Proposition de communication 16e colloque CREIS-Terminal

Alan Ouakrat (CARISM-IFP/Inria)

Une défiance justifiée envers la production automatisée de données sur ordiphones ? L'acceptabilité sociale des méthodes numériques pour étudier les usages des technologies numériques connectées.

Résumé

L'enquête Pratiques culturelles et usages de l'informatique connectée (PRACTIC) s'appuie sur une application Android, nécessitant l'installation sur les terminaux des enquêtés. Nous revenons dans cet article sur les difficultés de recrutement et d'implication de participants volontaires à une expérimentation sociologique sur les usages des ordiphones, menée à l'Inria. Le faible nombre d'utilisateurs recrutés ainsi que les obstacles rencontrés dans la phase d'expérimentation invitent à mutualiser les enseignements de cette recherche en matière de vie privée. *In fine*, se pose la question de l'acceptabilité de tels dispositifs d'enquête en sciences sociales, basé sur l'observation du comportement d'individus à l'aide d'une instrumentation informatique, automatisant la production de traces sur l'activité des utilisateurs, et installée sur leur ordiphone.

Mots-clés

Ordiphone, application, méthode numérique, acceptabilité sociale, production automatisée de données.

Introduction

L'essor de l'internet grand public tient dans le développement de services dont l'accès est proposée gratuitement à l'utilisateur-final. De la même façon, sur le mobile, la majeure partie des applications des ordiphones[1] sont téléchargées gratuitement[2]. Or, la plupart des éditeurs d'applications n'informent pas correctement les utilisateurs sur l'appropriation de leurs données personnelles, leur circulation et leur exploitation[3] (WSJ, 2010). Les éditeurs d'applications ne seraient parfois pas même au courant de l'information collectée et transmise sur leurs utilisateurs, ayant recours à des librairies publicitaires, collectant elles-mêmes directement des informations par le biais de serveurs-tiers (Achara *et al.*, 2013). La gratuité a donc un coût pour les consommateurs : leurs données personnelles sont échangées sur le marché publicitaire. Les conditions d'utilisations de ces services et applications peuvent apparaître comme insuffisamment explicites, voire comme un abus de la confiance des utilisateurs, car ces données sont appropriées sans l'obtention d'un consentement éclairé[4] (Tacma, 2013). Les entreprises privées sont les principales dépositaires et utilisatrices des données personnelles des utilisateurs du web 'gratuit'. Les risques liés à l'usage et au maniement de ces données sont donc plus élevés pour elles que pour les scientifiques. Ces derniers doivent cependant réfléchir collectivement à leurs pratiques en la matière. Une problématique – non traitée à notre connaissance dans la littérature – est celle de

[1] Nous préférons utiliser le terme « ordiphone » plutôt que l'anglicisme « *smartphone* », trop connoté. Ce dernier induit une vision méliorative de l'outil (*smart)*.

[2] Près de 90% des applications téléchargées seraient gratuites d'après une étude menée par Flurry : http://blog.flurry.com/bid/99013/The-History-of-App-Pricing-And-Why-Most-Apps-Are-Free.

[3] Une enquête menée par le *Wall Street Journal* en 2010 révélait que sur 101 applications d'ordiphones populaires 56 transmettaient l'identifiant utilisateur unique à d'autres entreprises sans que l'utilisateur n'en ait conscience ni n'exprime son consentement. 47 applications communiquaient la localisation géographique de l'utilisateur, 5 envoyaient l'âge, le genre et d'autres informations personnelles à des tiers (WSJ, 2010).

[4] Issue du domaine médical, la notion de consentement éclairé implique une présentation claire des risques afin que le patient puisse exercer un jugement informé dans sa prise de décision. Sur Internet, le consentement est informé par les conditions d'utilisation, cependant tout est fait par ces entreprises pour décourager la lecture de ces conditions d'utilisation. Il serait souhaitable que soient standardisées et simplifiées ces conditions d'utilisation par des indicateurs clairs et lisibles. Le site web *Terms of Services; Didn't Read* (tosdr.org) est une initiative visant à permettre une plus grande lisibilité des conditions d'utilisation aux utilisateurs des services web.

la fragilisation de la confiance des utilisateurs en raison des mauvaises pratiques des entreprises et des gouvernements en matière de vie privée. Notre enquête de terrain a coïncidé avec le scandale PRISM et les révélations d'une surveillance massive et généralisée des utilisateurs des technologies numériques connectées. Ce type d'événement et d'autres relatifs à la traçabilité numérique sont-ils susceptibles d'éroder la confiance des utilisateurs dans ces services et ces technologies ? Nous faisons l'hypothèse qu'ils peuvent avoir une influence négative sur l'acceptabilité de dispositifs scientifiques expérimentaux d'observations des usages. Et ce d'autant plus quand les scientifiques souhaitent recueillir le consentement éclairé des participants, en explicitant ce qu'ils mesurent, comment sont traitées les données et à quelles fins elles sont collectées. Cet article vise à proposer une réflexion sur les raisons explicatives des difficultés de recrutement et d'implication des utilisateurs, rencontrées lors d'une enquête de terrain, Pratiques culturelles et usages de l'informatique connectée (PRACTIC), menée à l'Institut National de la Recherche en Informatique et en Automatique (INRIA) depuis janvier 2013.

1. La circulation des données personnelles, rançon de la « gratuité » commerciale

La prise de conscience croissante de la surveillance numérique industrielle et gouvernementale pourrait affecter négativement l'expérience des utilisateurs des technologies numériques connectées, diminuant leur confiance dans les fournisseurs de services web gratuits, voire d'une manière plus générale dans les technologies numériques connectées. Les entreprises de services web ont accompagné la massification des usages de l'internet grand public. Ces dernières pratiquent une tarification asymétrique : les consommateurs bénéficient d'un accès gratuit aux services en échange de leur attention à la publicité et des données qu'ils renseignent et qui sont collectées à leur insu sur leur comportement. Le succès des services web tels que Google, Facebook ou Twitter a permis à ces entreprises américaines d'accumuler une masse considérable d'informations sur les utilisateurs. Le coût de ces services gratuits réside donc dans les données personnelles communiquées par les utilisateurs, volontairement ou non. Dans le maniement des grands ensembles de données, l'industrie a incontestablement une longueur d'avance sur la recherche publique. Avec l'essor du web 2.0, il y a eu une croissance exponentielle des traces d'activités comportementales des utilisateurs, en plus des informations directement renseignées par ceux-ci. Les données de navigation ou d'usage des applications sont les traces laissées par la navigation sur internet : historique, temps passé, sites visités, mots-clés tapés en requête dans un moteur de recherche, etc. La collecte de ces données peut être considérée comme problématique dans la mesure où elle intervient de façon invisible pour l'utilisateur. Ces données du « web invisible » (Castelluccia *et al.*, 2013) comprennent les données de suivi du comportement des utilisateurs, généralement échangées « dans le dos » de ces derniers. Ces mêmes informations servent à déduire par inférences statistiques les caractéristiques comportementales et psychosociologiques d'un individu, sans avoir à connaître son identité par le biais d'informations personnelles identifiantes (Castellucia, 2012 ; Kosinski *et al.*, 2013 ; Duhigg, 2012).

L'essentiel des traces produites par l'activité des utilisateurs est approprié par des entreprises privées, qui en font une source de création de valeur, par exemple par la revente à des tiers ou l'utilisation à des fins d'amélioration d'un service (Rieder, 2010). Ces entreprises peuvent décider d'y donner accès à des chercheurs en sciences sociales, mais bien souvent, il s'agit d'un accès restreint et partiel. L'accès peut être accordé à des tiers privés (sociétés), publics (Etat, surveillance) ou à des scientifiques. Dans ce dernier cas, on parle de « privilège de l'accès » (Boyd & Crawford, 2011). Il s'agit alors d'un usage de « seconde main » de « données d'occasion », c'est-à-dire produite par d'autres et à d'autres fins que la recherche scientifique. Avec les méthodes de recherche numérique, il existe un risque de perte d'indépendance de la recherche publique face aux grands acteurs du Web. Le rôle central pris par la technologie informatique dans la recherche numérique doit s'accompagner d'une réflexivité redoublée de la part des chercheurs envers les outils, les méthodes et les données qu'ils mobilisent. En ce sens l'appareillage critique

des sciences de l'information et de la communication vis-à-vis des méthodes quantitatives peut être utilement mis à contribution (Plantin et Monnoyer-Smith, 2012). De façon concomitante à notre enquête, les révélations d'Edward Snowden d'un scandale d'espionnage massif américain nommé PRISM[5], impliquant les grandes sociétés américaines de services web, a contribué à mettre à jour la vulnérabilité de la vie privée des utilisateurs des technologies numériques connectées. Le grand public prend progressivement conscience de l'enjeu de la sécurité des données personnelles à l'ère du numérique. Toutefois, ce qui constitue des informations hautement sensibles peut varier d'un individu à l'autre et en fonction des contextes d'exposition de ses données. Une donnée personnelle n'a pas de valeur dans l'absolu mais par rapport à un contexte d'usage particulier (Carrascal *et al.,* 2011 ; Rey, 2012). Dans quelle mesure un utilisateur peut-il exercer un jugement informé, concernant la suppression de ses données si celles-ci sont collectées de façon invisible et font l'objet d'un usage opaque, non documenté – notamment lorsqu'elles sont transmises à des tiers ? Le risque de confier des informations personnelles est mis en balance avec la gratification immédiate proposée par la relation marchande : « Le confort ressenti à recourir au quotidien à diverses applications technologiques font apparaître les risques perçus comme décorrélés de pratiques qui, de par leur fréquence, et l'attente d'immédiateté et de facilité qui les accompagne, ne souffrent pas d'être encombrées, ni différées, (presque) quel qu'en soit le motif » (Rey, 2012, p.138). Le fait de divulguer des informations personnelles intervient le plus souvent en accompagnement d'une autre action, et n'est pas un but en soi pour la personne qui poursuit un autre objectif. Ainsi, la décision concrète de divulguer ou non des informations personnelles est fondue dans un contexte et des préoccupations autres (*ibid*, p.170). Les données personnelles sont souvent le prix et la condition d'accès aux services gratuits du web. Le consentement des utilisateurs est alors acquis dans des conditions de liberté de choix discutables.

L'ordiphone, compagnon numérique de la vie quotidienne

En juin 2013, il y avait plus de 25,1 millions de possesseurs d'ordiphones en France, soit 46,2 % de la population française de 11 ans et plus[6]. L'ordiphone est fortement intégré à la vie quotidienne des individus : 7 personnes sur 10 ne l'éteignent jamais[7] (CNIL, 2011). L'usage du web serait même en repli par rapport aux applications mobiles[8]. A la fois « concentrateur et producteur de données personnelles », l'ordiphone est bien souvent « une boite noire pour son possesseur » (CNIL, 2012). 51% des utilisateurs pensent que leurs données ne peuvent être enregistrées ou transmises sans leur accord. Or, nombre de données sont transmises à l'insu des utilisateurs. La CNIL attribue cette méconnaissance aux conditions d'utilisation des services et applications qualifiées de « peu lisibles et compréhensibles », ainsi qu'au manque de transparence des acteurs économiques en ce domaine.

---

[5] Wikipédia : PRISM programme de surveillance
https://en.wikipedia.org/wiki/PRISM_%28surveillance_program%29
Le Monde, Prism, Snowden, surveillance : 7 questions pour tout comprendre :
http://www.lemonde.fr/technologies/article/2013/07/02/prism-snowden-surveillance-de-la-nsa-tout-comprendre- en-6-etapes_3437984_651865.html
Comprendre le programme « PRISM », le 11 juin 2013 :
http://www.lemonde.fr/international/infographie/2013/06/11/le-programme-prism-en-une-infographie_3427774_3210.html

[6] Baromètre trimestriel réalisé par la Mobile Marketing Association France, en partenariat avec comScore, GFK et Médiamétrie.

[7] Enquête réalisée début novembre 2011 à la demande de la CNIL par l'institut Médiamétrie auprès de 2315 utilisateurs français de smartphones de 15 ans et plus.

[8] *Zdnet*, 05/11/13, « France : l'usage du Web poursuit son repli face aux applications mobiles » : http://www.zdnet.fr/actualites/france-l-usage-du-web-poursuit-son-repli-au-profit-des-applications-mobiles-39795313.htm

Les usages de l'ordiphone s'avèrent délicat à observer en raison de la relation de l'usager à son terminal qui est souvent très personnelle, voire intime. Après l'adoption massive du téléphone mobile, les ordiphones[9] pourraient être des appuis de choix dans l'observation des usages informationnels et communicationnels des technologies numériques connectées. L'ordiphone permet de coordonner et de réaliser diverses d'activités tout en étant un outil compagnon de notre vie quotidienne et de nos déplacements. Cependant, l'observation de l'activité à partir de l'ordiphone implique une forme d'intrusion dans la vie personnelle des individus, pouvant être ressentie négativement. Le téléphone mobile, objet hautement personnel que la majeure partie des utilisateurs garde désormais avec soi en permanence (Martin, 2007), est un terminal particulièrement sensible en matière de vie privée tant il concentre d'informations sur son propriétaire.

2. Une expérimentation de production automatisée de données sur ordiphones, l'enquête PRACTIC

Il n'est pas évident de faire installer un dispositif d'observation de son comportement à un individu sur son téléphone en obtenant son consentement éclairé. Cependant, la production automatisée de données permet au sociologue d'adresser directement ses questions, sans faire un usage « détourné » de données existantes et produites à d'autres fins. Pour mener à bien notre étude, nous avons construit une méthode d'interprétation des traces d'activités produites. Cette dernière repose sur la confrontation du déclaratif et de l'observé. Les représentations des individus sont saisies par un questionnaire et des entretiens avec un certain nombre d'entre eux. Le questionnaire permet de collecter des informations sociodémographiques, sur l'ancienneté de l'équipement et la régularité des pratiques culturelles et médiatiques. Il s'agit ainsi de situer nos enquêtés dans une trajectoire biographique et sociale qui contextualise leur usage des technologies numériques connectées.

Nature, conservation et ouverture des données produites

L'enquête PRACTIC s'intéresse à la dimension temporelle des pratiques sur ordiphones. Les données collectées sont des informations horodatées relatives à l'usage de l'ordiphone dans la vie quotidienne. Nous n'enregistrons ni les contenus des échanges, ni l'identité des correspondants. Ces données concernent l'activation de l'écran de l'ordiphone (*screen*), le temps passé par les applications au premier plan (*appusage*), le réseau utilisé (3G, Edge, wifi) (*network*), le chargement de la batterie, la mise-à-jour ainsi que l'installation et la suppression d'applications (*appchange*), le nombre et la durée de SMS et d'appels émis et reçus (*sms*, *call*). Les données sont stockées sur des serveurs de l'Inria et du Laboratoire de Haute-Sécurité de Télécom-Nancy. Elles sont conservées de façon anonymisée, en étant liées à un identifiant unique par utilisateur et par terminal (*userID* et *deviceID).* Elles portent sur des utilisateurs-volontaires ayant exprimé un consentement éclairé et signé une charte de confidentialité. L'expérience dure plusieurs mois, l'utilisateur a la possibilité d'avoir accès à ses données et de les supprimer. Il peut aussi, à tout moment, arrêter l'application ou la supprimer. Nous avions fixé la durée de collecte des données à deux mois, puis nous l'avons étendue à quatre mois lors de l'étape nantaise, en raison du public étudié (les étudiants) et de la période de l'année (l'expérience a eu lieu en juin, période proche des vacances scolaires). Un des enjeux de la phase expérimentale est de repérer à partir de quand les données sont redondantes, le moment à partir duquel des données supplémentaires ne ferons que confirmer les régularités observées. Il s'agit ainsi de repérer la *durée utile de l'expérience* pour saisir les rythmes de vie et les habitudes d'usages d'un individu. Dans la phase exploratoire de l'enquête nous avons dû donner une certaine souplesse à ses règles en rendant possible la dés-anonymisation, en effectuant nous-mêmes la pseudonymisation des enquêtés et en établissant un fichier de correspondance des *userID* aux données de contact des participants (adresse email, nom et prénom). L'identification des

[9] Près de 4 français sur 10 possède un ordiphone (39% d'après CREDOC, 2013), le taux d'équipement est en forte progression.

individus fut nécessaire pour mener les entretiens et interpréter les traces par rapport aux réponses au questionnaire. Ces premiers traitements sur un petit nombre d'utilisateurs ont eu pour objectif de calibrer notre dispositif d'enquête et nos mesures, ainsi que d'en mesurer la perception. Désormais pour participer, il faut créer un compte sur la plateforme APISENSE. L'adresse email et le nom et prénom du participant sont décorrélées des données collectées. Les données sont pour l'heure réservées à l'équipe des chercheurs du projet. Ces données seront ensuite partagées à d'autres équipes de recherche exclusivement sous une forme anonymisée (usage de pseudonymes pour remplacer les *userID*) et agrégée.

Une régulation de la vie privée inscrite dans le droit et la technique

Les bases de données constituées par l'enquête PRACTIC ont fait l'objet de déclaration auprès du Correspondants Informatique et Libertés (CIL)[10] de l'Inria. Mais au-delà la contrainte juridique, la régulation peut être incorporée dans la technique, c'est-à-dire dès la conception du code informatique à partir duquel fonctionne l'outil technique. Remparts et gardes-fous contre les usages déviants en matière de vie privée, les applications de ce concept empêche des chercheurs de se saisir maladroitement de l'outil en exposant la vie privée de leurs enquêtés – parfois malgré eux. Toutefois, une certaine souplesse doit être mise en œuvre dans l'élaboration de ces dispositifs, à plus forte raison lorsqu'ils sont exploratoires et expérimentaux afin de pouvoir, par exemple, les données des individus à une échelle micro-sociale au cours d'enquêtes ethnographiques par exemple. En inscrivant la sécurisation de la vie privée dans les objets et les processus (Kessous, 2012, p.129), cette approche de *privacy-by-design* empêche techniquement l'identification des individus en décorrélant par exemple les données collectées des informations identifiantes relatives aux individus (Le Métayer, 2013). Ainsi, l'outil intègre ces principes de fonctionnement dès sa conception. Ceci ne doit toutefois pas laisser à penser aux scientifiques impliqués qu'ils seraient de-responsabilisés par cette intégration de la question de la vie privée à l'outil technique utilisé : doivent continuer à prévaloir les procédures d'anonymisation des données des participants, d'engagement de confidentialité et de non circulation des données auprès de tiers. Les sciences sociales ont historiquement acquises 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 dans leur approche de nouveaux objets tels que les usages des terminaux mobiles connectées. Avec l'appui et l'instrumentation de l'INRIA, nous avons pu produire nos propres données, ce qui nous a permis d'expérimenter une forme de transparence auprès des individus enquêtés sur les données produites et la manière dont elles sont utilisées. Conscients des dangers de la collecte automatisée de données à grande échelle (*big data*), l'outil conçu respecte la vie privée dès sa conception. Pour l'heure, le nombre de personnes utilisant l'application est relativement limité, mais le dispositif vise à être déployé largement (idéalement sur plusieurs centaines d'utilisateurs). Dans les premières phases de l'enquête, les observations portant sur un petit nombre de participants, ceci a été peu propice à inciter à la confiance les utilisateurs et a nécessité l'identification des individus (pseudonymisation, données non publiques dans cette phase de l'enquête).

Retour sur la phase d'expérimentation de l'enquête PRACTIC

Mesurer l'activité d'individus depuis leur ordiphone pose problème. Cela nécessite le consentement éclairé des individus. Il n'a pas été facile de convaincre les populations étudiantes d'accepter l'installation sur leur terminal de l'application. Les contreparties développées, formes de gratifications immédiates à l'usage de l'application, jouent un rôle essentiel dans l'acceptabilité du dispositif d'enquête. L'application doit ainsi pouvoir susciter de l'intérêt en elle-même et non

[10] Nouvel outil de régulation, le CIL permet à la structure qui le désigne d'être dispensée de déclaration tout en intégrant la question des données à caractère personnel et de la vie privée en son sein. Contenu dans la Directive Européenne que transpose la loi de 2004, le dispositif des CIL n'est pas une innovation française. L'un de ses grands rôles est de diffuser la culture « informatique et libertés » au sein de l'organisme l'ayant désigné, et de veiller au respect des dispositions légales en matière de traitement des données (Rey, 2012, p.81).

nécessairement quant à l'enquête sociologique dont elle est l'instrument. Ainsi, des retours doivent être faits à l'utilisateur sur ses propres pratiques, afin de l'aider à développer sa réflexivité. Les données sont basées sur le comportement de l'utilisateur et mesurent ses usages du terminal (temps d'usage par jour, applications les plus utilisées, etc.)

Les populations enquêtées ont été des étudiants, essentiellement en sciences de l'information et de la communication (1ère et 2nde année à l'IUT de Caen, le jeudi 11 avril 2013 ; bachelor et master1 à SciencesCom, à Nantes le jeudi 13 juin 2013), mais aussi en sciences et techniques de l'information et de la communication (Télécom Nancy, le 2 octobre 2013). Notre enquête ne s'appuie pas sur un échantillon représentatif de la population étudiante. Les participants ont dû compléter un questionnaire en ligne (méthode numérisée), qui visait à recueillir des informations sociodémographiques, ainsi que sur leur équipement et leurs pratiques numériques et culturelles. Ensuite, ceux qui disposaient d'un terminal Android pouvait, s'il le souhaitait et après signature d'une charte de confidentialité, installer l'application APISENSE. Certains ont été recontactés pour mener des entretiens.

Tableau 1. Récapitulatif des données chiffrées sur les populations étudiantes enquêtées

Critères de dénombrement de la population participante	Caen	Nantes	Nancy
Nombre d'étudiants contactés	86	130	300
Nombre d'étudiants présents	30	60	70
Nombre de questionnaires complétés	28	51	26/69
Nombre de possesseurs d'ordiphones	24	43	35
Nombre de possesseurs d'ordiphones Android	13	10	30
Nombre de participants volontaires pour l'installation d'APISENSE	12	3	21
Nombre d'applications installées	8 (2 fonctionnent le jour de l'expérience)	2	15
Nombre de participants ayant continué jusqu'au bout	2	1	12
Nombre de participants toujours actifs	2	1	10

La stratégie de diffusion initialement adoptée a été peu fructueuse, comme en atteste le tableau 1. Elle a supposé de contacter beaucoup d'individus pour n'en recruter que quelques-uns, disposant d'un ordiphone, équipé du système d'exploitation Android et volontaires pour participer à l'expérience PRACTIC.

Le dispositif d'enquête ne semble pas avoir été jugé acceptable par les potentiels participants, pour quelle(s) raison(s) ? Nous pouvons pointer, avec Ulrich Beck (2008), des raisons liées au fait que l'individualisme contemporain est marqué par une prise de conscience plus élevée des risques, un scepticisme réflexif, une défiance vis-à-vis des institutions, une plus grande confiance dans le marché et les promesses de la société de l'information. Si ces éléments nous mettent sur la voie, il nous faut aussi interroger des facteurs propres à notre démarche. La stratégie de dissémination initialement envisagée fut de présenter ce dispositif auprès de populations étudiantes réunies dans une salle de cours. Nous avons décrit explicitement aux enquêtés chaque type de donnée collectée, tout en les sensibilisant aux risques de la traçabilité numérique en matière de vie privée. Or, ce discours n'a pas porté ses fruits. Décrire les risques encourus en matière de traçabilité numérique puis assurer se comporter de façon vertueuse n'a pas suffi à convaincre. Avant d'installer l'application, chaque sujet a dû lire et signer un consentement détaillé

qui liste le type de données collectées, décrit la manière dont les données sont traitées et ce à quoi elles vont servir.

L'application a été perfectionnée pour proposer à l'utilisateur des rendus visuels sur ses traces d'activité. Cette contrepartie visuelle a pour ambition de rendre le dispositif plus séduisant et mieux accepté par l'utilisateur. Ceci étant, le risque est pris de perturber son activité ordinaire, dans la mesure où il sera en mesure d'interagir avec l'application. La production de traces publiques permet des effets de réflexivité. En ce sens, l'application peut être considérée comme ayant partie liée avec le champ de la quantification de soi (*quantified-self*)[11], puisque l'utilisateur a accès à des informations sur son comportement d'utilisation de l'ordiphone et peut donc potentiellement ajuster son comportement en conséquence (Pharabod *et al.*, 2013). Le phénomène de représentation de soi par la quantification peut mener à une conduite où l'on agit pour compter et pour contempler la trace chiffrée de son action. L'intégration de contreparties visuelles et du questionnaire à l'application fait le pari d'un dispositif d'étude scientifique des usages pouvant avoir de l'intérêt en lui-même pour les utilisateurs.

Conclusion

L'objectif de notre enquête exploratoire est de démontrer l'intérêt d'une connaissance publique et non marchande des usages des technologies numériques connectées, lorsqu'elle s'appuie sur la production automatisée de données issues d'ordiphones. Les sciences sociales comme la société civile pourraient utilement profiter de ce type de données. Les premiers résultats confirment la productivité scientifique d'une telle démarche. Il nous paraît ainsi souhaitable que progresse une forme de connaissance publique des usages des technologies numériques connectées, indépendamment des représentations véhiculées par le marché, mais respectueuse dans sa conception et sa mise en œuvre de la vie privée des participants.

Références bibliographiques

Achara J. P., J.-D. Lefruit, V. Roca, and C. Castelluccia (2013), "Detecting Privacy Leaks in the RATP App: how we proceeded and what we found", (to appear in Springer) *Journal of Computer Virolog and Hacking Techniques* (JCVHT).

Beck U. (2008), *La société du risque. Sur la voie d'une autre modernité*. Paris : Flammarion.

Boyd D. et K. Crawford (2011), "Six provocations for Big Data", paper presented at Oxford Internet Institute's "A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society", 21 septembre, http://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID1926431_code1210838.pdf?abstractid=1926431&mirid=1, consulté le 15 juin 2012.

Castelluccia C., M. A. Kaafar, and M.-D. Tran (2012), "Betrayed by Your Ads! Reconstructing User Profiles From Targeted Ads". – Information leakage through ads served in targeted advertising.

Castelluccia C., S. Grumbach and L. Olejnik (2013), "Data Harvesting 2.0: from the Visible to the Invisible Web", *The Twelfth Workshop on the Economics of Information Security*, Georgetown University, Washington, D.C., June 11-12.

[11] Il s'agit d'un phénomène auquel semble adhérer une partie des utilisateurs, puisque 19% des possesseurs d'ordiphones ont installé au moins une application dédiée au suivi de leur bien-être (CNIL, 2013). La mesure de soi incarnerait ainsi « un rapport nouveau entre les individus et leurs données », selon ce mouvement né en 2007 en Californie et impulsé par Gary Wolf, l'un des éditeurs de la revue technologique *Wired*.

Carrascal J.-P., C. Riederer, M. Cherubini, V. Erramilli, R. de Oliveira (2011), "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online", *International World Wide Web Conference*, 11p.

Duhigg C (2012), *The Power of Habit: Why We Do What We Do In Life and Business* (Random House, New York). – a major US retail network used customer shopping records to predict pregnancies of its female customers and send them well-timed and well-targeted offers.

Kessous E. (2012), *L'attention au monde. Sociologie des données personnelles à l'ère numérique*, 316p.

Kosinski M., D. Stillwell, and T. Graepel (2013), "Private traits and attributes are predictable from digital records of human behavior", *PNAS*.

Le Métayer D. (2013), "Privacy by Design: A Formal Framework for the Analysis of Architectural Choices", *ACM Conference on Data and Application Security and Privacy (CODASPY 2013),* pages 95-104.

Martin C., *Le téléphone portable et nous. En famille, entre amis, au travail,* coll. Communication et civilisation, L'Harmattan, Paris, 2007.

Pharabod A.-S., V. Nikolski et F. Granjon (2013), « La mise en chiffres de soi. Une approche compréhensive des mesures personnelles », *Réseaux*, 177, pp.97-129

Plantin J.-C. et L. Monnoyer-Smith (2012), « Pour une analyse critique de l'apport heuristique et méthodologique de la recherche numérique pour les SIC », *Congrès de la SFIC*.

Rey Bénédicte (2012), *La vie privée à l'ère du numérique*, Lavoisier, 297 p.

Rieder B. (2010), « Pratiques informationnelles et analyse des traces numériques : de la représentation à l'intervention », *Etudes de communication*, N°35, pp.91-104

Ressources complémentaires

Terms and conditions may apply (July 2013), http://tacma.net/

Terms of Service, Didn't Read: http://tosdr.org/

Wikipédia : PRISM programme de surveillance
https://en.wikipedia.org/wiki/PRISM_%28surveillance_program%29

Equipe Mobilitics - INRIA
http://planete.inrialpes.fr/~achara/mobilitics/
https://team.inria.fr/privatics/files/2013/09/Lemonde30Aout.pdf
https://team.inria.fr/privatics/paris-metro-tracks-and-trackers-why-is-the-ratp-app-leaking-my- private-data/

CREDOC, 2013, *La diffusion des technologies de l'information et de la communication dans la société française*, novembre, N°297, 288p. http://www.credoc.fr/pdf/Rapp/R297.pdf

CNIL (2011), « Smartphone et vie privée ». Enquête réalisée début novembre 2011 à la demande de la CNIL par l'institut Médiamétrie : http://www.cnil.fr/fileadmin/documents/La_CNIL/publications/DEIP/CNIL-Etude-SMARTPHONES-2011.pdf

CNIL (2012), « Smartphones et vie privée : pour une nouvelle vision de la protection des données ? », Lettre *Innovation & Prospective*, N°02, février

Wall Street Journal, 2010, "Your Apps Are Watching You!", edited December 17th, 2010, by Scott Thurm et Yukari Iwatani Kane.