à-dire des actions *contextualisées* qui lui permettent, par l'intermédiaire de différentes technologies, d'entrer en relation avec un espace donné (cf. Ruby, 2003). Cela est à mon sens indéniable. En revanche, mettre un trait d'égalité entre ces activités qui sont réalisées dans un contexte particulier et les données géosociales récupérées à partir d'APIs est un raccourci *fallacieux* générateur d'amalgames.

En effet, les activités produites par nos deux compères n'ont véritablement de sens que si l'on prend en compte les contextes dans lesquels elles s'inscrivent. Sans un minimum d'information contextuelle, on serait par exemple tenté de croire que le bistro B est plus populaire que le restaurant R alors qu'il s'agit en réalité de l'inverse. Si l'on se fie uniquement aux coordonnées géographiques, on pourrait penser que la photographie prise par Clovis (celle de la Tour Eiffel) est exclusivement associée à la place de la Concorde. Or, puisque le quantitatif prime en tout point sur le qualitatif dans la sphère des Big data, les données géosociales reposent sur deux principes fondamentaux que sont l'*anonymisation* et la *décontextualisation*. Les données massives étant toutes dépourvues de contexte, personne n'est réellement capable d'expliquer dans quelle(s) condition(s) et surtout pour quelle(s) raison(s) un check-in, un commentaire géolocalisé ou bien un tweet géotagué a été publié. Devant le flot ininterrompu de données, cela relève d'une gageure. Et c'est bien la raison pour laquelle Anderson (2008) affirme que « les nombres parlent d'eux-mêmes ».

L'enquête qualitative de Lindqvist et ses collaborateurs (2011) a pourtant montré que tous les check-ins publiés sur Foursquare (désormais Swarm) ne se valaient pas. Ils ont précisément identifié cinq motivations relatives à la publication de check-ins. La première correspond à la recommandation de nouveaux lieux par le LBSN : en publiant fréquemment des check-ins dans une catégorie particulière de lieu (p. ex. un bar), l'application propose en retour une liste appropriée de lieux à visiter (c.-à-d. d'autres bars). La deuxième motivation concerne l'aspect ludique du réseau social : les utilisateurs sont en compétition entre eux et sont constamment incités à publier des check-ins pour obtenir des badges et des titres de maires. La rencontre de personnes dont le profil est en adéquation avec le sien (en termes de check-ins) fait également partie des motivations identifiées. Certains interviewés ont déclaré publier des check-ins dans des endroits spécifiques afin de savoir qu'elles étaient les personnes qui partageaient les mêmes intérêts et habitudes de fréquentation. Conserver une trace personnelle des lieux visités est la quatrième motivation recensée par Lindqvist et al.

(2011). La dernière motivation est purement égocentrique : il apparaît qu'une bonne partie des interviewés publie des check-ins dans des lieux présentant une forte valeur spatiale collective afin d'améliorer leur propre image.

C'est notamment en s'intéressant à cette cinquième motivation que Rost et ses collaborateurs (2013) sont arrivés à la conclusion que l'acte de publier un check-in était en fait une forme de communication à part entière basée sur les lieux. Ils ont clairement démontré qu'un check-in n'équivalait pas forcément à la visite d'un lieu (au sens physique). Effectivement, une partie des utilisateurs de Swarm prétendent se trouver à un endroit *via* un check-in alors que les coordonnées géographiques du dispositif depuis lequel ce dernier a été émis indiquent une position située à des kilomètres du lieu mentionné. En réalité, l'objectif pour les utilisateurs de LBSNs n'est pas de contribuer à un quelconque enrichissement d'ordre géographique ; à l'inverse des néogéographes (cf. la plateforme cartographique OpenStreetMap par exemple). Il est de *communiquer*, tout simplement. Les autres plateformes du web social ne sont pas en reste pour autant. On sait par exemple qu'un volume difficilement quantifiable de tweets est généré automatiquement par des robots et non des humains (Boyd et Crawford, 2012).

En définitive, le fait que nous ne connaissions pas les conditions (incluant les motivations et le contexte) dans lesquelles les données géosociales ont été créées nous empêche tout bonnement d'accéder à cet espace social tant convoité. Étant a-contextuelles par essence, ces données ne nous permettent pas d'approcher le vécu spatial collectif ; et encore moins de quantifier le sens des lieux urbains. Elles permettent tout au plus de générer un indice moyen de *popularité* à l'échelle urbaine. Plutôt que de tenter en vain d'éclaircir les conditions dans lesquelles sont produites ces données (Fischer, 2012), je pense qu'il est d'abord essentiel de questionner la nature même de l'information géolocalisée partagée sur les réseaux sociaux : s'agit-il réellement de données géographiques à proprement parler ?

4.2. Des métadonnées géographiques avant tout

Le lieu est un construit théorique de la géographie (sociale) qui ne peut être exclusivement appréhendé de manière quantitative. Contrairement aux Big data, le qualitatif doit primer sur le quantitatif – sans pour autant l'exclure – lorsque l'on cherche à saisir toute la subtilité du concept de « lieu ». Spécifiquement, le lieu est une condition de l'expérience humaine (Entrikin, 2003) ; un espace au sein duquel la notion de distance n'est pas pertinente (Lévy, 2003b). Chaque lieu est doté d'un esprit et d'une personnalité qui ne peuvent être pleinement saisis qu'à travers la sensibilité de l'Homme (Tuan, 1977). Le lieu est donc un concept dont la précision des contours qui le définissent est laissée à la seule appréciation de la subjectivité de chacun. Pour espérer approcher le sens des lieux urbains, il nous est nécessaire d'explorer la sensibilité humaine. Faire fi de cette condition nous éloigne inévitablement de la (des) véritable(s) signification(s) du lieu. Or, comme je l'ai souligné dans la précédente section, le propre même des Big data – et donc des données géosociales – est d'annihiler le caractère sensible de l'Homme en le substituant par des nombres. Ce que les sociétés comme Google cherchent à « étudier » à travers les données massives, ce sont des comportements *répétitifs* – que l'on

peut donc *anticiper* – et non une somme d’expériences vécues au sens *phénoménologique* du terme. Aussi, je ne peux que m’inscrire en faux contre la position qui soutient que les données géosociales constituent le moyen tant attendu par lequel les géographes pourront enfin quantifier le sens des lieux urbains (cf. Jenkins et al., 2016). À mon sens, cette position relève d’un fantasme. Elle traduit peut-être un complexe d’infériorité qu’ont certains chercheurs en sciences humaines et sociales vis-à-vis des sciences dures et de l’objectivité qu’elles revendiquent. Ces mêmes individus y voient sans doute une occasion en or pour objectiver et circonscrire par le biais d’un arsenal quantitatif le concept de « lieu » ; pourtant furtif par essence.

En outre, j’ai tenté de montrer dans la section précédente que la dimension sociale des données géosociales n’était qu’un vaste écran de fumée qui, une fois dissipé, révèle un artefact considéré par plus d’un comme « géographique ». Cependant, l’adjectif « géographique » est selon moi beaucoup trop riche pour qualifier l’information *géolocalisée* qui émane des réseaux sociaux. Car la dimension géographique des données géosociales se résume en fait à une simple information de *localisation* définie par une paire de latitude et de longitude (ou de X et de Y le cas échéant). Ces coordonnées sont soit déterminées à partir du dispositif depuis lequel le contenu est publié, soit préenregistrées dans une base de données (cf. le système de check-in⁴). Ainsi, la « donnée géosociale » est comparable à une équation qui inclut trois variables au total : « où », « combien », et « quoi » ; la variable « qui » étant absente puisque les données subissent toutes un processus d’anonymisation. En ce sens, la donnée géosociale est interprétable non pas en termes de lieu mais de *localité* ; soit une association entre un phénomène donné et un point de repère nommé sur une étendue vide (Lévy, 2003b). Une « localité » ne constitue pas un « lieu » mais tout au plus une « première entrée dans l’univers du lieu » pour paraphraser Lévy (2003b ; p. 561).

La question suivante mérite dès lors d’être soulevée : est-il pertinent d’employer l’expression « donnée géographique » alors que la dimension géographique de l’information provenant des réseaux sociaux est en réalité réduite à une simple balise de géolocalisation ? Le terme « métadonnée » traduirait beaucoup mieux la réalité ; à savoir que la localisation vient tout simplement compléter une donnée centrée sur un phénomène quantifié. Parler de « données » génère en fait une grande confusion. Les travaux de recherche cités en exemple dans la troisième section de cet article envisagent pour la plupart la donnée géosociale comme une donnée qui demeure géographique avant tout. La localisation est par conséquent mise au premier plan alors qu’elle est en fait secondaire. Deux contenus numériques distincts qui disposent de la même information de localisation sont *de facto* jugés équivalents (Quesnot et Roche, 2015c). Ainsi, un check-in publié à partir de la Tour Eiffel se voit attribuer la même valeur qu’un tweet géolocalisé *automatiquement* à proximité du monument parisien alors que celui-ci

⁴ Lorsqu’un utilisateur publie un check-in, il fait appel à une base de données qui recense un ensemble de lieux. Celle-ci est enrichie continuellement par la communauté du réseau social. Si un lieu est absent de la base, les utilisateurs ont la possibilité de l’y ajouter en spécifiant *à minima* un nom et une information de localisation (*lat/long*).

évoque l'élection de Trump à la présidentielle américaine. Le *social media ingestor* développé par Stefanidis et ses collaborateurs (2013) illustre à merveille la primauté accordée à l'équivalence *positionnelle*. C'est bien la raison pour laquelle Crampton et ses collègues (2013) invitent les chercheurs travaillant sur le géoweb à dépasser le champ restreint de l'analyse et de la cartographie des données géotaguées.

4.3. Des données fabriquées de toute pièce par des sociétés privées

S'aventurer au-delà de la simple paire de coordonnées géographiques nous amène à la réalité algorithmique qui sous-tend la fabrication des « données » géosociales. Je ne fais pas uniquement allusion aux algorithmes dont le rôle principal est d'affecter une information de localisation à un contenu numérique créé par un internaute *lambda* (c.-à-d. l'acte de géotaguer). Je fais référence ici à l'ensemble des algorithmes qui sont déployés pour *fabriquer* de manière totalement opaque – et j'insiste sur le mot « fabriquer » – cette fameuse « donnée géosociale ». Car la géolocalisation de contenus n'est en fait qu'un des multiples leviers que les géants du numérique activent pour générer du profit. Cette réalité – c'est-à-dire l'intention commerciale – est trop souvent occultée par celles et ceux qui exploitent les données géosociales. En 2010, Proulx et Millerand qualifiaient déjà le contexte économique de l'époque de « capitalisme informationnel » (p. 17) :

L'émergence d'un capitalisme informationnel veut dire que l'information joue un rôle pivot dans la production de la valeur. [...] avec la décennie 2000 et l'apparition des plateformes collaboratives, les contributions des utilisateurs ordinaires ensuite agrégées en grand corpus de données et de métadonnées deviennent le matériau privilégié dans la construction de la valeur des entreprises géantes de l'Internet.

Dans cet esprit, Thatcher (2014) explique très clairement que faire des Big data la source principale de données de sa recherche, c'est accepter qu'elles soient fabriquées par des sociétés privées qui exercent un *contrôle total* sur leur accès. Dans le monde des Big data, « donnée » équivaut à « argent ». C'est en grande partie pourquoi ces sociétés appliquent des restrictions quant à la récupération et à l'usage de *leurs* données. À titre d'exemple, l'API de Foursquare n'autorise qu'un maximum de 50 réponses par requête. L'envoi de requêtes multiples est considéré comme du pillage. Cette pratique est interdite par les conditions d'utilisation de la plateforme. Du côté de Twitter, le quota alloué est fluctuant. L'API gratuite *Streaming* ne donne accès qu'à un faible pourcentage de l'ensemble des tweets produits. Ce chiffre obscur oscillerait vraisemblablement entre 1.5 et 3% ; la société Twitter restant toujours très vague à ce sujet. Il est toujours possible de négocier avec Gnip⁵ – moyennant finance cette fois-ci – l'accès au *decabose* de Twitter ; soit 10% du volume total de tweets produit en temps réel. Et parmi ce pourcentage non représentatif de la Twittosphère, seuls 1 à 2% des tweets sont géotagués.

⁵ Fondée en 2008, la société Gnip revend les données de certaines plateformes du web social ; dont Foursquare et Facebook. Elle a été rachetée par Twitter en avril 2014.

Dès lors, peut-on encore parler d'exhaustivité lorsque l'on fait référence aux Big data et en particulier aux données géosociales ? Par ailleurs, l'argument qui consiste à affirmer que ces données sont un gage d'objectivité est lui aussi contestable. Car sans compter le fait qu'elles soient fabriquées dans l'optique de générer des profits, le chercheur qui les exploite doit nécessairement sélectionner certains attributs qu'il veut privilégier et en exclure d'autres (Boyd et Crawford, 2012). Or, il s'agit là d'un processus purement subjectif.

En définitive, le chercheur qui exploite des données géosociales est donc à la merci d'organismes qui peuvent non seulement changer à tout moment les règles d'usage de leurs APIs, mais aussi modifier les algorithmes qui fabriquent et redistribuent ces données. À l'heure actuelle, aucun utilisateur de l'API de Foursquare ne sait par exemple sur quel(s) indicateur(s) se base la plateforme pour établir l'indice de popularité des lieux qu'elle a répertoriés. Il s'agit d'une opération totalement obscure. Qui plus est, les ingénieurs en informatique des plateformes du web social mettent régulièrement à jour tous ces algorithmes afin d'optimiser les performances de leur API. Comme l'ont très bien souligné Lazer et ses collaborateurs (2014), ces fréquentes mises à jour compromettent sérieusement la reproductibilité des données exploitées par le chercheur.

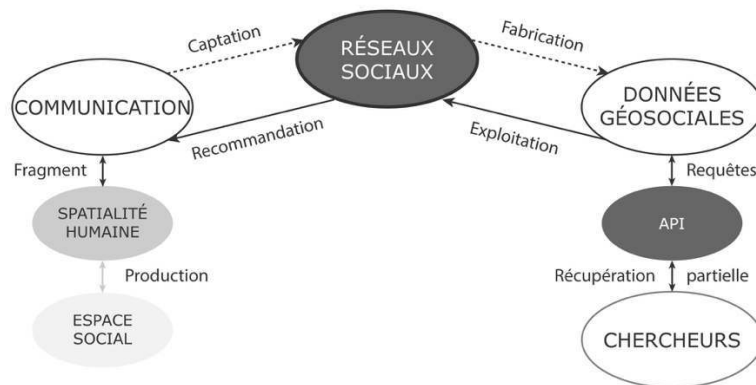


Figure 3 : Les données géosociales : un accès restreint aux pratiques spatiales.

Compte tenu de tous ces éléments, penser – ou plutôt croire – que les utilisateurs des plateformes du web social *produisent* des données (géographiques ou non) n'est ni plus ni moins qu'un *leurre*. Un contributeur d'OpenStreetMap produit de la donnée géographique alors qu'un utilisateur du web social *communique* à travers des commentaires, des check-ins, des mentions « J'aime » ou encore des émoticônes. Il ne raisonne pas en termes de latitude et de longitude. Les organismes propriétaires des réseaux sociaux et des autres plateformes du web social *subtilisent* toutes les informations dont elles ont besoin pour mieux *prédire* les comportements de leurs utilisateurs. Autrement dit, les données géosociales ne sont pas produites par les citoyens. Elles sont fabriquées de toute pièce par des organismes privés qui les redistribuent partiellement

via des APIs ou des sociétés partenaires comme Gnip. Les algorithmes exploités pour fabriquer et distribuer ces données sont totalement inaccessibles pour la personne qui les récupère. Au final, force est de constater qu'il existe une multitude de filtres entre les pratiques spatiales déployées sur la toile – soit l'objet d'étude principal des géographes – et les chercheurs eux-mêmes (cf. figure 3).

5. REMARQUES CONCLUSIVES

L'idée d'une donnée qui conjuguerait à elle seule les dimensions « géographiques » et « sociales » de l'expérience humaine est plus que séduisante pour un géographe ; en particulier pour celui qui ne réduit pas l'espace à un simple contenant des relations sociales. À l'image des Big data, les « données » « géo » – « sociales » sont souvent (re)présentées comme un puits sans fond alimenté en continu par les utilisateurs de réseaux sociaux. Cette caractéristique leur confère une certaine prétention ; et non des moindres : celle d'être exhaustives. Dans le *paradigme* des Big data, le réel numérisé se confond avec la « réalité ». En suivant ce principe, la « véritable » connaissance émerge nécessairement des Big data et de leur analyse. Par extension, les données géosociales permettent d'accéder à l'*espace social* lui-même ; celui que nous engendrons et perpétuons par nos représentations, nos usages et nos pratiques spatiales quotidiennes. J'ai tenté de démonter cette dernière assertion en mettant tout d'abord en avant le fait que ces « données » sont toutes – sans exception – dépourvues de *contexte*. Étant a-contextuelles par essence, elles ne peuvent contenir qu'un infime fragment de nos pratiques spatiales. Elles ne sont, pour ainsi dire, que l'expression de la spatialité humaine dans sa forme la plus simple. L'unique intérêt qu'elles apportent est de quantifier un phénomène géoréférencé. Aussi, dans la mesure où les réseaux sociaux constituent avant tout un moyen de *communiquer*, le phénomène géoréférencé en question n'est autre qu'une manière de s'exprimer qui se présente sous plusieurs formes (p. ex. des tweets ou des check-ins) et dans un certain *contexte*. En réalité, les « données » géosociales ne peuvent être considérées comme des données géographiques à part entière. Elles s'apparentent plus à des *métadonnées* puisque leur dimension « géographique » se résume à une balise de géolocalisation. Amplement suffisant pour des organismes qui cherchent à analyser spatialement les comportements humains (p. ex. les déplacements) afin de mieux les anticiper, ce géotag ne représente en vérité qu'une information très pauvre pour le chercheur qui souhaite approcher le vécu spatial collectif dans toute sa complexité. L'ultime argument que je défends réside dans la *production* de la donnée géosociale elle-même. Les utilisateurs de réseaux sociaux produisent-ils réellement de la donnée géosociale ? Cette question appelle selon moi une réponse définitivement négative. L'ère de l'anticipation et de la précognition ayant désormais succédé à celle de l'automatisation, l'information géolocalisée émanant des réseaux sociaux est exploitée pour générer du profit. La donnée géosociale est en réalité fabriquée de toute pièce par des sociétés privées (c.-à-d. les réseaux sociaux) afin de mieux prédire le comportement de leurs utilisateurs. L'objectif étant de leur fournir du contenu (p. ex. publicitaire) qui soit en accord avec leur profil (personnalisation algorithmique). En définitive, les chercheurs qui exploitent ces artefacts doivent *a*

minima passer par deux filtres : (1) la production de la donnée géosociale par les réseaux sociaux (comment est-elle exactement produite ?) et (2) sa diffusion (comment est-elle sélectionnée et renvoyée par l'API ?). L'argumentaire développé tout au long de cet article me pousse à affirmer que l'exploitation des données géosociales ne peut à elle seule nous permettre d'appréhender toute la profondeur de la spatialité humaine. Le recours à des données qualitatives (*small data*) est, me semble-t-il, primordial ; sinon inévitable.

BIBLIOGRAPHIE

- ANDERSON C. (2008), The end of theory: The data deluge makes the scientific method obsolete, *Wired Magazine*, vol. 16, n° 7. Repéré à <https://www.wired.com/2008/06/pb-theory/>
- BOYD D., CRAWFORD K. (2012), "Critical questions for big data", *Information, Communication & Society*, vol. 15, n° 5, pp. 662-679.
- CAILLY L. (2003), « Sociale », in : Lévy, J. & Lussault, M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 852-855.
- COUTANT A., STENGER T. (2009), « Les configurations sociotechniques sur le web et leurs usages : Le cas des réseaux sociaux numériques », in : *7ème Colloque du chapitre français de l'ISKO Intelligence collective et Organisation des connaissances*, Lyon, pp. 27-34.
- CRAMPTON J., GRAHAM M., POORTHUIS A., SHELTON T., STEPHENS M., WILSON M. W., ZOOK A. M. (2013), "Beyond the geotag: Situating 'big data' and leveraging the potential of the geoweb", *Cartography and Geographic Information Science*, vol. 40, n° 2, pp. 130-139.
- CRANSHAW J., SCHWARTZ R., HONG J. I., SADEH N. (2012), "The livelihoods project: Utilizing social media to understand the dynamics of a city", in: *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*, USA, AAAI Press, pp. 58-65.
- EGENHOFER M. J., CLARKE K. C., GAO S., QUESNOT T., FRANKLIN W. R., YUAN M., COLEMAN D. (2016), "Contributions of GIScience over the past twenty years", in: ONSRUD H. & KUHN W. (eds.), *Advancing Geographic Information Science: The Past and the Next Twenty Years*, USA, GSDI Association Press, pp. 9-34.
- ENTRIKIN J. N. (2003), « Lieu », in : LEVY J. & LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 557-560.
- FISCHER F. (2012), "Vgi as big data. A new but delicate geographic data-source", *Geoinformatics*, vol. 15, n° 3, pp. 46-47.
- GOODCHILD M. F. (1992), "Geographical information science", *International Journal of Geographical Information Systems*, vol. 6, n° 1, pp. 31-45.
- GOODCHILD M. F. (2007), "Citizens as sensors: The world of volunteered geography", *GeoJournal*, vol. 69, n° 4, pp. 211-221.
- GOODCHILD M. F. (2013), "The quality of big (geo)data", *Dialogues in Human Geography*, vol. 3, n° 3, pp. 280-284.

- GRAHAM M., SHELTON T. (2013), "Geography and the future of big data, big data and the future of geography", *Dialogues in Human Geography*, vol. 3, n° 3, pp. 255-261.
- HARVEY F. (2013), "To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information", in: SUI D. S., ELWOOD S. & GOODCHILD M. F. (eds.), *Crowdsourcing Geographic Knowledge*, Netherlands, Springer, pp. 31-42.
- HEY T., STEWART T., & K., T. (2009), *The fourth paradigm: Data-intensive scientific discovery*, USA, Microsoft Research, 284 p.
- JENKINS A., CROITORU A., CROOKS A. T., STEFANIDIS A. (2016), "Crowdsourcing a collective sense of place", *PLoS ONE*, vol. 11, n° 4, e0152932.
- JURDAK R., ZHAO K., LIU J., ABOUJAOUDE M., CAMERON M., NEWTH D. (2015), "Understanding human mobility from Twitter", *PLoS ONE*, vol. 10, n° 7, e0131469.
- KELLEY M. J. (2013), "The emergent urban imaginaries of geosocial media", *GeoJournal*, vol. 78, n° 1, pp. 181-203.
- KITCHIN R. (2013), "Big data and human geography Opportunities, challenges and risks", *Dialogues in Human Geography*, vol. 3, n° 3, pp. 262-267.
- KITCHIN R. (2014), "Big data, new epistemologies and paradigm shifts", *Big Data & Society*, vol. 1, n° 1, pp. 1-12.
- KWAN M.-P. (2016), "Algorithmic Geographies: Big Data, Algorithmic Uncertainty, and the Production of Geographic Knowledge", *Annals of the American Association of Geographers*, vol. 106, n° 2, pp. 274-282.
- LAZER D., PENTLAND A., ADAMIC L., BARABÁSI A.-L., BREWER D., CHRISTAKIS N., VAN ALSTYNE M. (2009), "Life in the network: the coming age of computational social science", *Science*, vol. 323, n° 5915, pp. 721-723.
- LAZER D., KENNEDY R., KING G., VESPIGNANI A. (2014), "The parable of Google Flu: Traps in big data analysis", *Science*, vol. 343, n° 6176, pp. 1203-1205.
- LEVY J. (2003a), « Géographie », in : LEVY J., LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 399-401.
- LEVY J. (2003b), « Lieu », in : LEVY J., LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 560-561.
- LINDQVIST J., CRANSHAW J., WIESE J., HONG J. I., ZIMMERMAN J. (2011), "I'm the mayor of my house: Examining why people use foursquare - a social-driven location sharing application", in: *Proceedings of the Conference on Human Factors in Computing Systems*, Vancouver, ACM, pp. 2409-2418.
- LUSSAULT M. (2003a), « Lieu », in : LEVY J., LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 561-563.
- LUSSAULT M. (2003b), « Identité spatiale », in : LEVY J., LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 480-481.
- LUSSAULT M. (2009), *De la lutte des classes à la lutte des places*, Paris : Grasset, 220 p.

- MILLER H. J., GOODCHILD M. F. (2015), "Data-driven geography", *GeoJournal*, vol. 80, n° 4, pp. 449-461.
- NOULAS A., SCELLATO S., MASCOLO C., PONTIL M. (2011), "An Empirical study of geographic user activity patterns in foursquare", in: ADAMIC L. A., BAEZA-YATES R. A., COUNTS S. (eds.), *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, USA, AAAI Press, pp. 570-573.
- PRENSKY M. (2009), "H. Sapiens Digital: From digital immigrants and digital natives to digital wisdom", *Innovate: Journal of Online Education*, vol. 5, n° 3, Article 1.
- PROULX S., MILLERAND F. (2010), « Le web social, au carrefour de multiples questionnements », in : MILLERAND F., PROULX S., RUEFF J. (eds.), *Web social : mutation de la communication*, Québec, Presses de l'Université du Québec, pp. 13-30.
- QUESNOT T., ROCHE S. (2015a), "Measure of landmark semantic salience through geosocial data streams", *ISPRS International Journal of Geo-Information*, vol. 4, n° 1, pp. 1-31.
- QUESNOT T., ROCHE S. (2015b), "Quantifying the significance of semantic landmarks in familiar and unfamiliar environments", in: FABRIKANT S. I., RAUBAL M., BERTOLOTO M., DAVIES C., FREUNDSCHUH S., BELL S. (eds.), *Spatial Information Theory*, LNCS, vol. 9368, Switzerland, Springer, pp. 468-489.
- QUESNOT T., ROCHE S. (2015c), "Platial or locational data? Toward the characterization of social location sharing", in: BUI, T. X., RALPH H., SPRAGUE J. (eds.), *Proceedings of the 48th Annual Hawaii International Conference on System Sciences*, Washington, IEEE Computer Society Press, pp. 1973-1982.
- RELPH E. (1976), *Place and placelessness*, London: Pion, 156 p.
- ROST M., BARKHUUS L., CRAMER H., BROWN B. (2013), "Representation and communication: Challenges in interpreting large social media datasets", in: *Proceedings of the 2013 conference on Computer Supported Cooperative Work*, San Antonio, ACM, pp. 357-362.
- ROUVROY A. (2014), *Des données sans personne : le fétichisme de la donnée à caractère personnel à l'épreuve de l'idéologie des Big Data*. Repéré à https://www.academia.edu/7754350/Des_donnees_sans_personne_le_fetichisme_de_la_donnee_a_caractere_personnel_a_lepreuve_de_lideologie_de_s_Big_Data.
- RUBY C. (2003), « Pratique », in : LEVY J., LUSSAULT M. (dir.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris : Belin, pp. 740.
- SADIN E. (2015), *La vie algorithmique. Critique de la raison numérique*, Paris : Éditions l'échappée, 278 p.
- STEFANIDIS A., CROOKS A., RADZIKOWSKI J. (2013), "Harvesting ambient geospatial information from social media feeds", *GeoJournal*, vol. 78, n° 2, pp. 319-338.
- STEIGER E., PORTO DE ALBUQUERQUE J., ZIPF A. (2015), "An advanced systematic literature review on spatiotemporal analyses of Twitter data", *Transactions in GIS*, vol. 19, n° 6, pp. 809-834.

- STOCK M. (2006), « Construire l'identité par la pratique des lieux », in : DE BIASE A., ALESSANDRO C. (dir.), *Chez nous. Territoires et Identités dans les mondes contemporains*, Paris : Éditions de la Villette, pp. 142-159.
- THATCHER J. (2014), "Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data", *International Journal of Communication*, vol. 8, 1765-1783.
- TUAN Y. (1977), *Space and place: The perspective of experience*, Minneapolis: University of Minnesota Press, 235 p.
- VARELA F., THOMPSON E., ROSCH E. (1993), *L'inscription corporelle de l'esprit : Sciences cognitives et expérience humaine*, Paris : Seuil, 377 p.
- WINTER S., KUHN W., KRÜGER A. (2009), "Does Place have a place in geographic information science?", *Spatial Cognition & Computation*, vol. 9, n° 3, pp. 171-173.