Présentation

Imaginer la sociologie numérique

Presentation

Imagining digital sociology

Nicolas Baya-Laffite et Bilel Benbouzid

Volume 49, numéro 2, automne 2017

La sociologie numérique
Digital Sociology

URI : <https://id.erudit.org/iderudit/1054272ar>

DOI : <https://doi.org/10.7202/1054272ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0038-030X (imprimé)

1492-1375 (numérique)

[Découvrir la revue](#)

Citer ce document

Baya-Laffite, N. & Benbouzid, B. (2017). Présentation : imaginer la sociologie numérique. *Sociologie et sociétés*, 49(2), 5–32. <https://doi.org/10.7202/1054272ar>

Résumé de l'article

S'il est admis que les techniques d'enquête en sociologie — et les modes de raisonnement qui leur sont associés — sont étroitement liées à leurs contextes institutionnels et intellectuels d'apparition, comment le développement des écosystèmes numériques transforme-t-il actuellement les manières de savoir sur le social ? Dans cet article introductif, nous observons la façon dont les sciences sociales computationnelles (SSC) et les humanités numériques mettent en tension la sociologie. D'un côté, les sciences sociales computationnelles concurrencent la sociologie en traitant de ses objets les plus classiques, mais dans une perspective prédictive qui ne lui est pas familière. D'un autre côté, les humanités numériques s'imaginent comme le chapiteau des sciences sociales, mais avec une offre numérique qui s'inscrit dans la tradition méthodologique de l'exégèse textuelle de laquelle les sociologues cherchent à se distancier. La sociologie se trouve comme prise dans l'étau des sciences sociales computationnelles et des humanités numériques. Nous avons observé trois types de réactions face à cette situation — la protection, la conservation et l'adaptation — qui montrent à quel point le numérique pénètre progressivement la discipline sociologique, depuis son propre cœur. Pour que celle-ci devienne pleinement numérique, il convient néanmoins de faciliter les conditions d'accès aux données numériques, notamment en discutant le cadre juridique, économique et technique des écosystèmes numériques de la recherche publique.

Présentation

Imaginer la sociologie numérique

NICOLAS BAYA-LAFFITTE

Université de Lausanne – Laboratoire d'étude des sciences et des techniques (STS-Lab)
Courriel : Nicolas.BayaLaffite@unil.ch

BILEL BENBOUZID

Université Paris Est – Laboratoire Interdisciplinaire, Science, Innovation et Société (LISIS)
Courriel : bilel.benbouzid198@gmail.com

RÉSUMÉ

S'il est admis que les techniques d'enquête en sociologie — et les modes de raisonnement qui leur sont associés — sont étroitement liées à leurs contextes institutionnels et intellectuels d'apparition, comment le développement des écosystèmes numériques transforme-t-il actuellement les manières de savoir sur le social ? Dans cet article introductif, nous observons la façon dont les sciences sociales computationnelles (SSC) et les humanités numériques mettent en tension la sociologie. D'un côté, les sciences sociales computationnelles concurrencent la sociologie en traitant de ses objets les plus classiques, mais dans une perspective prédictive qui ne lui est pas familière. D'un autre côté, les humanités numériques s'imaginent comme le chapiteau des sciences sociales, mais avec une offre numérique qui s'inscrit dans la tradition méthodologique de l'exégèse textuelle de laquelle les sociologues cherchent à se distancier. La sociologie se trouve comme prise dans l'étau des sciences sociales computationnelles et des humanités numériques. Nous avons observé trois types de réactions face à cette situation — la protection, la conservation et l'adaptation — qui montrent à quel point le numérique pénètre progressivement la discipline sociologique, depuis son propre cœur. Pour que celle-ci devienne pleinement numérique, il convient néanmoins de faciliter les conditions d'accès aux données numériques, notamment en discutant le cadre juridique, économique et technique des écosystèmes numériques de la recherche publique.

S'IL EST ADMIS QUE LES TECHNIQUES D'ENQUÊTE en sociologie — et les modes de raisonnement qui leur sont associés — sont étroitement liées à leurs contextes institutionnels et intellectuels d'apparition, comment le développement des écosystèmes numériques transforme-t-il actuellement les manières de savoir sur le social ?

Depuis plus d'une dizaine d'années, de nombreuses recherches et innovations se sont appuyées sur les écosystèmes numériques. Les flux des signaux émis par les êtres humains au gré de leurs actions produites au moyen des téléphones portables, des messageries électroniques, des moteurs de recherche, des réseaux sociaux, des sites de *microblogging*, des applications Web et des objets connectés alimentent des bases de données, mais aussi des technologies d'intelligence artificielle, qui en retour peuvent être utilisées pour organiser l'information, produire des savoirs et optimiser la performance des individus et des collectifs.

Alors qu'il s'agit sans doute de l'un des enjeux les plus importants de ce début du XXI^e siècle, les sociologues sont peu équipés et formés pour tirer profit de ces écosystèmes numériques, laissant aux informaticiens et aux physiciens l'apanage de l'organisation et de la production des connaissances sur de nombreux phénomènes sociaux, historiques et culturels. Des milliers d'études sur des objets classiques de la sociologie (le crime, les organisations, le travail, la sociabilité, l'art, l'innovation, etc.) s'appuient sur ces écosystèmes numériques, mais leurs résultats sont moins souvent valorisés dans les revues de sociologie que dans les revues scientifiques généralistes (*Science* et *Nature*, notamment). Les conférences des informaticiens et des physiciens (celles par exemple des différentes sociétés de l'influent « Institute of Electrical and Electronics Engineers » [IEEE]) traitent désormais d'objets communs avec la sociologie, mais les techniques développées ne sont pas associées aux questionnements de la discipline. Des espaces spécialement dédiés à la production de savoirs sociaux à partir des écosystèmes numériques se sont constitués dans les grandes entreprises du numérique (comme Facebook, Google, Microsoft), mais la présence des sociologues y reste marginale. En Amérique du Nord et en Europe, les agences publiques de financement de la recherche, généralement dédiées à l'innovation technologique, financent régulièrement des projets pour analyser les données issues des écosystèmes numériques, mais elles exigent des solutions pratiques et s'inspirent de modèles d'affaires qui ne correspondent pas à la manière dont la sociologie s'est historiquement constituée en savoir critique et d'émancipation. Comment dès lors s'appuyer sur les écosystèmes numériques pour produire des savoirs sur le social qui soient propres à la sociologie ?

Dans cet article introductif, nous observons la façon dont deux domaines connexes de la sociologie — les sciences sociales computationnelles (SSC) et les humanités numériques — mettent en tension la discipline (1). D'un côté, les sciences sociales computationnelles concurrencent la sociologie en traitant de ses objets les plus classiques, mais dans une perspective prédictive qui ne lui est pas familière. D'un autre côté, les humanités numériques s'imaginent comme le chapiteau des sciences sociales, mais avec une offre numérique qui s'inscrit dans la tradition méthodologique de l'exégèse textuelle de laquelle les sociologues cherchent à se distancier. La sociologie se

trouve comme prise dans l'étau des sciences sociales computationnelles et des humanités numériques. Nous avons observé trois types de réactions face à cette situation — la protection, la conservation et l'adaptation — qui montrent à quel point le numérique pénètre progressivement la discipline sociologique, depuis son propre cœur (2). Pour que celle-ci devienne pleinement numérique, il convient néanmoins de faciliter les conditions d'accès aux données numériques, notamment en discutant le cadre juridique, économique et technique des écosystèmes numériques de la recherche publique (3).

L'ÉTAU DES SCIENCES SOCIALES COMPUTATIONNELLES ET DES HUMANITÉS NUMÉRIQUES

C'est à l'extérieur du champ sociologique que les promesses numériques les plus alléchantes ont été formulées, par les sciences sociales computationnelles d'une part, et par les humanités numériques de l'autre. Observons-les tour à tour, en analysant la tension que ces deux domaines d'études créent en sociologie, en premier lieu autour des enjeux méthodologiques des écosystèmes numériques.

Les nouvelles méthodes des écosystèmes numériques

En 2009, quinze chercheurs américains publient *Life in the network: the coming age of computational social science* (Lazer et al., 2009). Cet article, publié dans la revue *Science*¹, est programmatique et devient rapidement notoire. Alors que les SSC remontent aux simulations sociales des années 1960, l'article annonce sans réserve l'apparition d'un nouveau domaine et une révolution scientifique en marche par la modélisation des comportements humains à partir de l'analyse des données numériques massives produites par les individus dans le cours de leur vie quotidienne². Forts d'une dizaine d'années d'expérience en matière d'analyse de données numériques, les auteurs voient dans les écosystèmes numériques des avenues de recherches inédites en sciences sociales : si la physique des particules subatomiques dispose du Centre européen pour la recherche nucléaire (CERN) et l'astronomie du télescope Hubble, les SSC ont désormais les écosystèmes numériques pour parvenir à reconstituer la toile du réseau des relations entre les humains et les choses. Comment ces écosystèmes numériques transforment-ils la manière de penser les méthodes pour l'analyse des phénomènes sociaux ?

Les données disponibles grâce aux écosystèmes numériques ont atteint une échelle de population sans précédent, fournissant aux chercheurs des données brutes et permettant d'observer de façon synchrone et évolutive l'activité de millions d'individus.

1. *Science* est une revue généraliste de référence. Publiant les travaux des scientifiques les plus renommés, *Science* est considérée comme l'une des plus prestigieuses revues scientifiques.

2. Si les chercheurs qui se revendiquent des SSC ne viennent pas des sciences sociales, mais de la physique et de l'informatique, tout porte à prendre au sérieux leur ambition de renouveler et même remplacer les sciences sociales existantes. Les plus importants d'entre eux occupent des places dans des universités prestigieuses (MIT, Stanford, Columbia, etc.) et collaborent directement avec les grandes plateformes du numérique, d'où des recherches en SSC qui s'inscrivent souvent dans des logiques d'entreprise.

En 2007, Leskovec et Horvitz (Leskovec et Horvitz, 2008) observent par exemple pour la première fois le réseau mondial des échanges de 180 millions d'utilisateurs de la messagerie Microsoft Messenger. Ils montrent l'existence d'une homophilie dans les interactions selon les tranches d'âge et d'une hétérophilie pour les variables de genre, et, à l'aide d'une topographie du réseau mondial des échanges, ils vérifient l'hypothèse des six degrés de séparation de Stanley Milgram (nous revenons plus bas sur l'analyse de réseaux). En 2015, avec une masse de données tout aussi volumineuse, Bakshy et ses collaborateurs analysent un échantillon de 10 millions de comptes Facebook américains, pour comprendre la manière dont l'information est partagée sur ce réseau. En pleine crise relative au phénomène de « bulle de filtre³ », Bakshy et ses collaborateurs (Bakshy *et al.*, 2015) cherchent à comprendre la manière dont les internautes sont exposés à des opinions contraires sur le réseau et comment ils sélectionnent l'information qui apparaît sur leur fil d'actualité. En comparant l'exposition produite par l'algorithme avec ce que les internautes sélectionnent réellement, les auteurs parviennent à montrer que les utilisateurs régulent eux-mêmes leur fenêtre de visibilité, plus que ne le fait l'algorithme en filtrant l'information. Aussi volumineuses que soient les données analysées, il est reproché à ces études de ne pas maîtriser la représentativité des flux capturés (Venturini *et al.*, 2014b), limitant de fait les possibilités de montée en généralité. Les chercheurs des SSC se défendent pourtant de cette critique : selon eux, la représentativité est nécessaire si l'objectif du chercheur est de parler au nom d'une population dans son ensemble (les Québécois, les étudiants, etc.). Or les très gros volumes de données permettent plutôt de saisir des *groupes* spécifiques (les utilisateurs d'une messagerie, des blogueurs, etc.) et une grande hétérogénéité de *situations* (Salganik, 2017).

Les écosystèmes numériques peuvent également être utilisés pour collecter les données d'activité des individus en temps réel (*reality mining*) (Pentland, 2015) et en permanence (*always-on data*) (Salganik, 2017). Il s'agit dans cette approche d'utiliser des capteurs de toutes sortes (téléphones portables, badges, comptes de réseaux sociaux, etc.) afin de collecter les données des activités à grande échelle et d'enregistrer l'ensemble de l'activité des personnes. Pentland et ses collaborateurs du MIT Medialab considèrent les données produites par ces capteurs comme des signaux sociaux à partir desquels il est possible d'analyser les comportements humains (Pentland, 2015). Par exemple, dans le cadre d'une analyse des données de mobilité de centaines d'utilisateurs à partir de leurs téléphones portables (l'état du téléphone, les comptes rendus des appels et des messages textuels, etc.), les chercheurs du MIT (Eagle *et al.*, 2009) démontrent que ces données d'activité produisent des résultats plus satisfaisants que les données d'enquête (déclaration des listes d'amis, estimation spatiale et temporelle de ces amitiés, etc.) pour prédire des relations d'amitié ou la satisfaction dans les relations de travail. Plus fiables que les déclarations des enquêtés, ces *signaux honnêtes* émis

3. Il s'agit de la dénonciation de l'enfermement des internautes dans une bulle informationnelle construite par les algorithmes des plateformes du Web (Pariser, 2012).

par les individus dans le quotidien de leur interaction sont particulièrement appréciés des SSC : elles ne « tromperaient » pas l'analyste.

Dans cette perspective expérimentale inédite, les écosystèmes numériques ont permis le développement d'une série de recherches utilisant le Web pour créer des « laboratoires virtuels ». Le Web rend en effet possible la création d'expérimentations sociologiques en ligne. L'expérience notoire en la matière est celle de Salganik et ses collaborateurs (Salganik et Watts, 2009), qui ont proposé à un échantillon de 2930 participants d'écouter et d'évaluer 48 chansons d'artistes inconnus. Seule une partie des groupes d'internautes constitués pour l'expérience étaient informés de la notation et du nombre de fois que d'autres membres de leur groupe avaient téléchargé chaque chanson. L'expérience avait pour objectif d'évaluer si l'influence sociale produit des effets, ou non, sur les préférences individuelles pour telle ou telle chanson. Les auteurs constatent dans cette expérience que les préférences musicales des individus ont été modifiées lorsqu'ils ont été exposés à de l'information sur les préférences des autres. Plus important encore, ils observent que l'étendue de l'influence sociale a des conséquences importantes pour les résultats collectifs qui émergent. Plus l'influence sociale est grande, plus les résultats collectifs deviennent inégaux et imprévisibles. La popularité des chansons varie lorsque les individus s'influencent les uns les autres et il devient alors plus difficile, à mesure que l'influence des uns sur les autres augmente, d'anticiper quelles chansons deviendront les plus populaires. De tels résultats auraient bien sûr été difficiles et laborieux à obtenir sur des petites populations ou des échantillons réduits. Les écosystèmes numériques et le dispositif d'expérimentation en conditions contrôlées qu'ils rendent possible permettraient donc de développer, en sociologie, des expérimentations aussi rigoureuses qu'en psychologie expérimentale (Watts, 2013). Plus encore, l'utilisation de sites d'externalisation ouverte (*crowdsourcing*) pour recruter et payer des sujets⁴ offre à la sociologie une situation analogue à la longue tradition du recrutement des étudiants privilégié par les sciences comportementales (Mason et Watts, 2009).

Quand les sciences sociales computationnelles revisitent les objets classiques de la sociologie

Les SSC se sont approprié les objets d'études classiques de la sociologie par l'analyse de réseaux tirés des écosystèmes numériques — des réseaux d'une nature différente et d'une taille largement supérieure à celle de ceux habituellement traités par les sociologues (Mercklé, 2016). La contribution des SSC se situe de prime abord sur le plan des mathématiques appliquées par la création de deux nouveaux modèles de graphe : d'abord, en 1998, la création par Watts et Strogatz d'un modèle de graphe de type réseaux « petit monde », dont les caractéristiques principales sont une faible distance moyenne entre toutes les paires de nœuds du réseau et des nœuds très connectés à leur

4. Pensons par exemple au service Mechanical Turk d'Amazon, qui vise à faire travailler une myriade d'êtres humains rémunérés à la tâche.

voisinage immédiat, ce qui les différencie des graphes aléatoires connus des mathématiciens jusqu'alors, dont l'ensemble des nœuds du réseau ont à peu près le même degré (Watts et Strogatz, 1998)⁵; puis en 1999, la création par Barabási et Albert d'un modèle de graphe dit à invariant d'échelle (*scale-free network*), dont la distribution des degrés suit une loi de puissance, c'est-à-dire quelques nœuds très fortement connectés, et un très grand nombre de nœuds très faiblement connectés (Barabási et Albert, 1999). Mais aussi sophistiqués soient-ils, les modèles de graphes développés par les SSC ne parviennent pas à mettre en évidence d'autres phénomènes que ceux déjà connus en sociologie. Les réseaux « petit monde » de Watts ont été découverts par Milgram dans les années 1960 au cours de sa célèbre expérience du « petit-monde », dont la vulgate a retenu l'énoncé (toujours discuté) que chaque personne est connectée à n'importe qui à travers cinq intermédiaires au maximum (Travers et Milgram, 1969). Les réseaux invariants d'échelle de Barabási et Albert (1999) peuvent être assimilés à l'*effet Matthew* de Merton (Merton, 1968), décrivant les mécanismes par lesquels les plus favorisés tendent à accroître leur avantage en captant la majorité des ressources⁶.

Il est généralement considéré que les recherches des SSC n'ont pas été suivies de transformation théorique majeure du point de vue de la sociologie générale et de la sociologie des réseaux en particulier. À la lecture des travaux de référence des SSC, on recense plutôt des contributions empiriques et méthodologiques qui revisitent, à l'appui de méthodes originales et de données nouvelles, des questions et des « découvertes » classiques de la sociologie. Ainsi, lorsque Pentland s'intéresse à la circulation des habitudes et des croyances qui se propagent au travers des réseaux (Pentland, 2015), il aborde une des questions premières de la sociologie, mais à la manière d'un physicien : les habitudes et les croyances sont pour lui des flux d'idées qui se traduisent dans des changements de comportements, en analogie avec les flux d'énergie qui génèrent des changements de mouvement (d'où la réactivation de la notion abandonnée par les sociologues de « physique sociale »).

Quantifiés à partir de données sociométriques (en particulier les fréquences des interactions entre individus d'une organisation), les flux d'idées correspondent à la proportion d'utilisateurs susceptibles d'adopter une nouvelle idée introduite dans le réseau social. Les flux d'idées sont modélisés comme un modèle d'influence qui tient compte de « la structure du réseau, de la force de l'influence sociale et de la susceptibilité des individus aux nouvelles idées » (Pentland, 2015)⁷ — autant d'éléments (structure, influence) qui servent à soutenir l'idée que l'intelligence est collective. Pentland montre que la production, l'échange et la diffusion des idées varient selon la composition des

5. Le degré d'un nœud correspond au nombre de liens adjacents à ce nœud.

6. Pour une bonne analyse de l'apport limité des SSC à l'analyse des réseaux sociaux, on peut consulter les cours de Laurent Beauguitte à l'ENSAE (Beauguitte, 2015).

7. Les flux d'idées circulent selon deux mécanismes, l'engagement et l'exploration. L'exploration désigne les moments où les personnes parlent à d'autres personnes plus ou moins éloignées de leur réseau de proximité. L'engagement est le processus d'appropriation des nouvelles idées collectées au cours de l'exploration.

réseaux dans lesquels les individus sont insérés. Ceux qui obtiennent le maximum de production, d'échange et de diffusion ne sont pas exclusivement en contact avec les individus qui leur ressemblent. Ils savent tirer profit d'une diversité optimale. Pentland montre aussi qu'il y a un maximum au-delà duquel le bénéfice de la diversité se dissipe⁸.

Demandez à Pentland pourquoi les flux d'idées circulent de cette façon, il vous répondra que la réponse existe déjà dans la documentation scientifique (Pentland, 2013). En effet, pour rendre compte de ses observations, Pentland s'appuie sur un article de 1999 de Kelley, chercheur en sciences de gestion, qui a étudié la réussite exceptionnelle de certains chercheurs des *Bell Labs* (les laboratoires de la société de téléphonie AT&T aux États-Unis). Alors que les *Bell Labs* recrutaient les meilleurs ingénieurs formés dans les universités américaines les plus prestigieuses, Kelley observe que seuls quelques-uns d'entre eux deviennent des contributeurs remarquables (des « stars performers »), et que la plupart n'apportent aucune contribution décisive (les « solids performers »). Plutôt solitaires, les « solids performers » accumulent du savoir de leur côté, alors que les « stars performers » développent de nombreux réseaux de relations diversifiés, ce qui les aide à adopter des points de vue variés grâce aux membres de leur réseau qui jouent eux-mêmes un ensemble plus divers de rôles de travail.

Dans son cours au Collège de France du 27 janvier 2017, Pierre-Michel Menger mobilise les résultats des travaux de Pentland comme « une manière de retourner en partie l'argument d'une individualisation égocentrée des talents [...], d'ajouter aussi une dimension de caractérisation à ceux qui dans une situation de compétition et de travail à forte intensité concurrentielle [...] se relie[n]t à leur environnement de manière productive, non pas pour le simple plaisir, mais un plaisir qui peut avoir un retour sur investissement en quelque sorte [...]. On n'est pas dans un monde d'altruisme pur et parfait où tout serait désintéressé au dernier degré » (Menger, 2014). Alors que l'interprétation de Menger cherche à comprendre les logiques de concurrence et de production des inégalités dans les transformations du travail, Pentland propose lui de se servir de ses résultats pour développer des outils de monitoring des échanges à partir d'un réseau de capteurs, permettant de retracer, dans les environnements de travail, qui parle à qui et comment, afin d'optimiser la performance des individus et des organisations (Pentland, 2012). Ainsi, quand Menger met côte à côte le vocabulaire de la physique sociale avec celui de la sociologie (« différences et inégalités ») dans le sous-titre de son cours — *Qu'est-ce que le talent ? Éléments de physique sociale des différences et des inégalités* —, c'est pour montrer tout le contraste entre l'innovation méthodologique de Pentland et l'objet de la discipline sociologique dont le projet fondamental porte sur l'analyse des sociétés humaines du point de vue des changements qui affectent leur organisation interne, des différences et des inégalités qui les ordonnent et des cadres de pensée qui les structurent (Karsenti et Lemieux, 2017).

Les SSC semblent néanmoins évoluer ces dernières années en cherchant à davantage collaborer avec la sociologie. Les travaux de Watts et de ses doctorants en sont un

8. Nous reprenons ici la synthèse que fait Pierre-Michel Menger (2014) du travail de Pentland.

bon exemple. Watts a toujours montré un intérêt plus grand pour cette discipline que la plupart des chercheurs des SSC⁹. Travaillant sur les réseaux « petit monde », il manifeste une certaine réserve quant à l'« intérêt social » de sa découverte : « the real issue is that there is a big difference between two people being connected by a short path and their being able to find it [...]. But even if it is true that everyone can be connected to everyone else in only six degrees of separation, so what? How far is six degrees anyway? From the point of view of getting a job, locating information, or getting yourself invited to a party, anyone more distant than a friend of a friend is for all intents and purposes, a stranger. So [...] anything more than two degrees might as well be a thousand » (Watts, 2004 mentionné par Beauguitte (2015)). Cette réserve n'exclut pas, selon Watts, un usage des SSC dans la perspective de produire des connaissances plus générales et fondamentales sur les phénomènes sociaux. Dans une allocution programmatique sur les SSC en 2013, Watts rappelle qu'en dépit de plusieurs milliers d'articles publiés dans le domaine des SSC, peu de progrès ont été réalisés relativement à la compréhension des systèmes financiers, l'évolution des organisations complexes, la dynamique des mouvements sociaux, etc. (Watts, 2013).

Selon Watts, la contribution des SSC à la sociologie passe d'abord par l'apprentissage d'un certain usage de la prédiction, distinct des analyses des mécanismes de causalité interprétables travaillés par les sciences sociales traditionnelles. Les limites théoriques de l'exactitude prédictive (*accuracy*) dans les systèmes sociaux complexes devraient selon lui être mieux caractérisées, ce qui permettrait d'établir des attentes à l'égard de ce qui pourrait être prédit ou expliqué. L'exactitude prédictive et l'interprétabilité devraient être reconnues comme des compléments, et non comme des substituts¹⁰, lors de l'évaluation des explications, ce qui mènerait « à des sciences sociales meilleures, plus reproductibles et plus utiles » (Hofman *et al.*, 2017). Même lorsqu'elles présentent un fort intérêt pour la sociologie, les SSC remettent en question le régime de scientificité propre aux sciences sociales, fondamentalement interprétatives (Passeron, 2006), en le faisant cohabiter avec les modalités d'évaluation des énoncés scientifiques des physiciens.

Une sociologie fongible dans les humanités numériques ?

Longtemps limitées aux domaines des arts, des lettres et des religions, les humanités numériques ambitionnent d'encapsuler toutes les sciences humaines et sociales (SHS) : « Pour nous, les *digital humanities* concernent l'ensemble des sciences humaines et

9. Formé comme physicien et anciennement sociologue des réseaux à Columbia University, Watts est actuellement chercheur principal au sein du *Computational social science group* à Microsoft Research.

10. La question de l'usage de l'exactitude prédictive se pose dans un contexte où les méthodes de l'apprentissage automatique (*machine learning*) se généralisent dans tous les domaines de recherche, notamment dans les sciences sociales. En matière de régression ou de classification, le *machine learning* implique une différence épistémique fondamentale avec les méthodes paramétriques classiques (les modèles économétriques de régression linéaire par exemple) : le modèle est une sortie du calcul, et non plus une entrée. Pour une confrontation empirique des usages des deux types d'approches, on peut consulter l'excellente analyse de Boelaert et Ollion (Boelaert et Ollion, à paraître).

sociales, des arts et des lettres», peut-on lire dans le *Manifeste des digital humanities* rédigé en 2010 à Paris, cosigné par plus de 250 chercheurs et 10 institutions (Dacos, 2011). L'ambition des humanités numériques est cependant différente de celle des SSC : c'est le numérique lui-même, et notamment le Web comme plateforme de développement de projets, qui apparaît comme un modèle et une ressource pour repenser les modes d'organisation et de structuration des données, de l'information et de la connaissance dans toutes les SHS :

Le numérique comme instrument de recherche ; le numérique comme outil de communication ; le numérique comme objet de recherche. C'est de ce complexe-là que les humanités numériques se saisissent et c'est pour cette raison qu'elles représentent bien plus qu'un mouvement de mode passager et superficiel, quoi qu'en disent les mauvaises langues ; un véritable mouvement de fond appelé à redéfinir l'ensemble des champs de la recherche en SHS (Dacos et Mounier, 2014).

L'idée de mimer la logique du Web, et les technologies informationnelles qui lui sont associées, vise à mettre en numérique les parties du monde et du patrimoine qui n'existent pas encore dans cet état. La tâche est immense et Google a ouvert la voie, en numérisant tous les jours une portion un peu plus grande de notre environnement physique et culturel. C'est dans ce même esprit qu'a été développé le projet ambitieux de la *Venice Time Machine* qui propose « une modélisation multidimensionnelle de Venise et de son empire méditerranéen. Son ambition consiste à rendre interopérables des données concernant l'histoire environnementale (évolution de la lagune), urbaine (morphogenèse de la ville), humaine (démographie et circulation) et culturelle (politique, commerce, évolution artistique) » (Kaplan, 2013).

Contrairement à la sociologie, les humanités ont un avantage positionnel concernant ce travail de numérisation. Dès les années 1940, elles ont développé des méthodes de traitement automatique de la langue puis, dans les années 1980, des pratiques d'encodage de l'information pour l'archivage des corpus. Plus récemment, elles se sont tournées vers les méthodes récentes d'extraction de connaissances (text mining) largement utilisées dans les sciences du Web et en intelligence artificielle. Elles ont donc fait du texte leur spécialité. Il est dès lors assez évident que les humanités se tournent vers les innovations qui permettent de numériser les textes, de les indexer et d'en extraire de l'information.

La situation est beaucoup moins évidente en sociologie où l'analyse textuelle n'apparaît pas dans le canon des méthodes d'enquête. Cointet et Parasio (Cointet et Parasio, à paraître) soulignent que les sociologues mobilisent diverses méthodes de l'analyse textuelle, comme la modélisation thématique (*topic modeling*), les analyses de réseaux, la lexicométrie, l'analyse de sentiments, les analyses stylistiques et de plongement de mots (*words embedding*). Ils identifient trois types d'usage de ces méthodes : la détection, pour thématiser la multiplicité des cadres interprétatifs liés à un évènement, une situation ou une pratique ; l'identification des stratégies des individus et des organisations au travers des textes qu'ils produisent (textes scientifiques, communication sur les réseaux sociaux et par voie de presse), par l'usage de méthodes d'intelli-

gence artificielle variées, en particulier la détection automatique d'entités nommées; enfin, la modélisation de « l'énonciation comme acte social », en identifiant « plusieurs catégories de termes — selon qu'ils désignent des agents, des actes, des scènes ou des motifs par exemple — entre lesquelles ils reconstituent les relations. Cela leur permet de mettre au jour non pas des ensembles thématiques ou des stratégies, mais un ensemble d'actions accomplies par le locuteur lorsqu'il prend la parole » (*ibid.*). Aussi fascinantes de telles analyses textuelles soient-elles sur le plan méthodologique, leur apport reste très limité au regard des connaissances qu'elles produisent, hormis pour l'analyse des stratégies d'acteur. De plus, l'usage de l'analyse textuelle dans l'enquête sociologique est débattu au sein même de la discipline. Lorsque les données sont collectées en ligne (sur des forums par exemple), les chercheurs ont très peu d'information sur les locuteurs. Ils se trouvent devant des matériaux abondants, mais dont l'épaisseur sociale présente de sérieuses lacunes (Beuscart, 2017). Si les données textuelles sont massives, le sociologue est contraint de déléguer à la machine le travail interprétatif et, partant, d'avoir recours à des méthodes qui font émerger automatiquement des catégories d'analyse, ou qui produisent une modélisation plus ou moins grossière de l'énonciation (Cointet et Parasie, à paraître).

Ce qui distingue *in fine* les humanités numériques de la sociologie, c'est que celles-là envisagent en grande partie le numérique comme un outil à leur service pour améliorer les connaissances des significations intrinsèques des documents, alors que celle-ci tente d'aborder le texte à partir de ses usages sociaux, et considère les énonciations comme une activité sociale. La sociologie ne pourra trouver de place confortable dans le grand chapiteau des humanités numériques¹¹ qu'à la condition d'y trouver autre chose qu'un « épistémocentrisme de la théorie herméneutique de la lecture [automatique de texte] » (Bourdieu, 1997).

IMAGINER LA SOCIOLOGIE NUMÉRIQUE ENTRE PROTECTION, CONSERVATION ET ADAPTATION

Entre l'émergence des SSC qui entendent se substituer à la sociologie et des humanités numériques qui proposent de l'intégrer dans un socle commun aux sciences humaines et sociales, plusieurs sociologues ont pris position relativement à ce que les transformations numériques entraînent pour la discipline. Bien avant les premiers ouvrages dont les titres font explicitement référence à l'avènement d'une sociologie numérique (Jessie *et al.*, 2016; Lupton, 2015; Marres, 2017; Orton-Johnson et Prior, 2013), les prises de position ont été multiples. Rendre compte de toutes les nuances de ces débats dépasse cet article. Nous nous limiterons donc ici à en proposer une schématisation selon trois types de réactions : la méfiance des revendications des SSC; l'intégration du numérique au niveau des objets de la sociologie tout en adoptant, sur le plan méthodologique, une posture prudente; l'adaptation épistémologique, le numérique étant

11. Notons que les travaux de Moretti apparaissent comme une source d'inspiration singulière qui permet de faire exister la sociologie dans les humanités numériques (Moretti, 2013).

vu comme l'opportunité de surmonter des problèmes durables de la discipline sur le plan empirique et théorique. C'est au milieu de ces réactions diverses que l'on commencera à entrevoir la possibilité d'une sociologie numérique loin d'être stabilisée.

Protection réflexive

Articulé comme une réponse à une menace venant de l'extérieur, et en particulier des SSC, ce premier type de réaction appelle à la mise en garde de la sociologie. C'est la position que l'on observe notamment chez Andrew Abbott, directeur de l'*American Journal of Sociology*, utilisateur critique des méthodes quantitatives et l'un des plus fins analystes des transformations actuelles de la discipline. Lors de la 37^e Conférence Marc Bloch dans le grand amphithéâtre de la Sorbonne en juin 2015, Abbott lance une attaque foudroyante contre les SSC :

Les imbéciles venus de la physique et de l'informatique qui ont imaginé pouvoir répondre à toutes les questions posées par les sciences sociales et les humanités avec leurs modèles de graphes exponentiels aléatoires et leurs téraoctets de corpus de textes vont finir dans le même dépotoir que les sociométristes et les physiciens sociaux des années 1930, les sociobiologistes des années 1960, les théoriciens des jeux des années 1950 et des années 1990, et ainsi de suite. Eux aussi sont des Turcs, et quand la météorologie intellectuelle tournera au mauvais temps, ils vont comme leurs prédécesseurs se retirer des portes de Vienne, redescendre à la hâte le Danube et passer les Portes de Fer avant que l'hiver n'arrive. Ils n'ont pas vraiment d'intérêt pour les questions importantes des humanités ou des sciences sociales. Ils sont une cause de contrariété, et même un peu plus que cela (Abbott, 2016, p. 580).

Si les physiciens du social ont le vent en poupe, et si Abbott les attaque avec autant de mépris, c'est parce qu'ils profitent de l'état de crise dans laquelle les possibilités d'enquêtes nouvelles liées aux écosystèmes numériques ont mis la sociologie. Les sociologues sont désormais contraints (ou quasi forcés, à lire Abbott entre les lignes) d'opérer des choix judicieux, de crainte que l'habitus intellectuel qui les anime disparaisse d'ici quelques dizaines d'années. Selon Abbott, l'une des grandes transformations contemporaines qui créent les conditions d'un changement et affectent les sciences sociales de l'intérieur est « la déconnexion sans cesse grandissante entre la sophistication des démarches empiriques, d'une part, et la simplicité, voire la naïveté, des raisonnements normatifs qui sous-tendent les sciences sociales, d'autre part » (Abbott, 2016, p. 580). Comme le souligne Etienne Ollion (2015), dans son dernier ouvrage, Abbott soutient même que l'actuelle multiplication de données transforme le savoir (Abbott, 2014). Partant de son expérience d'enseignant et de directeur de recherche, il soutient que les étudiants ont développé des compétences certaines pour aller chercher de l'information, mais que l'articulation entre questions de recherches et enquête est progressivement remplacée par un hyperempirisme qui néglige trop souvent les exigences de la construction de l'objet. Si la sociologie doit se transformer, ce n'est pas seulement du côté des méthodes numériques, mais en repensant le projet normatif autour duquel la discipline s'est constituée.

Intégration conservatrice

Un deuxième positionnement, moins protectionniste, mais relativement conservateur, prône une intégration du numérique sur le plan des objets de la sociologie, mais invite à une incorporation maîtrisée de nouvelles techniques numériques dans l'approche de l'enquête. C'est ce que Beuscart, Dagiral et Parasie laissent entendre dans un ouvrage de synthèse proposant le (re)positionnement d'Internet comme « objet total » de la discipline (Beuscart *et al.*, 2016). En examinant un vaste ensemble de travaux réalisés depuis 20 ans sur les manières dont Internet pénètre et bouleverse différentes sphères de la vie collective — sociabilité, politique, économie, travail, culture, etc. —, les auteurs définissent les contours d'une sociologie capable de rendre compte des spécificités des actions sociales en ligne, par l'enquête, alimentée par nombre d'acquis sociologiques portant sur les logiques d'action hors ligne. Devenu incontournable dans les existences quotidiennes des individus, Internet ne peut plus être abordé comme s'il s'agissait d'un domaine séparé de la vie sociale, car il reconfigure en grande partie les objets d'enquête. De plus, une connaissance générale des mécanismes sociaux est devenue indispensable si l'on veut comprendre le phénomène Internet en profondeur. Or, une sociologie qui intègre pleinement Internet ne signifie aucunement, au prétexte d'un renouveau méthodologique qui serait associé à une « sociologie numérique » en émergence, un abandon des méthodes éprouvées servant à la discipline pour produire des connaissances, telles que l'entretien, le questionnaire et l'analyse traditionnelle de corpus textuels. Les méthodes classiques s'adaptent au partage en ligne/hors ligne qui structure ces travaux et des techniques nouvelles sont intégrées au sein des approches d'enquête existantes, à condition de ne pas perturber les fondements épistémologiques de la discipline. Cette approche reste prudente à l'heure d'utiliser les traces numériques comme indicateurs des comportements hors ligne, les extrapolations depuis les observations en ligne étant découragées par les problèmes que posent les traces à l'analyse sociologique (représentativité difficile à établir, manque d'épaisseur sociologique des acteurs et des actions, difficulté d'isoler ce qui est propre au média de ce qui est propre au phénomène étudié, etc.). Malgré cette prudence, les traces numériques restent dans cette perspective un matériau précieux pour l'enquête et sont l'occasion d'un dialogue avec les sciences informatiques susceptible d'enrichir l'analyse sociologique (Beuscart, 2017).

La recherche récente de Parasie et Cointet (Parasie et Cointet, à paraître) sur la formation des publics dans le contexte de transformations liées à la numérisation dans le monde de la presse illustre bien cette approche prudente, quoiqu'innovante, des traces numériques. Prenant pour cas la plateforme cartographique d'exploration de données du crime du *Los Angeles Times*, les auteurs examinent la façon dont des internautes s'agrègent autour de ce dispositif pour s'informer, et se demandent si cet agrégat peut être considéré comme un « public ». Cette question sociologique classique — les ressorts de la formation des publics —, examinée dans le cadre des transformations numériques d'un domaine précis — la presse —, conduit les auteurs à mettre au point une méthode *ad hoc* d'analyse quantitative du texte. Dans une démarche non

intrusive et ascendante, la coexistence de trois modalités de « faire public », est examinée à partir d'un corpus de 28 364 commentaires d'internautes issus de la plateforme, liés à la totalité des 3153 victimes d'homicide rapportées sur une période de 6 ans (2010-2016). L'approche adoptée par les auteurs leur permet de coder les commentaires tout en conservant le contrôle sur le processus de construction des catégories d'analyse, afin d'identifier des types de discours visant à donner sens aux occurrences d'homicides. Si le manque d'informations sociodémographiques contraint l'analyse, d'autres informations sur les modes et contextes de prise de parole peuvent être utilisées, comme la distribution de celle-ci et son ancrage par quartier de la ville de Los Angeles. En incorporant les informations sociodémographiques des quartiers et en distinguant les locuteurs ancrés dans un seul quartier et ceux qui interviennent au sujet d'homicides dans plusieurs quartiers, les chercheurs redonnent aux locuteurs et à leurs situations une certaine épaisseur sociale. La méthode originale utilisée illustre bien les efforts réalisés pour concilier les approches algorithmiques et celles de la sociologie interprétative, sans s'en remettre aveuglément aux premières.

Adaptation refondatrice

Un troisième type de positionnement considère avec enthousiasme la traçabilité numérique comme l'occasion d'un renouveau majeur pour la sociologie. En 2007, Mike Savage et Roger Burrows annoncent une crise à venir pour la sociologie empirique, liée à l'obsolescence proche des méthodes de collecte de données qui ont fait la spécificité de la discipline (les enquêtes de population, les suivis longitudinaux, les entretiens, les reconstitutions de réseaux sociaux, etc.) — un diagnostic que les deux sociologues ont confirmé par la suite avec l'avènement du *big data* (Burrows et Savage, 2014). Les méthodes d'enquêtes coûteuses et souvent laborieuses des sociologues, quantitatives ou qualitatives, seraient progressivement « déclassées » par l'accès supposément facilité à un nombre important de données de transactions sociales enregistrées au quotidien par les acteurs publics et, de plus en plus, par les acteurs de l'économie marchande et d'Internet :

In an age of knowing capitalism, sociologists have not adequately thought about the challenges posed to their expertise by the proliferation of « social » transactional data which are now routinely collected, processed and analysed by a wide variety of private and public institutions (Savage et Burrows, 2007).

Pour survivre à un contexte de concurrence accrue entre différents types de savoir sur le social, la sociologie doit, selon eux, s'adapter aux évolutions numériques en marche. La même année, Roger Burrows, cette fois avec David Beer, publie un article (Beer et Burrows, 2007) sur les orientations que pourrait prendre cette adaptation. Ils proposent de passer d'une sociologie du Web à une sociologie qui, pour ainsi dire, se love dans le Web. Selon les auteurs, les transformations numériques ne doivent plus uniquement intéresser les sociologues des technologies de l'information et des communications, mais elles interpellent l'ensemble de la sociologie. Or, rompant avec les

approches d'une sociologie d'Internet, les dispositifs du Web, dont, par exemple, les réseaux sociaux, doivent dépasser le statut « d'objet de recherche » pour devenir des instruments de recherche à part entière. Beer et Burrows voient les réseaux sociaux comme des archives ouvertes sur la vie quotidienne des individus, une sorte de « codification continue des habitus », par l'enregistrement systématique des préférences, des choix, des points de vue, des caractéristiques physiques, des situations géographiques, des parcours éducatifs, des situations professionnelles, etc. Cependant, si en 2007 les données du Web paraissent encore facilement accessibles, près de dix ans plus tard, ce n'est plus le cas.

Dans la continuité du programme ébauché en 2007 par Beer et Burrows, un certain nombre de centres de recherche universitaires, croisant à proportion variable études des médias, sociologie, informatique et design de l'information, ont vu le jour à travers le monde. Parmi ces expériences, celle de la *Digital Methods Initiative* (DMI) de l'Université d'Amsterdam, créée en 2007 par Richard Rogers, et celle du médialab de Sciences Po, fondé en 2009 par Bruno Latour, nous intéressent particulièrement par le caractère audacieux de leurs propositions méthodologiques et théoriques.

L'approche développée par Richard Rogers et ses collègues au sein de la DMI a prôné une vision dans laquelle les approches de numérisation des méthodes classiques ou des données existantes ne sont plus satisfaisantes. Selon eux, l'enjeu est de développer une épistémologie adaptée à l'utilisation du Web comme ressource. Le cœur du programme consiste en la conception de méthodes « nativement numériques » permettant de détourner (« to repurpose ») de façon opportuniste les dispositifs en ligne afin d'exploiter leurs données, elles aussi « nativement numériques », aux fins de la recherche sociologique. Dans les termes de Rogers, il s'agit de faire émerger :

a new era in Internet research, one that no longer concerns itself with the divide between the real and the virtual. It concerns a shift in the kinds of questions put to the study of the Internet. The Internet is employed as a site of research for far more than just online culture. The issue no longer is how much of society and culture is online, but rather how to diagnose cultural change and societal conditions by means of the Internet. The conceptual point of departure is the recognition that the Internet is not only an object of study but also a source (Rogers, 2013, p. 21).

Dans cette perspective, les méthodes numériques visent à se servir des traces numériques pour étudier les phénomènes en ligne, mais aussi la culture et la société en général. Les dispositifs du Web sont « repurposed », détournés, fonctionnant comme autant de *proxies*, c'est-à-dire comme indicateur par procuration, pour l'étude des phénomènes qui les traversent, mais qui s'étendent au-delà d'eux. Mue par cette ambition, la DMI a produit depuis sa création toute une panoplie d'outils *open source*¹² conçus pour « détourner » différentes plateformes (Rieder et Röhle, 2012) comme Facebook (Rieder, 2013), Twitter (Borra et Rieder, 2014), ou YouTube (Rieder *et al.*, 2018), par exemple pour l'étude de la formation des problèmes (« issues ») et de leurs

12. <https://wiki.digitalmethods.net/Dmi/ToolDatabase>

publics. C'est aussi autour de la cartographie numérique des controverses dans le cadre du projet européen MACOSPOL¹³ qu'a pris forme la collaboration entre la DMI et le médialab de Sciences Po.

Avec la création du médialab, Bruno Latour et ses collègues ont développé une vision théorique particulière de la sociologie numérique qui reflète bien les espoirs suscités par les développements dans les domaines des humanités numériques et des SSC, annonçant l'avènement d'une « science sociale de troisième génération » (Boullier, 2017). Cette vision est présentée dans un texte notoire (Latour *et al.*, 2013). Les auteurs y décrivent comment, grâce aux traces numériques massives et à la navigation des réseaux de relations entre des « profils », les dichotomies classiques qui ont guidé une bonne partie des discussions sociologiques entre niveau individuel et niveau de la structure, ainsi qu'entre approches qualitative et quantitative, peuvent être dépassées par une approche à un seul niveau. Les traces numériques ouvriraient la voie pour une articulation inédite, sur le plan de la recherche empirique, entre la sociologie de l'acteur-réseau développée par Latour et ses collègues au Centre de sociologie de l'innovation (CSI) depuis les années 1980 (Callon, 2013), et une réinterprétation de la théorie sociale de Gabriel Tarde développée à la fin du XIX^e (Latour, 2010b).

Misant sur la navigation des traces numériques comme moyen de suivre les associations d'entités, et par là la société en train de se faire, le médialab et la DMI ont collaboré dans une série de projets très ambitieux. Au cœur de ceux-ci se trouvait la cartographie des controverses (Latour, 2010a; Venturini, 2010, 2012). Dans cette perspective, le projet Contropedia¹⁴ s'est servi de Wikipédia comme *proxy* pour l'étude des controverses sociotechniques. Visant à développer un dispositif de détection et exploration des controverses et à produire des énoncés valables hors ligne, ce projet a cherché à détourner des données telles que le volume des historiques d'édition et la profondeur des discussions dans les pages de discussion associées pour opérationnaliser le concept de « controversialité » (Borra *et al.*, 2014). En cours de recherche, les responsables du projet réalisent cependant l'effet spécifique de la nature du proxy — Wikipédia comme espace de discussion très structuré par une culture et des normes qui lui sont propres — sur l'observation des controverses. Si les tensions propres aux débats sur un sujet discuté en société apparaissent dans l'encyclopédie en ligne, il est cependant difficile d'affirmer que ce que l'on observe reflète la dynamique hors ligne (Venturini *et al.*, 2018). Contraints par cet obstacle, les chercheurs ont été amenés à proposer une approche moins audacieuse de *repurposing* qui limite la portée des conclusions tirées de l'analyse des traces numériques (Weltevrede et Borra, 2016).

C'est à l'aune de ce type d'expériences que se pose la question épistémologique du *repurposing* en sociologie numérique (Marres, 2017, chap. 3; Venturini *et al.*, 2018), dans ce que l'on peut identifier comme une *réflexion proxilogique*, autrement dit un questionnement sur le rapport d'adéquation entre l'objet d'étude et les sources des

13. www.mappingcontroversies.net/

14. <http://contropedia.net/>

traces numériques, appréhendées comme *proxies*¹⁵. Cette réflexion devient une partie essentielle du travail en sociologie numérique, consistant à répondre à la question de la portée des énoncés sociologiques produits au travers de ces proxies numériques. En ceci, toute sociologie numérique suppose une sociologie du numérique et une étude des médias dont on se sert dans l'enquête, au point que la distinction entre « numérique/Internet comme objet » et « numérique/Internet comme moyen » (Hargittai et Sandvig, 2015) devient en quelque sorte obsolète.

IMAGINER LE MOUVEMENT OPEN TRACES

Aussi différentes que soient les manières d'envisager les études numériques, tous les chercheurs s'accordent sur le problème général de l'accès aux traces. Imaginer la sociologie numérique, c'est aussi imaginer les conditions d'ouverture des écosystèmes numériques pour la recherche publique. Trois types de plateformes fournissent aujourd'hui ce type de données susceptibles d'alimenter la sociologie numérique : les plateformes publiques, les plateformes éditoriales et les plateformes privées, en particulier les plateformes du Web. L'accès aux traces numériques pose des problèmes spécifiques qui se distinguent en partie de ceux posés par les mouvements *Open data* et *Open process* qui composent le mouvement plus général de l'*Open science*. Pour marquer cette différence, nous appelons dans ce texte *Open traces* le mouvement émergeant autour de l'ouverture des diverses traces numériques

Open traces n'est pas open data

Commençons par l'ouverture des données des plateformes publiques administratives, pour lesquelles le cadre juridique a beaucoup évolué ces dernières années dans de nombreux pays. Force est de constater que si les conditions d'accessibilité aux données de ces plateformes se sont largement améliorées avec l'*Open data*, elles correspondent encore mal aux exigences de qualité de la recherche en sociologie. En effet, si la démarche d'ouverture est déjà largement engagée, avec la publication en ligne de très nombreux jeux de données par les gouvernements et les entreprises qui assurent des missions de service public, « on trouve parfois des données à la provenance comme à la qualité douteuses [...]. Les données produites par des administrations enjointes par les pouvoirs publics à "l'ouverture" font que les bases de différents services sont rapidement agrégées afin de répondre à cette commande politique. Cela se fait souvent sans les habituelles procédures de contrôle et d'harmonisation que mettent en place les chercheurs dans leurs travaux, et souvent sans même une information sur les conditions de la production des données » (Ollion, 2015).

Ainsi, le portail data.gouv.fr ne permet pas de documenter les *datasets* issus de la recherche selon les standards internationaux nécessaires au monde académique,

15. Dans un texte récent, Venturini et ses collègues ont proposé une série de questions servant d'appui à la réflexion sur le recours aux plateformes comme proxies en méthodes numériques (Venturini *et al.*, 2018).

notamment la participation des scientifiques à la sélection des données collectées. La sociologie numérique exige une granularité des données assez fine qui augmente le risque, même si elles sont rendues anonymes, d'une « réidentification » des personnes en croisant les variables de la base avec d'autres jeux de données.

Lorsque le risque de réidentification ne peut être suffisamment réduit, d'autres possibilités que les données accessibles en ligne sur les portails des gouvernements peuvent être envisagées. En France, c'est ce qu'imaginent Antoine Bozio, directeur de l'Institut des politiques publiques (IPP), et Pierre-Yves Geoffard, directeur de l'École d'économie de Paris, dans leur rapport sur l'accès des chercheurs aux données administratives, remis en 2017 à Axelle Lemaire, secrétaire d'État auprès du ministre de l'Économie et des Finances chargée de l'industrie, du numérique et de l'innovation (Bozio et Geoffard, 2017). Dans ce rapport, préparé à l'occasion de l'article 36 de la loi pour une République numérique visant un nouveau cadre d'accès des chercheurs aux données publiques dont la publication est restreinte (données personnelles, secret professionnel, etc.), ils soulignent la nécessité de pouvoir exploiter des données administratives pour la recherche, mais démontrent aussi que l'accès pratique à ces données reste difficile, préconisant alors la création d'une instance de concertation pour unifier les procédures et favoriser les échanges entre administration et recherche.

Les recommandations esquissées dans le rapport — notamment la création d'un vade-mecum rappelant l'état du droit applicable à destination des chercheurs et des producteurs de données et plusieurs modèles économiques pour minimiser les coûts d'accès à ces données — montrent bien que les questions de l'accès aux données administratives pour la recherche soulèvent des problèmes spécifiques qui ne correspondent pas à ceux du mouvement *Open data* tel que porté par le gouvernement. Il s'agit d'ailleurs de poser les jalons d'un mouvement d'ouvertures des *traces administratives*, car ce sont ces traces que les chercheurs ont besoin de collecter.

Le cas de la Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS), dont le système de gestion enregistre chaque dépense de soins pour plus de 60 millions d'assurés vivant en France, met en lumière cette différence entre les mouvements *Open data* et *Open traces*. Les jeux de données du CNAMTS sur le portail data.gouv.fr ne permettent pas l'analyse des parcours de soins, l'étude des inégalités de santé et de l'accès aux soins. Pour ce type d'étude, les chercheurs doivent se tourner vers le Système national d'information interrégime de l'Assurance Maladie (Sniiram), qui a été conçu de manière à être accessible à des utilisateurs externes à l'Assurance Maladie, notamment les chercheurs, au sein de l'Institut des données de santé (IDS) (Caillol, 2015). Les données sont regroupées par la CNAMTS et sont restituées aux organismes autorisés par un arrêté, sous forme de bases thématiques agrégées ou d'un échantillon concernant les données individuelles. Comme le souligne Caillol (2015 : 62), rares sont les chercheurs à être autorisés à accéder à la base complète, mais les organismes de recherche à but non lucratif peuvent être destinataires d'une extraction *ad hoc* des données du Sniiram, dans le cadre d'une procédure d'approbation par l'Institut des données de santé et d'autorisation de la Commission Nationale de l'Informatique et des Libertés (CNIL).

Le cas du CNAMTS le montre bien : rendre accessibles les données d'une administration nécessite un accompagnement spécifique des chercheurs pour résoudre les problèmes techniques liés à l'interopérabilité avec les outils internes, apporter les outils informatiques et de traitement des données adaptés à un usage en dehors de l'institution et fournir une documentation et un accompagnement (Caillol, 2015). Le CNAMTS est plus une exception que la règle. Pour chaque secteur public, il reste à imaginer les conditions de possibilité pour que la recherche et l'administration puissent coconstruire ensemble les infrastructures de stockage distribuées en réseaux.

Open traces n'est pas Open access

Concernant les plateformes éditoriales, l'enjeu est de donner l'accès aux textes et données inclus ou associés aux écrits scientifiques en autorisant et en facilitant l'usage des techniques d'analyse automatisée, Text & Data Mining (TDM), visant à analyser des textes et des données sous forme numérique (Battisti et Schöpfel, 2017). Une part importante de la sociologie numérique tient à la possibilité d'exploiter ce type de plateforme avec des dispositifs logiciels adaptés à la nature et à la masse des données numériques qu'elles enregistrent. Or, en France, les grands éditeurs qui détiennent la majeure partie des publications scientifiques peuvent proscrire, par des solutions contractuelles, la fouille de textes et de données aux chercheurs, même si les abonnés disposent par ailleurs d'un accès légal à l'ensemble des publications scientifiques comprises dans les bases de données. Cette interdiction se réfère notamment au droit *sui generis* des bases de données qui protège les producteurs de bases de données qui ont pris le risque de l'investissement dans l'établissement de la base.

Depuis les débats de la consultation publique autour de la loi pour une République numérique (en particulier sur l'article 38), il a été imaginé que les chercheurs puissent utiliser les techniques de TDM en vue de l'analyse de données massives de corpus scientifique, dans le cadre de projets à but non lucratif. Mais ce qui a été envisagé introduit une nouvelle exception dans la loi française en dehors du cadre européen actuel, un niveau où les lobbies des grandes plateformes éditoriales maintiennent la pression, d'où l'avis négatif rendu par le Conseil d'État sur l'article 38. Quand bien même le décret relatif à cet article serait publié, le problème de l'accès aux données des plateformes éditoriales pour la sociologie numérique ne serait que partiellement résolu.

L'accessibilité des données des plateformes n'est pas suffisante pour conduire des recherches en sociologie des sciences et des techniques. Si grâce aux plateformes éditoriales, la scientométrie a l'énorme avantage de travailler sur un ensemble de données normalisées (références aux auteurs, aux articles, aux revues, aux institutions et aux mots clés eux-mêmes), les chercheurs ont besoin d'infrastructures *transplateformes* qui permettent de suivre — et donc de *retracer* — les parcours d'innovation par-delà les laboratoires. Pour ce faire, il leur faut relier les données des plateformes éditoriales aux données qui décrivent ce qui se passe en amont et en aval de la production textuelle de la recherche.

Le projet *Research infrastructure for science and innovation studies* (RISIS), coordonné par Philippe Larédo, montre bien l'ampleur de cette tâche. Porté par un collectif

de chercheurs issus de 13 institutions partenaires à travers 10 pays européens, RISIS rassemble des ressources de nature différente provenant d'une multitude de plateformes existantes relatives aux publications scientifiques, aux performances académiques, aux financements de la recherche, aux enquêtes sur la mobilité européenne, aux dynamiques de l'innovation des entreprises, aux entreprises européennes de taille moyenne à croissance rapide, aux établissements d'enseignement supérieur européens et aux rapports d'évaluation des politiques publiques. L'objectif est de faire converger ces sources éparses par l'harmonisation, la construction d'ontologies communes, etc.

La création de ce type d'infrastructure dépasse le cadre du mouvement *Open Access stricto sensu*. Il s'agit de rassembler les traces produites au gré de l'activité des scientifiques, mais aussi des différents acteurs qui financent, évaluent et commercialisent l'innovation.

Open traces, ou la question nouvelle de l'accès aux données du Web

La question de l'ouverture des plateformes des données privées a été posée plus tardivement. Par conséquent, le cadre juridique est encore embryonnaire, mais les débats sont riches et les idées ne manquent pas. En France, c'est autour de la notion de « données d'intérêt général (DIG) » (Pailhes, 2018) que l'accès aux données des entreprises privées est imaginé. Dans un rapport déposé en 2015 dans le cadre de la préparation de la loi pour une République numérique, l'Inspection générale des finances (IGF), le Conseil général de l'économie, de l'industrie, de l'énergie et des technologies (CGEIET) et des membres du Conseil d'État s'intéressent aux organismes privés détenant des données dont la publication peut se justifier en raison de leur intérêt public et, partant, pour la recherche publique (CGEIET et IGF, 2015). Ce projet prend pour point de départ l'analyse économique de Louis-David Benyayer et Simon Chignard (2015), qui ont identifié trois facettes de la valeur des données : quand elles sont revendues par ceux qui les collectent, les produisent ou les enrichissent, les données prennent une forme de matière première ; quand elles sont utilisées en interne dans les entreprises, pour réduire les coûts, optimiser la production et propulser l'innovation, elles prennent une forme de levier ; enfin, quand elles servent à la défense et la conquête d'une position concurrentielle, elles prennent une valeur d'actif stratégique.

Partant de ce constat, le rapport propose de réguler le marché des données privées en imposant « l'ouverture de certaines données détenues par des personnes privées [qui] pourrait concourir à l'intérêt général » (*ibid.*, p. 2). La mission imagine par exemple de permettre l'accès de la statistique publique aux bases de données privées s'il y a motif d'intérêt général, tout en renforçant les garanties pour les personnes concernées ; ou bien encore, de ne pas imposer dans certains cas de figure la gratuité de l'accès, mais de réguler la rémunération perçue par les détenteurs des données, ce qui permettrait d'empêcher les plateformes de pratiquer des prix prohibitifs pour la recherche publique. Toute la difficulté d'imposer l'accès tient à la nécessité de maintenir une concurrence loyale entre les entreprises qui ouvrent et celles qui ferment l'accès à leurs données. Le rapport conclut que la grande diversité des secteurs et des données

concernés rend impossible la mise en place d'un régime juridique unique des données d'intérêt général. La mission imagine une approche sectorielle dans la démarche d'ouverture de ces données privées, ce qui signifie que le débat sur les conditions d'accès aux données privées reste encore à imaginer et dépend de la proactivité des chercheurs eux-mêmes.

Pour le secteur des plateformes du Web, les premières initiatives émergent du côté de la recherche. Les restrictions d'accès aux données sur les interfaces de programmation applicative¹⁶ (API), ces plateformes soulèvent régulièrement des indignations publiques de la part des chercheurs. Par exemple, dans une lettre ouverte adressée à Facebook, après la quasi-fermeture de son API due à l'affaire Cambridge Analytica, Axel Bruns et ses collaborateurs (Bruns, 2018) réclament des politiques simples d'accès aux données pour la recherche et demandent d'accepter l'entrepôt des données de la plateforme à des fins de partage, ainsi qu'un engagement ouvert et transparent avec la communauté des chercheurs. Il arrive que les plateformes rendent leurs données disponibles au moyen des appels à projets mettant en concurrence des équipes de recherche (par exemple Twitter en 2013, LinkedIn et Facebook récemment), mais elles en limitent l'accès à une poignée d'individus dont elles estiment les questionnements pertinents. Les données des plateformes restent donc des actifs stratégiques, même lorsqu'il s'agit de les partager avec la recherche publique.

Dans la communauté des chercheurs, il est désormais admis que les administrateurs des plateformes ne peuvent décider seuls du partage des traces produites. Sur quel principe peut-on en exiger l'accès? Les plateformes du Web peuvent revendiquer un certain degré de propriété des données, mais ces données appartiennent aussi aux utilisateurs des applications en ligne. En effet, depuis la mise en application du règlement général sur la protection des données (RGPD), le renforcement du droit à l'oubli et la portabilité des données, les utilisateurs ont complètement accès à leurs données. Le texte empêche dans une certaine mesure l'appropriation unique des données par les plateformes.

À qui appartiennent ces données? La controverse est toujours ouverte et convoque deux positions différentes. D'un côté, une conception libérale développée par Lanier (Lanier, 2014), et récemment reprise en France par le think tank GénérationLibre (2018), conçoit la monétisation des données personnelles qui doivent être protégées à travers un droit de propriété privée dévolu aux individus. D'un autre côté, l'accent est mis sur la dimension collective des données personnelles et la nécessité de penser leur protection à travers le prisme du droit social (Cardon et Casilli, 2015 ; Maurel et Aufrère, 2018)¹⁷.

Par-delà l'accessibilité des données du Web, il convient aussi de mettre en conformité le secteur de la recherche publique avec le règlement général sur la protection des

16. Les API des plateformes sont les façades qui servent à des programmes informatiques à interroger les plateformes pour collecter les données.

17. Nous devons cette cartographie de la controverse à Lionel Maurel. Voir les nombreux billets sur son blogue S.I.Lex pour comprendre plus en détail ce débat : <https://scinfolex.com/>

données pour ce qui relève de la protection des personnes concernées. Les modalités d'encadrement des traitements du Web pour la recherche publique présentent des risques spécifiques sur ce sujet : dans la plupart des cas, le fondement légal du consentement des personnes concernées, la première base du RGPD, n'est pas l'appui légal approprié pour une recherche scientifique numérique. Pour celle-ci, le RGPD prévoit une série de dérogations spécifiques dans son article 5 qui édicte les principes relatifs au traitement des données à caractère personnel. Il reste à déterminer la portée exacte de ces dérogations en ce qui concerne l'accès aux données du Web pour la recherche publique. C'est dans cette perspective que la Commission Nationale de l'Informatique et des Libertés devra travailler à élaborer les règles relatives à la réutilisation de données publiquement accessibles « en ligne », en particulier sur des services de réseaux sociaux, à des fins de recherche scientifique publique.

Comme le laisse supposer le cas français, les solutions sur les conditions de l'accès aux données du Web sont donc émergentes. Elles se trouvent à la croisée des chemins, entre le droit des données personnelles, le droit de propriété, le droit à la concurrence et le droit social.

LES CONTRIBUTIONS DES ARTICLES DE CE NUMÉRO

Les sept contributions que nous réunissons dans ce numéro offrent des entrées de problématisation diverses à l'heure d'imaginer une sociologie numérique. Nous les avons agencées dans un parcours de réflexion qui nous permet d'explorer les manières dont la sociologie est affectée par les transformations numériques, tant du fait de la complexification du déploiement des relations sociales dans les écosystèmes numériques, qu'en raison des défis méthodologiques et épistémologiques que présentent les traces numériques.

Nouveaux agencements algorithmiques : sujets humains dans le monde de leurs artifices

Un premier ensemble d'articles s'intéresse à la manière dont le déploiement d'écosystèmes numériques confronte la sociologie à la question du « sujet ».

Francis Jauréguiberry explore l'agencement entre l'individu et les systèmes algorithmiques de recommandation des plateformes, les applications d'assistance personnelle et les pratiques d'autosuiivi (*self-tracking*) au moyen des divers objets connectés qui font fonctionner la machine de profilage du big data. Au croisement d'une sociologie de la modernité et d'une sociologie des usages des technologies de l'information et de la communication, Jauréguiberry interroge ce qu'il advient de la confrontation de l'individu avec l'image instrumentale de soi que produisent ces systèmes, ce qu'on peut rapprocher du « *doppelgänger* algorithmique » (Harcourt, 2015) ou du « data double » (Haggerty & et Ericson, 2000). L'auteur propose une réponse optimiste et provocatrice, en laissant entendre que cette confrontation peut conduire le sujet à un ressaisissement de soi dans sa quête de cohérence individuelle. Cherchant à nuancer les propositions critiques qui se bornent à indiquer les dérives résultant de la prévision

algorithmique, dont la perte supposée d'autonomie individuelle et collective, Jauréguiberry souligne l'importance d'un processus de prise de conscience, ainsi que la nécessité de maîtriser les flux de données. La diversité des stratégies raisonnées de déconnexion volontaire, et les formes diverses de résilience face au big data, laissent supposer l'existence d'un sujet capable de dédoublement réflexif, nonobstant la réalité des inégalités qui structurent les situations individuelles.

Olivier Glassey explore un autre agencement actancier humain/machine algorithmique, qui résulte de notre interaction avec les socialbots, ou robots sociaux, programmes informatiques qui peuplent désormais les espaces numériques en ligne sous des formes diverses — par exemple des profils d'utilisateurs ou agents conversationnels algorithmiques sur les plateformes de réseaux sociaux. Il s'agit là d'un phénomène dont la portée suscite des inquiétudes politiques — comme on l'a vu avec la mobilisation de socialbots sur Twitter, dans le cadre de la campagne pour les présidentielles de 2016 aux États-Unis — et épistémologiques — lorsqu'il s'agit de penser la nature « sociale » des espaces numériques, où humains et bots interagissent dans des situations difficiles à démêler. Dans la continuité des travaux sur le rapport de l'homme à son environnement technique, Glassey analyse les discours interprétatifs et réflexifs que suscite l'organisation de la vie numérique avec ces entités logicielles anthropomorphisées, ce qui le conduit à s'interroger sur ce qui fait le propre du comportement humain dans la sociabilité numérique.

L'épistémologie des traces numériques : héritages, tensions, collaborations, et adaptations entre sociologie et sciences sociales computationnelles

Un second ensemble d'articles pose la question de l'épistémologie de la sociologie numérique, confrontée sur son terrain à l'avancée des physiciens et des informaticiens. Les auteurs analysent les héritages et les tensions épistémiques — ainsi que les possibilités d'articulation — entre l'épistémologie de la sociologie interprétative et celle des sciences sociales computationnelles (SSC).

Boris Beade part de l'idée de crise de la sociologie avancée par Savage et Burrows (Savage et Burrows, 2007) il y a plus d'une décennie. Si la sociologie est entrée en crise, explique-t-il, celle-ci relève moins d'un problème d'empirie lié à un déclasserement méthodologique, que d'un souci épistémologique lié au resurgissement, par le big data et sous l'égide de la physique sociale, du projet positiviste duquel elle a historiquement pris ses distances. Pour étayer cette hypothèse, Beade examine une série d'enjeux qui traversent la sociologie quant à la façon dont la traçabilité numérique généralisée affecte et complexifie les relations sociales, leur observation et la production de connaissance et de sens à partir d'elles. La traçabilité qui résulte de la médiation numérique est marquée par un phénomène d'hypercentralité : une poignée de grandes entreprises du numérique disposent d'une capacité hors norme de contrôle des médiations (Beade, 2014). Les acteurs qui maîtrisaient les médiations antérieures perdent leur position privilégiée. C'est le cas des sociologues académiques envers des physiciens et des informaticiens qui, compétents pour la maîtrise des données massives, sont

désormais sollicités pour résoudre des questions sociales. On assiste alors, selon Beaude, à la résurgence de tensions épistémologiques qui ont historiquement marqué les partages classiques, d'une part entre sciences de la nature et sciences de la société, et d'autre part entre sciences de la société interprétatives et sciences de la société positivistes. Le problème, pour la sociologie, ne serait pas tant celui de l'accès aux traces numériques, que celui des difficultés à s'y fier pour produire des connaissances sur le monde social, sans pour autant renoncer au prisme interprétatif qui lui est propre. Une telle ambition implique des collaborations d'un nouveau genre, entre des chercheurs qui maîtrisent la collecte et le traitement des traces numériques, et ceux qui en maîtrisent les significations potentielles.

Dominique Boullier aborde quant à lui ces questions épistémologiques en examinant la manière dont les SSC adoptent des « points de vue » du social dans leur approche des traces numériques massives. Or, propose-t-il, ces points de vue impliquent des manières spécifiques d'accorder des capacités d'agir à des entités sociales distinctes héritées de la tradition sociologique. En prenant appui sur une analyse historique de l'instrumentation en sciences sociales (Boullier, 2015a, 2015b), l'auteur propose également de distinguer trois entités sociales alternativement privilégiées dans les recherches : la focale sur la société et les effets des structures, associée aux recensements exhaustifs ; la focale sur les individus et leurs préférences, associée aux sondages représentatifs ; et la focale sur les répliques et leurs effets d'émergence et de propagation, associée à la traçabilité numérique (Boullier, 2017). Par une analyse de la façon dont ces trois points de vue sont mobilisés d'abord dans différents courants de la sociologie classique et contemporaine, puis dans un vaste corpus de travaux en SSC sur Twitter, Boullier montre que les focales sur les structures sociales ou les préférences individuelles, privilégiées dans la tradition sociologique, prédominent toujours dans les façons dont les informaticiens cherchent à rendre compte des phénomènes observés sur Twitter. La focale sur les entités circulantes, sur les répliques et sur leurs effets d'émergence propre d'une approche tardienne du social restant rare, les renouvellements méthodologiques liés à l'exploitation des traces numériques de Twitter en SSC n'impliquent pas forcément un renouvellement épistémologique en termes des points de vue sur le monde social.

Francis Chateauraynaud et David Chavalarias abordent ces questions épistémologiques en examinant, par un retour d'expérience, les défis que pose l'hybridation entre le mode interprétatif de raisonnement de la sociologie et le mode de raisonnement axiomatique propre aux SSC. L'enjeu, pour eux, est de réussir à articuler une approche instrumentale nouvelle à même d'offrir des prises critiques (Bessy et Chateauraynaud, 1995) sur la manière dont les processus sociaux — et notamment, ici, les controverses — se déploient dans les espaces numériques. Twitter est à nouveau l'exemple privilégié, en ce qu'il constitue un espace central dans la vie des controverses dont il convient d'analyser les dynamiques. Cette perspective de recherche met en lumière les tensions fondamentales qui ont historiquement marqué les efforts de formalisation en sciences sociales, entre d'une part l'épistémologie de la sociologie interprétative, fondée sur

la narration et la description, et d'autre part celle des sciences sociales computationnelles, fondée sur des approches formelles du code et de la modélisation mathématique. Contre ceux qui considèrent ces approches comme irréconciliables, les auteurs défendent le projet d'une sociologie numérique qui les rend complémentaires. Les auteurs entendent alors dessiner l'espace d'une épistémologie qui suppose un nouveau régime de preuve, permettant de surmonter, par des opérations interprétatives, les incomplétudes des systèmes formels de mesure, les incertitudes des conventions qui donnent signification aux observations et les irréductibilités des expériences dans lesquelles elles prennent sens.

Nous concluons cette section avec la traduction d'un article de Duncan Watts, « Common Sense and Sociological Explanations », initialement publié dans l'*American Journal of Sociology* (Watts, 2014). L'article aborde ce qui serait la condition épistémologique d'une sociologie renonçant à la production de sens au moyen d'explications interprétatives, fondées sur le sens commun et sans validité causale, pour se concentrer sur des explications scientifiquement valables en vertu de leur force prédictive. Ceci devient possible, selon Watts, en concevant la prédiction de manière ample, selon la définition la plus large de la validation croisée, comme un mécanisme qui permet à la fois de faire des prédictions probabilistes et des prédictions sur des faits stylisés ou des modèles de résultats. Accepter cette définition ample implique pourtant d'abandonner trois idées qui résument, selon Watts, la manière étroite et erronée dont les sociologues ont pensé la prédiction, à savoir : 1) que la prédiction est forcément déterministe ; 2) qu'elle concerne nécessairement l'avenir ; et 3) qu'elle ne peut être faite qu'au sujet d'événements ou de résultats précis. Fondant cette proposition sur une critique incisive des explications interprétatives qui sont au cœur de la tradition sociologique, le texte n'a pas été sans susciter de réponses. Catherine Turco et Ezra Zuckerman, deux sociologues du MIT (Turco et Zuckerman, 2017), ont par exemple reproché à Watts d'attribuer à la méthode interprétative en sociologie des problèmes qui relèvent en fait de la nature réflexive des sociétés humaines et des limites qu'elle pose à la prédiction en sciences sociales. Illustrant le type de propositions qui alimente les craintes d'une « colonisation » des sciences sociales par les sciences informatiques (McFarland *et al.*, 2016), l'article de Watts invite les sociologues à s'engager dans une réflexion sur les chemins à prendre pour imaginer la sociologie numérique, face à l'avancée inéluctable des épistémologies propres aux physiciens et aux informaticiens.

Donner accès aux données : les hackathons et *data sprints* dans la chaîne longue de la fabrique des données

Dans le dernier article du numéro, Celya Gruson-Daniel et Constance de Quatrebarbes se penchent sur l'accès aux données numériques et leur valorisation dans le cadre de la recherche en sciences sociales. Elles analysent une arène particulière de travail sur les données : les *data sprints* et les hackathons. Importés en sciences humaines et sociales, ces dispositifs revêtent un intérêt particulier pour réfléchir à l'engagement des chercheurs auprès des divers acteurs, publics ou privés, qui participent à la fabrique des

données. Ces dispositifs, réunissant pour quelques jours des personnes aux profils très divers, académiques et non académiques, constituent un locus de la recherche collaborative, interdisciplinaire et expérimentale en études numériques. Ils offrent, en cela, une entrée intéressante pour penser les enjeux relatifs à la fabrique des données. Gruson-Daniel et Quatrebarbes reviennent sur un hackathon organisé autour des données issues du site Web d'une consultation publique concernant le projet de loi pour une République numérique. En observant le hackathon en train de se faire, les auteures rendent compte du travail invisible nécessaire à l'obtention et la mise à disposition des données : enrôlements, négociations, choix d'extraction, etc. Leur travail met l'accent sur la manière dont les choix techniques et organisationnels qui s'opèrent en amont de l'évènement traduisent un engagement politique typique des formes d'activisme propres à la culture du hack. Les auteures promeuvent la nécessité d'une réflexivité sur les choix méthodologiques et techniques destinés à l'obtention des données et à leur utilisation. Cette réflexivité s'avère cruciale pour l'avènement d'une sociologie numérique, en ce moment critique où le cadre juridique de l'accès aux traces numériques est encore en cours de constitution.

BIBLIOGRAPHIE

- ABBOTT, A. (2014), *Digital Paper: A Manual for Research and Writing with Library and Internet Materials*, Chicago, University of Chicago Press.
- ABBOTT, A. (2016), « L'avenir des sciences sociales : entre l'empirique et le normatif », in *Annales HSS*, vol. 71, n° 3, p. 575-596.
- BAKSHY, E., S. MESSING et L. ADAMIC (2015), « Exposure to ideologically diverse news and opinion on Facebook », *Science*, vol. 348, n° 3, 6239, p. 1130-1132.
- BARABASI, A.-L. et R. ALBERT (1999), « Emergence of Scaling in Random Networks », *Science*, vol. 286, n° 5439, p. 509-512.
- BATTISTI, M. et J. SCHÖPFEL (2017), « Quel paysage juridique pour l'exploration de données? », *I2D — Information, données & documents*, vol. 54, n° 2, p. 25-26.
- BAYA-LAFFITE, N. (2017), Moving beyond the digitalised and natively digital divide. The case of the mapping of climate policy debates in multiple spaces, 3rd International Conference on Public Policy (ICCP3), Singapour.
- BEAUDE, B. (2014), *Les fins d'Internet*, Limoges, FYP.
- BEAUGUITTE, L. (2015), ENSAI Cours 4 — L'analyse des réseaux complexes, *groupe fmr*, <https://groupefmr.hypotheses.org/3644>, consulté le 21 janvier 2018.
- BEER, D. et R. BURROWS (2007), « Sociology And, of and in Web 2.0: Some Initial Considerations », *Sociological Research Online*, vol. 12, n° 5, p. 1-13.
- BESSY, C. et F. CHATEAURAYNAUD (1995), *Experts et faussaires. Pour une sociologie de la perception*, Paris, Métailié.
- BEUSCART, J.-S. (2017), « Des données du Web pour faire de la sociologie... du Web? », in Menger, P.-M. et S. PAYE (dir.), *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, p. 141-161.
- BEUSCART, J.-S., É. DAGIRAL et S. PARASIE (2016), *Sociologie d'internet*, Paris, Armand Colin.
- BOELAERT, J. et E. OLLION (à paraître), « The Great Regression Machine Learning, Econometrics, and the Future of Quantitative Social Sciences », *Revue française de sociologie*.
- BORRA, E. et B. RIEDER (2014), « Programmed method: developing a toolset for capturing and analyzing tweets », *Aslib Journal of Information Management*, vol. 66, n° 3, p. 262-278.

- BORRA, E., E. WELTEVREDE, P. CIUCCARELLI, A. KALTENBRUNNER, D. LANIADO, G. MAGNI, ... et T. VENTURINI (2014), «Contropedia — the Analysis and Visualization of Controversies in Wikipedia Articles», *Proceedings of The International Symposium on Open Collaboration*, p. 34: 1.
- BOULLIER, D. (2015a), « Les sciences sociales face aux traces du big data », *Revue française de science politique*, vol. 65, n° 5, p. 805-828.
- BOULLIER, D. (2015b), « Vie et mort des sciences sociales avec le big data », *Socio. La nouvelle revue des sciences sociales*, n° 4, p. 19-37.
- BOULLIER, D. (2017), « Pour des sciences sociales de troisième génération (SS3G) : Des traces numériques aux répliques », in Menger P.-M. et S. PAYE (dir.), *Big data et traçabilité numérique : Les sciences sociales face à la quantification massive des individus*, Paris, Collège de France, p. 163-184.
- BOURDIEU, P. (1997), *Méditations pascaliennes*, Paris, Le Seuil.
- BOZIO, A. et P.-Y. GEOFFARD (2017), *L'accès des chercheurs aux données administratives : État des lieux et propositions d'actions*, Paris, Conseil national de l'information statistique.
- BRUNS, A. (2018), Facebook shuts the gate after the horse has bolted, and hurts real research in the process, *Internet Policy Review*, <https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786>, consulté le 28 mai 2018.
- BURROWS, R. et M. SAVAGE (2014), « After the crisis? Big Data and the methodological challenges of empirical sociology », *Big Data & Society*, vol. 1, n° 1, p. 1-6.
- CAILLOL, H. (2015), « Ouverture des données de santé : l'expérience de l'Assurance maladie », *Informations sociales*, n° 191, p. 60-67.
- CALLON, M. (2013), « Sociologie de l'acteur réseau », in AKRICH M. et B. LATOUR (dir.), *Sociologie de la traduction : Textes fondateurs*, Paris, Presses des Mines, p. 267-276.
- CARDON, D. et A. CASILLI (2015), *Qu'est-ce que le Digital Labor?*, Paris, Institut national de l'audiovisuel.
- CGET et IGF (2015), Rapport relatif aux données d'intérêt général : <https://www.economie.gouv.fr/files/files/PDF/DIG-Rapport-final2015-09.pdf>, consulté le 28 mai 2018.
- CHIGNARD, S. et L.-D. BENYAYER (2015), *Datanomics Les nouveaux business models des données*, Limoges, FYP.
- COINTET, J.-P. et S. PARASIE (à paraître), « Ce que le big data fait à l'analyse sociologique des textes. Un panorama critique des recherches contemporaines », *Revue française de sociologie*.
- DACOS, M. (2011), Manifeste des Digital humanities, *THATCamp Paris*, Billet.
- DACOS, M. et P. MOUNIER (2014), Humanités numériques. État des lieux et positionnement de la recherche française dans le contexte international, *Institut français*.
- EAGLE, N., A. PENTLAND et D. LAZER (2009), « Inferring friendship network structure by using mobile phone data », *Proceedings of the National Academy of Sciences*, vol. 106, n° 36, p. 15274-15278.
- GENERATIONLIBRE (2018), *Patrimonialité des données numériques*, <https://www.generationlibre.eu/data-a-moi/>, consulté le 20 juin 2018.
- HAGGERTY, K. D. et R. V. ERICSON (2000), « The surveillant assemblage », *The British Journal of Sociology*, vol. 51, n° 4, p. 605-622.
- HARCOURT, B. E. (2015), *Exposed: Desire and Disobedience in the Digital Age*, Cambridge, Harvard University Press.
- HARGITTAI, E. et C. SANDVIG (2015), « How to think about digital research », in : HARGITTAI E. et C. SANDVIG (dir.), *Digital Research Confidential: The Secrets of Studying Behavior Online*, MIT Press, p. 1-28.
- HOFMAN, J. M., A. SHARMA et D. J. WATTS (2017), « Prediction and explanation in social systems », *Science*, vol. 355, n° 6324, p. 486-488.
- JESSIE, D., G. KAREN et T. MCMILLAN COTTON (dir.) (2016), *Digital sociologies*, Bristol, Policy Press.
- KAPLAN, F. (2013), Lancement de la « Venice Time Machine », <https://fkaplan.wordpress.com/2013/03/14/lancement-de-la-venice-time-machine/>, consulté le 20 décembre 2017.
- KARSENTI, B. et C. LEMIEUX (2017), *Socialisme et sociologie*, Paris, Éditions de l'École des hautes études en sciences sociales.

- LANIER, J. (2014), *Who Owns the Future?*, London, Simon and Schuster.
- LATOUR, B. (2010a), *Cogitamus: six lettres sur les humanités scientifiques*, Paris, La Découverte.
- LATOUR, B. (2010b), « Tarde's idea of quantification », in CANDEA M. (dir.), *The Social after Gabriel Tarde: Debates and Assessments*, London, Routledge, p. 145-162.
- LATOUR, B., P. JENSEN, T. VENTURINI, S. GRAUWIN et D. BOULLIER (2013), « Le tout est toujours plus petit que ses parties », *Réseaux*, n° 177, p. 197-232.
- LAZER, D., A. PENTLAND, L. ADAMIC, S. ARAL, A.-L. BARABASI, D. BREWER, ... M.V. ALSTYNE (2009), « Computational Social Science », *Science*, vol. 323, n° 5915, p. 721-723.
- LESKOVEC, J. et E. HORVITZ (2008), Planetary-Scale Views on an Instant-Messaging Network, *arXiv: 0803.0939 [physics]*, <https://arxiv.org/abs/0803.0939>, consulté le 20 décembre 2017.
- LUPTON, D. (2015), *Digital Sociology*, New York, Routledge.
- MARRES, N. (2017), *Digital Sociology: The Reinvention of Social Research*, London, John Wiley & Sons.
- MASON, W. et D. J. WATTS (2009), « Financial Incentives and the « Performance of Crowds » », in: *Proceedings of the ACM SIGKDD Workshop on Human Computation*, ACM, New York, p. 77-85.
- MAUREL, L. et L. AUFRERE (5 février 2018), Pour une protection sociale des données personnelles, – *S.I.Lex* – <https://scinfolex.com/2018/02/05/pour-une-protection-sociale-des-donnees-personnelles/>, consulté le 3 avril 2018.
- MC FARLAND, D. A., K. LEWIS et A. GOLDBERG (2016), « Sociology in the Era of Big Data: The Ascent of Forensic Social Science », *The American Sociologist*, vol. 47, n° 1, p. 12-35.
- MENGER, P.-M. (2014), *Le travail, sa valeur et son évaluation*. <https://www.college-de-france.fr/site/pierre-michel-menger/course-2013-2014.htm>, consulté le 3 avril 2018.
- MERCKLE, P. (2016), *Sociologie des réseaux sociaux*, Paris, La Découverte.
- MERTON, R. K. (1968), « The Matthew Effect in Science: The reward and communication systems of science are considered », *Science*, vol. 159, n° 3810, p. 56-63.
- MORETTI, F. (2013), *Distant Reading*, London et New York, Verso.
- OLLION, É. (2015), « L'abondance et ses revers. Big data, open data et recherches sur les questions sociales », *Informations sociales*, n° 191, p. 70-79.
- ORTON-JOHNSON, K. et N. PRIOR (2013), *Digital Sociology: Critical Perspectives*, Springer.
- PAIHES, B. (2018), « Comment définir et réguler les « données d'intérêt général » ? Enjeux numériques », n° 2, *Annales des Mines*, <http://www.annales.org/enjeux-numeriques/2018/en-2018-02/EN-2018-06-9.pdf>, consulté le 15 juin 2018.
- PARASIE, S. et J.-P. COINTET (à paraître), « Comment se forment les publics d'une carte de crimes ? Une analyse computationnelle de traces textuelles », *Réseaux*.
- PARISER, E. (2012), *The Filter Bubble: What The Internet Is Hiding From You*, London, Penguin.
- PASSERON, J.-C. (2006), *Le raisonnement sociologique: Un espace non poppérien de l'argumentation*, Paris, Albin Michel.
- PENTLAND, A. (2015), *Social Physics: How Social Networks Can Make Us Smarter*, New York, Penguin.
- PENTLAND, A. (1^{er} avril 2012), The New Science of Building Great Teams, *Harvard Business Review*, <https://hbr.org/2012/04/the-new-science-of-building-great-teams?referral=00134>, consulté le 15 juin 2018.
- PENTLAND, A. (1^{er} novembre 2013), Beyond the Echo Chamber, *Harvard Business Review*, <https://hbr.org/2013/11/beyond-the-echo-chamber>, consulté le 15 juin 2018.
- RIEDER, B. (2013), « Studying Facebook via Data Extraction: The Netvizz Application », in: *Proceedings of the 5th Annual ACM Web Science Conference*, ACM, New York, p. 346-355.
- RIEDER, B., A. MATAMOROS-FERNANDEZ et Ö. COROMINA (2018), « From ranking algorithms to « ranking cultures »: Investigating the modulation of visibility in YouTube search results », *Convergence*, vol. 24, n° 1, p. 50-68.
- RIEDER, B. et T. RÖHLE (2012), « Digital methods: Five challenges », in: *Understanding Digital Humanities*, London, Palgrave MacMillan, p. 67-84.
- ROGERS, R. (2013), *Digital Methods*, Cambridge, MIT Press.
- SALGANIK, M. J. (2017), *Bit by Bit: Social Research in the Digital Age*, Princeton, Princeton University Press, 448 p.

- SALGANIK, M. J. et D. J. WATTS (2009), « Web-Based Experiments for the Study of Collective Social Dynamics in Cultural Markets », *Topics in Cognitive Science*, vol. 1, n° 3, p. 439-468.
- SAVAGE, M. et R. BURROWS (2007), « The Coming Crisis of Empirical Sociology », *Sociology*, vol. 41, n° 5, p. 885-899.
- TRAVERS, J. et S. MILGRAM (1969), « An Experimental Study of the Small World Problem », *Sociometry*, vol. 32, n° 4, p. 425-443.
- TURCO, C. J. et E. W. ZUCKERMAN (2017), « Verstehen for Sociology: Comment on Watts », *American Journal of Sociology*, vol. 122, n° 4, p. 1272-1291.
- VENTURINI, T. (2010), « Diving in magma: How to explore controversies with actor-network theory », *Public Understanding of Science*, vol. 19, n° 3, p. 258-273.
- VENTURINI, T. (2012), « Building on faults: how to represent controversies with digital methods », *Public Understanding of Science*, vol. 21, n° 7, p. 796-812.
- VENTURINI, T., N. BAYA-LAFFITE, J.-P. COINTET, I. GRAY, V. ZABBAN et K. D. PRYCK (2014a), « Three maps and three misunderstandings: A digital mapping of climate diplomacy », *Big Data & Society*, vol. 1, n° 2, p. 1-19.
- VENTURINI, T., L. BOUNEGRU, J. GRAY et R. ROGERS (2018), A reality check(list) for digital methods, *New Media & Society*, <http://journals.sagepub.com/doi/abs/10.1177/1461444818769236>, consulté le 15 juin 2018.
- VENTURINI, T., D. CARDON et J.-P. COINTET (2014b), « Présentation », *Réseaux*, n° 188, p. 9-21.
- WATTS, D. J. (2004), *Six Degrees: The Science of a Connected Age*, New York, W. W. Norton & Company.
- WATTS, D. J. (2013), « Computational social science: Exciting progress and future directions », *The Bridge on Frontiers of Engineering*, vol. 43, n° 4, p. 5-10.
- WATTS, D. J. (2014), « Common Sense and Sociological Explanations », *American Journal of Sociology*, vol. 120, n° 2, p. 313-351.
- WATTS, D. J. et S. H. STROGATZ (1998), « Collective dynamics of « small-world » networks », *Nature*, vol. 393, n° 6684, p. 440-442.
- WELTEVREDE, E. et E. BORRA (2016), « Platform affordances and data practices: The value of dispute on Wikipedia », *Big Data & Society*, vol. 3, n° 1, p. 1-16.