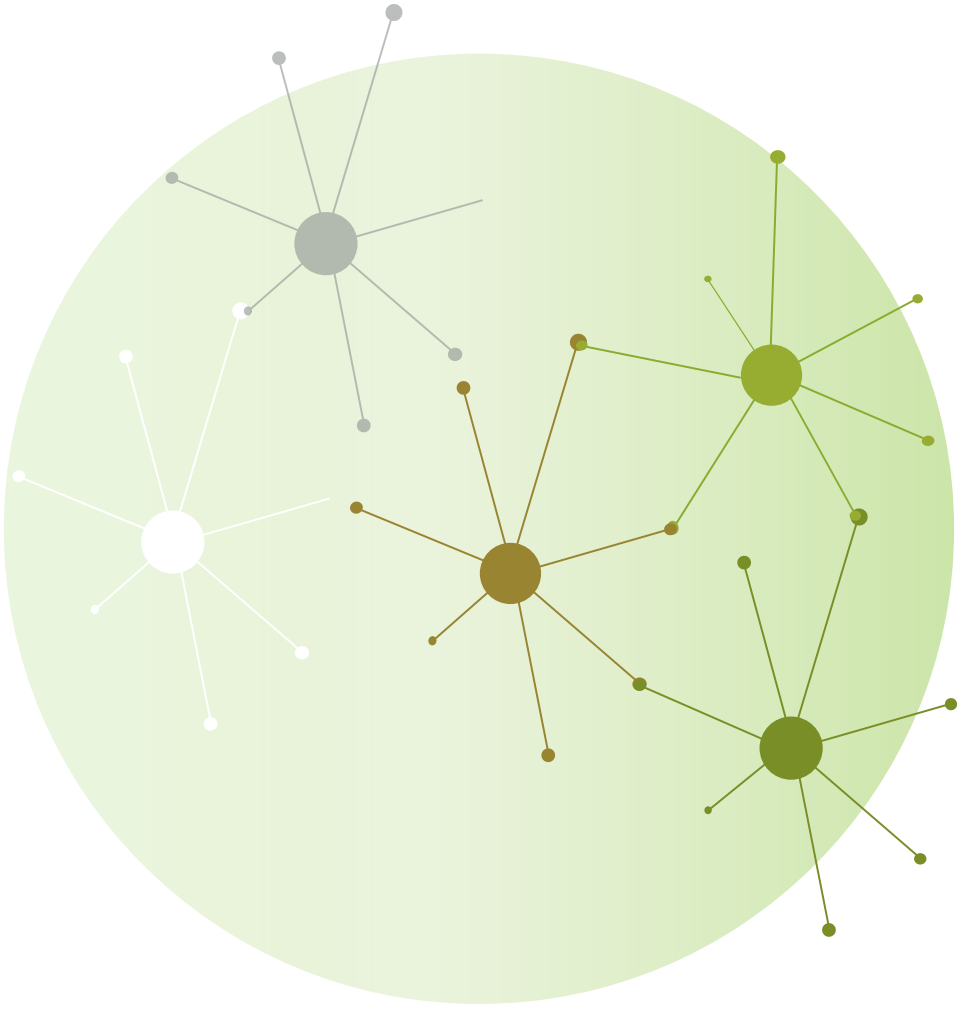
DONNÉES ET SERVICES

DIRECTION DE LA PROSPECTIVE ET DU DIALOGUE PUBLIC

DONNÉES PERSONNELLES : ENJEUX, LIMITES, PERSPECTIVES

Synthèse de Hubert GUILLAUD, Marine ALBARÈDE et Manon MOLINS

Octobre 2016



Fing

GRANDLYON la métropole

Données personnelles : enjeux, limites, perspectives

Dossier de synthèse coordonné par Hubert Guillaud, Marine Albarède et Manon Molins pour la Fing. 07/102016

Les données que les systèmes techniques recueillent sur leurs utilisateurs sont au cœur des évolutions des rapports entre les organisations et leurs usagers. Leurs collectes, leurs traitements, leurs croisements, leurs analyses sont appelés à façonner les relations à venir entre entreprises et clients, entre administrations et usagers, entre politiques et citoyens.

Malgré toutes les limites que l'on peut souligner dans la métaphore des données comme le nouvel or noir[1] de l'économie numérique, celles-ci sont incontestablement appelées à en être le carburant, dont la valeur repose plus dans la circulation et le traitement que dans le stockage. Elles ne sont pas qu'un patrimoine à faire fructifier, elles sont les objets mêmes des échanges. "Les données - en particulier les données personnelles - sont au cœur de tous les modèles d'affaires de l'économie numérique[2]". Or, la définition des modalités qui vont présider à cet échange a un impact total sur l'économie et les formes relationnelles qui seront possibles demain. La manière dont nous définissons les données personnelles, dont nous les traiterons, dont nous les ouvrirons ou les partagerons auront des conséquences radicalement différentes sur le futur des rapports entre les gens et les organisations.

En ce sens, la place des collectivités et des administrations - et par-delà elles, de la politique - dans la façon dont elles vont définir les modalités d'accès et de partage des données, représente un enjeu stratégique primordial. Les modèles, les visions, les stratégies qu'elles vont porter auront un impact direct sur l'économie du futur, sur les rapports à venir entre les citoyens et leurs institutions et les citoyens entre eux.

Tel est l'enjeu de cette note de synthèse : permettre de comprendre et cerner les enjeux à venir pour les collectivités dans le rapport à ces nouveaux objets que sont les données personnelles.

[1] Voir notamment : http://www.henriverdier.com/2013/03/non-les-donnees-ne-sont-pas-du-petrole.html et https://donneesouvertes.info/2013/10/04/la-donnee-une-matiere-premiere-bien-etrange/

[2] Rapport dit Colin et Collin, "Mission d'expertise sur la fiscalité des données numériques", 2013 : http://www.economie.gouv.fr/files/rapport-fiscalite-du-numerique_2013.pdf

SOMMAIRE

1. Qu'est-ce qu'une donnée personnelle ?

Au sens de la loi[3], une donnée personnelle est une information qui permet d'identifier une personne physique, directement ou indirectement.

"Cela concerne aussi bien les informations directement nominatives (le nom et le prénom), que les informations qui permettent d'identifier, indirectement, une personne physique. C'est le cas d'un numéro de téléphone (qui permet d'identifier le titulaire d'une ligne téléphonique), d'un numéro de plaque minéralogique (qui renvoie au titulaire de la carte grise). C'est également le cas d'éléments du corps humain tels que l'empreinte digitale ou l'ADN d'une personne."[4] C'est bien sûr également le cas d'une adresse e-mail, d'un numéro de carte bancaire, d'une adresse physique, du numéro de sécurité sociale...

Comme l'explique l'économiste Fabrice Rochelandet[5], les données personnelles sont des informations qui sont propres à chacun et qui permettent de distinguer une personne au sein d'un groupe social donné.

On dit qu'une personne est identifiée, lorsque son nom apparaît directement dans un fichier et qu'elle est identifiable lorsqu'un fichier comporte des informations permettant indirectement son identification : comme son adresse IP[6], son numéro d'immatriculation, sa photographie, sa résidence, sa profession, son sexe, son âge, son numéro de carte de paiement, son numéro de sécurité sociale, un enregistrement de sa voix…

1.1 Les données personnelles

Ce que montre cette définition très large des données personnelles, c'est qu'il n'y a pas une seule forme de données personnelles, clairement catégorisable, identifiable et délimitable, mais une multitude de données à caractère personnel.

Données objectives, subjectives, sensibles, "anonymes"...

---

[3] "*Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne.*" Article 2 de la loi modifiée du 6 janvier 1978 dit "Informatique et Libertés" : https://www.cnil.fr/fr/loi-78-17-du-6-janvier-1978-modifiee#Article2

[4] http://www.cil.cnrs.fr/CIL/spip.php?article1586

[5] Rochelandet (Fabrice), *Économie des données personnelles et de la vie privée*, "Repères", La Découverte, 2010.

[6] La question de savoir si l'adresse IP d'un utilisateur est une donnée personnelle est encore débattue. La CNIL considère que, parce qu'elle permet d'identifier tout ordinateur connecté au réseau (et donc la personne physique titulaire de la ligne), elle doit être considérée comme une donnée à caractère personnel.

Dans son ouvrage, Fabrice Rochelandet distingue des données objectives (données d'identification et de contact, données physiques et de signalement, données socio-démographiques, données juridiques et données financières et foncières) des données subjectives (préférences et centre d'intérêt, opinions et activités politiques, religieuses ou syndicales, données comportementales, données géographiques et données relationnelles). La CNIL, elle, distingue plutôt les données personnelles des données sensibles. Les données sensibles étant celles qui fournissent des informations sur l'origine ethnique des individus, leurs opinions politiques ou religieuses, leur orientation sexuelle ou leur état de santé. Par principe, leur collecte et traitement sont interdits sauf autorisation explicite de la personne concernée et traitements justifiés par l'intérêt public et autorisés par la CNIL. La CNIL distingue d'ailleurs d'autres données à risque, comme les données génétiques, les données relatives aux infractions pénales et aux condamnations, les données comportant des appréciations sur les difficultés sociales des personnes, etc. Les données personnelles portent donc différents niveaux d'information que la société décrète comme plus ou moins problématiques.

Le fait qu'elles soient "anonymes" (c'est-à-dire qu'elles ne comportent pas de données d'identification ou de contact ou de données identifiables) ne signifie pas que les données ne soient pas personnelles. Des données considérées comme anonymes peuvent constituer des données à caractère personnel si elles permettent d'identifier une personne indirectement ou par recoupement d'information. L'historique des requêtes d'utilisateurs distincts sur un moteur de recherche peut ainsi permettre d'identifier indirectement des utilisateurs, comme le rappelait la célèbre affaire des logs d'AOL[7] qui a permis d'identifier des utilisateurs en analysant simplement les requêtes qu'ils tapaient sur le moteur de recherche d'AOL.

> Les journalistes du *New York Times*[8] sont par exemple parvenus à identifier l'utilisateur "4417749", une veuve de 62 ans, grâce à la liste de ses requêtes (par exemple : *"chien qui fait pipi partout"*, *"taxe foncière de Harrisburg, Virginie"*, *"solitude"*, *"Paranoïa"*, *"Thé pour une bonne santé"*, etc.).

L'identification peut également se faire par simple recoupement quand l'indication d'une profession et d'un code postal ou d'une date de naissance et d'une commune de résidence peuvent permettre d'identifier une seule et même personne par exemple, comme le notaire de Trifouillis-les-oies.

Par nature, les technologies de l'information et de la communication génèrent des enregistrements distincts, et donc de nombreuses "traces informatiques" ou métadonnées

[7] Les logs sont le journal des enregistrements des utilisateurs sur un site internet, où les enregistrements de chaque utilisateurs sont séparés les uns des autres : http://www.internetactu.net/2006/09/07/a-qui-appartiennent-mes-logs/

8 http://www.nytimes.com/2006/08/09/technology/09aol.html?i=5090&en=f6f61949c6da4d38&ex=1312776000&adxnnl=1&partner=rssuserland&emc=rss&adxnnlx=1155096385-A78mK3ngRfE6mUfYTIcDcA

facilement exploitables, qui peuvent devenir autant de données personnelles (un appel passé par un téléphone portable, une connexion à Internet, un retrait ou un paiement via une carte bancaire...). La définition très large des données personnelles permet de considérer une donnée anonyme ou anonymisée[9] comme personnelle si cet anonymat est réversible, c'est-à-dire si les traces permettent de remonter jusqu'à la personne en question, comme c'est le cas de la personne 4417749 dans les logs d'AOL.

Des données personnelles aux métadonnées

Les métadonnées sont des informations liées à une donnée : l'horaire d'un appel téléphonique est l'une des métadonnées liées à un appel téléphonique par exemple. Ces "traces" correspondent à tous les éléments autres que le contenu d'un message. Si le texte de votre e-mail est le contenu, les métadonnées sont toutes les informations qui permettent à cet e-mail d'être acheminé d'un expéditeur à son destinataire : ce qui comporte notamment l'horaire d'envoi, l'adresse de l'expéditeur et du destinataire, le titre, le format, etc. Les métadonnées des appels téléphoniques mobiles par exemple fournissent nombre d'indications très personnelles sur les gens avec qui vous êtes en relation, les heures où vous les avez appelés ou bien les heures où vous avez été appelé, votre localisation liée à l'antenne où s'est allumé votre téléphone, etc. Un simple tweet pouvant comporter un message de 140 caractères, porte avec lui plus d'une trentaine de métadonnées[10].

Pour la CNIL, ces métadonnées sont des données personnelles[11].

[9] Les techniques d'anonymisation des données reposent sur 3 critères cumulatifs : l'impossibilité d'isoler une personne (les données doivent porter sur des groupes de personnes) ; l'impossibilité de faire des liens (les données d'un même jeu ou de plusieurs jeux portant sur le même groupe ne doivent pas pouvoir être reliées entre elles) ; l'impossibilité d'inférer (aucune information ne doit pouvoir être déduite à partir de la connaissance des données d'un ou plusieurs jeux).
https://fr.wikipedia.org/wiki/Anonymat

[10] http://readwrite.com/2010/04/19/this_is_what_a_tweet_looks_like/

[11] http://lexpansion.lexpress.fr/high-tech/loi-renseignement-la-cnil-dement-cazeneuve-sur-le-recueil-de-donnees-anonymes_1671961.html

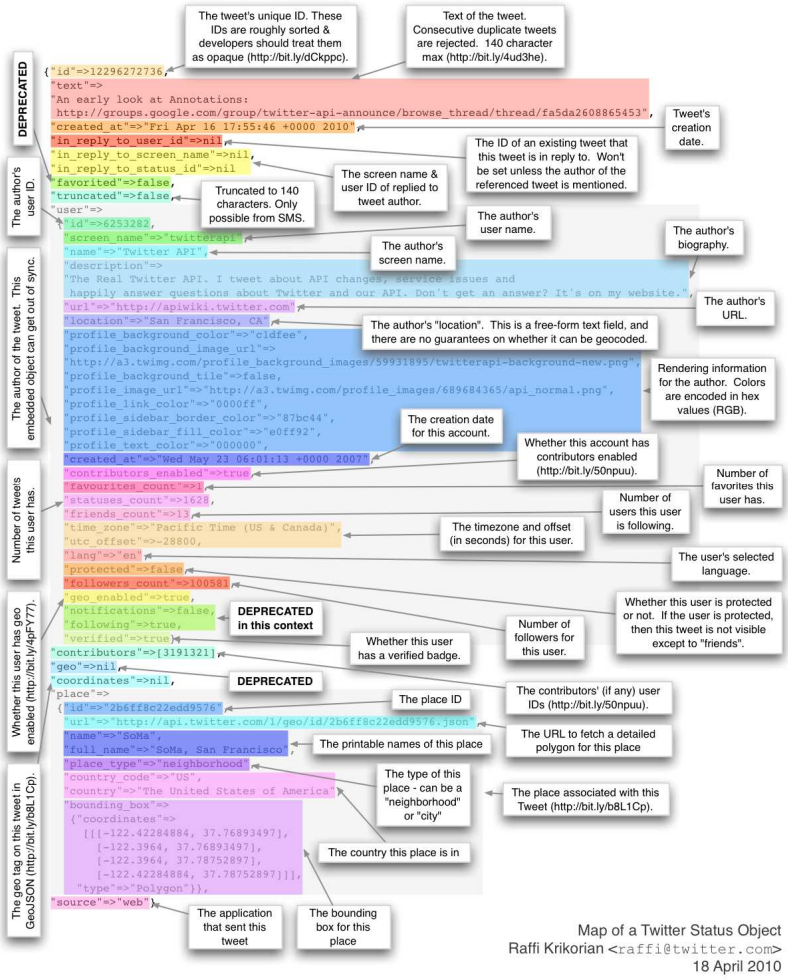
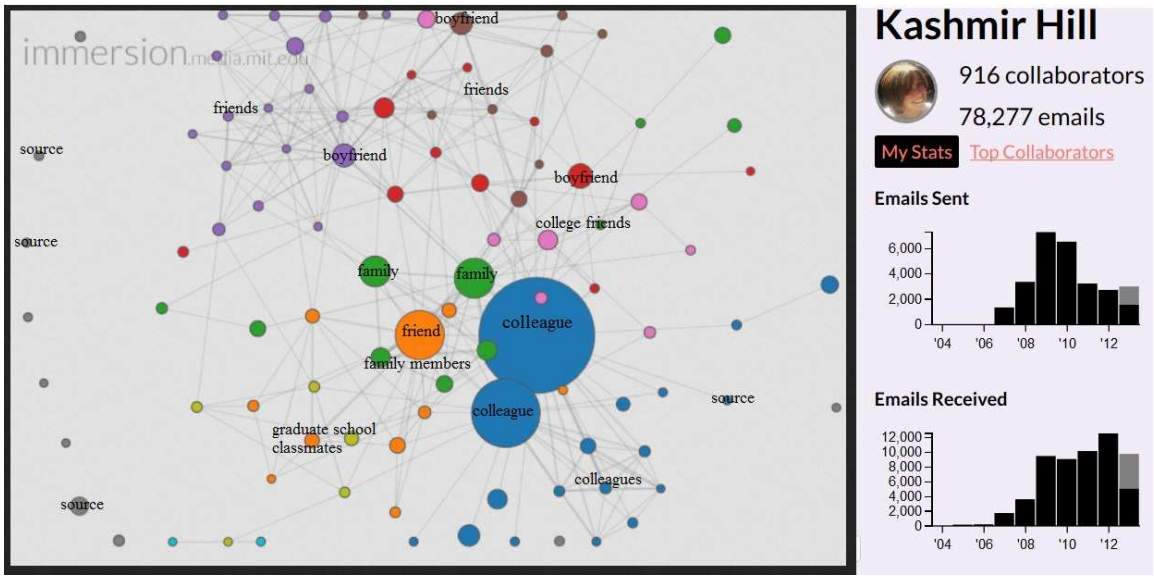


Image : détail des métadonnées d'un tweet par Rafi Krikorian, 2010.



Immersion : le MIT a développé un logiciel sécurisé qui vous permet de regarder les métadonnées de l'ensemble de vos e-mails, vous permettant par exemple de voir avec quels interlocuteurs vous échangez le plus ces dernières semaines ou depuis plusieurs années : https://immersion.media.mit.edu

Impossibles taxonomies des traces

Dans son livre, *Data and Goliath*, le spécialiste de la sécurité, Bruce Schneier distingue 6 types de traces selon la manière dont elles ont été créées :

- les données de service (les informations que vous fournissez pour obtenir un service, par exemple toutes les données que détient un commerçant à votre inscription);
- les données révélées (les données que vous partagez consciemment sur le web en en réglant les modalités d'accès);
- les données confiées (les données que l'on poste sur une plateforme comme Facebook ou Twitter et qu'on ne contrôle pas nécessairement);
- les données annexes (les données partagées par d'autres qui nous identifient ou nous mentionnent, à l'image d'une photo publiée par un tiers et vous identifiant nommément);
- les données de comportement (qui sont créées par notre interaction entre nous et nos machines et services : typiquement vos comportements de navigation sur un site web ou vos comportements d'achats);
- les données dérivées (qui sont des données nous concernant résultant d'autres données, à l'image des profils que dressent de nous les sociétés publicitaires depuis toutes les autres données récoltées).

Fabrice Rochelandet quant à lui tente d'inventorier tous les modes de collecte (formulaires, informations divulguées publiquement, traces collectées à l'insu de la personne, informations divulguées par des tiers, données obtenues par recoupement) et d'exploitation des données personnelles (exposition de soi, recherche d'information et surveillance, revente de données, personnalisation, offre contextualisée, exploitation marketing, utilisations illégales…), qui ne peuvent être qu'incomplètes à mesure que les techniques s'affinent et se développent.

Implications en terme de régulation

Tous les dispositifs connectés, par nature, émettent des traces et génèrent donc des données.

Les instances de régulation, notamment la CNIL, insistent sur plusieurs principes :

- Les collectes et traitement doivent être déclarés aux autorités de contrôle.

- Le consentement des personnes concernées par le traitement des données personnelles doit être recueillies préalablement à ce traitement. Or, celui-ci se résume trop souvent à une acceptation de très longues et factuelles conditions générales d'utilisation - CGU -, ce qui n'est pas, comme on nous le présente souvent, un compromis résultant d'un arbitrage équitable, librement consenti et mutuellement avantageux, mais d'une résignation du public comme l'expliquait la remarquable étude[12] de Joseph Turow, Michael Hennessy et Nora Draper. Ce recueil du consentement nécessite donc une transparence sur les traitements auxquels seront soumis les données concernant les personnes. Cela va au-delà des droits d'accès, de rectification et d'opposition dont disposent les utilisateurs, puisque l'on devrait notamment demander aux utilisateurs le droit d'utiliser leurs données pour toute nouvelle forme d'utilisation.

- Les entreprises, services et objets qui procèdent à des collectes de données doivent limiter la collecte et la conservation des données de leurs utilisateurs aux volumes et catégories nécessaires à l'exercice de leurs activités. C'est ce qu'on appelle la "loyauté de la collecte" ou "principe de finalité", qui doit être proportionnée et pertinente par rapport à l'usage prévu. Le vrai problème est de parvenir à définir la pertinence et la proportion et d'en laisser l'appréciation aux seules entreprises qui opèrent le service[13].

- Enfin, les responsables des collectes et traitements doivent prendre toutes les mesures pour préserver la sécurité des données, empêcher qu'elles soient déformées, endommagées ou que des tiers non autorisés y aient accès. Le problème est qu'à mesure que les systèmes se numérisent, que les données sont collectées dans des bases de données connectées, les affaires de fuite et de piratage de données de plus en plus massives se démultiplient[14].

---

[12] http://www.internetactu.net/2015/06/11/donnees-personnelles-limpuissance-nest-pas-le-consentement/

[13] Une étude récente menée par le journal Vocativ sur la collecte des données par les applications pour smartphone montrait que plus de la moitié des 25 applications les plus populaires de Google Play demande l'accès à vos contacts. Un tiers demandent l'accès aux autorisations les plus alarmante à savoir, l'accès aux contacts, l'accès aux SMS, l'accès au journal d'appel et l'accès au microphone… http://www.internetactu.net/a-lire-ailleurs/107395494784/ Pour se moquer de ces principes, un magazine danois s'était amusé à faire une caméra cachée pour montrer ce que donnerait ces demandes d'accès si elles étaient le fait d'un boulanger à ses clients : https://www.youtube.com/watch?v=xYZtHIPktQg

[14] http://www.internetactu.net/a-lire-ailleurs/103452338364/

Dans le domaine des objets connectés, les législateurs ont introduit les notions de *privacy by design* (protection des données dès la conception) et *privacy by default*[15] (sécurité par défaut). Le premier suppose que les questions de respect de la confidentialité et de protection des données soient intégrées dès la conception d'un objet ou service connecté. La *privacy by default* signifie que par défaut le service doit être réglé sur le niveau le plus protecteur pour l'utilisateur, à l'image de systèmes de personnalisation sans identification[16].

Image : les 7 principes de la sécurité par défaut mise en image par la CNIL australienne[17] : à savoir : la vie privée intégrée dès la conception, le respect de l'utilisateur, la visibilité et la transparence, des fonctionnalités pleines et entières, une sécurité de bout en bout, la vie privée comme fonctionnalité par défaut, et la vie privée comme principe pro-actif et préventif.

Reste que la mise en œuvre de ces principes n'est pas si simple, pour les organisations comme pour les collectivités. Cela suppose d'éclairer les enjeux de la collecte de données,

[15] via http://lexpansion.lexpress.fr/actualite-economique/droit-des-donnees-le-paradoxe-des-objets-connectes_1795051.html

[16] http://www.internetactu.net/2007/10/22/personnalisation-sans-identification/

[17] http://prov.vic.gov.au/government-recordkeeping/privacy-by-design-a-new-approach-to-privacy-in-government

d'identifier les mauvaises pratiques, de rappeler les principes, de sensibiliser et de former aux règles à respecter. "La sensibilisation est un travail de longue haleine", expliquait Marie-Laure Baron, correspondante informatique et libertés au Conseil départemental de la Charente-Maritime qui a mis en place un dispositif de ce type[18]. Comme l'explique Charles Nepote[19], responsable du programme Infolab de la Fing[20], dans un monde où la donnée est partout, il faut se préoccuper de la culture de la donnée de tous.

Le règlement européen sur la protection des données[21] adopté le 14 avril 2016 et qui sera applicable en 2018 dans tous les pays membres de l'Union européenne renforce les droits des citoyens, leur donne plus de contrôle sur leurs données et offre un cadre juridique unifié pour les entreprises. Pour les organisations, ce règlement nécessite qu'elles se mettent à jour notamment :

- en introduisant un consentement explicite et positif des utilisateurs ;
- il introduit un droit à l'effacement (droit à l'oubli) ;
- un droit à la portabilité des données (c'est-à-dire que l'utilisateur puisse les récupérer dans un format exploitable pour s'en servir lui-même ou auprès d'autres services) ;
- un droit de refus du profilage ;
- elle renforce la *privacy by design* et *by default* ;
- introduit un droit de notification des utilisateurs en cas de fuite de données…
- ainsi que la réalisation d'études d'impacts pour les responsables des traitements
- et renforce les sanctions.

Diminuer la collecte ou renforcer la protection ?

Des associations de défense des libertés civiles comme l'Electronic Frontier Fondation[22] prônent des politiques encore plus radicales : diminuer la collecte de données[23] ; les chiffrer pour renforcer leur confidentialité et rendre les échanges plus difficiles ; limiter la durée de conservation et prôner un effacement instantané[24] ; rendre la collecte transparente pour les humains et pas seulement pour les machines pour que les utilisateurs puissent mieux comprendre ce qui s'échange et mieux s'y opposer ou le réguler.

---

[18] http://www.lagazettedescommunes.com/457312/charente-maritime-les-agents-sensibilises-a-la-protection-des-donnees-a-caractere-personnel/

[19] http://www.usine-digitale.fr/article/organisations-pourquoi-vous-devez-vous-preoccuper-de-la-culture-data-de-vos-collaborateurs.N435337

[20] http://fing.org/infolab et http://infolabs.io

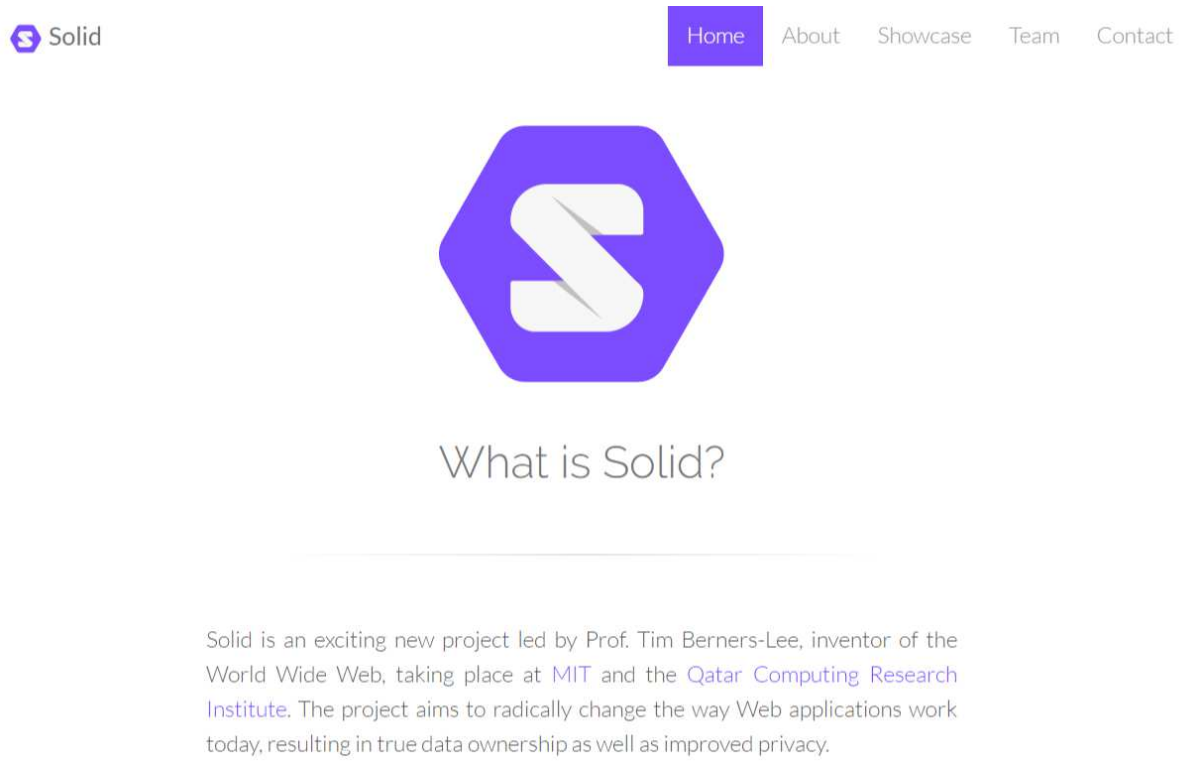
[21] Voir https://www.cnil.fr/fr/adoption-du-reglement-europeen-par-le-parlement-europeen-un-grand-pas-pour-la-protection-des-donnees et https://fr.wikipedia.org/wiki/R%C3%A8glement_g%C3%A9n%C3%A9ral_sur_la_protection_des_donn%C3%A9es

[22] http://www.internetactu.net/2009/10/26/critiques-du-web%c2%b2-44-que-faire-face-a-la-puissance-des-donnees/

[23] Comme par exemple par le biais des systèmes de personnalisation sans identification : http://www.internetactu.net/2007/10/22/personnalisation-sans-identification/ Ou les méthodes d'obfuscation consistant à altérer les données : http://www.internetactu.net/2014/10/22/lobfuscation-larme-du-faible/

[24] Comme par exemple l'ajout de dates de péremption pour les données : http://www.internetactu.net/2015/10/20/faut-il-une-date-de-peremption-pour-les-donnees/

Pour autant, peut-on se passer des données ? Les entreprises, les administrations et les chercheurs : toutes les composantes de la société estiment que le progrès de la société est lié à l'exploitation des données[25]. Pour autant, même les plus ardents défenseurs de l'exploitation des données sont conscients de la nécessité de formes de régulation s'appuyant sur un usage plus respectueux des données[26] qu'elles ne le sont actuellement. De nombreux projets de recherche tentent d'esquisser des solutions solides pour remédier à la prédation des données des utilisateurs, à l'image de protocoles et de standards d'échanges de données comme openPDS[27], l'OpenMustard Seed[28], MaidSafe[29], ou plus encore Solid[30], un projet du MIT mené par Tim Berners-Lee, l'inventeur du web lui-même. Solid est un projet qui vise à séparer les données des applications et des serveurs qui les utilisent, afin que les utilisateurs retrouvent le contrôle de leurs données. Si ces projets sont encore des ébauches, ils permettent d'envisager de changer l'internet en apportant des technologies qui n'étaient pas disponibles à l'origine, permettant d'offrir à la fois la commodité d'usage et la décentralisation des données, c'est-à-dire d'utiliser les services d'aujourd'hui, leur permettre d'utiliser les données des utilisateurs, sans qu'ils puissent se les accaparer[31]. Derrière toutes ces solutions, la question du chiffrement des données se pose avec insistance comme une solution sérieuse à la maîtrise.



[25] Voir notamment Alex Pentland, *Social Physics*, 2014 : http://www.internetactu.net/2014/05/20/big-data-vers-lingenierie-sociale/
[26] http://www.internetactu.net/2013/07/03/dautres-outils-et-regles-pour-mieux-controler-les-donnees/
[27] http://openpds.media.mit.edu/
[28] http://idhypercubed.org/wiki/
[29] http://maidsafe.net/
[30] https://solid.mit.edu/
[31] http://www.internetactu.net/a-lire-ailleurs/internet-3-0-peut-on-reprendre-le-controle-des-geants/

1.2 Données en réseau

Si dans la question des données personnelles nous ne regardons que les données, il nous manque une partie de l'équation permettant de comprendre les transformations en cours. Les possibilités de recoupement et de traitement (les techniques de croisement et d'exploitation) ne cessent d'évoluer et élargissent encore le champ de ce qui relève des informations personnelles.

L'impact de la modélisation

Reprenons l'exemple des métadonnées de votre téléphone mobile. Nous avons vu qu'elles donnent des informations sur votre réseau social : qui vous appelle et qui appelez-vous… permettant de tracer votre graphe social[32], c'est-à-dire à la fois de visualiser l'ensemble de vos correspondants et ceux avec lesquels vous échangez le plus, donc ceux dont vous êtes le plus proche. L'exploitation des informations géographiques liées aux antennes auxquelles se connecte votre téléphone mobile pour joindre le réseau téléphonique donne le détail des horaires de vos déplacements et trace vos parcours géographiques avec une grande précision. Reliées aux mêmes métadonnées de vos correspondants, ces informations permettent d'inférer des rencontres avec des personnes ou des lieux.

Mais l'exploitation de ces métadonnées de connexion peut indiquer encore bien plus. Le MIT a créé un modèle croisant les métadonnées de connexion et un test psychologique[33], permettant de déduire des caractéristiques psychologiques depuis des données de mobilités. Leur modèle utilise 36 indicateurs provenant de l'analyse des métadonnées d'usage des téléphones (localisation, usage du téléphone, régularité, diversité des contacts, activité - comme le temps mis à répondre à un texto…) permettant de prédire les caractéristiques comportementales de n'importe quel abonné. Le modèle est relativement fiable : il est capable à partir des données de mobilité de prédire votre profil psychologique selon des tests éprouvés… Cela signifie qu'à partir d'un simple profil d'usage de votre téléphone depuis les métadonnées récupérées d'une énorme base de données où chacun paraît protégé par la masse, on peut en déduire finement vos caractéristiques psychologiques, comme votre performance au travail ou la capacité à prendre des décisions d'achat… c'est-à-dire des choses qui n'ont rien à voir avec l'usage de votre mobile a priori.

> Votre personnalité se dévoile dans le moindre de vos comportements et à l'heure où tous nos comportements sont enregistrés, nos personnalités sont dans toutes les traces de nos activités. Déplacements, transactions, horaires, appréciations… L'activité est désormais une alternative à l'identité.

[32] http://www.internetactu.net/2007/09/28/comprendre-le-graphe-social/
[33] http://www.internetactu.net/2013/12/05/en-quoi-les-big-data-sont-elles-personnelles/

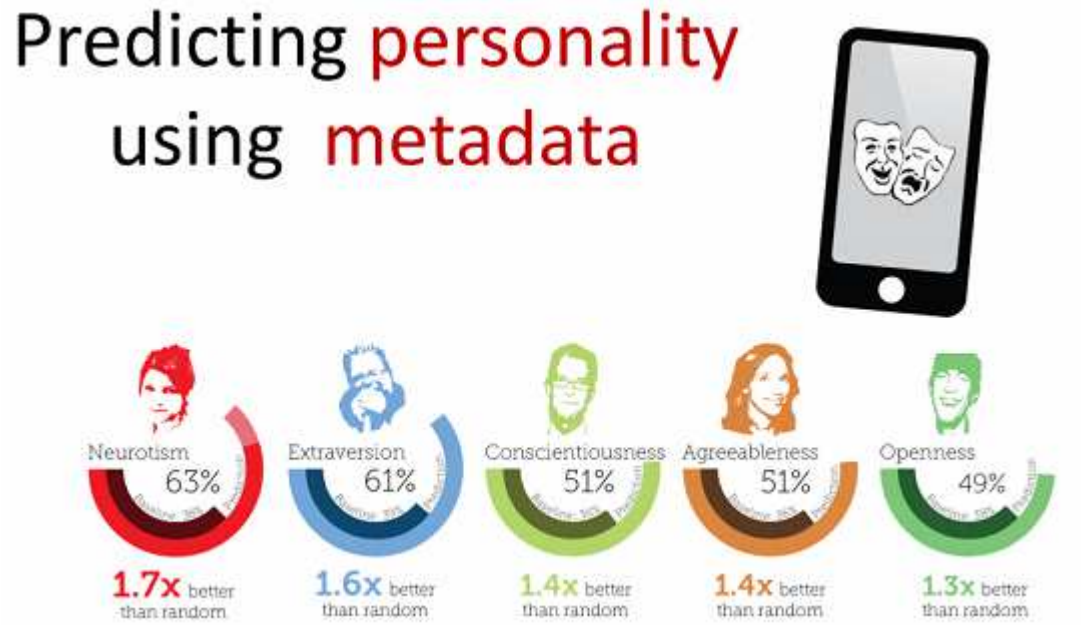


Image : Les 5 caractéristiques psychologiques du test BFI et leur niveau de corrélation avec des données de mobilité, via le poster de l'étude « Qu'est-ce que votre téléphone dit de vous ? »[34].

Cet exemple montre combien il est difficile d'anonymiser les données. Qu'enlever les numéros de téléphone ou les noms des abonnés ne suffit pas à rendre les bases de données sûres. Et que de telles bases disent bien plus que les déplacements qui sont les nôtres ou les réseaux relationnels auxquels on appartient. L'analyse de nos traces peut faire glisser des données anonymes à des données personnelles et des données personnelles à des données sensibles ou à risque. Des modèles de traitement savent extraire des traces de vos habitudes et comportements ou de vos publications des informations psychologiques, d'orientation sexuelle, de santé ou politiques[35]…

De la simple vitesse de réponse à un e-mail ou à un SMS, un système technique pourrait établir un profil psychologique complet d'un utilisateur, simplement en se référant à des modèles établis depuis des échantillons, avec un taux de fiabilité légèrement supérieur à 50%.

Aussi imparfaits que soient les modèles, on peut désormais déduire des appréciations sur vous depuis le moindre de vos comportements enregistrés. Et il suffit de bien peu de données finalement pour le faire… Il suffit de construire des modèles et pour cela, un échantillon d'utilisateurs qui donnent accès à leurs données et remplissent un test permettent de créer des modèles qui pourront être utilisés sur d'autres données similaires.

[34] http://web.media.mit.edu/~yva/InfographicPersonality.png et http://www.demontjoye.com/projects.html
[35] http://www.zdnet.fr/actualites/facebook-ce-que-vos-like-disent-de-vous-39788152.htm

Ces exemples montrent également que toute donnée, quand elle subit des traitements, peut être transformée en données sensibles. Quand il y a suffisamment de données permettant d'inférer des modèles, des données anodines peuvent devenir des données à risque. Aujourd'hui, le fait d'acheter un nouveau produit dans un grand magasin suffit à faire croire à celui-ci que vous êtes enceinte[36] et à vous adresser des programmes publicitaires spécifiques. L'enjeu pourtant n'est pas circonscrit au marketing, hélas. Les traitements, la mise en réseau des données, leur analyse impactent tous les domaines de la vie courante[37] : éducation, emploi, consommation, police…

Bien sûr, ces modèles ne sont pas sans défauts. Leur fiabilité et leurs conclusions sont d'autant plus aléatoires que les données qu'ils utilisent sont rares, mais celles-ci s'améliorent avec l'ajout de données. Votre vitesse de réponse à un e-mail ou à un SMS couplé au graphe social des personnes avec qui vous communiquez permet d'établir un profil psychologique plus précis par exemple[38]. Cela ouvre la question de la fiabilité et de l'opacité des modèles utilisés pour nous profiler, nous calculer, nous prédire, nous catégoriser… C'est ce que dénonce la mathématicienne Cathy O'Neil dans son livre *Weapons of math destruction*[39] : en pointant par exemple le grave dysfonctionnement des systèmes qui attribuent des scores de récidives aux détenus[40] et qui conduisent des gens à demeurer en prison sur des critères qui ne sont pas corrélés au risque de récidive.

L'impact des croisements de données

Toutes les données sont en passe de devenir des données à caractère personnel. Certes "toutes" les données ne sont pas devenues personnelles (les horaires de transports, les données de capteurs environnementaux, certains documents administratifs… par exemple), mais une masse considérable de données qui ne l'étaient pas a priori sont en passe de le devenir. D'autant plus que leur couplage, chaque jour plus facile, peut à chaque moment faire basculer des données "sans valeurs" en données à caractère personnel.

L'internet et les télécommunications ont rendu plus facile l'échange de données. Plus accessibles, elles sont désormais également plus échangeables, commercialisables ou récupérables. On peut désormais croiser des informations pour construire des modèles : croiser des informations météorologiques avec des informations d'achats dans un magasin pour pouvoir créer des modèles permettant de prédire les réassorts selon les prévisions

[36] http://rue89.nouvelobs.com/2014/03/11/donnees-persos-europeens-lisez-bien-petite-histoire-pere-americain-250588 et http://rue89.nouvelobs.com/2014/10/21/etre-efficace-publicite-doit-faire-semblant-rater-cible-255585

[37] http://www.internetactu.net/2016/01/13/nos-systemes-pour-une-retroingenierie-des-systemes-techniques/

[38] C'est ce que propose, d'une certaine manière la startup Crystal Knows, le correcteur comportemental qui vous suggère des corrections quand vous devez écrire un mail selon la personnalité de votre correspondant : http://www.internetactu.net/2015/05/11/vers-des-technologies-de-lempathie/

[39] Voir notamment : http://www.internetactu.net/2014/11/18/ouvrir-les-modeles-pas-seulement-les-donnees/ et http://www.internetactu.net/2016/06/29/il-est-plus-que-temps-que-le-big-data-evalue-ses-impacts/

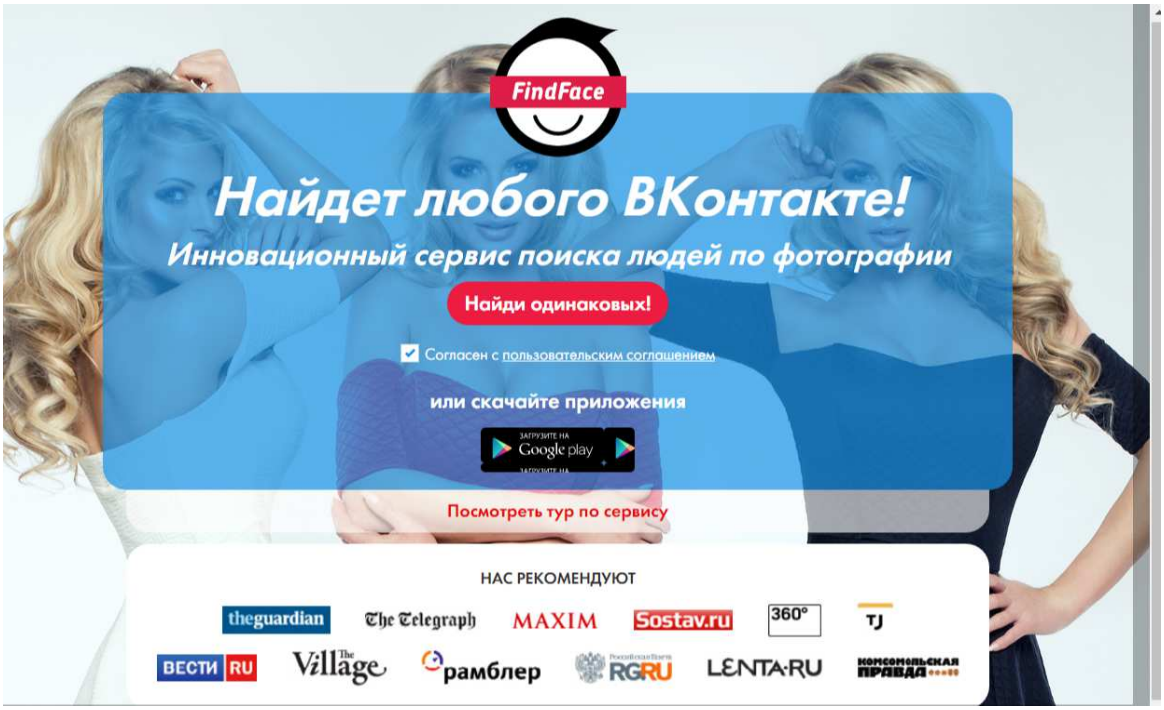
[40] "Un type qui violente un enfant tous les jours pendant un an obtiendra peut-être un score de risque faible parce qu'il a un boulot. Alors qu'un type arrêté pour ivresse publique obtiendra un score élevé parce qu'il est sans domicile fixe. Les facteurs de risque ne vous disent pas si une personne doit aller en prison ; ils vous disent surtout quels sont les bons critères fixer pour une mise à l'épreuve." http://www.internetactu.net/a-lire-ailleurs/144849747113/

météo à 10 jours. On peut également croiser les données de ses clients avec celles d'autres sites ou d'autres plateformes, comme le pratique le modèle de la publicité en ligne. L'essentiel du marketing digital consiste à trouver des clients et des revenus par le croisement de données. Les démarchages publicitaires reposent désormais sur l'achat massif de bases de données concurrentes ou qualifiées pour adresser sa publicité en ligne ou par e-mail. Ou à enrichir ses données de celles d'autres bases de données pour améliorer le profilage de ses propres clients.

Demain, des logiciels permettront d'identifier qui a écrit un texte en observant tout simplement la masse des écrits du web et proposeront des correspondances qui reposeront sur le style, le vocabulaire, la grammaire ou les tics de langages de chacun… De même, combien de temps les visages anonymes sur les photos que nous échangeons ? Le resteront-ils, quand on regarde les progrès de la reconnaissance faciale…

RECONNAISSANCE FACIALE : LA RÉIDENTIFICATION EN QUESTION ?

Les logiciels de reconnaissance faciale les plus évolués à ce jour sont ceux du FBI, de Facebook (Deepface) et de Google (FaceNet)[41] qui permettent d'identifier des visages avec une précision allant (du premier au dernier) de 85% à 99,63% (mais si le système du FBI est moins performant, c'est surtout parce que les images à analyser sont souvent de moins bonne qualité que celles postées sur les réseaux sociaux). Si en Europe, la reconnaissance des visages est strictement encadrée[42], ce n'est pas le cas dans le reste du monde. L'application russe FindFace[43] permet d'identifier un inconnu à partir d'une photographie en analysant les images des 350 millions de comptes de vk.com, le principal réseau social russophone. En Russie l'application permettrait d'identifier 70% des jeunes pris en photo (le taux d'identification diminue en âge du fait de la moindre présence des plus âgés sur les réseaux sociaux). Aux États-Unis, les associations de défense des consommateurs se sont retirées des discussions de groupes de travail sur l'utilisation équitable des technologies de reconnaissance faciale pour le commerce parce qu'elles n'arrivaient pas à obtenir des droits minimums pour les consommateurs, à savoir un consentement explicite des consommateurs. Le problème qui se pose, et qui nous concerne nous aussi Européens malgré nos dispositions législatives protectrices, c'est le risque d'un usage permissif généralisé de la reconnaissance faciale, facilité par des applications dé-territorialisées exploitant des images de profils qui peuvent s'avérer très volatiles. Et bien sûr, le risque d'un couplage de la vidéosurveillance avec la reconnaissance faciale, permettant d'identifier nommément demain, quiconque se promène dans la rue.



41 http://alireailleurs.tumblr.com/post/121737023394/reconnaissance-faciale-avons-nous-droit-%C3%A0-la

42 http://www.journaldunet.com/ebusiness/expert/57861/les-risques-juridiques-des-logiciels-de-reconnaissance-faciale.shtml

43 https://findface.ru/ et http://rue89.nouvelobs.com/2016/05/05/findface-appli-capable-didentifier-peu-trop-monde-263943

Poussée à son acmée, le principe du libre croisement des données permis par un environnement totalement dérégulé conduit également à la disparition des données qui ne seraient pas personnelles.

La mise en réseau des données conduit à démultiplier les possibilités de croisement. Le développement de modèles de calculs et de traitements permet de démultiplier l'interprétation et le sens de n'importe quelle donnée.

L'anonymisation est-elle encore possible ?

"Au milieu des années 90, une commission d'un groupe d'assurance américain décidait de publier des données médicales anonymisées d'employés de l'État du Massachusetts. Un étudiant en informatique, Latanya Sweeney, en demanda une copie et travailla à leur "réidentification". Le gouverneur du Massachusetts assurant que l'organisme d'assurance avait protégé chaque patient en effaçant tous les identifiants nominatifs, Sweeney utilisa les listes de votants de la petite ville où habitait le gouverneur et croisa les deux bases de données. Seulement 6 personnes dans cette ville partageaient les mêmes dates de naissance, seulement 3 étaient des hommes et un seul partageait le même code postal… que le gouverneur. L'informaticien envoya au gouverneur tout son dossier médical."

Cette histoire que raconte Nate Anderson pour Ars Technica[44] montre bien que l'anonymisation à une époque où les données sont démultipliées n'est plus si simple. Quelques années plus tard, Latanya Sweeney démontra[45] d'ailleurs que 87 % des Américains pouvaient être identifiés uniquement à partir de 3 informations : le code postal, la date de naissance et le sexe.[46]"

Depuis, les études n'ont cessé d'alerter la société sur les limites de l'anonymisation des données. L'étude "Unique dans la foule"[47] a montré que 4 points géographiques étaient suffisants pour ré-identifier 95% des utilisateurs d'une base de données de métadonnées d'appels de 1,5 million d'abonnés au téléphone mobile. *"Nos données de déplacements sont encore plus personnelles que nos empreintes digitales.[48]" Nos routines journalières sont tellement uniques que nul ne peut se cacher dans la foule.*

Le fait que très peu de données suffisent à identifier un individu, à le discrétiser comme disent les spécialistes des bases de données, pose la difficulté qu'il y a de rendre les données anonymes. Nettoyer les identifiants les plus évidents ne suffit plus. Ré-identifier ou dé-

[44] http://arstechnica.com/tech-policy/2009/09/your-secrets-live-online-in-databases-of-ruin/

[45] http://dataprivacylab.org/projects/identifiability/paper1.pdf

[46] http://www.internetactu.net/2009/09/21/critique-du-web%C2%B2-34-toutes-les-donnees-sont-devenues-personnelles/

[47] http://www.internetactu.net/2013/12/05/en-quoi-les-big-data-sont-elles-personnelles/ et http://www.internetactu.net/2013/06/25/peut-on-fouiller-les-donnees-des-telephones-mobiles-en-respectant-la-vie-privee/

[48] Dans *Les preuves de l'identité*, Edmond Locard, le fondateur de la police scientifique, explique qu'il suffit d'utiliser 12 points de références pour être sur et certain d'identifier les empreintes digitales d'un individu.

anonymiser des données se fait le plus souvent avec une étonnante facilité. Pour le spécialiste du sujet, Paul Ohm, de l'université du Colorado, on ne peut garantir à la fois l'utilité maximale des données et une confidentialité maximale. Les informations permettant l'identification sont une catégorie en expansion constante. De simples critiques de films par des internautes sur un site ont par exemple suffit à réidentifier des utilisateurs[49] en les comparant à d'autres critiques publiées ailleurs en ligne. La distinction binaire traditionnelle entre données à caractère personnel et les autres devient difficile à maintenir.

L'avènement des Big Data ne rend pas seulement la protection de la vie privée beaucoup plus difficile, estiment Viktor Mayer-Schönberger et Kenneth Cukier[50] (et c'est un euphémisme, l'informaticien Arvind Narayanan de l'université de Princeton, estime[51] déjà qu'à l'heure des Big Data, *"l'anonymat est devenu algorithmiquement impossible"*), elle présente aussi de nouvelles menaces, comme la justice ou la police prédictive… Le principal risque des Big Data ne porte pas tant sur la vie privée que sur le risque que nous soyons amenés à juger les gens non pas sur leur comportement réel, mais sur leur propension à avoir le comportement que les données leur prêtent. Le monde social n'est pas né avec les Big Data, mais celles-ci pourraient bien nous y enfermer[52].
Dans un monde de données ambiantes, la tentation de tout savoir devient presque irrésistible. Pourtant, face aux résistances sociales et psychologiques que cela ne va pas manquer d'introduire, il s'agit bien d'en comprendre l'essence et non pas de les minorer. En s'insinuant dans des transactions dont elle était absente, l'intelligence ambiante va bouleverser notre rapport à notre environnement et à l'information qui émane de nous même. Pour y répondre, il va certainement falloir offrir toujours plus de garanties à l'individu et décider d'un vrai bond en avant dans la protection de l'intimité. En échange de la collecte des données collectives que l'informatique omniprésente va libérer, nous ne pouvons pas céder nos données personnelles. Au contraire[53].

Souvent, les systèmes de capteurs conservent des données qui ne semblent pas indispensables à leur fonctionnement courant. Faut-il que notre système de télépéage ou notre pass Navigo conserve des données nominatives, permettant de savoir qui passe à tel ou tel portique ? A-t-il besoin de conserver nos lieux d'entrée et de sortie (alors que d'autres compteurs sont en place, comme les compteurs de passage aux barrières et aux portes) ? L'important, pour ces systèmes, c'est de savoir qu'un titulaire de droit à la

[49] Le fait de poster de simples avis sur des films peut aussi permettre de vous identifier, expliquent Arvind Narayanan et Vitaly Shmatikov. Quand Netflix, le loueur de films par internet américain, a rendu disponible sa base de données de recommandations de films – anonymisée la encore – pour lancer son concours d'amélioration de son moteur de recommandation, des scientifiques ont combiné ces données avec d'autres données de recommandation sur internet leur permettant de réidentifier un grand nombre de recommandations.
[50] Cukier (Kenneth), Mayer-Schöneberger (Viktor), *Big Data, la révolution des données est en marche*, Robert Laffont, 2014 et http://www.internetactu.net/2013/05/14/big-data-nouvelle-etape/
[51] http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/
[52] http://www.internetactu.net/2013/05/14/big-data-nouvelle-etape/
[53] http://www.internetactu.net/2007/12/21/comment-proteger-notre-vie-privee-dans-un-monde-ou-la-tracabilite-explose/

possibilité de franchir la barrière, pas qui il est, ni à quel endroit il passe, ni où il va[54]. Et surtout que se passe-t-il quand ces données-là peuvent être croisées avec des milliers d'autres ?

La dissymétrie des données

Les questions de diminution de la collecte et du renforcement de la protection que nous évoquions précédemment se doublent désormais de questions de régulation du croisement des données et d'intelligibilité des modèles utilisés[55]. A mesure que les collectes, les traitements et les croisements se développent, via des outils d'apprentissage automatisés de plus en plus ingénieux, les questions de protection des utilisateurs deviennent plus primordiales et plus complexes, car les fronts sur lesquels avancer se démultiplient et se complexifient.

Les données sont devenues dissymétriques. Pour tout un chacun, un simple code postal n'est qu'une information sans grande importance, d'autant plus qu'il ne révèle pas son lieu d'habitation précis. Pour ceux qui le recueillent, ce code postal dit bien d'autres choses. Il permet d'inférer un niveau de revenu, un âge médian, une couleur politique… auxquels s'associent des profils commerciaux, psychologiques… qui infèrent à leur tour un niveau d'employabilité, de maladie, de consommation, de criminalité, de pauvreté… Nous sommes là confrontés à deux niveaux d'information totalement inégaux et finalement bien peu équitables.

[54] http://www.internetactu.net/2009/09/21/critique-du-web%C2%B2-34-toutes-les-donnees-sont-devenues-personnelles/

[55] Voir notamment le programme NosSystèmes, lancé par la Fing : http://www.internetactu.net/2016/01/13/nos-systemes-pour-une-retroingenierie-des-systemes-techniques/

2. À qui appartiennent les données personnelles ?

"Les données personnelles des individus ne sont pas considérées comme leurs propriétés, mais comme un droit attaché à la personne humaine et au respect qui lui est dû"[56]. Les données personnelles des individus échappent ainsi au droit d'auteur et à la propriété intellectuelle. Elles appartiennent par défaut (pourrait-on dire par analogie) au domaine public... Comme l'expliquait Isabelle Falque- Pierrotin, présidente de la CNIL, "Ceux qui pensent être propriétaires de mes données se trompent. J'ai toujours la possibilité d'exercer mes droits et notamment le droit d'accès, de rectification et de suppression même si mes données sont traitées par d'autres. Ce que l'on confond encore trop souvent, c'est la propriété sur un fichier, laquelle existe effectivement, et les données personnelles qui, elles, relèvent d'un droit fondamental, le droit à la vie privée, lequel ne peut être cédé." [57] Si les organisations ont un droit de propriété sur le fichier qu'elles ont créé, elles n'en ont pas sur l'information qu'elles ont collectée, qui appartient à leur titulaire. Mon adresse, mon téléphone, mon e-mail, mon nom que les entreprises collectent m'appartiennent.

Les données personnelles ne sont donc pas un bien individuel. Elles sont un droit, mais pas une propriété.

2.1 L'impossible (mais irrésistible) privatisation

Nombreux sont ceux qui militent pour leur "patrimonialité", c'est-à-dire l'instauration d'une propriété privée sur les données, permettant aux utilisateurs d'en faire l'usage qu'ils souhaitent. Leur permettant de les céder ou non, contre rémunération ou non, de manière définitive ou temporaire, partielle ou intégrale. Cette logique, souvent défendue, a pourtant de nombreuses fois été rejetée…

Par le Conseil d'État dans son rapport sur le numérique[58], par le Conseil national du numérique dans son rapport sur la Neutralité des plateformes[59]… Les arguments pour repousser cette vision libérale du droit des données personnelles tiennent au fait que cela renforcerait l'individualisme sans prendre en compte le rapport de force entre les individus et les organisations. En outre, accorder un droit de propriété individuelle sur ses données

[56] https://scinfolex.com/2014/06/19/le-cnnum-sest-prononce-contre-linstauration-dun-droit-de-propriete-privee-sur-les-donnees-personnelles/
[57] http://www.lesechos.fr/25/11/2014/lesechos.fr/0203937716964_isabelle-falque-pierrotin-----ceux-qui-pensent-etre-proprietaires-de-nos-donnees-se-trompent--.htm
[58] http://www.conseil-etat.fr/Decisions-Avis-Publications/Etudes-Publications/Rapports-Etudes/Etude-annuelle-2014-Le-numerique-et-les-droits-fondamentaux
[59] http://cnnumerique.fr/plateformes/

risque de renforcer les inégalités entre les citoyens capables de gérer, protéger ou monétiser leurs données et ceux qui abandonneraient ces fonctions au marché. Mais surtout, admettre la patrimonialité des données implique que l'État devrait renoncer à sa "logique de protection" de l'individu. Les données étant considérées comme une extension de notre organisme, appliquer un droit de propriété aux données personnelles équivaudrait à appliquer un droit de propriété à notre propre corps.

Comme le souligne le juriste Lionel Maurel[60], "le propre de la propriété est d'être cessible et transférable". Elle permet d'organiser "une chaîne de transferts des droits au profit d'intermédiaires économiques". Pour lui, "Loin d'aboutir à une « souveraineté » retrouvée, le détour par la propriété mènerait les individus sous la coupe d'autres intermédiaires et précipiterait encore davantage le mouvement de marchandisation des données personnelles." Comme il le souligne[61] en analysant des plateformes de marchandisation de ses données comme Datacoup[62], ces plateformes servent surtout à transférer les droits des individus à des tiers, sans expliciter clairement aux utilisateurs ce qu'il sera fait des données par les entreprises auxquelles elles sont revendues.

Des données en Communs

Si le régime de propriété des données semble s'exclure à mesure qu'on en étudie la possibilité, d'autres pistes peuvent être ouvertes, comme nous y invitait Valérie Peugeot[63]. Notamment l'idée de développer une sphère de données en Communs, c'est-à-dire des données considérées comme ressource collective, réutilisables sous certaines conditions fixées par la communauté qui en a la gestion et veille à leur protection[64]. Ou encore, avance la chercheuse, imaginer un régime de "faisceau de droits" permettant autour de mêmes ressources d'identifier différents droits (posséder, utiliser, gérer, monétiser, transmettre, modifier…). C'est ce que met en œuvre le programme MesInfos[65] lancé par la Fing autour de la notion de Self data, du retour des données aux utilisateurs, consistant à permettre aux individus de récupérer et accéder plus facilement aux données dont disposent les organisations sur eux, mais aussi à leur permettre d'accéder à de nouveaux services mieux adaptés à leurs besoins.

60 https://scinfolex.com/2014/06/19/le-cnnum-sest-prononce-contre-linstauration-dun-droit-de-propriete-privee-sur-les-donnees-personnelles/
61 https://scinfolex.com/2014/10/01/le-miroir-aux-alouettes-de-la-revente-des-donnees-personnelles/#more-7887
62 https://datacoup.com
63 http://vecam.org/archives/article1289.html
64 Voir également sur ce sujet : http://www.internetactu.net/2012/06/22/design-your-privacy-pour-une-licence-de-partage-des-donnees-personnelles/
65 http://mesinfos.fing.org

Reste que comme le pointe très justement le juriste Lionel Maurel[66], sans y apporter une réponse, "Les licences restent fondamentalement des instruments de gestion individuelle". Une gouvernance en communs des données personnelles nécessite une réponse d'ordre institutionnel, et pas seulement juridique. Comme le souligne encore Lionel Maurel[67], l'interconnexion des données personnelles entre elles nécessite une réponse collective. A l'heure des données liées, à l'heure où les informations relatives à un individu renseignent aussi sur d'autres personnes, la gestion des données personnelles de manière personnelle est certainement un paradigme à dépasser. A l'image de vos données de contacts auxquelles accèdent les applications que vous autorisez, votre acceptation d'un service a un impact sur ceux qui ne l'acceptent pas. Les machines de Gmail lisent vos e-mails, mais ont également accès à tous les messages qui sont adressés par tous ceux qui l'utilisent, même les messages de ceux qui ne veulent pas utiliser Gmail ! Si les données personnelles sont désormais en réseau, la gestion de ce réseau comme biens communs est une question ouverte.

Au fond, c'est ce qu'il nous semble ressortir de plus fort dans les réflexions du groupe de travail : l'idée que la protection de la vie privée, conçue comme un édifice juridique fonctionnant par défaut et pour tous, doit désormais se compléter de dispositifs de "maîtrise", plus complexes et mouvants, qui permettent aux individus – dans des limites à mieux définir – d'organiser à leur manière ce qu'ils veulent défendre, ce qu'ils veulent exposer et ce qu'ils sont prêts à négocier. Et aussi, de dispositifs collectifs capables d'exercer des formes de pression que l'État ne parvient pas (ou plus) à exercer."[68]

[66] https://scinfolex.com/2014/09/01/une-gouvernance-en-communs-des-donnees-personnelles-est-elle-possible/#more-7750 qui s'appuie sur les travaux du chercheur Antonio Casilli : http://www.casilli.fr/2014/08/27/quatre-theses-sur-surveillance-de-masse-et-vie-privee/

[67] https://scinfolex.com/2016/01/15/eriger-le-reseau-des-donnees-personnelles-en-bien-commun/

[68] http://www.internetactu.net/2009/04/06/vie-privee-ou-sont-les-regulateurs-ou-sont-les-regulations/

2.2 L'accaparement par défaut

On a vu que beaucoup de données et que nos métadonnées se révélaient très personnelles. Que les possibilités de croisement et de traitement étaient démultipliées par l'interconnexion et les progrès des capacités d'analyse et de calcul. En attendant que des solutions collectives s'imaginent, s'esquissent ou se mettent en place, la réalité à laquelle nous sommes confrontés, c'est celle de la prédation de nos données. Que ce soit par de nouvelles "baronnies"[69] que sont les Gafa et autres Natu, qui s'accaparent les données des utilisateurs pour les exploiter, des ententes[70] entre les spécialistes du ciblage publicitaire pour se partager les données qu'ils accumulent sur les utilisateurs afin d'améliorer sans cesse leurs performances. Et surtout, la démultiplication des vols et piratages de données, très organisés sur le darknet[71]…

Dans la réalité, comme le souligne l'avocat Alain Bensoussan[72], le respect des droits fondamentaux des utilisateurs en matière de données (propriété, intimité, sécurité et souveraineté) est rarement respecté. L'utilisateur est la plupart du temps démuni pour s'assurer que les données qu'il confie à un service ne seront pas cédées à un tiers, volées, utilisées par-devers lui.

QUE VALENT NOS ADRESSES QUAND NOUS SIGNONS UNE PÉTITION ?

La journaliste italienne Stefania **MAURIZI**[73] a enquêté sur la manière dont Change.org, la plateforme pétitionnaire, monétise les adresses e-mails qu'elle collecte. Elle ne vend pas que des bases d'adresses, mais vend surtout des profils depuis les données qu'elle accumule sur nous.
« Si vous avez signé une pétition sur les droits des animaux, l'entreprise sait que vous avez une probabilité 2,29 fois supérieure d'en signer une sur la justice. Et si vous avez signé une pétition sur la justice, vous avez une probabilité 6,3 fois supérieure d'en signer une sur la justice économique, 4,4 d'en signer une sur les droits des immigrés et 4 fois d'en signer une autre encore sur l'éducation". Le problème, souligne la journaliste italienne*, c'est que les utilisateurs comprennent fort mal ce à quoi ils s'engagent : "lorsqu'ils signent une pétition sponsorisée, il suffit qu'ils laissent cochée la mention "Tenez-moi informé de cette pétition" pour que leur adresse électronique soit vendue par Change.org à ses clients qui ont payé pour cela ».*

Pour Alain Bensoussan, la création de profils va déboucher immanquablement sur la monétisation des données personnelles. "Chacun va devenir le trader de l'exploitation commerciale de ses propres données personnelles"[74]. A moins que, en fait, les utilisateurs demeurent les manipulés. Les systèmes et services à qui ils confient leurs données et qui génèrent des données sur eux semblent surtout les utiliser par-devers les utilisateurs. Les entreprises qui récoltent des données sur les utilisateurs s'en réservent les traitements, les croisements, les modélisations. L'accaparement des données est de plus en plus la règle que l'exception.

69 http://www.internetactu.net/2013/05/14/big-data-nouvelle-etape/
70 http://www.numerama.com/tech/154171-adobe-lance-une-cooperative-de-donnees-personnelles-pour-vous-tracer-partout.html
71 https://www.letemps.ch/economie/2016/10/02/vos-emails-mots-passe-se-monnaient-internet
72 https://www.alain-bensoussan.com
73 https://framablog.org/2016/07/20/ce-que-valent-nos-adresses-quand-nous-signons-une-petition
74 http://www.archimag.com/vie-numerique/2014/03/26/donnees-personnelles-chacun-trader-propres-donn%C3%A9es

2.3 De la restitution des données aux utilisateurs à la collectivisation des données : l'administration face au défi des données personnelles

S'il existe d'autres initiatives, comme MesInfos[75] portée par la Fing, consistant à promouvoir une restitution des données aux utilisateurs (l'initiative anglaise MiData, les initiatives américaines de Blue et Green Button, midata.coop en Suisse et l'initiative Datamixer en Belgique[76]), afin de renforcer leur capacité de contrôle, force est de constater qu'elles sont encore à contre-courant des pratiques du secteur. Si le nouveau règlement européen introduit des règles renforçant le respect les données des utilisateurs, il reste encore à développer des pratiques et des systèmes qui l'implémentent vraiment.

Les administrations et collectivités doivent s'engager dans le retour des données aux utilisateurs. Elles doivent s'engager plus avant dans la protection des données. Elles doivent imaginer avec des entreprises des services pour les exploiter d'une manière responsable et respectueuse[77]. Elles doivent piloter la transformation numérique de l'exploitation des données en impulsant une réflexion et une action sur de nouveaux systèmes de gouvernance des données personnelles par des collectifs, en développant la participation citoyenne à des plateformes sous forme de coopératives de données[78], imaginer de nouvelles formes de services publics participatifs dont la gouvernance serait assurée co-portée avec et par les utilisateurs eux-mêmes. Des plateformes responsables capables d'organiser les Communs. Telle est la prochaine étape. Si les données sont devenues le socle des services de la ville, comme l'explique le sociologue Bruno Marzloff[79], les opérateurs privés à qui les services publics confient ces systèmes ne peuvent pas être les seuls bénéficiaires de ces données. Le risque est que la planification et la régulation des villes par les données cessent d'être une affaire publique. Or, si "le digital est un levier de transformation de nos cités plus considérable que ne le furent l'électricité et la voiture", il va falloir trouver les modalités d'une économie fonctionnelle autour de l'exploitation des données personnelles. La donnée nécessite une gouvernance adaptée. Le développement de plateformes privées qui impactent directement le fonctionnement des villes, comme Uber et Airbnb, nécessite que les acteurs publics trouvent de nouvelles modalités de régulation, d'un nouvel espace de délibération[80]. Les acteurs publics doivent impulser une nouvelle politique de la ville numérique.

---

[75] http://mesinfos.fing.org/

[76] Voir notamment le dossier sur la réutilisation des données personnelles : http://www.internetactu.net/2012/06/19/reutilisation-des-donnees-personnelles-14-rendre-leurs-donnees-aux-utilisateurs/

[77] Le professeur de droit Jonathan Zittrain expliquait récemment que les professions fiduciaires, comme les médecins ou les avocats, avaient l'interdiction d'utiliser l'information de leurs clients pour leurs propres intérêts, et appelait à ce qu'on applique des dispositions similaires aux grandes entreprises technologiques : http://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/

[78] http://www.internetactu.net/2015/11/25/leconomie-solidaire-necessite-un-internet-de-la-solidarite/

[79] http://abonnes.lemonde.fr/idees/article/2016/06/08/smart-cities-attention-a-l-appropriation-des-data-personnelles-par-le-prive_4942828_3232.html

[80] http://www.internetactu.net/a-lire-ailleurs/plateformes-et-metropoles/

Conclusion :

De cette synthèse, nous pouvons tirer 3 pistes d'actions pour les Métropoles :

- **Formation, médiation** : Sensibiliser agents, administrations et publics aux enjeux des données personnelles pour développer des pratiques permettant de limiter la collecte, les croisements et les traitements et respecter les utilisateurs.
- **Symétrie d'information** : Après avoir recueilli le consentement, avoir proposé des droits d'accès, de rectification et d'opposition aux données personnelles des utilisateurs, l'avenir consiste à préparer les systèmes techniques à rendre les données aux utilisateurs et demain à leur proposer une symétrie d'information des données que l'administration collecte et exploite des utilisateurs.
- **Expérimenter des formes de gouvernance des données** : Identifier des terrains d'expérimentation pour imaginer de nouvelles formes de gouvernances des données, notamment sous la forme de coopératives qui rendraient plus de pouvoir et de droits à leurs détenteurs.

RETROUVEZ
TOUTES LES ÉTUDES SUR

WWW.MILLENAIRE3.COM

MÉTROPOLE DE LYON
DIRECTION DE LA PROSPECTIVE ET DU DIALOGUE PUBLIC
20 RUE DU LAC
CS 33569 - 69505 LYON CEDEX 03