Les traces de l'activité humaine dans le numérique

Melanie Dulong de Rosnay

Melanie Dulong de Rosnay, Les traces de l'activité humaine dans le numérique, in Mokrane Bouzeghoub & Rémy Mossery (dir.), *Les Big Data à Découvert*, CNRS Editions, mars 2017, p. 80.

Selon une typologie du Forum Économique Mondial, les traces de l'activité humaine dans le numérique désignent les données créées par et à propos des personnes. Elles comprennent : les données volontaires, créées et partagées explicitement par les individus, par exemple les commentaires et les contributions sur les réseaux sociaux ; les données observées, issues de l'enregistrement involontaire des activités, par exemple les données de géolocalisation des utilisateurs à partir de leur téléphone, l'historique de navigation ; et les données inférées à propos des individus, basées sur l'analyse de nos données volontaires ou observées, par exemple l évaluation de la solvabilité, la déduction de l'âge et de l'origine sociale à partir du prénom.

Le traitement de nos traces par les plateformes prédatrices

Le traitement des traces de l'activité humaine mène à une logique d'encodage de règles dans les dispositifs numériques et algorithmiques, qui peuvent tirer des conclusions et prendre des décisions affectant nos vies (cf. VIII.15), sur la base des informations que nous laissons, consciemment ou non, sur les plateformes, les réseaux, les objets connectés et les villes intelligentes. Collectées à des fins de marketing et de profilage (cf. VII.2), nos données rémunèrent les plateformes, en échange d'une utilisation gratuite de leurs services.

Le droit européen requiert le consentement explicite et informé des citoyens avant tout traitement de données personnelles, et pose des conditions d'accès, de rectification et de retrait de ces données (cf. VIII.6). Les informations communiquées doivent être transparentes et compréhensibles pour les utilisateurs. Mais la régulation algorithmique* renforce le pouvoir des plateformes sur les utilisateurs, qui n'ont pas toujours la possibilité de faire appel des décisions ou de comprendre le processus de traitement. Nos choix et activités réduisent le champ des articles proposés par les réseaux sociaux dans la curation des liens qui nous sont présentés sur Facebook, renforçant nos préférences. Les données sur notre santé ou notre conduite automobile issues des objets connectés (cf. VIII.7), sportifs ou domestiques, pourront renseigner les assureurs et influer sur le montant de nos cotisations (cf. VII.5). L'impact du croisement des données est méconnu et sous-estimé ; c'est parfois l'agrégation de deux sources de données anodines de manière isolée qui révélera une information et rendra un individu vulnérable à la discrimination.

Les réseaux sociaux, les applications mobiles et les objets connectés ne sont pas seuls à recueillir et analyser nos données personnelles. Les administrations utilisent des plateformes pour calculer l'impôt, orienter les étudiants, affecter les enseignants, effectuer des statistiques, offrir des données ouvertes... Les règles d'interprétation doivent être portées à la connaissance du sujet d'une décision qui en ferait la demande, mais si le code source de l'algorithme est fermé, les chercheurs en informatique et les militants des droits numériques ne pourront pas mener d'expertise indépendante. Les chercheurs en sciences sociales se retrouvent donc avec une masse de données issues des plateformes publiques et privées disponibles en ligne, mais ils ne peuvent pas les exploiter sans l'autorisation formelle des contributeurs, ni dans le cas de conditions contractuelles restrictives (comme Twitter). Les potentialités en terme de modélisation des comportements, des relations, de la formation de l'opinion publique sont immenses, et les conséquences pour la vie privée peu maîtrisées, dans la mesure où des données même anonymisées pourraient être agrégées avec d'autres et permettre la ré-identification des individus (cf. VIII.5). Ainsi, la mise à disposition par Wikipédia d'une interface de programmation (API*) permettant de télécharger l'historique des contributions est très utile pour les chercheurs, car elle facilite le traitement de l'ensemble des contributions à un article et l'analyse des controverses. Cependant, un effet secondaire de cette fonctionnalité est, par le traitement automatisé d informations certes publiques, de révéler les centres d'intérêts et les opinions politiques des contributeurs.

Vue d'ensemble des utilisations possibles des traces laissées par l'utilisateur sur sa vie privée.
G. PETKOS et al., Deliverable D6.1, USEMP privacy scoring framework, 2015, p. 19.

http://www.usemp-project.eu/wp-content/uploads/2015/05/usemp_deliverable_d6.1.pdf

#	Type d'information	Description	Utilisations possibles	Valeur (pour les annonceurs)
A	Démographie	Données personnelles, telles que le sexe, l'âge, la nationalité, l'origine ethnique…	Discrimination. Il s'agit du type d'information le plus réutilisé	**Haute.** Pour les annonceurs qui veulent cibler des utilisateurs avec des critères démographiques spécifiques
B	Traits psychologiques	Définis par les psychologues (extraversion, ouverture…)	Discrimination, par exemple dans lors d'un entretien d'embauche	**Basse**
C	Profil sexuel	Statuts maritaux, préférences, habitudes	Discrimination, par exemple dans le milieu du travail, de l'éducation, du logement	**Haute.** Pour les annonceurs qui veulent cibler les consommateurs en fonction de leur statut marital et de leur mode de vie sexuel
D	Opinions politiques	Soutien à des politiques, adhésion à un parti, militantisme	Discrimination, par exemple sur le lieu de travail ou lors d'un entretien d'embauche	**Haute**. Pour les annonceurs qui veulent cibler les consommateurs en fonction de leurs opinions politiques
E	Croyances religieuses	Religion déclarée (le cas échéant) et les croyances	Discrimination, par exemple pour la vente ou la location de logements, pour l'embauche ou sur le lieu de travail	**Modérée.** Pour les annonceurs qui veulent cibler les consommateurs en fonction de leurs croyances religieuses et culturelles
F	Éléments de santé	Addictions (tabagisme, consommation d'alcool…), état de santé, maladies, condition physique (exercices sportifs)	Discrimination, comme le refus d'octroi d'une couverture sociale ou la pratique de prix discriminatoires	**Haute.** Pour les annonceurs qui veulent cibler les consommateurs en fonction de leurs habitudes
G	Localisation	Caractéristiques de l'endroit et traces des anciens lieux d'habitation	Discrimination, par exemple pour l'assurance habitation, harcèlement	**Haute.** Pour les annonceurs qui veulent cibler les consommateurs en fonction de leur localisation ou de leur lieu d'habitation

H	Profil consommateur	Produits et marques préférés	Publicité ciblée et discrimination dans l'affichage des prix en ligne	**Haute.** Pour les annonceurs qui veulent cibler les consommateurs en fonction de leurs habitudes de consommation, et des objets qu'ils utilisent pour se connecter à Internet

Les réponses de la société civile

Face à ces risques, l'utilisateur peut adopter des outils d'anonymisation développés par la société civile pour masquer les traces. Des plug-ins pour les navigateurs bloquant la traçabilité aux moteurs de recherche distribués et applications en pair-à-pair, les alternatives techniques renforçant la privacy-by-design sont nombreuses (cf. III.10). Cependant, elles peuvent être difficiles d'accès, leurs interfaces sont peu ergonomiques et elles demandent d'accomplir une démarche pour sortir des plateformes prédatrices, choix pouvant mener à une exclusion de certains services liés.

Des projets de recherche tentent de sensibiliser les utilisateurs et de développer une littératie numérique et une culture de la vie privée. L'éducation et le développement de connaissances en matière de vie privée (figure) peuvent prendre la forme de sites illustrant comment les réseaux sociaux peuvent réutiliser les données inférées en les contextualisant. Ainsi, l'application Databait du projet européen USEMP montre aux utilisateurs qui donnent accès à leur profil quelles informations peuvent être extraites à partir des données laissées sur les plateformes, et permet de prendre conscience de l'étendue des traces révélées et de la valeur des données personnelles.

En l'absence d'une régulation étatique protectrice et de mécanismes de réappropriation collective, l'autorégulation des plateformes et l'éducation du public restent des solutions limitées face à l'impact du traitement de nos traces numériques sur nos vies et notre libre-arbitre.

Références bibliographiques

M. DULONG DE ROSNAY – « Vie privée et données personnelles », in D. FRAU-MEIGS et A. KIYINDOU (dir.), *La diversité culturelle à l'ère du numérique : glossaire critique*, La Documentation Française, 2014. <halshs-01078704>
M. DULONG DE ROSNAY – *Les Golems du numérique. Droit d'auteur et Lex Electronica*, Presses des Mines, 2016. <halshs-01303291>
L. MERZEAU – « L'intelligence des traces », *Intellectica*, 2013. <halshs-01071211>
T. VENTURINI – « Great expectations. Méthodes quali-quantitative et analyse des réseaux sociaux », in J.-P. FOURMENTRAUX (dir.), *L'ère postmedia. Humanités digitales et cultures numériques*, Hermann, 2012. <hal-01064258>

Glossaire

API. Interface de programmation applicative. Porte d'accès à un logiciel ou une application offrant des fonctionnalités provenant d'autres programmes.

Régulation algorithmique. Prise de décision automatisée sur la base du traitement de données personnelles agrégées

Pair-à-pair. Modèle informatique basé sur une architecture distribuée, répartissant les tâches entre les pairs sans passer par un serveur central

Littératie numérique. Compétences et compréhension du monde numérique. A comparer avec alphabétisation.