



HAL
open science

Re-socialiser les traces d'activités numériques : une proposition qualitative pour les Sciences de l'Information et de la Communication

Alan Ouakrat, Julien Mésangeau

► To cite this version:

Alan Ouakrat, Julien Mésangeau. Re-socialiser les traces d'activités numériques : une proposition qualitative pour les Sciences de l'Information et de la Communication. *Revue française des sciences de l'information et de la communication*, 2016, Humanités Numériques et Sciences de l'Information et de la Communication, 8. halshs-01294749

HAL Id: halshs-01294749

<https://shs.hal.science/halshs-01294749>

Submitted on 29 Mar 2016

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

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



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

Resocialiser les traces d'activités numériques : une proposition qualitative pour les SIC

Alan Ouakrat, Julien Mésangeau

Alan Ouakrat est docteur en Sciences de l'Information et de la Communication, chercheur post-doctoral au Centre de Recherches et d'Études des Médiations (CREM) de l'Université de Lorraine. Il est aussi chercheur associé au Centre d'Analyse et de Recherche Interdisciplinaire Sur les Médias (CARISM) de l'Université Panthéon-Assas (Institut Français de Presse).

alan.ouakrat@gmail.com

Julien Mésangeau est Maître de Conférences en Sciences de l'Information et de la Communication à l'Université Sorbonne-Nouvelle. Il y est également chercheur au sein de l'équipe Médias, cultures et pratiques numériques (MCPN) du laboratoire Communication, information, médias (CIM).

mesangeaujulien@gmail.com

Résumé

Les méthodes de recherche numérique suscitent une attraction forte chez les chercheurs en sciences de l'information et de la communication. Au-delà de la nouveauté et de l'« effet de mode », l'illusion de la facilité semble également caractériser cet attrait. Nous proposons un retour réflexif sur deux expériences de recherches empiriques menées à partir de données numériques : une enquête sur les usages du *smartphone* d'une part et sur les réseaux sociaux professionnels d'autre part. Nous soulignons l'intérêt d'une combinaison des méthodes dans l'interprétation des traces numériques afin de ne pas négliger le rôle des pratiques non numériques des enquêtés et leur profil sociodémographique dans l'explicitation des comportements numériques observés.

Mots-clés : méthode, épistémologie, interdisciplinarité, numérique, trace, activité.

Abstract

The digital research methods elicit a strong attraction among researchers in communication and media studies. Beyond the novelty and the "fad" the illusion of ease also seems to characterize this appeal. We offer a reflexive return to two empirical research experiences from digital data: a survey on the use of smartphone on one hand and one on professional social networks on the other hand. We highlight the value of a combination of methods in the interpretation of digital traces in order not to overlook the role of non-digital practices of respondents and their socio-demographic profile in the explanation of the digital behaviors observed.

Keywords: digital methods, epistemology, interdisciplinarity, digital, trace, activity.

Introduction

Les technologies numériques jouent un rôle croissant dans les pratiques sociales que sont les activités culturelles et médiatiques. Une des caractéristiques centrales de ces technologies est la traçabilité des interactions entre l'homme et la machine. Pour saisir les usages du numérique plusieurs types d'outils peuvent être mobilisés, qu'ils émanent du monde industriel ou de la recherche (Beauvisage, 2013, p.188). Les données d'audience issues des instituts de mesure tel que Médiamétrie produisent une connaissance générale utile, par exemple, pour analyser les usages de l'actualité en ligne et sur le mobile (Ouakrat, 2013). Il est aussi possible d'exploiter des données extraites d'interface de programmation (API) de services web tel que LinkedIn pour creuser l'usage des réseaux sociaux numériques professionnels (Mésangeau, 2014), ou encore de mobiliser une sonde logicielle pour enregistrer l'activité d'utilisateurs de terminaux mobiles connectés (Ouakrat, 2015). Dans ces exemples, le numérique est à la fois l'objet d'étude et l'instrument d'analyse. Il offre des possibilités de capture, de traitement et de visualisation des données (Plantin et Monnoyer-Smith, 2013, p.42 ; Diminescu et Wievorka, 2015, p.9).

Dans cet article, nous nous appuyons sur deux recherches empiriques - l'une sur les usages du *smartphone*, l'autre sur les réseaux sociaux numériques professionnels -, pour démontrer qu'une combinaison de méthodes s'avère utile au dépassement des limites posées par les données numériques. Certes, ces dernières constituent potentiellement une source nouvelle de compréhension des activités humaines. Cependant, elles demeurent une traduction très partielle et limitée des pratiques. Elles décrivent des liens sociaux de façon restreinte et circonstanciée et ne représentent qu'un pan de l'activité des individus. De plus, elles ont tendance à effacer le contexte social et biographique dans lequel s'inscrivent les individus pour se concentrer sur une forme particulière de l'activité, celle qui peut être enregistrée. Les traces captées et rendues disponibles aux chercheurs en sciences de l'information et de la communication (SIC) sont construites et re-définies par les modalités de prélèvement et de mise en forme de l'information. Ces données doivent donc être resocialisées *a posteriori* pour y attacher les coordonnées sociales et biographiques des individus étudiés et donner lieu à des interprétations dans une perspective sociologique.

Les recherches empiriques menées reposent sur une logique méthodologique similaire : passer d'un ensemble d'individus à un petit nombre de cas pour creuser plus finement les comportements observés, en faire apparaître les raisons explicatives. Ainsi, la cohérence d'une posture épistémologique et praxéologique particulière dans l'exercice de la recherche en sciences sociales est défendue. Elle insiste sur la nécessité de situer socialement les traces numériques analysées, celles produites sciemment par l'individu (comme les éléments de profils d'utilisateurs sur un réseau social numérique professionnel) ou résultant d'une interaction avec un terminal numérique enregistrée par une sonde logicielle (l'horodatage d'actions effectuées par l'utilisateur sur son *smartphone*, la mesure des applications au premier plan par exemple). Malgré le caractère relativement lourd et complexe de la mise en œuvre d'une telle démarche méthodologique, c'est à ce prix qu'une forme d'« épaisseur sociale » (Jouët, 2011)

peut être ré-introduite pour saisir la dimension collective des comportements – y compris lorsqu'ils sont réalisés seuls (Lahire, 2013). Quel lien les traces d'activités numériques entretiennent-elles avec le monde social ? Dans quelle mesure et à quelles conditions les variables sociodémographiques s'avèrent-elles pertinentes pour approcher, dans une perspective qualitative, les traces d'activités numériques quantifiées ?

En premier lieu, nous abordons la redéfinition des pratiques de la recherche liée à l'émergence des méthodes numériques. En effet, ces dernières semblent exercer une forte attractivité sur les chercheurs en SIC. Nous revenons ensuite sur les recherches empiriques menées pour faire le bilan de l'apport et des limites des méthodes numériques telles qu'elles sont mobilisées dans nos recherches respectives. Enfin, nous soulignons la capacité réflexive permise par les visualisations de données numériques et nous montrons comment elles peuvent être mobilisées pour appréhender certains régimes de l'action individuelle et collective, moins évidents à saisir par d'autres méthodes. Nous discutons ainsi de l'apport des méthodes numériques aux SIC.

Le fantasme de la totalité : les traces et leur rapport au social

Les traces constituent un matériau d'étude qui peut apparaître riche par sa précision et son volume, mais qui, en réalité, se révèle fortement périssable (les usages évoluent en permanence) et à contextualiser (car fortement lié et dépendant des conditions d'extraction des données et de réalisation de l'enquête)¹. Les généralisations issues de ce type de données ne permettent donc pas de donner une explication sociologique satisfaisante des pratiques individuelles et collectives. Ces données sont toujours liées à un contexte de production donné et un cadre très particulier de définition de ce qui est quantifié.

Tout voir, tout capturer ?

Les espaces numériques sont caractérisés par la traçabilité des actions des utilisateurs. L'interactivité engendre la production de traces qui permettraient de saisir la vie collective sur ces espaces (Mitchell, 2009 ; Venturini, 2012). Cependant, interpréter ces traces n'est pas aussi simple qu'il n'y paraît. De plus, il convient de résister à la tendance des recherches numériques à réduire les individus et les collectifs aux traces produites (Jouët, 2011, p.80). En effet, pour certains chercheurs en sciences humaines et sociales (SHS), les traces s'apparenteraient à un dispositif d'observation « total » révolutionnant les méthodes classiques et effaçant les clivages entre *micro* et *macro*, *qualitatif* et *quantitatif*. Ainsi, « *la quantité et la richesse des données laisseraient penser que l'on peut passer, par une sorte de « zoom », de l'ensemble à l'individu et de la moyenne à l'idiosyncrasie* » (Rieder, 2010, p.91). Les méthodes numériques permettraient de maintenir le focus et l'étendue des observations (Venturini et Latour, 2012). Avec la traçabilité numérique, les chercheurs ne seraient désormais plus obligés de choisir entre la précision (qualitatif) et l'ampleur de leurs observations (quantitatif). Pour les chercheurs du Medialab de Sciences Po Paris par exemple, certains phénomènes sociaux pourraient être suivis

¹ Pour une typologie des données mesurées et collectées de façon automatisée sur le web, voir Beauvisage (2013).

dans chacun de leurs plis grâce à une chaîne de recherche entièrement numérisée, qui analyse le tissu social par la description du tissage de chacun de ses fils. Or, et nous rejoignons ici la position de Roger Bautier, la manipulation de ces outils devrait, nous semble-t-il, appeler en parallèle des enquêtes ethnographiques resituant l'ancrage social de telles pratiques. De plus, il convient d'être attentif à l'orientation de tels outils, susceptibles de « naturaliser » les rapports de pouvoir (Bautier, 2007). Plus largement, il s'agit aussi de se méfier d'un « *fantasme de l'objectivité des chiffres et d'une « science totale » autorisée par le traitement informatisé de grands corpus de données* » (Plantin et Monnoyer-Smith, 2013). La collecte et le traitement automatisés des données repousseraient « une étape plus loin » les dangers de l'erreur humaine et du jugement subjectif dans les sciences sociales, en rapprochant ces dernières des mathématiques axées sur la rationalité froide, l'objectivité et le caractère essentialiste des sciences naturelles (Rieder et Rhöle, 2012). En réalité, les traces numériques concernent une partie relativement restreinte des activités culturelles et sociales des individus, bien que cette part soit croissante. Considérer qu'il est possible de comprendre l'ensemble de l'activité d'un individu à partir de ses traces numériques, et seulement de ces dernières, est trompeur et renvoie à une idée très réductrice du social.

Des corrélations sans causalité ?

L'interrogation fondamentale formulée par Plantin et Monnoyer-Smith semble ici trouver sa place: « *assiste-t-on à l'émergence d'un champ de recherche dans lequel les hypothèses seraient tributaires de la nature des données disponibles* » (Plantin et Monnoyer-Smith, 2013, p.61) ? La principale critique qui peut être adressée aux méthodes numériques est celle d'une « *certaine superficialité des indicateurs et d'un manque d'approfondissement qualitatif de l'analyse [...] dus aux nécessités de systématisation et d'automatisation de l'observation à grande échelle* » (Rebillard, 2011, p.355). Ainsi, le plus dommageable semble être le « *déficit plus général de problématisation amont de la recherche, voire un manque d'intégration de la dimension sociologique [qui] peut conduire à des interprétations assez sommaires* » (Ibid.). Les raisons explicatives des comportements observés sont souvent effacées au profit du volume des répétitions numériques et de vastes modélisations.

Les données sont aujourd'hui abondamment disponibles. Ceci entraîne un renversement de la logique habituelle de la science, basée sur un modèle hypothético-déductif (les hypothèses et la problématique sont formulées en amont, *a priori*), vers un modèle inductif (les hypothèses et la problématique émergent des données collectées, *a posteriori*). La facilité de collecte des données invite, d'abord, à collecter les données avant de poser les questions (Croll, 2012). Le savoir produit par le traitement de ces traces d'activités numériques est un savoir purement prédictif, qui s'écarte d'un savoir des causes et des justifications des comportements : « *Rompant avec les ambitions modernes de la rationalité déductive reliant les phénomènes observables (c'est-à-dire les phénomènes préalablement sélectionnés comme objets d'observation et d'analyse en fonction de critères d'intérêt explicites ou implicites) à leurs causes, la rationalité statistique suit une logique inductive bien particulière dès lors qu'elle tire sa force du traitement automatisé d'informations dont la seule qualité est l'aspect massif : indifférente aux causes des phénomènes, cette rationalité s'ancre dans l'observation purement*

statistique de corrélations (indépendantes de toute logique) » (Rouvroy et Berns, 2010, p.91). Tout au contraire, le savoir causal produit par les SHS cherche à expliquer le monde social. Il tente d'attribuer des causes aux phénomènes observés, d'en comprendre l'origine et le développement. Ainsi, il participe à identifier les facteurs et les variables sociales ou démographiques qui déterminent, au moins en partie, certains comportements.

La pertinence limitée des méthodes numériques pour saisir le social

Les visualisations issues des données numériques sont séduisantes visuellement mais très réductrices et limitées en termes d'explicitation du social. A l'inverse d'une position qui consisterait à replier les observations des sciences sociales sur des « *points de focalisation d'attention* » et des « *entités circulantes* » (Boullier, 2015), car ces objets sont observables à partir des traces d'activités numériques, nous pensons qu'il faudrait au contraire enrichir les traces d'activités numériques par les coordonnées sociales, démographiques et biographiques des individus étudiés. Il s'agirait ainsi pour les sciences sociales de ne pas adopter une forme de syncrétisme abscond lié aux objets étudiés mais de réaffirmer la cohérence des ensembles étudiés par ces dernières : les groupes, les sociétés, les cultures, les organisations et les individus. Dès lors, les sciences sociales n'abdiqueraient plus face à leur volonté de comprendre le monde et les ensembles sociaux et ne se replieraient pas sur ce qu'il est simplement possible d'étudier à partir des traces numériques. Certes, le web est un monde de flux, mais cela ne signifie pas pour autant qu'il faille se résigner à traiter les problématiques ouvertes par nos disciplines dans ce domaine uniquement de cette manière. Ce serait alors déposséder les sciences sociales de leur fonction première qui est d'aider à produire des connaissances sur le monde social.

Retours d'expériences de terrain concernant les méthodes numériques

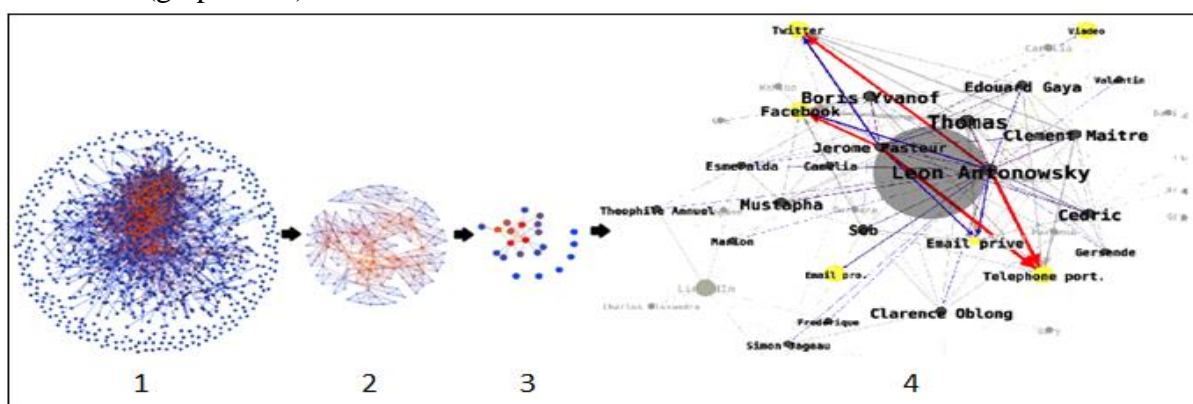
Les données ne sont jamais « brutes » mais « forcées » dans des catégories ontologiques définies par des conventions (Strasser, 2012). Les sciences des données, passées et présentes, n'ont jamais été limitées à l'analyse de données. Elles ont toujours reposé sur une combinaison d'hypothèses et de méthodes de recueil des données. Nous proposons de revenir sur les moyens méthodologiques mis en œuvre dans nos recherches empiriques pour dépasser l'incomplétude des données numériques à laquelle nous avons chacun été confronté.

Données numériques et étude d'usages : une mutation de la fonction herméneutique du graphe

L'analyse de réseaux sociaux est une méthodologie d'étude du social qui permet à la recherche en SHS de disposer de graphes autorisant l'étude conjointe des niveaux micro (l'individu), méso (les relations interpersonnelles de cet individu) et macrosociologique (le réseau plus vaste dans lequel ces relations sont insérées). Cette méthode est exploitable pour assurer l'analyse d'un groupe constitué au terme d'une période d'ethnographie prolongée. En revanche, l'analyse de réseau requiert des adaptations non-négligeables lorsqu'elle est exploitée pour analyser des données numériques.

Un retour sur l'étude des pratiques de réseautage sur LinkedIn

Dans le cadre d'une recherche consacrée à l'étude des pratiques de réseautage, nous avons pu associer l'analyse de réseaux conventionnelle et l'étude d'un graphe extrait du web (Mésangeau, 2012). Nous avons profité de cette capacité de l'analyse de réseau à permettre un jeu sur les échelles sans, pour autant, exploiter le versant mathématique de ces opérations. L'image ci-dessous décrit l'évolution de notre recherche en quatre graphes. Elle illustre le passage des 1200 participants du réseau (graphe n°1) aux 127 individus membres de cliques sollicités pour un entretien (graphe n°2) et, enfin, l'ensemble des 22 utilisateurs finalement rencontrés (graphe n°3).



Les 22 individus présents dans le graphe 3 devaient remplir un questionnaire pour renseigner la régularité et le contenu des échanges. Sur la base de ces données, nous avons pu reconstituer un graphe plus élaboré - graphe 4 - permettant de faire figurer la variété des médiations concourant à la constitution et à l'entretien d'un réseau social médiatisé par les TIC. Dans un second temps, l'utilisateur se voyait présenter un graphe représentant une partie de son réseau de contacts *LinkedIn*. La confrontation à ces graphes a permis d'avoir un retour sur l'histoire des relations et leurs médiations. Nous nous sommes alors rendu compte que le graphe n'était pas une représentation signifiante pour tout utilisateur. Pour certains, ce dernier ne représentait pas adéquatement leurs relations *LinkedIn*, lesquelles sont plutôt vues comme membres d'une liste : des noms, des fonctions, présents dans les répertoires et sollicités si besoin. Pour d'autres, les relations *LinkedIn* sont à penser en termes de cercles, le « cercle » *LinkedIn* comprenant de plus petits cercles, chacun lié à une affiliation sociale précise (*alumna*, entreprise, etc.). Enfin, quelques personnes développaient une lecture structurale du support proposé. Cette observation nous permet de présenter le graphe comme outil d'investigation, comme moyen de formaliser des processus complexes qui pourront, dans un second temps, être analysés avec les utilisateurs concernés, de façon qualitative.

Le volume et la variété des données accessibles sur internet interrogent la permanence de critères hérités des sciences sociales telles que l'exhaustivité et la représentativité (Boullier, 2015). Il est vrai que ces données permettent des choses jusqu'ici inimaginables, l'étude d'un

réseau de millions de nœuds par exemple. Néanmoins, les données des corpus numériques sont par nature incomplètes, tant au niveau quantitatif que qualitatif. D'un point de vue quantitatif, les limitations techniques font que les corpus extraits des réseaux socionumériques sont imparfaits, la conséquence est que certains acteurs pourtant centraux n'apparaissent pas dans les graphes que nous avons reconstitués. D'un point de vue qualitatif, même sur un outil ciblant certains domaines, le lien social médiatisé a une qualité qui ne peut être affirmée *a priori*, un lien sur *LinkedIn* étant équivoque (il peut être un lien strictement amical, bien que présent dans le réseau de contact sur un outil professionnel, et vice versa). L'analyse de graphe peut donc être mobilisée pour sa capacité à représenter, situer et rendre visible certains processus relationnels et communicationnels autrement invisibles. A titre de précaution, le chercheur peut se baser sur une « hypothèse nulle » selon laquelle les « relations professionnelles » contractées sur *LinkedIn*, de même que les « relations amicales » constituées sur *Facebook*, sont, au mieux, des relations « *actives techniquement* », similaires à ce que Caroline Haythornthwaite nomme les « liens latents » (Haythornthwaite, 2002). Partant, c'est à l'étude sociologique des usages de déterminer si ces mêmes relations sont également, pour reprendre les propos de l'auteure, « *actives socialement* ». Par ailleurs, il est essentiel de considérer le rôle structurant de la plateforme dans les usages, les prescriptions du dispositif sociotechnique utilisé.

Cette démonstration permet de poser deux observations. La première est que l'analyse de réseau basée sur des données relationnelles issues des réseaux socionumériques (RSN) peut être riche en enseignements, à condition de ne pas transposer trop rigidement cette méthode héritée d'une sociologie économique peu portée sur l'étude des médiations matérielles de l'action humaine. Ce faisant, on passe d'analyses positives des réseaux à une lecture plus exploratoire et heuristique, déployant des lectures aux niveaux macro, méso et micro, à l'instar de ce que Beuscart et Couronné proposaient, en 2009, dans leur étude des liens de citation entre utilisateurs de *Myspace* (Beuscart et Couronné, 2009). Difficilement interprétables en l'état, de tels supports appelleront un travail d'étayement basé sur une étude qualitative des usages et des trajectoires biographiques des répondants. Notre seconde observation est qu'il n'est pas envisageable d'étudier les pratiques des utilisateurs d'un RSN en circonscrivant la recherche à l'interprétation des structures des réseaux de contacts élaborés en ligne (Mésangeau, 2014). Le passage du troisième au quatrième graphe présenté dans l'encadré ci-dessus suppose de changer de cadre théorique : nous passons d'un individualisme méthodologique résumant l'acteur et le lien social à une dimension problématisant l'environnement social et technique des individus. En conséquence, ce graphe constitue également une rupture d'échelle et de méthode qu'il nous a fallu assumer lorsque nous avons souhaité illustrer comment les relations, en partie représentées sur *LinkedIn*, étaient intrinsèquement liées à l'histoire de groupes d'appartenance, aux dynamiques relationnelles et à l'usage des médias sociaux. Lors de cette dernière étape, le graphe occupa une fonction narrative cruciale, permettant un compte-rendu de la recherche où l'auteur pouvait aisément circuler de l'analyse d'une modalité du lien social à une autre, c'est-à-dire d'une « version professionnelle » de la relation à une modalité plus intime, « version amicale » de celle-ci. Ce graphe, qui représentait des liens sociaux et la médiation opérée par les outils de communication, aura permis d'étoffer l'analyse des réseaux de contacts des enquêtés.

Données numériques et étude d'usages : Combiner les méthodes pour approfondir le questionnement

Les données numériques peuvent être considérées comme une première appréhension d'un terrain, dont il s'agira de poursuivre le questionnement par d'autres méthodes, notamment des questionnaires et des entretiens pour contextualiser ce matériau numérique, lui rendre son inscription dans un contexte social.

Lors de l'enquête Pratiques culturelles et usages de l'informatique connectée (PRACTIC) menée à l'Institut National de la Recherche en Informatique et en Automatique (INRIA), nous avons eu recours à des questionnaires, des traces d'activités numériques et des entretiens avec des utilisateurs de *smartphones*. Il s'agissait ainsi de ne pas limiter l'analyse aux données numériques collectées, constituées par des traces d'activités des enquêtés. Seules, ces données se révèlent problématiques à intégrer dans une perspective qualitative en sciences sociales. Pour illustrer ce point, nous pouvons donner l'exemple de la fréquence des contacts : prise indépendamment, elle s'avère être un indicateur relativement pauvre du degré d'implication d'un individu dans une pratique ou une relation. De la même façon, l'intentionnalité et l'intensité vécue des pratiques ne se matérialisent que très imparfaitement dans les traces d'activité numériques et le temps passé sur les sites ou les applications mobile (Ouakrat, 2015 ; Ouakrat, 2013). Les traces d'activités numériques ne sont pas suffisantes, il est nécessaire de les contextualiser dans un ensemble plus large de pratiques, comme les pratiques culturelles et médiatiques des individus (Rebillard, 2011). Cet élargissement du champ d'observation permet ainsi de saisir « *le contexte et le sens social données aux activités en ligne* » (Jouët, 2011) par chaque utilisateur.

L'un des principaux intérêts des matériaux numériques est de repérer des régularités ou des irrégularités dans les comportements, et ce à différentes échelles (individu, groupe, etc.) et avec une profondeur temporelle adaptée aux phénomènes étudiés (de la microseconde à une période de plusieurs années, etc.). La capacité de zoomer finement permet l'observation d'un niveau de détail non accessible à l'œil nu et le tri quantitatif des comportements le repérage de phénomènes pertinents dont on peut ensuite creuser l'analyse par d'autres méthodes. C'est ainsi que nous avons pu mettre en valeur des catégories dans les comportements d'usage du *smartphone*, en fonction de l'intensité d'usage (fréquence et temps passé), à partir desquelles nous avons pu sélectionner des utilisateurs pour approfondir avec eux l'analyse qualitative de leurs usages du terminal au cours d'entretiens semi-directifs (Ouakrat, 2015). De plus, c'est muni des représentations visuelles de leur activité numérique que nous sommes allés les rencontrer, enrichissant ainsi l'échange d'un décentrement du regard de l'enquêté sur ses propres pratiques qu'il pouvait avoir tendance à naturaliser du fait pour lui de leur apparente banalité.

Un retour sur l'enquête PRACTIC

La méthodologie de l'enquête a consisté à associer un questionnaire auto-administré (méthode numérisée) à une application de collecte automatisée de données centrées sur l'utilisateur (méthode numérique) et des entretiens semi-directifs en face-à-face (méthode classique de la sociologie). Le questionnaire a permis de recueillir des informations sociodémographiques sur l'équipement et les pratiques médiatiques et culturelles. Les entretiens visaient à saisir plus finement les logiques d'usage du *smartphone* et les représentations des enquêtés. Enfin, l'application-sonde installée sur le *smartphone* des enquêtés s'inscrivait dans la continuité d'études quantitatives centrées sur l'utilisateur (*user-centric*) comme cela a pu être fait à partir de données de navigation web (*web usage mining*) (Beauvisage, 2007). Les informations furent collectées du côté de l'utilisateur par une sonde logicielle. Ainsi s'agissait-il de tracer son comportement en situation naturelle par le recueil automatique de données d'usages et l'enregistrement horodaté d'actions typées. L'application intervenait au niveau du système d'exploitation mobile Android de Google pour enregistrer l'usage des applications. Les données recueillies gagnent en précision ce qu'elles perdent en couverture : elles portent sur un nombre restreint d'utilisateurs et concernent une période limitée. Au total, 97 enquêtés se sont portés volontaires pour utiliser l'application, 260 questionnaires ont été complétés et 9 entretiens semi-directifs réalisés. 40 individus, qui ont à la fois laissé fonctionner l'application plus de 14 jours et répondu au questionnaire, ont été retenus. Ainsi avons-nous pu confronter les propos de ces enquêtés à la représentation visuelle de leur activité. Les traces correspondent à une capture à un instant T d'une action effectuée dans un contexte donné. Souvent incomplètes, elles sont des visions partielles de l'activité, pas nécessairement métonymiques. Nous refusons également d'inscrire notre travail dans une perspective comportementaliste, qui essaierait de « faire parler » le social à partir du comportement numérique individuel. Toutefois, il convient de ne pas sous-estimer les difficultés méthodologiques qu'a comporté notre démarche. Les données observées sont dé-corrélées *temporellement* des réponses aux questionnaires (qui visent plutôt à recueillir une opinion, un avis général de l'utilisateur, ses représentations). De plus, les entretiens ne permettent pas d'explicitier les traces, car ces dernières sont trop précises, il s'agit de micro-actions situées. La difficulté peut néanmoins être contournée en s'intéressant aux habitudes (pratiques récurrentes) et aux régularités (sur une semaine, un mois).

L'enquête PRACTIC n'est qu'un cas d'application très spécifique des méthodes de recherche numérique, qui renvoient à un territoire bien plus étendu. Les enjeux identifiés du recours à ces méthodes sont une compréhension plus fine du comportement numérique des usagers grâce à la possibilité de les observer directement sans interférer avec leurs pratiques. De plus, les comportements peuvent être étudiés à grande échelle et à moindre frais, comparativement à une observation qui implique un suivi physique des individus et un risque plus grand de perturbation de l'activité des enquêtés². Les limites identifiées résident dans le fait que les

² Les méthodes numériques posent néanmoins deux problèmes. Le premier est celui de l'acceptabilité sociale de telles méthodes qui peuvent être problématiques si elles sont perçues comme intrusives ou

traces numériques d'activités ne suffisent pas, prises indépendamment, à produire une compréhension fine du comportement des individus. Les données ainsi produites manquent de contextualisation situationnelle, sociale et biographique.

Discussion de l'apport des méthodes numériques aux Sciences de l'Information et de la Communication

Là où le numérique était censé offrir des facilités, il impose aussi ses propres contraintes et limites qui alourdissent le processus de recherche et déforme aussi bien les réponses obtenues que les questions posées. Il est possible - et selon nous nécessaire - de profiter pleinement de l'interdisciplinarité constitutive des sciences de l'information et de la communication pour associer aux données numériques des informations sociodémographiques et relatives à la trajectoire biographique des individus.

L'apport spécifique des méthodes numériques à la compréhension du social

Si les méthodes numériques peuvent apparaître limitées et nécessiter d'être complétées par d'autres méthodes telles que les questionnaires et les entretiens, dans une perspective qualitative – ce qui est l'argument de cet article –, elles peuvent aussi se révéler précieuses pour faire apparaître des actions non-perceptibles à l'enquêté lui-même et donc difficile à objectiver lors d'un entretien ou dans un questionnaire. Pour illustrer cette idée, nous pouvons faire référence aux différents régimes de l'action individuelle et collective notamment définis dans les travaux de Laurent Thévenot (2006). Ces régimes de l'action individuelle et collective nous permettent d'aborder les usages des individus autrement que par le prisme de l'action calculée orientée en finalité – cette dernière renvoyant à ce que Thévenot nomme le « *régime du plan* ». Or, nous constatons que les visualisations des traces d'activités numériques peuvent être utilisées pour représenter des actions non-dicibles par l'enquêté ou celles qui, procédant d'un grand nombre d'usages, lui sont difficilement perceptibles. Nous présentons d'abord l'action en familiarité avant de nous intéresser à l'action en commun théorisée par Pierre Livet et Laurent Thévenot (1994).

Les méthodes numériques peuvent aider à mettre en visibilité des points d'observation qui ne seraient pas accessibles autrement tels que les « actions en familiarité » décrites par Thévenot, à un niveau intrapersonnel, parfois inconscient, notamment dans l'usage du *smartphone* lorsqu'il s'agit, par exemple, d'« actions-réflexes » fortement incorporées à la vie quotidienne

sensibles en matière de vie privée (nous en avons fait l'expérience avec PRACTIC lorsqu'il s'est agi de demander aux individus d'installer sur leur propre téléphone mobile personnel l'application destinée à enregistrer leurs pratiques) (cf. Ouakrat, 2014). Le second problème, plus large, concerne le recrutement web. Il n'est d'ailleurs pas propre à la recherche en sciences sociales sur le numérique, mais se pose aussi dans un contexte industriel (par exemple pour le panel Médiamétrie//Net Ratings, composé aux deux tiers de recrutements web et qui nécessite des redressements statistiques importants). La difficulté posée est alors celle du profil des participants à l'étude qui s'auto-désignent et ont généralement un profil particulier, non représentatif de la population générale (Ouakrat, 2013).

de l'enquête (Ouakrat, 2015). Les données numériques sont finalement assez peu loquaces sur le « pourquoi ». En effet, les raisons explicatives des actions observées doivent souvent être recherchées par d'autres méthodes. En revanche, les données numériques peuvent être très fines et précises sur le « comment » et le « quand », voire permettre d'accéder à un niveau de connaissance de certains phénomènes qui n'étaient pas accessibles à l'enquêteur auparavant.

Outre la possible inexactitude des informations fournies par les répondants avec des méthodes déclaratives de recueil d'informations, de nombreuses motivations, représentations et pratiques ne peuvent être verbalisées. La distinction proposée par Thévenot entre les registres de l'action *en plan* (une action rationalisée, donc dicible) et de l'action *en familiarité* (une action ressentie, donc non dicible) peut être mobilisée. D'après cet auteur, l'action calculée – dite *en plan* – ordonne le monde en assignant de la valeur aux hommes et objets concernés par le projet de l'individu (Thévenot, 2006, p. 110). L'action familière, de son côté, n'a pas besoin d'être exprimée puisqu'elle s'éprouve à un niveau « intrapersonnel », dans l'expérience de l'aisance, de la commodité. En conséquence, la mise en commun de ces régimes (leurs niveaux de communicabilité) est inégale (Mésangeau, 2014, p.238). Au final, le plan est une action qui est communicable, s'exprimant notamment dans le « *jugement critique portée sur l'accomplissement d'un plan et ses défaillances* », tandis que la familiarité est une expérience indicible, s'éprouvant dans l'« appréciation du geste familier et des ratages ou dérapage » (*Ibid.*, 239). Plus largement, et bien que cela ne soit pas le sujet de cet article, l'action en familiarité interroge le contrôle et la maîtrise de l'individu sur ses propres actions, en somme les conditions de l'exercice de sa rationalité face à la technique lorsqu'elle devient banale et fortement intégrée à la vie quotidienne.

Une autre logique d'action peut être identifiée lorsque l'on interroge ces données numériques et les visualisations qu'elles permettent de produire. Il s'agit d'actions qui obéissent à une norme tacite, déduite des comportements que l'utilisateur peut rencontrer, et qu'il cherchera à imiter. Nous empruntons à Pierre Livet et Laurent Thévenot le concept d'« action à plusieurs » (Livet et Thévenot, 1994) pour mieux définir ces usages négociés à l'échelle des groupes. Cette action procède de « *l'agrégation d'actions individuelles* » où, sans qu'une norme n'ait été négociée, l'individu a néanmoins conscience que son action est en partie contrainte – et rendue possible – par celle d'autres individus. L'« action à plusieurs », à l'échelle des usages des réseaux socionumériques professionnels, est d'autant plus présente qu'elle est selon nous constante pour la majeure partie des usages. En effet, la gestion du profil d'utilisateur, tout autant que celle des relations, se fait sans règle ou norme explicite (propre à des *alumni* ou une communauté de métier, ...), chaque action se retrouve, en conséquence, orientée par l'« *agrégation des actions individuelles* » (Livet, 1994, p.254) qui tendent, par exemple, à cadrer les choix de rédaction des profils (manière de se présenter) ou les pratiques de réseautage développées sur le RSN.

L'élaboration d'une analyse profitant des données numériques et de l'étude qualitative des usages offre la possibilité de représenter et d'analyser certains traits de ce phénomène. Nos efforts visant à produire des représentations, en combinant données numériques et données recueillies par des entretiens ont permis d'analyser les conventions d'usage de ces différentes

technologies de communication; des conventions tacites, produits complexes des trajectoires des personnes, des relations et de l'appropriation progressive des possibilités offertes par la plateforme ou le terminal. Les méthodes numériques offrent ainsi les moyens de caractériser des processus sociaux particulièrement complexes à appréhender autrement, tels que les phénomènes non-dicibles, sortant habituellement du cadre de la recherche qualitative.

Une épistémologie en construction

Aucune méthode n'étant exempte de biais. Il appartient au chercheur de faire preuve d'honnêteté intellectuelle concernant l'explicitation des conditions de réalisation de l'enquête, de composition de l'échantillon et les difficultés rencontrées notamment dans le recrutement, en donnant à la communauté scientifique la capacité de reproduire les expériences empiriques menées ou *a minima* de les comprendre et de les situer. Les sciences sociales ont historiquement acquis une tradition d'analyse critique et réflexive en termes de précautions méthodologiques face aux populations enquêtées. Cette réflexivité méthodologique pourrait être utile aux sciences informatiques. Avec les méthodes numériques, la recherche est traversée par une quantophrénie qui, au détriment de l'effort de théorisation, privilégie une « *nouvelle croyance idéologique qui voudrait que les données existent en soi, indépendamment des modèles théoriques et méthodologiques qui les ont fait se constituer* » (Burrows et Savage, 2014 cité in Proulx, 2011).

L'appréhension statistique est souvent décevante en raison des limites de l'outil utilisé, des savoir-faire ou compétences des chercheurs pour traiter les données, ou encore d'une volonté d'affiner les résultats que ne permettent pas les statistiques en elles-mêmes. Il peut aussi y avoir des incompréhensions entre les disciplines lors d'une recherche interdisciplinaire, sachant que les concepts n'ont pas le même sens d'une discipline à l'autre et les centres d'intérêts ou les questionnements ne sont pas nécessairement les mêmes. Le risque est alors pris d'être face à une « importation décontextualisée » des concepts et des méthodes, une technique d'emprunt qui est la forme la moins élaborée du dialogue entre les disciplines (Bessière et Gojard, 2004, p.67). Ceci est d'ailleurs un phénomène observable dans la pratique de la recherche en analyse de réseau où de nombreux travaux en *computer sciences* abordent des données massives à l'aide d'indicateurs pensés à l'origine pour qualifier une dynamique relationnelle à l'échelle de quelques individus (Backstrom *et al.*, 2006 ; Ahn *et al.*, 2007; Ferrara et Fiumara, 2011). Les SIC sont une discipline carrefour où peuvent se rencontrer différentes approches méthodologiques et théoriques. Cette interdisciplinarité permet de convertir un outil informatique *a priori* peu pertinent pour expliquer des phénomènes sociaux en un outil d'exploration, de découverte et susceptible de générer la réflexivité des chercheurs comme des enquêtés.

Le rôle pris par la technologie informatique dans la recherche numérique doit ainsi s'accompagner d'une réflexivité renforcée de la part des chercheurs envers les outils, les méthodes et les données qu'ils mobilisent. Sans cette vigilance, le risque est grand selon nous de voir les humanités numériques perdre leur capacité critique d'interroger le fonctionnement

des mondes sociaux. Les chercheurs en SIC, investissant des travaux aujourd'hui étiquetés humanités numériques, sont confrontés à de nombreuses difficultés méthodologiques et épistémologiques. Le coût d'entrée dans la recherche numérique pourrait paraître condamner la démarche. Pour autant, nous avons tenté, à travers cet article, d'en souligner la possible pertinence.

Conclusion

Les données numériques ne sont généralement pas conçues pour être exploitées par des chercheurs en sciences sociales. Elles peuvent néanmoins receler d'une pertinence épistémologique et d'une productivité scientifique, à condition de parvenir à les adapter aux questions traitées. Dans nos recherches empiriques, nous avons fait la proposition d'intégrer les données numériques correspondant à l'usage d'un service (RSN professionnel) ou d'un terminal (le *smartphone*) à une démarche qualitative. Cette dernière semble avoir aujourd'hui tous les torts face à un contexte d'attractivité forte exercée par les méthodes et les données numériques sur l'activité des chercheurs en SIC. Elle demande du temps et les résultats qu'elle produit sont moins spectaculaires et séduisants visuellement. Il semble néanmoins souhaitable que continuent d'être développées des enquêtes sur les pratiques numériques qui tentent d'interroger et de comprendre l'activité des individus en tant que telle mais aussi au sein d'ensembles sociaux plus larges (tels que des groupes ou des populations), même si elles ne peuvent prétendre à la représentativité³.

Par ailleurs, nous soulignons le fait que les outils utilisés pour faire de la recherche numérique ne sont pas neutres. En effet, ils contribuent à transformer ce qui est mesuré et à déplacer le questionnement. En ce sens, nous refusons d'abandonner aux outils de visualisation et aux plateformes commerciales des géants du web, la capacité à maîtriser seuls les logiques de regroupement (*clusters*) et d'interprétation de ces données. De plus, nous récusons une acception socialement plate des humanités numériques, où les dimensions sociologique et communicationnelle des phénomènes étudiés se résumeraient à un graphe extrait de données collectées sur le net.

Comme toute technologie nouvelle, l'analyse de données numériques propose, pour reprendre les termes de Madeleine Akrich, un « *nouvel arrangement des choses et des gens* », porteurs de pouvoirs et de responsabilités (Akrich, 2006). Alors que ce nouvel arrangement est en cours de définition, nous avons souhaité faire une proposition méthodologique appuyée par nos expériences empiriques. Elle consiste en la re-socialisation des données numériques à partir d'une approche qualitative, tout en exploitant la réflexivité que peuvent produire les visualisations de ces données pour l'enquêteur comme pour l'enquêté.

³ Ce problème est lié aux moyens dont disposent la recherche publique en France ainsi qu'aux difficultés posées par le recrutement web telles que le biais d'auto-sélection des participants.

Références bibliographiques

AKRICH Madeleine, « La description des objets techniques », dans Madeleine Akrich, Michel Callon et Bruno Latour (sous la dir. de), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'ENSMP, 2006, p.159-178

AHN Yong-Yeol, Seungyeop HAN, Haewoon KWAK, Sue MOON and Hawoong JEONG, “Analysis of Topological Characteristics of Huge Online Social Networking Services”. *Proceedings of the 16th international conference on World Wide Web*, Canada, 2007. <http://www2007.org/papers/paper676.pdf> (dernière consultation le 15/02/16)

BAUTIER Roger, « Les réseaux de l'internet : des artefacts bien trop vivants », *Les Enjeux de l'information et de la communication*, 2007, communication au colloque MEOTIC, les 7 et 8 mars, <http://lesenjeux.u-grenoble3.fr/2007-meotic/Bautier/index.html> (dernière consultation le 08/02/16).

BACKSTROM Lars, Dan HUTTELOCHER, Jon KLEINBERG and Lan XIANGYANG, “Group formation in large social networks: membership, growth, and evolution”, *Proceeding of the KDD'06*, August 20–23, 2006, Philadelphia, Pennsylvania, USA, 2006.

BEAUVISAGE Thomas, « Mesurer l'activité sur ordinateur. Enjeux et méthodes », *Réseaux*, 2007, n°145-146, p. 45-80.

BEAUVISAGE Thomas, « Compter, mesurer et observer les usages du web : outils et méthodes », dans Barats Christine (sous la dir. de), *Manuel d'analyse du web en Sciences Humaines et Sociales*, Paris, Amand Colin, 2013, p.188-205.

BESSIERE Céline et Séverine GOJARD, « A propos des journées économie et sciences sociales. Décloisonnements disciplinaires, transferts et malentendus », *Economie Rurale*, 2004, n°282, p.61-73

BEUSCART Jean-Samuel et Thomas COURONNE, « La distribution de la notoriété en ligne. Une étude quantitative de MySpace », *terrains et travaux*, 2009, n°15.

BOULLIER Dominique, « Vie et mort des sciences sociales avec le big data », *Socio*, 2015, N°4, p.19-37.

BURROWS Roger et Mike SAVAGE, « After the crisis? Big Data and the methodological challenge of empirical sociology », *Big Data & Society*, 2014, April-June, p. 1-6.

CROLL Alistair, “Big data is our generation's civil rights issue, and we don't know it: What the data is must be linked to how it can be used”, O'Reilly Radar, 2012. Source : <http://www.cc.gatech.edu/~beki/cs4001/big-data.pdf> (dernière consultation le 13/02/16).

DIMINESCU Dana et Michel WIEVIORKA, « Le défi numérique pour les sciences sociales », *Socio*, 2015, Vol.4, p.9-17.

FERRARA Emilio and Giacomo FIUMARA, “Topological Features of Online Social Networks”, *Communications in Applied and Industrial Mathematics*, N°2, Vol. 2, 2011, p.1-20.

HAYTHORNTHWAITE Caroline, “Strong, weak and latent ties and the impact of the new media”, *The Information Society*, 2002, n°18.

JOUËT Josiane, « Des usages de la télématique aux Internet Studies », In Denouel Julie et Fabien Granjon (sous la dir. de), *Communiquer à l'ère numérique. Regards croisés sur la sociologie des usages*. Paris, Presses des Mines, 2011.

LAHIRE Bernard, *Dans les plis singuliers du social. Individus, institutions, socialisations*, Editions La Découverte, 2013, 171 p.

LIVET Pierre, *La communauté virtuelle, action et communication*, Paris, Editions de l'Eclat, 1994

LIVET Pierre et THEVENOT Laurent, « Les catégories de l'action collective », in: ORLEAN André (sous la dir. de), *Analyse économique des conventions*, Paris : PUF, 1994, pp.139-167.

MESANGEAU Julien, *Figures du “réseautage en ligne” sur les réseaux socionumériques professionnels: le cas d'un groupe d'anciens sur LinkedIn*, Thèse de doctorat en Sociologie, document non-publié, 2012, <https://halshs.archives-ouvertes.fr/tel-00783392/>

MESANGEAU Julien et Arnaud POVEDA, « Analyser l'adoption des réseaux socionumériques professionnels : entre approche pragmatiste et étude de la dimension symbolique des usages », *Etudes de communication*, 2013, n°41, Varia.

MESANGEAU Julien, « Articuler graphes et représentations d'utilisateurs d'un réseau socionumérique : retours sur une méthodologie d'entretien », *Sciences de la société*, 2014, n°92, p.143-159.

MITCHELL Tom M., “Mining our reality”, *Science, Computer Science*, 2009, 18 December, Vol. 326, p.1644-1645 http://www.cs.cmu.edu/~tom/pubs/Science2009_perspective.pdf (dernière consultation le 15/02/16).

OOUAKRAT Alan, « La consommation d'information en ligne : un cadrage quantitatif », in Josiane Jouët et Rémy Rieffel (sous la dir. de), *S'informer à l'ère numérique*, Presses Universitaires de Rennes, 2013, p.159-192.

OUAKRAT Alan, « Une défiance justifiée envers la collecte automatisée de données sur smartphones ? L'acceptabilité sociale des méthodes numériques pour étudier les usages des technologies numériques connectées », Actes du 16e colloque CREIS-Terminal, Nantes, 2014, 3-4 avril, 9p. Source : <https://hal.archives-ouvertes.fr/hal-01220292/document> (dernière consultation le 08/02/16)

OUAKRAT Alan, « Du rythme d'usage du smartphone au rythme de vie : les normes temporelles informelles des pratiques d'une population étudiante », *Questions de communication*, 2015, septembre, N°27, p.301-321.

PLANTIN Jean-Christophe et Laurence MONNOYER-SMITH, « Ouvrir la boîte à outils de la recherche numérique. Trois cas de redistribution de méthodes », *Tic & Société*, 2013, Vol. 7, N°2, p.39-66.

PROULX Serge, « La sociologie des usages, et après ? », *Revue française des sciences de l'information et de la communication* [En ligne], 2015, N°6.

Source : <http://rfsic.revues.org/1230> (dernière consultation le 08/02/16)

REBILLARD Franck, « L'étude des médias est-elle soluble dans l'informatique et la physique ? À propos du recours aux *digital methods* dans l'analyse de l'information en ligne », *Questions de communication*, 2011, N°20, p.353-376.

RIEDER Bernhard, « Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention », *Etudes de communication*, 2010, N°35, p.91-104.

Source : <https://edc.revues.org/2249> (dernière consultation le 08/02/16)

RIEDER Bernhard and Theo RÖHLE, "Digital methods: five challenges" in David M. Berry (Ed.), *Understanding digital humanities*, Palgrave Macmillan, 2012, p. 67-84.

ROUVROY Antoinette et Thomas BERNS, « « Le nouveau pouvoir statistique ». Ou quand le contrôle s'exerce sur un réel normé, docile et sans événement car constitué de corps « numériques »... », *Multitudes*, 2010, n°40, Vol. 1, p.88-103.

STRASSER Bruno J., "Data-driven sciences: From wonder cabinets to electronic databases", *Studies in History and Philosophy of Biological and Biomedical Sciences*, 2012, N°43, p.85-87.

VENTURINI Tommaso et Bruno LATOUR, « le tissu social. Traces numériques et méthodes qualiquantitatives », 2012, Source (dernière consultation le 08/02/16) :

http://www.medialab.sciences-po.fr/publications/Venturini_Latour-Le_Tissu_Social.pdf

VENTURINI Tommaso, « Great expectations: méthodes quali-quantitative et analyse des réseaux sociaux », in *L'Ere post-média. Humanités digitales et cultures numériques*, Jean-Paul

Fourmentaux (sous la dir. de), Editions Hermann, collection « Cultures numériques », 2012, Paris, p.39-51. Source (dernière consultation le 08/02/16) :

<http://spire.sciencespo.fr/hdl:/2441/11qiboags09v98jje54m1fvqh/resources/venturini-great-expectations.pdf>