

Usages des traces numériques en géographie : potentiels heuristiques et enjeux de recherche

Boris Mericskay, Matthieu Noucher, Stéphane Roche

▶ **To cite this version:**

Boris Mericskay, Matthieu Noucher, Stéphane Roche. Usages des traces numériques en géographie : potentiels heuristiques et enjeux de recherche. L'Information géographique, 2018, 82 (2), pp.39-61. 10.3917/lig.822.0039 . halshs-01809615

HAL Id: halshs-01809615
https://shs.hal.science/halshs-01809615

Submitted on 6 Jun 2018

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.

[VERSION AUTEUR AVANT MISE EN PAGE DE L'EDITEUR]

Boris Mericskay, Matthieu Noucher, Stéphane Roche, « Usages des traces numériques en géographie : Potentiels heuristiques et enjeux de recherche », *L'Information géographique* (2/2018), pp. 39-61, Armand Colin. Disponible sur : http://www.revues.armand-colin.com/geographie/linformation-geographique/linformation-geographique-22018/usages-traces-numeriques-geographie-potentiels-heuristiques-enjeux-recherche

Usages des traces numériques en géographie : potentiels heuristiques et enjeux de recherche.

Boris Mericskay, maître de conférences, UMR ESO, Université Rennes 2
boris.mericskay@univ-rennes2.fr

Matthieu Noucher, chargé de recherche au CNRS, UMR Passages, Bordeaux
matthieu.noucher@cnrs.fr

Stéphane Roche, professeur titulaire, Centre de recherche en géomatique, Université Laval, Québec, stephane.roche@scg.ulaval.ca

Résumé. Le rapport entre trace et espace est abordé ici sous l'angle des traces numériques intégrant une dimension spatiale qui, avec le développement des usages d'Internet, tendent à se démultiplier. Après avoir défini et caractérisé les différents types de traces géonumériques pour montrer toute la diversité de leur registre de fabrique, nous proposons d'analyser leur potentiel inédit pour les sciences qui s'intéressent à la dimension spatiale du social, tout en insistant sur les nombreuses questions - méthodologique, épistémologique et éthique - soulevées par leur emploi.

Mots-clés. Traces géonumériques ; information géographique ; spatialité ; Internet ; géoweb

Abstract. In this paper, the relationship between trace and geographical space is discussed from the digital footprint perspective by integrating the spatial dimension. Indeed, within the growing use of the Internet the variety of digital footprint tend to widely multiply. We firstly define and characterize the different types of geodigital traces in order to emphasize on their various way of production. We secondly provide an analysis of their unique potential as food for thoughts in sciences which deal with the spatial dimensions of social dynamics and facts. Epistemological, methodological and ethical issues regarding the use of spatial digital footprint are finally considered.

Keywords. Geodigital footprint ; geographical information ; spatiality ; Internet ; geoweb

"Caminante, son tus huellas / el camino, y nada más;
caminante, no hay camino, / se hace camino al andar. [...]"[1]
(Antonio Machado, Canto XXIX Proverbios y cantares, Campos de Castilla, 1917)

Début 2018, l'application de fitness Strava était au coeur d'une polémique concernant ses données de localisation. Installée sur des smartphones ou des bracelets électroniques, cette application offre à ses utilisateurs la possibilité d'enregistrer des informations liées à leurs courses sportives (course à pieds, cyclisme, nage, ski...), telles que la vitesse moyenne sur un trajet, les calories brûlées ou encore la géographie du parcours. Les données centralisées de ces 27 millions d'utilisateurs constituent potentiellement une source inédite pour analyser les pratiques sportives. Ainsi, la carte de chaleur (*heatmap*), mise en ligne en septembre 2017, permet d'explorer la compilation de l'ensemble des données publiques soit 1 milliard d'activités enregistrées représentant 27 milliards de km de traces recueillies (figure 1).

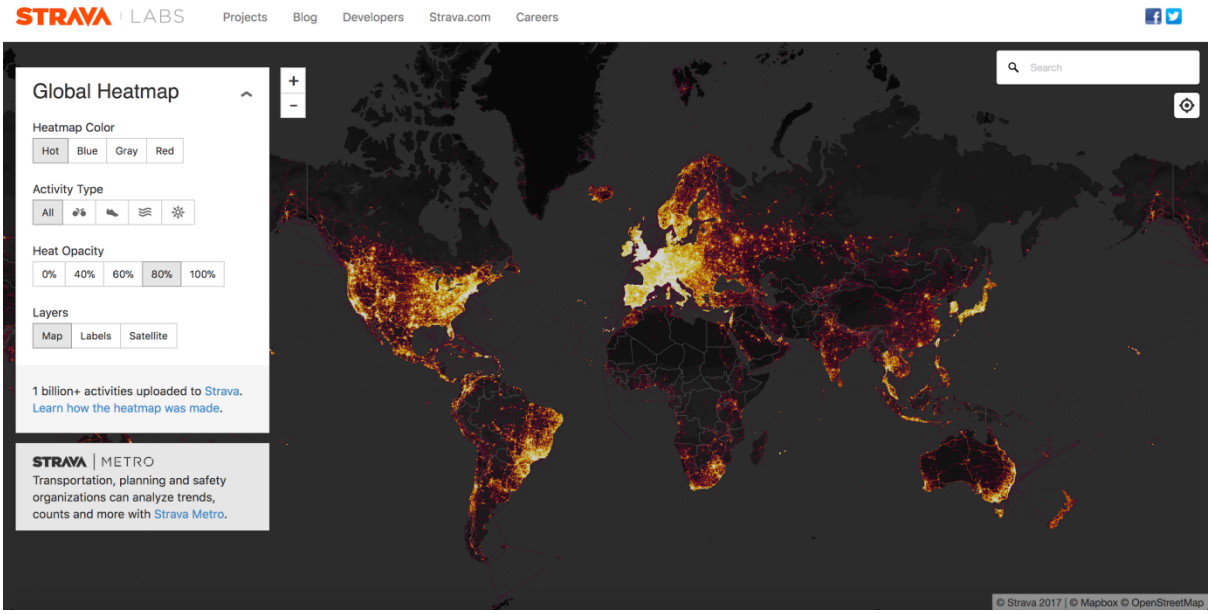


Figure 1. La carte de chaleur de l'agrégation des traces issues de l'application Strava
Source : https://labs.strava.com/heatmap

Fin janvier 2018, un étudiant de Sydney, Nathan Ruser, a accusé cette application de révéler par erreur la localisation de plusieurs bases militaires secrètes. En effet, à partir de la carte de chaleur, ce jeune australien a expliqué qu'il était facile de repérer des zones sensibles dans des lieux peu fréquentés. La cartographie de Strava permet ainsi de connaître les parcours effectués par les militaires sur des terrains d'opération (Mali, Syrie, Afghanistan, Irak...) utilisant des bracelets connectés dont la géolocalisation est activée avec synchronisation a posteriori (figure 2). Ces données peuvent facilement servir ensuite pour localiser des bases militaires ou effectuer une reconnaissance des tours de gardes. L'armée américaine avait recommandé l'usage des bracelets connectés, allant jusqu'à en acheter 2500

[1] Marcheur, ce sont tes traces / ce chemin, et rien de plus ; / Marcheur, il n'y a pas de chemin, / Le chemin se construit en marchant.

pour ses troupes afin de stimuler leurs pratiques sportives. Elle se retrouve aujourd'hui contrainte de lancer un plan d'encadrement pour la gestion de la confidentialité des traces personnelles générées par ces mêmes bracelets.



Figure 2. La cartographie Strava comme ici près du fleuve Niger au Mali permet d'identifier très facilement une activité humaine.

Cet exemple peut sembler anecdotique, il est pourtant symptomatique des conflits d'usages qui risquent de se multiplier dans les années à venir au rythme de la géonumérisation exponentielle de nos pratiques quotidiennes et du monde qui nous entoure (Kitchin, 2014 ; Abitboul et Peugeot, 2017). Avec l'avènement de l'Internet mobile et la démultiplication des objets connectés (téléphones, bracelets, voitures, mobiliers urbains...) la géolocalisation des biens, des personnes comme des contenus numériques semble devenir incontournable. L'indexation spatiale quasi-systématique de nos empreintes numériques (volontairement ou non), si elle offre de nombreuses opportunités de repenser la compréhension des territoires, pose également un lot d'interrogations d'ordre épistémologique, méthodologique et éthique.

Dans cet article, nous proposons de revenir sur la question des traces géonumériques à la fois comme constructions sociotechniques, nouvelles sources de compréhension des territoires et objet de recherche émergent. Dans un premier temps, la notion de trace est présentée et sa déclinaison géonumérique est envisagée selon quatre grandes perspectives permettant sa caractérisation (flux / marque ; volontaire / involontaire ; implicite / explicite ; individuelle / agrégée ; géographique par nature / géographique dérivée). Ces nuances permettent de souligner les contextes fortement diversifiés au sein desquels les traces numériques sont générées. Ce faisant, cette fine caractérisation permet d'éviter les discours trop englobants sur un objet aux multiples facettes qui est loin de constituer un ensemble homogène. La deuxième partie illustre, à partir d'exemples, le potentiel heuristique des traces numériques en géographie en insistant sur leur capacité à rendre visible les spatialités ou encore à repenser les lieux. Enfin, dans la dernière partie sont présentés les grands enjeux méthodologiques, épistémologiques et éthiques associés aux usages des traces géonumériques.

1. De la trace à la trace géonumérique

1.1 Retour sur la notion de trace numérique

La trace peut se définir, *a minima*, comme l'empreinte, la marque d'une action. Terme polysémique, notre acceptation de la « trace » rejoint le paradigme indiciaire de Ginzburg (1980) qui la caractérise, comme le souligne Alexandre Serres (2002) selon trois grands aspects : l'individualité de l'objet d'analyse (la trace renvoie à des pratiques singulières), le caractère indirect de son déchiffrement (la connaissance qui s'opère par trace ne passe ni par l'expérimentation, ni par l'observation directe) et le caractère conjoncturel de son interprétation (il existe une marge irréductible d'aléatoire et d'incertitude). Or, depuis quelques années, le développement du web engendre un engouement pour la déclinaison numérique de cette notion (Galinon-Mélenec et Zlitni, 2013).

Les traces numériques peuvent ainsi se définir comme des données personnelles, descriptives de l'activité ou de l'identité d'un individu. Elles sont produites et stockées soit de manière automatique (après un consentement plus ou moins explicite), soit par le biais d'un dépôt intentionnel. A la manière d'une trace de pas, elles constituent des empreintes numériques mais à la différence des traces de pas qui sont éphémères, les traces numériques sont archivables et exploitables parfois très longtemps après leur apparition. Les traces numériques sont omniprésentes, nous en produisons individuellement des centaines voire des milliers au quotidien par le biais de nos courriels, recherches sur le web, achats en ligne, déplacements en transport en commun ou encore paiements avec une carte de crédit... Prises isolément, elles n'ont guère de sens. Mais regroupées, traitées et combinées, elles constituent des ressources informationnelles significatives et à grande valeur ajoutée potentielle.

1.2 Vers une définition des traces géonumériques

Dans le contexte d'une société toujours davantage empreinte de numérique, la dimension spatiale occupe une place grandissante tant nos pratiques sont de plus en plus connectées et géolocalisées (mobilité, communication, tourisme...). L'usage exponentiel des objets connectés à l'image des smartphones conduit à une multiplication et à une diversification des données géographiques personnelles (Kitchin, 2014). Ces dernières se caractérisent par de nouvelles modalités de production tant au niveau de leur source, de leur volume que de leur précision. On peut définir les traces numériques comme les empreintes laissées derrière soi lorsque l'on navigue sur Internet ou lorsque l'on recoure à un service à partir d'un terminal numérique. Elles sont dites géonumériques, si des données (ou méta-données) géographiques autorisant leur spatialisation, leur sont associées. Il peut s'agir de coordonnées géographiques, d'un toponyme, d'un numéro de borne Wifi, d'une cellule de réseau de téléphonie mobile ou encore d'un nom de lieu.

Si pendant longtemps l'analyse des traces a relevé essentiellement de l'exploitation de dispositifs dédiés à cet usage (*dataloggers*) ou d'applications mobiles pour smartphones (Drevon et al., 2014), désormais de nombreux projets mettent en œuvre des partenariats avec des opérateurs de télécommunication ou développent des protocoles de collecte pour extraire des données en "circulation" sur le web. C'est ainsi que le *web scraping*, technique d'extraction

du contenu de sites web via un script permettant d'automatiser l'opération, devient une pratique de plus en fréquente, à l'image des travaux d'Anne-Cécile Mermet (2017) sur le « syndrôme Airbnb à Helsinki ». Les traces existantes, issues de sources non académiques, peuvent alors provenir :

- des données relatives aux téléphones portables. C'est ainsi, par exemple, que Julie Fen Chong (2015) a tenté de révéler les mobilités ordinaires des touristes en Ile-de-France à partir des données issues de la téléphonie mobile fournies par l'opérateur Orange ;
- des données GPS issus de dispositifs embarqués ou d'objets connectés à l'image de l'anecdote sur l'application Strava évoquée en introduction ;
- des données fournies par les opérateurs de transports en commun ou de vélos en libre-service (Vogel et al., 2011 ; Wang et al., 2015 ; Wergin et Buehler, 2017) ;
- des métadonnées associées aux contenus (textes, photos ou vidéos) qui circulent sur les réseaux sociaux. L'analyse du contenu de Twitter ou de Foursquare a, par exemple, permis à des équipes de recherche d'identifier et de redessiner des "quartiers de vie" de quelques grandes villes nord-américaines pour les comparer aux quartiers administratifs (Frias Martinez et Frias Martinez, 2014) ;
- à une échelle plus fine encore, des données issues de bornes Bluetooth peuvent être analysées pour suivre précisément le parcours de touristes, par exemple, dans des espaces muséographiques (Yoshimura et al., 2014).

Si l'ensemble de ces recherches contribue significativement : "*à une meilleure lisibilité de l'environnement urbain et plus précisément de son inégale pratique par les individus*" (Beaude, 2015 : 137), l'hétérogénéité des sources et des conditions de fabrication de ces traces, doit être regardée de près avant toute tentative de généralisation de leur interprétation.

1.3 Essai de qualification des traces géonumériques

De par leur grande diversité, il est difficile de dresser une typologie exhaustive des traces géonumériques tant les modes de production, les technologies mobilisées comme les logiques sous-jacentes sont composites. En réponse à cette hétérogénéité, nous proposons de distinguer les traces géonumériques selon quatre grandes dimensions qui renvoient à la fois à leur nature (flux / marque), leur dimension géographique (implicite / explicite), leur intentionnalité (volontaire / involontaire) et leur niveau d'appréhension (individuel / agrégé). Cette classification s'inspire en partie des distinctions et catégories d'information géographique volontaire (*Volunteered Geographic Information*) et de données géospatiales issues de processus d'externalisation ouverte (*crowdsourcing*) proposées dans l'ouvrage collectif (Sui et al., 2013).

Traces flux / traces marques

Une première distinction entre flux et marques permet d'appréhender les traces géonumériques selon leur mode d'inscription dans l'espace. Benjamin Steck (2011) distingue en ce sens la *trace-flux* qui serait "mouvement sur le territoire" de la *trace-marque* qui serait "cristallisation du mouvement sur la carte". Appliquées aux données spatiales, d'un côté les *traces-flux* matérialisent l'enregistrement d'une pratique sous la forme d'une ligne. Il peut s'agir d'un itinéraire de courses à pied enregistré via une application ou d'un déplacement en

voiture où l'individu mobilise un assistant de navigation. De l'autre, les *traces-marques* sous forme ponctuelle, matérialisent une action (comme un *check-in*), la publication d'un contenu (comme un tweet) ou encore un signalement. Il est important de différencier ces deux modes d'inscription spatiale dans la mesure où ils ne renvoient pas aux mêmes contextes de production et surtout aux usages qui en découlent en termes de traitement et d'analyse.

Traces volontaires / traces involontaires

Michael Goodchild a proposé dès 2007 la notion de « Volunteered Geographic Information » (VGI) pour évoquer l'externalisation croissante de la production de contenus géographiques par les citoyens devenus des capteurs de leur environnement (par exemple, avec le développement de plate-forme de publication et d'échange de photographies géoréférencées) ou de leur propre vie (par exemple, avec le développement des outils de « mesure de soi »). Mais plus récemment, plusieurs auteurs (Fisher, 2012 ; Caquard, 2014) ont détourné cette notion pour mettre en évidence le développement de dispositifs de capture involontaire de données géographiques à travers la notion de « inVoluntereed Geographic Information » (iVGI) à l'image de l'exploitation des traces numériques laissées par l'emprunt de vélos en libre-service, la publication d'un tweet ou encore l'utilisation d'une application mobile de navigation.

Ainsi, dans les processus de production de traces géonumériques, les individus peuvent ou non être conscients de l'enregistrement des traces qu'ils génèrent. L'activation subreptice des différentes fonctions de localisation (GPS, wifi, GSM, UMTS) au sein des dispositifs mobiles et plus spécifiquement à travers les applications est régulièrement dénoncée. Des travaux de l'INRIA ont, par exemple, mis en avant que les smartphones utilisant le système d'exploitation Android (de Google) continuaient à émettre des émissions Wi-Fi, même lorsque le signal était désactivé (Matte et al., 2017). Une enquête récente du journal Le Monde a de son côté révélé que l'application Izly, imposée aux étudiants pour le paiement de leur repas dans les restaurants universitaires gérés par le CNOUS, géolocalisait ses utilisateurs et transmettait ces informations de géolocalisation à des sociétés publicitaires (Untersinger, 2017). Dans la même veine, une enquête bien documentée de Corentin Durand (2017) met en lumière les nouvelles modalités de collecte et de monétisation des données personnelles de localisation via des applications tierces.

L'opposition volontaire/involontaire dans les processus de production de traces géonumériques apparaît ainsi un peu trop binaire tant dans le monde numérique tout est trace (Pelissier, 2015) et tant la prise de conscience de l'activation des fonctions de géolocalisation n'est pas toujours évidente à prouver ou à réfuter. Nombre d'applications mobiles incitent l'usager à activer la géolocalisation sans qu'il soit en capacité d'identifier la raison (pourquoi ?) et la fréquence (quand ?) de son déclenchement (figure 3). Cette tendance est d'ailleurs largement amplifiée par l'intégration des fonctionnalités de partage de position (aux proches et aux amis) que proposent désormais les réseaux sociaux comme Facebook ou SnapChat.



Figure 3. Exemples d'applications ayant accès aux données de localisation du téléphone

Traces implicitement géolocalisées / traces explicitement géolocalisées

La plupart des traces géonumériques ne sont pas des données explicitement géolocalisées c'est-à-dire qu'elles ne possèdent pas de coordonnées associées à leur contenu. Lorsque les traces sont générées depuis des applications qui permettent d'observer son environnement (inventaire faunistique ou floristique) ou son activité physique (tableau de bord sur ses pratiques sportives), la dimension spatiale est souvent centrale. La collecte de traces passe alors obligatoirement par l'activation des services de géolocalisation et la restitution des données se fait alors souvent sur un fond de plan qui va permettre de localiser (une espèce, un parcours) ou de visualiser la densité (des observations ou des activités).

Lorsque les traces sont implicitement géolocalisée, leur géoréférencement est dérivé, issu d'un traitement, plus ou moins complexe. Le processus de géoréférencement des traces, lequel autorise leur spatialisation, s'appuie sur un ensemble de techniques permettant de géolocaliser des données par des manières détournées à l'image du géocodage (localisation à l'adresse) ou du *geoparsing* (extraction de mots clés dans un texte et localisation à partir de dictionnaires de lieux). C'est le cas, par exemple, des photographies déposées sur des sites de partage comme Flickr ou Google Photos (ex Picasa). L'indexation spatiale des photographies qui permet d'y accéder à partir de l'exploration d'une carte se fait soit par l'utilisation des coordonnées géographiques intégrées dans les métadonnées du fichier soit par une extraction des mots-clés (tags) géographiques qui caractérisent la photo (par exemple, *tour Eiffel, Paris*).

Traces individuelles / traces agrégées

Enfin, les traces numériques peuvent également être envisagées individuellement ou de manière agrégée. Les données individuelles sont des résultats d'enregistrements issus d'une personne (ou d'un compte utilisateur) unique. Elles peuvent êtres nominatives (identité ou pseudonyme de la personne) ou dans certains cas anonymisées pour permettre leur utilisation dans le respect des législations. Les traces agrégées sont quant à elles regroupées, classées par catégorie ou traitées statistiquement. Les méthodes d'agrégation sont multiples (spatiales, statistiques, sociales, temporelles, etc.) et de plus en plus en plus complexes. Le passage de données individuelles à des unités spatiales modifiables et surtout arbitraires (Openshaw, 1981) doit être analysé de près pour éviter tout risque de sur-interprétation.

2. Potentiels heuristiques des traces numériques en géographie : quelques exemples

2.1 Visualiser les spatialités

Comme le souligne Boris Beaude (2015 : 134) : « *l'enjeu principal des traces numériques est un enjeu de visibilité* ». La traces géonumériques permettent, en effet, de saisir et de donner à voir des pratiques spatiales avec une granularité (l'individu) et des échelles spatio-temporelles difficilement accessibles avec d'autres méthodes.

À travers sa pléthore de services, Google enregistre au quotidien des milliers de traces de chacun de ses utilisateurs. Depuis quelques mois, la firme de Palo Alto permet à ses usagers de visualiser, de télécharger et de contrôler leurs différentes traces laissées sur la toile (recherche, communication, mobilité). La plate-forme *takeout* qui centralise toutes ses données personnelles propose notamment d'accéder à son historique de position et de le visualiser sur Google Maps. Une rapide exploration de ces traces permet d'appréhender de manière concrète le volume et le contenu de certaines de nos traces géonumériques comme l'illustre la figure 4 qui présente les traces d'un des auteurs. La carte de gauche présente les lieux identifiés par Google, il peut s'agir de commerces, de restaurants ou de gare que l'on a fréquenté. Dans cette logique de qualification automatique des lieux récurrents, Google localise également le domicile et le lieu de travail. La carte de droite illustre de son côté les traces brutes issues de l'historique de position.

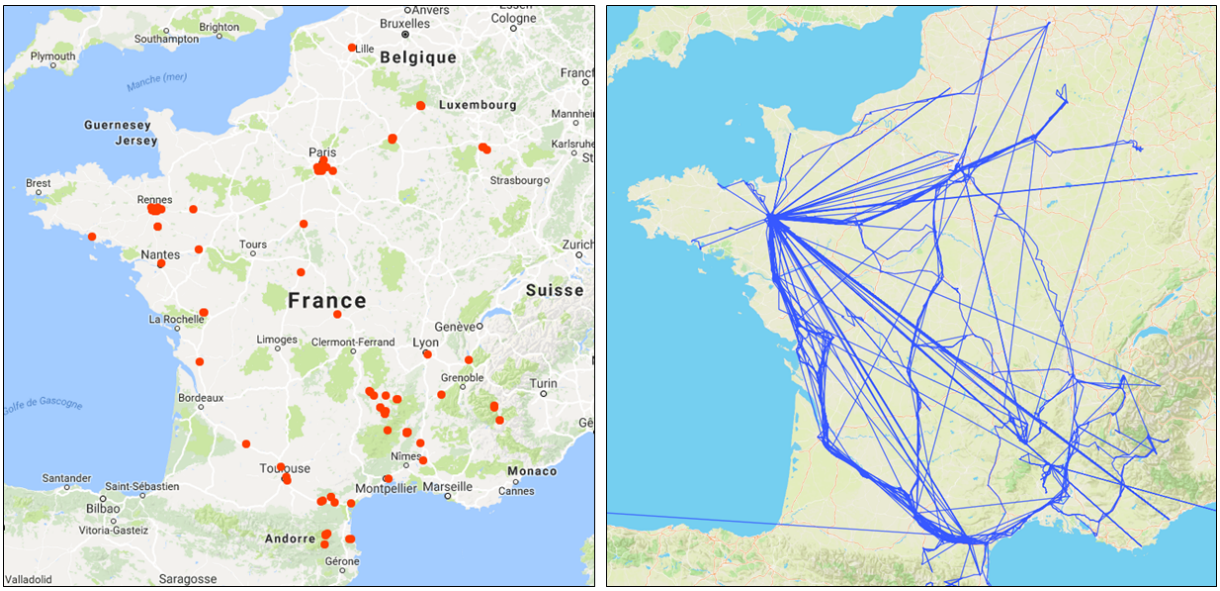


Figure 4. Traces géonumériques issues d'un compte Google (lieux et trajets)

Certaines plateformes qui centralisent des millions de données agrègent des traces individuelles pour proposer des géovisualisations à l'esthétique particulièrement soignée. C'est le cas, par exemple de la collection de cartes *Local versus Tourist* (https://www.flickr.com/photos/walkingsf/sets/72157624209158632/) qui distingue pour des dizaines de métropoles les zones les plus photographiées par des touristes de celles préférées par les habitants. Pour ce faire, Éric Fisher a collecté les métadonnées de milliers de photos présentes sur la plateforme Flickr, en distinguant touristes et locaux selon la fréquence des prises dans la ville en question.

Les données relatives à l'utilisation des services vélos en libre services donnent également lieu à de multiples travaux autour de l'exploitation de ces traces (Rouquier et Bournat, 2014). L'intérêt pour cette thématique s'explique en partie par la bonne accessibilité aux données qui sont souvent disponibles librement sous des formes structurées (*via* des API en temps réel ou des lots de données archivés). Les informations compilées, agrégées et traitées (horaires, occupations des stations, localisations des stations de départ et d'arrivée, durée du trajet, profil de l'usager...) permettent de visualiser sous formes de cartes et de graphiques comment les usagers se comportent dans le temps et dans l'espace. Le projet *Bike Sharing Atlas* mené par une équipe de l'Université de Vienne propose par exemple l'analyse en temps réel de 400 services de vélos en libre-accès à travers le monde (http://bikesharingatlas.org).

Au-delà de la simple visualisation, de nombreuses études explorent davantage la dimension analytique à des fins d'optimisation de gestion (Vogel et al., 2011 ; Wang et al., 2015 ; Wergin et Buehler, 2017). Les travaux d'Etienne Côme et Latifa Oukhellou (2014) autour de l'utilisation des Vélib's proposent par exemple d'appréhender les pratiques sous des formes complexes basées sur une palette d'analyses (récurrences, profils d'usages similaires, matrices Origine Destination dynamiques, typologie des stations).

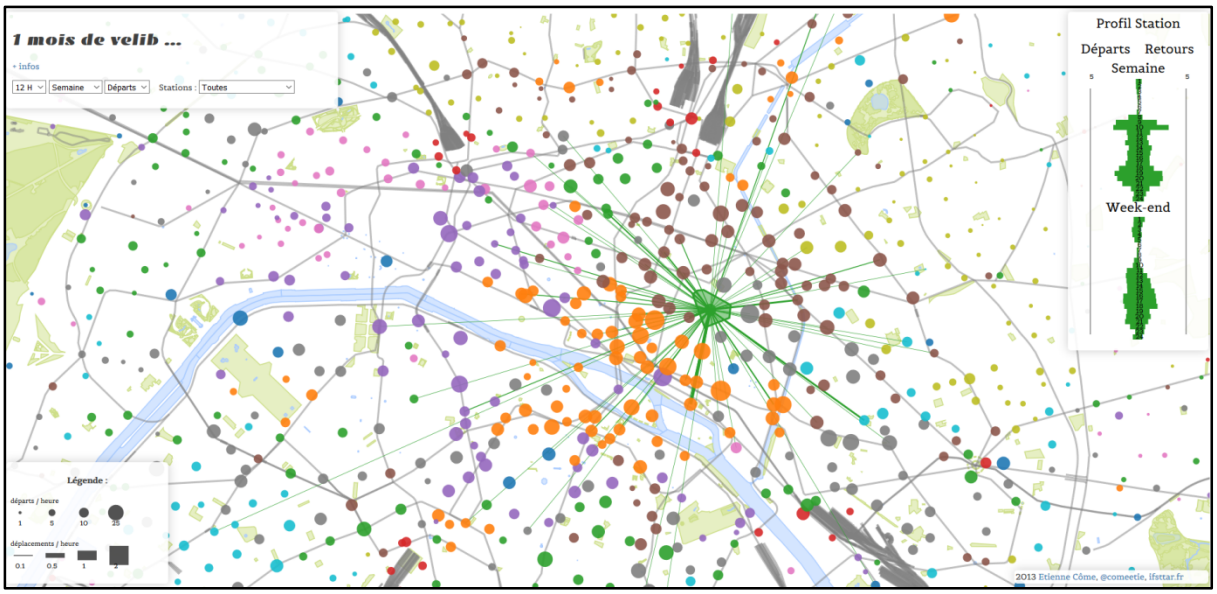


Figure 5. Interface de l'application *1 mois de velib.*
(http://www.comeetie.fr/galerie/velib/index.html)

Le domaine de la mobilité et des transports en général est un grand pourvoyeur de traces géonumériques. Il existe de nombreuses applications permettant de visualiser en temps réel la localisation des avions (https://www.flightradar24.com), des trains (http://raildar.fr), des bus (https://www.livemap24.com/) ou encore des bateaux (https://www.marinetraffic.com/). Autre thématique au cœur de questionnements actuels, les nouvelles formes de mobilité comme le covoiturage. Il est ainsi possible via l'API de l'entreprise BlaBlaCar de récupérer, d'analyser et de cartographier la multitude de trajets de cette communauté (figure 6).

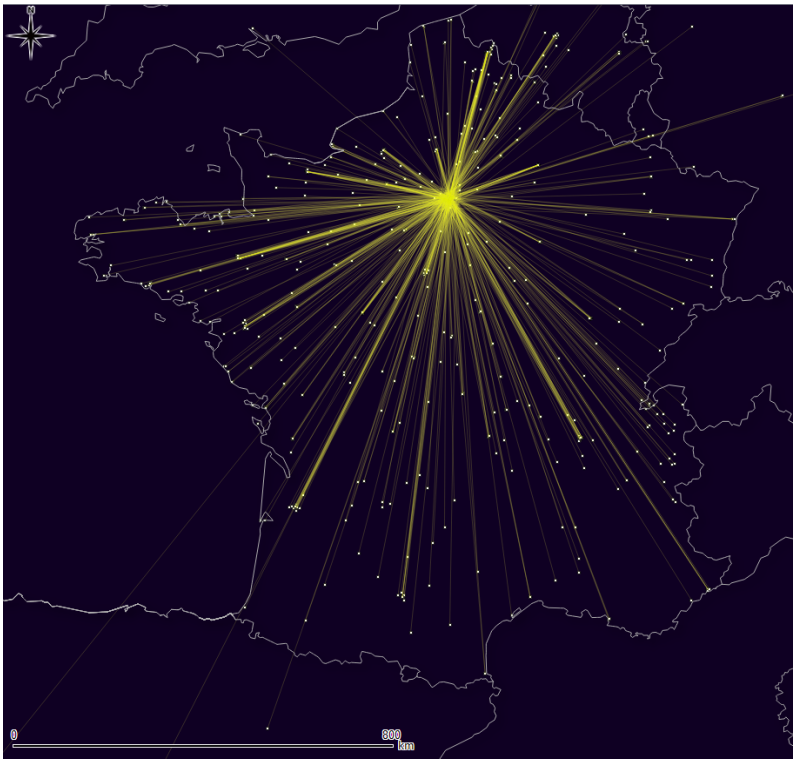


Figure 6. Carte des trajets BlaBlaCar du 01/03/2018 au départ de Paris

De leur côté, les réseaux sociaux et notamment Twitter font l'objet de plus en plus d"analyses et essais de géovisualisations selon différentes perspectives. La plateforme TweetMap (figure 7) propose par exemple l'exploration géographique de millions de tweets selon différents paramètres (hashtag, langue, date, lieux...). L'application Trendsmap présente sous forme d'une carte dynamique les sujets les plus populaires (https://www.trendsmap.com). Enfin, le service Twitmap permet de visualiser en temps réel les messages envoyés à proximité de sa propre localisation (https://twimap.com).

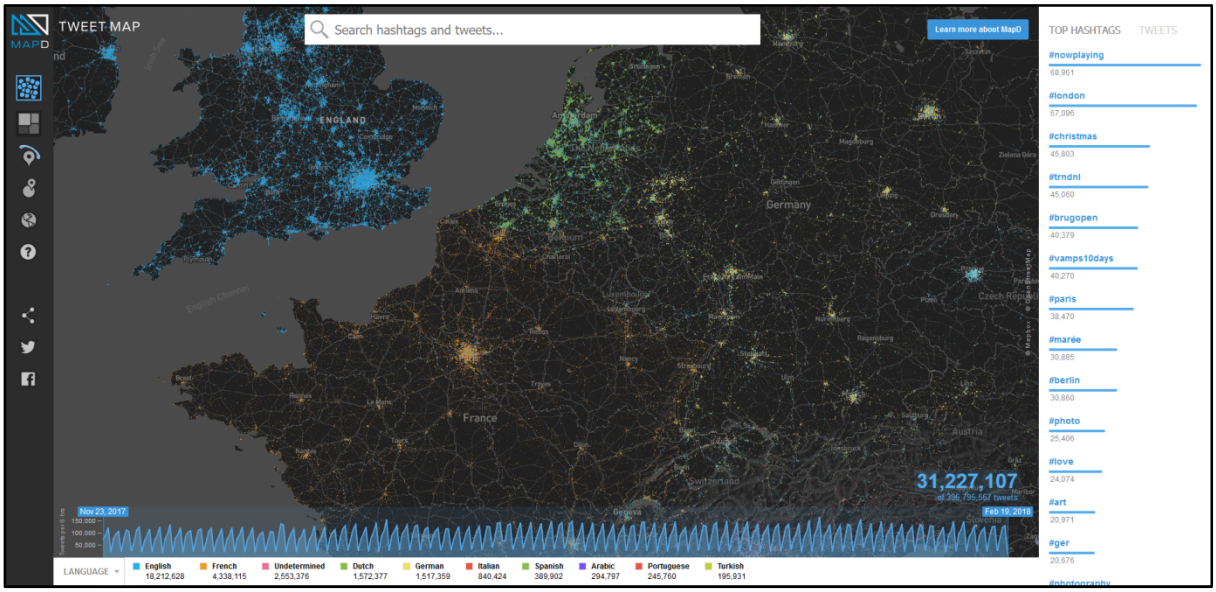


Figure 7. Interface de la plateforme TweetMap.
(https://www.mapd.com/demos/tweetmap)

Tous ces travaux d'artistes, de designers web ou de géographes permettent d'entrevoir le potentiel heuristique des traces numériques à travers leur capacité à rendre visible autrement les spatialités par la dimension inédite des données et surtout leur volume. Mais ces visualisations esthétisantes ne sont pas toujours facilement reproductibles et leur rigueur méthodologique peut être discutées (cf. 3.1).

2.2 Repenser les lieux

Sur un plan plus technique, l'exploitation des *traces-marques* en particulier, permet d'envisager de nouvelles voies de caractérisation de la saillance (géométrique, structurelle voire sémantique) des points de repères spatiaux (Quesnot and Roche, 2015). La saillance (ou prégnance) est un concept qui en psychologie correspond à la propension (ou qualité) d'un objet à se démarquer des autres éléments présents aux alentours (Raubal and Winter, 2002). Ces points de repères (*landmarks*) constituent l'épine dorsale des systèmes de planification d'itinéraires et d'aide à la navigation, de plus en plus utilisés dans le contexte des mobilités urbaines actives et équipées. La constitution automatique ou quasi-automatique de bases de données de "landmark" est un enjeu réel, et en l'espèce, les traces géonumériques constituent certainement les principales sources de données utilisables.

Mais au-delà de la spatialisation de ces traces qui permet leur usage comme source de données et leur visualisation, le recours aux traces numériques en géographie peut

encourager, par de multiples aspects, un renouvellement des réflexions sur la notion de “lieu”. Le “lieu” est un concept majeur en géographie. Il en existe différentes acceptions que les dictionnaires de géographie illustrent abondamment. Aussi il n’est pas question ici d’en dresser un état de l’art complet. Mais si l’on admet que le lieu est en quelque sorte la conjonction de la fonction d’un objet géographique (un évènement, un objet matériel, une idée...), d’une localisation géographique et d’un nom associé (Roche, 2016), alors le recours aux traces géonumériques ouvre une nouvelle voie d’exploration des lieux. Les traces individuelles donnent accès aux lieux “en train de se faire”. Elles sont en quelques sorte les porte-voix informationnels de celles et ceux qui font les lieux. En une localisation géographique cohabitent dans une synchronicité temporelle ou en décalage (lieux éphémères, lieux récurrents…) une multitude de lieux. L’épaississement numérique (réalité augmentée, objets connectés et réactifs…, mais aussi personnalisation algorithmique des lieux physiques) des lieux matériels rajoutent encore à la complexité de cette réalité (Zook and Graham, 2007). Jamais il n’a été aussi “facile” d’accéder à de l’information produite au rythme même des dynamiques spatiales qu’elle décrit. Une nouvelle forme d’intelligence spatiale (bien dans le sens étymologique de capacité à comprendre, à créer des liens pertinents entre les composantes d’un tout), l’intelligence des lieux (Roche, 2017), est alors à portée de main ou presque.

L’exploitation des données émises, par les capteurs dont on a précédemment mis en évidence la grande diversité, peut ainsi permettre de faire émerger des réseaux fonctionnels et/ou hiérarchisés, des polarités spécifiques, centrales, secondaires ou dérivationnelles. La configuration des réseaux de traces peut ainsi contribuer à révéler des concurrences ou des complémentarités en matière d’attractivité et de fonctionnalités entre des quartiers. Les travaux de Gwiazdzinski et Wu (2017) sur Shanghai à partir des traces issues du réseau social *SINA Micro-blog* permettent, par exemple, de repérer les pôles d’animation nocturnes et de visualiser leur évolution temporelle.

L’exploration des métadonnées (tags) de localisation des informations indexées par les internautes peut aussi permettre de comparer les toponymes officiels et les toponymes d’usages (Jones *et al.*, 2008). Ces traces sémantiques se révèlent alors riche de sens pour comprendre comment les habitants s’approprient - font - certains lieux ou certains géosymboles de la ville (Noucher, 2014). Les travaux autour de l'agrégation spatiale de traces pour redessiner les dynamiques locales urbaines illustrent bien les nouveaux rapports aux territoires qui se constituent (Arribas-Bel, 2014). L’application Livehoods (figure 8) mobilise par exemple des données de Foursquare et Twitter pour proposer des découpages de quartier basés sur les pratiques, la proximité et les temporalités des utilisateurs de ces réseaux sociaux (Cranshaw et al., 2012).

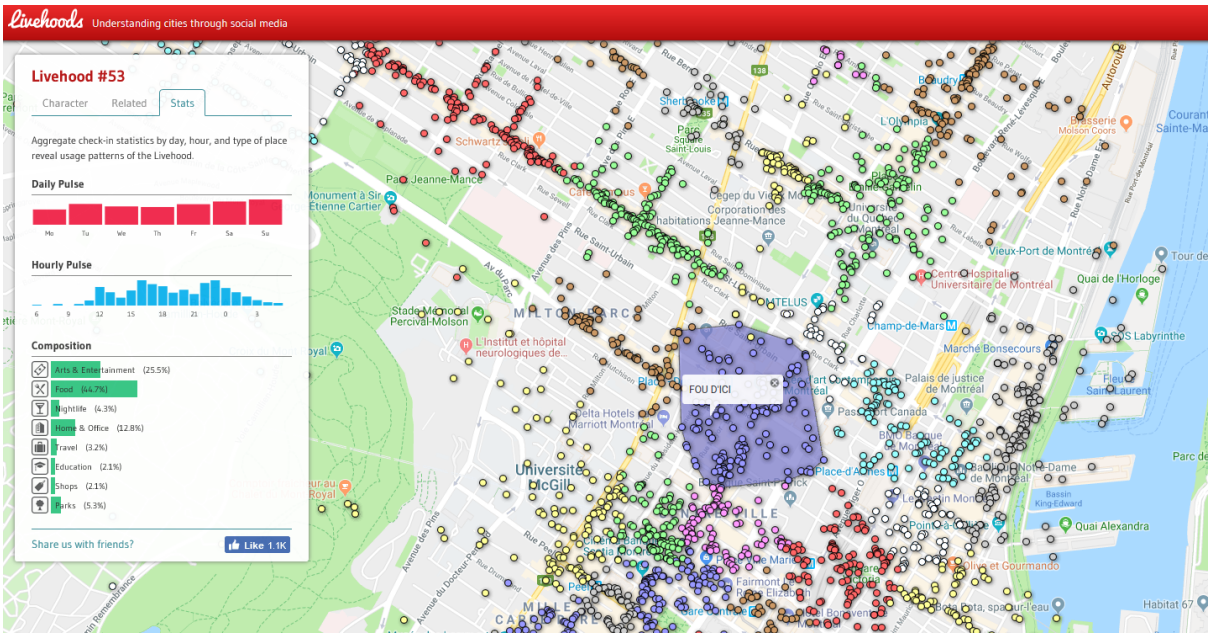


Figure 8. Interface de l'application Livehoods/Montréal.
(http://livehoods.org/maps/montreal)

3. Enjeux méthodologiques et éthiques des traces géonumériques

3.1 Enjeux techniques et méthodologiques

Beaucoup de travaux scientifiques mettent en garde sur les limites des solutions clés en main qui proposent des analyses des traces numériques (Fournier et Jacquot, 2014; Vece, 2014; Vidal, 2015). Essayer de faire parler les traces numériques c'est aujourd'hui se confronter aux techniques – multiples – de fabrique des informations géographiques en cherchant à ouvrir les boites noires algorithmiques dont elles sont issues. Celles-ci se caractérisent par une opacité croissante de leurs processus de collecte, d'agrégation, de traitement et de mise en circulation.

Il faut dire que beaucoup de traces possiblement récupérables sur la toile sont dans les faits de plus en plus verrouillées par les géants du numérique (Google, Facebook, Twitter, Airbnb, etc.), désireux de contrôler, à des fins commerciales ces données, éminemment stratégiques (Quesnot, 2016). Dans les faits, les API (*Application Programming Interface*) qui représentent des portes d'entrée vers les données sont de plus en plus verrouillées. Ainsi, l'étude de plateformes commerciales dont les traces numériques sont plus difficilement accessibles ou pour lesquelles les données accessibles sont loin d'être représentatives ne peut être envisagée sans un risque réel de sur-interprétation de données non représentatives. A titre d'exemple, seul 1% des messages du réseau social Twitter est géolocalisé (Hawelka et al. 2013) et l'API grand public ne donne accès qu'à 1% du trafic total (Morstatter et al. 2013). L'accès à un flux contenant la totalité des tweets peut se faire, via le flux de diffusion *Firehose* de Twitter, mais nécessite alors d'établir un partenariat (financier) avec la firme de San Francisco.

Une fois les données collectées, leur traitement relève bien souvent de la gageure. D'une part, les contenus produits sont rarement structurés pour être croisés et analysés. Les données sont bien souvent peu documentées et l'interopérabilité technique (format d'échanges) et sémantique (informations associées) demeure encore limitée. D'autre part, il n'existe pas encore de véritables solutions logicielles dédiées aux traitements de ces données d'un type particulier (volumineuses, hétérogènes et en temps réel). Quelques outils permettent, certes, des opérations basiques, mais les fonctionnalités de transformation et d'analyse spatiale, sont assez limitées par rapport aux systèmes d'information géographique (SIG). Quant aux nouvelles méthodes autour de la fouille de donnée, l'apprentissage ou l'analyse prédictive autour des données géographiques, elles restent encore très exploratoires et essentiellement développées par des informaticiens et non par des géographes ou des géomaticiens.

3.2 Enjeux éthiques

L'exploitation des traces numériques en géographie - comme dans tous les autres domaines d'usage - pose en particulier la question de leur propriété et de leur divulgation dans le domaine public (Lucchini et Elissalde, 2014). Les enjeux éthiques (Cliche et al., 2016) de l'usage de ces données doivent par conséquent demeurer au cœur de toute pratique de recherche qui viserait à les remobiliser.

L'un des premiers enjeux éthiques évoqué rapidement plus haut concerne l'effet boite noire des traces numériques. Compte tenu de leur mode de fabrique, il est extrêmement difficile de s'assurer que les moyens de production, de captation, d'analyse et de transformation même demeurent en phase avec les fins pour lesquelles elles ont été fabriquées (quand d'ailleurs ces fins sont définies, ce qui n'est pas toujours le cas). Dès lors, les détournements d'usage engendrent des risques d'ordre éthiques, mais aussi juridiques, dont la durée de vie est difficile à prévoir. En effet, la collecte de ces traces individuelles en particulier, ouvre la porte à une multitude d'utilisations futures potentielles qu'il est impossible de prévoir aujourd'hui (Sadowski et Pasquale, 2015). À quoi consent donc réellement l'usager-producteur lorsqu'il produit des traces et qu'il les diffuse ?

La question des enjeux éthiques étroits et des enjeux larges se posent aussi en matière de traces géonumériques. En effet, quand bien même des règles éthiques et juridiques viennent encadrer la production et l'usage des traces à une échelle individuelle, trace par trace, ces règles ne garantissent en aucun cas la prise en compte des enjeux éthiques associés à un niveau d'agrégation supérieur. En d'autres termes, les enjeux éthiques doivent être envisagés de manière distincte pour chacune des catégories de traces évoquées plus haut.

La re-matérialisation des traces numériques grâce (ou à cause de) la géolocalisation constitue également une autre source d'enjeux éthiques majeurs. Il y a en effet désormais une correspondance quasi-parfaite entre les traces numériques et les traces physiques (i.e. les lieux géographiques décrits par les traces), ou au moins une capacité accrue à mettre en relation ces deux ensembles de données. Les conséquences en matière de contrôle et de droit à la vie privée sont importantes. Dans ce contexte, l'idée que l'on puisse se cacher derrière une identité numérique est entièrement remise en question. Notre identité numérique

est consubstantielle à notre identité physique. Dès lors, les risques de surveillance, de catégorisation et de marchandisation des individus sont réels (Kitchin, 2014; Commission de l'éthique en sciences et en technologies, 2017).

La collecte massive de traces permet en effet d'envisager la définition de patrons comportementaux de mobilité et de déplacements types. La côte de confiance associée à chaque citoyen chinois en est un exemple. En effet, le gouvernement chinois travaille à la mise en œuvre d'un indicateur qui permettrait, en exploitant un ensemble de traces individuelles, de déterminer une côte de confiance et ainsi de classer chaque citoyen de manière ensuite à ajuster en fonction de cette côte, la nature et l'ampleur des services auquel chaque individu peut prétendre (c'est un peu ce que font déjà les agences d'évaluation en attribuant à chaque État une côte de crédit). La définition de normes et les triages socio-spatiaux deviennent dès lors possibles. Cette tendance risque d'être considérablement amplifiée par une utilisation normative des patrons issus de l'analyse des données (algorithmes et intelligence artificielle). À partir des patrons développés, la tentation pourrait être grande de prescrire de manière normative des comportements types, des normes d'urbanité imposées, en fonction des lieux, des jours de la semaine, des heures, et de les justifier par la « science de la donnée ».

La seconde conséquence concerne la personnalisation algorithmique des lieux physiques. Les sites web se reconfigurent déjà au rythme des recherches, des consultations, des achats en ligne. À la faveur des traces numériques, les espaces du web (Amazon en est un bon exemple, mais Google Maps également), apprennent des comportements et proposent, par personnalisation algorithmique, des environnements les plus proches des attentes. Demain, dans des environnements urbains instrumentalisés de capteurs dits intelligents, d'objets connectés communicants, chacun sera détecté et identifié, comme sur le Web, dès son arrivée dans un lieu (c'est déjà le cas dans bon nombre de centres commerciaux). La personnalisation algorithmique des lieux – une programmatique des lieux urbains en quelque sorte – pourrait alors entrer en action et les panneaux d'affichages, l'information ambiante (musique, publicité, etc.) ainsi que les notifications sur les téléphones ou tablettes, pourraient se reconfigurer au rythme de l'évolution des comportements urbains et en fonction des personnes présentes à un moment donné.

Conclusion : pour une analyse critique des usages des traces numériques en géographie

Penser le territoire par le prisme des traces numériques et de leur cartographie soulève de multiples interrogations tant éthiques que méthodologiques sur le statut des connaissances produites par ces nouvelles formes de comptage et de mesure. En effet, comme les exemples évoqués tout au long de cet article l'ont illustré, l'agrégation de ces données personnelles suggère une double rupture : d'une part, leur quantité permettrait de tendre vers une exhaustivité et une fréquence inédites ; d'autre part, leur point d'entrée par l'individu permettrait de « désagréger le social » en évitant les catégories parfois pensées comme réductrices.

Ainsi, en promettant aujourd'hui de saisir les mobilités quotidiennes, les pratiques touristiques et d'une manière générale le « pouls » de la ville, de façon immanente, nombre de travaux de recherche dont les analyses sont guidées par ces traces ne proposent plus de se référer à un quelconque référentiel, à une quelconque normalité, à de quelconques catégories (Rouvroy et Berns, 2013). On parle ainsi de *data driven research* ou de pilotage par la donnée pour évoquer ces recherches qui revendiquent des analyses fondées uniquement sur l'exploration de données accumulées et rejettent les approches hypothético-déductives (Plantin et Russo, 2016). La force du calcul et la quantité des données semblent permettre de prendre en compte ce qu'il y a de plus particulier chez les individus, ce qu'il y a de plus éloigné à la moyenne. Rouvroy et Berns (2013) analysent très bien comment cet effacement des classes d'équivalence (Desrosières, 1993) désamorce les questions éthiques relatives à l'utilisation de ces traces numériques individualisées par un évitement de la question du sujet conscient et réflexif : les données collectées paraissent insignifiantes car infra-personnelles (par exemple le nombre de km parcouru pendant une balade en vélo) et agrégées en profils grossiers et en catégories supra-personnelles, ad hoc, temporaires et non stigmatisantes (Gardin et al., 2016).

Dès lors, ces logiques d'agrégation et de mise en circulation rapide sur Internet de traces numériques doivent nous inciter à réactualiser la question de l'intentionnalité que les géographes anglo-saxons tenants de la « cartographie critique » avaient déjà largement travailler dans les années 80 (Harley, 1989). En effet, si le risque d'un détournement d'usage de données cartographiques pouvant être mobilisées en dehors de leur contexte de production ou en décalage avec leurs objectifs initiaux n'est pas nouveau, le caractère malléable des traces numériques qui peuvent être agrégées, désagrégées ou combinées avec d'autres sources rend toute tentative de décryptage de leur intentionnalité particulièrement périlleuse. Dès lors que s'opère une déconnexion entre la source des données (celui qui laisse volontairement ou non les traces) et l'utilisateur des données (celui qui met en carte les traces), il devient plus délicat d'identifier l'intention et l'intérêt, d'origine économique, politique, idéologique, voire même scientifique, qui vont orienter le dépôt des informations puis leur présentation cartographique. Ces préoccupations sont aujourd'hui renforcées, d'une part, par la quantité et l'instantanéité de la production des traces mais aussi, voire surtout, par la grande difficulté d'assurer leur traçabilité dès lors que des opérations d'agrégation ont été initiées. Si leur généalogie est difficile à établir, elle est pourtant indispensable pour envisager une analyse critique susceptible d'interroger finement la mécanique de ces processus. Il en dépend notre capacité à pouvoir choisir individuellement et collectivement les modes de classement, d'organisation et de mises en visibilité de nos spatialités.

Bibliographie

Abiteboul S., Peugeot V. (2017), *Terra Data*. Editions Le Pommier, Paris.

Arribas-Bel D. (2014), Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45-53.

Beaude B. (2015), « Spatialités algorithmiques », In Severo M. et Romele A. (dir.), *Traces numériques et territoires*, Paris, Presses des Mines, p. 135-162.

Caquard, S. (2013), « Collective cartographies in the social media era », *Progress in Human Geography*, 38(1), 141-150.

Cliche D., Roche S. et Turmel P. (2016), Les enjeux éthiques de la ville « intelligente » : données massives, géolocalisation et gouvernance municipale, *Ethica*, 20(1): 223-248.

Côme E., Oukhellou L. (2014), Model-based count series clustering for bike sharing system usage mining: a case study with the Vélib'system of Paris. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(3), 39.

Commission de l'éthique en sciences et en technologies (2017), "La ville intelligente au service du bien commun : lignes directrices pour allier l'éthique au numérique dans les municipalités au Québec", 110 p., 20 octobre 2017, URL: http://www.ethique.gouv.qc.ca/fr/publications/ville-intelligente.html

Cranshaw J., Schwartz R., Hong J. I. et Sadeh N. M. (2012), « The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City », *6th International Conference on Web and Social Media*.

Desrosières A. (1993), *La politique des grands nombres. Histoire de la raison statistique.* Paris, La Découverte, 440p.

Drevon G., Jambon F., Chardonnel S., Christophe S., André-Poyaud I., Davoine P.A., Lutoff C. (2014), « Évaluation comparée de l'apport de l'assistance GPS aux enquêtes de mobilité », *Netcom*, 28-1/2, pp. 13-34.

Durand C. (2017), Enquête : comment les apps Figaro, L'Équipe ou Closer participent au pistage de 10 millions de Français, *numerama,* 23/08/2017.

Fen-Chong J. (2015), « Révéler les mobilités ordinaires et touristiques en Ile-de-France à partir de données de téléphones portables » In Marie-Flore Mattei et Denise Pumain (dir.) *Données urbaines* 7, Economica, pp.67-76.

Fischer F. (2012), « VGI as big data: A new but delicate geo-graphic data-source », *GeoInformatics*, april/may, 46-47

Fournier C., Jacquot S. (2014), Les traces numériques des touristes – Un renouvellement de l'observation touristique ?, *Cahier « Big Data, Traces numériques & observation », Espaces tourisme et loisirs*, n°316, Editions Touristiques Européennes, pp. 66-72

Frias-Martinez V. et Frias-Martinez E. (2014), Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence*, *35*, 237-245.

Galinon-Melenec B. et Zlitni S. (dir.) (2013), *Traces numériques. De la production à l'interprétation*, Paris, CNRS éditions, série L'Homme-trace - tome 2, 274 p.

Gardin J., Didier S., Quentin A. (2016), « Liberté, Egalité, Computer », *justice spatiale | spatial justice*, n°10, Juillet 2016, http://www.jssj.org

Ginzburg C. (1980), « Signes, traces, pistes » Racines d'un paradigme de l'indice, *Le Débat*, n° 6, p. 3-44. DOI : 10.3917/deba.006.0003

Goodchild M. (2007), « Citizens as sensors: Web 2.0 and the volunteering of geographic information », *GeoFocus*, 7, 1-10.

Harley, B. (1989), « Deconstructing the map », *Cartographica* , 26, 1-20.

Hawelka B., Sitko I., Beinat E., Sobolevsky S., Kazakopoulos P., Ratti C. (2014), Geo-located Twitter as the proxy for global mobility patterns, *Cartography and Geographic Information Science*, 41, 260-271.

Hu W., Gwiazdzinski L. et Wan W. (2016), « Les nuits de Shanghaï », *Netcom* [En ligne], 30-3/4 | URL : http://journals.openedition.org/netcom/2519

Jones C., Purves R., Clough P., Joho H. (2008), «Modelling vague places with knowledge from the Web». *International Journal of Geographic Information Science*, n°22/10, p. 1045-1065. doi:10.1080/13658810701850547

Kitchin B. (2014), The real-time city? Big data and smart urbanism, *GeoJournal*, 79(1): 1-14.

Lucchini F. et Elissalde B. (2014), « Qui » et « Où » ? Les enjeux des données géolocalisées, *20e Biennale de géographie d'Avignon - Géopoint 2014* - « Controverses et géographies ».

Matte C., Cunche M. et Toubiana V. (2017), *Does disabling Wi-Fi prevent my Android phone from sending Wi-Fi frames?* (Doctoral dissertation, Inria-Research Centre Grenoble–Rhône-Alpes; INSA Lyon).

Mermet A.-C. (2017), « Airbnb and tourism gentrification: critical insights from the exploratory analysis of the "Airbnb syndrome" in Reykjavík », In Gravari-Barbas M. et Guinand S. (dir.) *Tourism and Gentrification in Contemporary Metropolises* , Routledge, coll. "Contemporary Geographies of Leisure, Tourism and Mobility", p. 52-74.

Morstatter F., Pfeffer J., Liu H., et Carley K. M. (2013), Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *Proceedings of ICWSM*.

Noucher M. (2015), « De la trace à la carte et de la carte à la trace : pour une approche critique renouvelée des nouvelles sources de fabrique cartographique », In Severo M. et Romele A. (ed.), *Traces numériques et territoires*, Paris, Presses des Mines, 213-223.

Noucher M. (2014), À Bordeaux, les cartes font le pont. *M@ppemonde*, n°115. https://mappemonde-archive.mgm.fr/num43/articles/art14301.html

Openshaw S. (1981), « Le problème de l'agrégation spatiale en géographie », *L'Espace géographique*, 10-1, pp.15-24.

Pélissier D. (2015), « Trace, vous avez vraiment dit Trace ? », *Présence numérique des organisations*, http://presnumorg.hypotheses.org/76.

Plantin J.C., Russo F. (2016), « D'abord les données, ensuite la méthode ? », *Socio*, 6 | 2016, 97-115.

Quesnot T. (2016), L'involution géographique : des données géosociales aux algorithmes, *Netcom* [En ligne], vol. 30, n°3/4, URL : http://netcom.revues.org/2545

Quesnot T. et Roche S. (2015), Measure of Landmark Semantic Salience through Geosocial Data Streams. *ISPRS International Journal of Geographical Information*, 4:1-31, URL : http://www.mdpi.com/2220-9964/4/1/1

Raubal M. et Winter S. (2002), *Enriching wayfinding instructions with local landmarks*. In M. J. Egenhofer & D. M. Mark (Eds.), Geographic Information Science. LNCS, vol. 2478 (pp. 243-259). Berlin: Springer.

Roche S. (2017), Spatial Thinking, Interfaces and Algorithmic Urban Places - Toward Smart Cities. *Progress in Human Geography*, 41(5): 657-666, online first July 9th 2016, DOI: 10.1177/0309132516650352.

Roche S. (2016), Less Spaces more Places in Smart Cities. *Progress in Human Geography*, 40(4): 565-573, online first May 19th 2015, DOI: 10.1177/0309132515586296.

Rouquier J.B., Borgnat P. (2014), « Cartographie des pratiques du Vélo'v : le regard de physiciens et d'informaticiens », *Revue Sciences/Lettres*, n°2
http://journals.openedition.org/rsl/487

Rouvroy A., Berns T. (2013), « Gouvernementalité algorithmique et perspectives d'émancipation : le disparate comme condition d'individuation par la relation? » *Politique des algorithmes. Les métriques du web. RESEAUX*, Vol.31, n.177, pp. 163-196.

Sadowski J. et Pasquale F. (2015), The spectrum of control: A social theory of the smart city, *First Monday*, 20(7), [en ligne], URL : http://firstmonday.org/article/view/5903/4660

Serres A. (2002), Quelle(s) problématique(s) de la trace ? *Séminaire du CERCO*, Rennes, le 13 décembre 2002.

Severo M., Romele A. (2015), *Traces numériques et territoires*, Paris, Presses des Mines, Territoires numériques.

Shelton T., Poorthuis A., Zook M. (2015), Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142, 198-211.

Steck B. (2011), « Flux et territoires : de la trace à la marque, une question de distance », In Galinon-Melenec, B. (éd.), *L'Homme trace. Perspectives anthropologiques des traces contemporaines*, Paris: CNRS Editions, pp 249-270.

Sui D., Elwood S. et Goodchild M. (eds.) (2013), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, Springer Netherlands, 396 p.

Untersinger M. (2017), Izly, l'appli du Cnous qui géolocalise des étudiants et renseigne des sociétés publicitairesl*, Le Monde*, 20.10.2017.

Vece P. (2014), Observation des données de la téléphonie mobile – Des perspectives prometteuses qui restent à confirmer, *Cahier « Big Data, Traces numériques & observation », Espaces tourisme et loisirs*, n°316, Editions Touristiques Européennes, pp. 98-104

Vidal P. (2015), Tracer sa route, en toute intransparence numérique ? *Netcom* [En ligne], 29-1/2 | URL : http://journals.openedition.org/netcom/1841

Vogel P., Greiser T. et Mattfeld D. C. (2011), Understanding bike-sharing systems using data mining: Exploring activity patterns. *Procedia-Social and Behavioral Sciences*, *20*, 514-523.

Wang X., Lindsey G., Schoner J. E., Harrison, A. (2015), Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development*, 142(1), 04015001.

Wergin J., Buehler R. (2017), Where Do Bikeshare Bikes Actually Go? Analysis of Capital Bikeshare Trips with GPS Data. *Transportation Research Record: Journal of the Transportation Research Board*, (2662), 12-21.

Yoshimura Y., Sobolevsky S., Ratti C., Girardin F., Carrascal J. P., Blat J. et Sinatra R. (2014), « An analysis of visitors' behavior in The Louvre Museum: a study using Bluetooth data », *Environment and Planning B: Planning and Design*, 41(6), 1113-1131.

Zook M. et Graham M. (2007), Mapping DigiPlace: Geocoded Internet Data and the Representation of Place, *Environment and Planning B: Urban Analytics and City Science* 34(3): 466-482. URL : https://doi.org/10.1068/b3311