Pierre Bellanger

Principes et pratiques des données personnelles en réseau

http://pierrebellanger.skyrock.com/3231110655-Principes-et-pratiques-des-donnees-personnelles-en-reseau.html

Contribution de Pierre Bellanger à l'étude 2014 du Conseil d'État : Technologies numériques et libertés et droits fondamentaux

Les données personnelles sont les informations renseignant, directement ou indirectement, sur un individu identifié. Cette définition établit un droit singulier de nature personnelle de l'individu concerné sur ses données, droit destiné à protéger, notamment, sa vie privée. Les réflexions juridiques et institutionnelles en cours tendent à vouloir renforcer et confirmer ce droit exclusif de chacun sur les données qui lui sont relatives. Les directions envisagées vont d'un droit autonome à en déterminer le recueil et l'usage, à une faculté indépendante d'administration - par la copie, la modification, le transfert ou la disparition partielle - jusqu'à, enfin, un droit de propriété privée par chacun de ses propres données personnelles.
Chacune de ces avancées a ses avantages et ses aléas. Mais correspondent-elles à la réalité des données personnelles d'aujourd'hui ?
En effet, ces dispositifs se fondent sur le présupposé que la donnée personnelle est autonome, ne renseigne que sur un seul individu, bref considère la donnée personnelle comme granulaire, indépendante et formant une entité en soi, soumise au droit d'un seul.
Cette conception, pertinente jadis au temps des fichiers du XXe siècle, ne correspond plus à la réalité. Aujourd'hui, les données personnelles ne sont plus isolées, elles sont en réseau. Elles forment un réseau de données, certes chacune demeure personnelle, mais elles sont désormais organisées en une totalité indissociable.
Et cela pour six raisons :

- les données personnelles ne sont pas isolables en pratique : donner accès à sa liste de contacts, à ses photos, à son agenda, à son courrier, à sa position, engage mécaniquement, de fait, les données personnelles d'autrui sur lesquelles on ne dispose d'aucun droit ;
- les données personnelles renseignent sur d'autres personnes : les algorithmes de corrélation, ces programmes informatiques qui permettent de déduire, par probabilité, des informations par le traitement prédictif de masse de données sans rapport direct avec l'information inférée, font que chaque donnée personnelle renseigne indirectement sur autrui. Par exemple : les données personnelles de clients bancaires croisées avec leur défaut de paiement vont servir à déterminer le risque d'impayé de nouveaux clients par la comparaison de leurs comportements. Par exemple encore : les données corrélées entre cancer du côlon et consommation en supermarché d'un groupe d'individus vont permettre de prédire le risque cancérogène d'une personne, sans relation avec le groupe témoin, et cela à partir de ses seuls tickets de caisse ;
- les données personnelles sont une extension de la personne : à la manière du sang, c'est un soi hors de soi. Engager le transfert ou la cession de données personnelles d'autrui, indissociables ou déductibles des siennes, en échange de l'accès « gratuit » à un service s'apparente par conséquent au trafic d'organes ;
- le contrôle individuel par accord de gré à gré devient impossible : les collecteurs de données personnelles sont destinés à se multiplier sous forme de capteurs disséminés partout et intégrés dans la plupart des objets ; tandis, qu'à son tour, chaque individu devient collecteur de

données sur lui-même et sur autrui. L'autorisation individuelle réfléchie à chaque captation, déjà aléatoire, n'est plus possible dans les faits. Le monde qui vient est un monde où tout incorpore de l'intelligence informatique, captatrice et communicante, par laquelle tout se traite et transite, afin de transmettre un flux permanent de données. La brosse à dent, la cafetière, l'automobile, le réfrigérateur, la montre, les lunettes, les vêtements, les chaussures captent et se connectent. L'objet muet deviendra l'exception, l'environnement aveugle disparaît ;

- la constitution d'un monopole des données personnelles : l'effet réseau s'applique aux données personnelles : la valeur d'une donnée est proportionnelle au carré du nombre de données auxquelles elle est reliée. En effet, la valeur d'une donnée provient de son contexte, apporté par des données supplémentaires.

Par exemple : l'achat d'une poussette renseigne sur la future consommation d'un foyer. Cette donnée unique a de la valeur pour tout annonceur publicitaire de produits destinés à la petite enfance. Mais une seconde donnée pourrait apprendre que l'achat de la poussette est un cadeau pour les voisins.

La donnée se vérifie, prend du sens et donc devient connaissance par son agrégation intelligente à d'autres. En conséquence, il n'y a pas valeur absolue de la donnée unitaire.

En revanche, le plus gros détenteur de données peut surenchérir sans cesse pour en acquérir de nouvelles, en numéraire ou en services gratuits, puisque c'est pour lui que les données ont le plus de valeur et que, de surcroît, chaque acquisition nouvelle accroît la valeur de l'ensemble déjà collecté, jusqu'à ce que, par cette logique, il en détienne le monopole ;

- l'encadrement juridique de la modélisation informatique du réel : la collecte globale et considérable de données - dont les données personnelles -constitue au final une tramage quantitatif de la réalité elle-même. L'appropriation par quelques entreprises de la reconstitution informationnelle du réel est source d'asymétries de concurrence dévastatrices et ne peut être empêchée par une somme de droits individuels.

Par exemple : la connaissance directe ou prédictive par un seul acteur du type de conduite de chaque automobiliste lui donne un avantage décisif et sans concurrence pour établir des tarifs d'assurance auto sur mesure et au meilleur prix, sélectionnant ses clients pour ne laisser à sa compétition que les conducteurs qu'il a détecté comme non rentables. Ainsi, une somme d'acceptations personnelles sans conséquence immédiate pour les individus concernés pourrait mettre un terme au secteur de l'assurance tel que nous le connaissons, fonder de ce fait un nouveau monopole qui ne tarderait pas à renchérir l'assurance pour tout le monde.

Ainsi, les données personnelles ne sont plus granulaires mais réticulaires, c'est-à-dire organisées en réseau. Les données personnelles ne sont plus séparées mais liées. Cette intrication forme le réseau des données personnelles qui se substitue, en fait, aux données personnelles isolées du passé.

Comment se représenter le réseau des données personnelles ? L'hologramme est une bonne analogie : il provient d'une plaque photographique éclairée qui produit une image tridimensionnelle. Chaque morceau de la plaque contient l'image entière à moindre définition. De la même manière pour le réseau de données : la totalité des données reproduit le réel et chaque donnée renseigne sur l'ensemble.

Prenons un exemple : un seul grain de sable renseigne sur toute la plage car la ressemblance entre la majorité des grains est forte. En revanche, un article provenant du rez-de-chaussée d'un grand magasin - un bracelet fantaisie - apporte peu d'information sur les articles de décoration ou de jardinage qui sont en étage. La plage est holonome : on peut déterminer l'information globale (la plage) à partir de l'information locale (le grain). Le grand magasin est autonome : chaque article ne détermine que lui-même.

Pour ce qui concerne l'être humain, il partage plus de 99 % de son génome avec les autres membres de son espèce et son comportement, selon une étude de la revue *Science*, est à 93 % prévisible. Les données personnelles sont de fait holonomes.
Bien entendu, cette similitude et ces homogénéités de comportements n'ôtent rien au caractère unique de chaque humain - qui s'exprimera par d'infinies variations et par des marges surprenantes - ni à sa liberté, car son libre arbitre préserve à chaque instant son improbabilité. Il n'en demeure pas moins que cette singularité s'exprime en relation à une forte conformité à la moyenne, à une sorte de barycentre comportemental.
Ainsi la vision des données comme indépendantes et fondamentalement séparées les unes des autres est une abstraction qui n'est plus pertinente. Les données personnelles se déterminent mutuellement et forment unréseau organique.
De plus, ce réseau est dynamique. Le volume de données collectées double tous les 18 à 24 mois. Les données, jadis discrètes et donc isolables, deviennent des flux continus d'informations captées et quantifiées à chaque instant, liant en temps réel les données de sources individuelles multiples. Enfin les liens logiques reliant les données entre elles se multiplient de manière exponentielle. Le réseau de données forme désormais une totalité animée en croissance permanente.
Par commodité, on appellera le réseau de données personnelles, le RDo.
Quelle est la nature juridique du RDo ? Il s'agit d'un objet sur lequel toutes les personnes, dont les données sont maillées, disposent de droits mais qui ne peut être matériellement divisé entre eux. Il est ni dissociable, ni individualisable par nature car chaque donnée personnelle renseigne sur les autres. C'est donc une forme d'indivision qui concerne toute la population.
Par ailleurs, les informations provenant du RDo sont d'un intérêt général majeur pour la collectivité, notamment, en matière de santé, de transports, de consommation, d'environnement ou encore de compétitivité économique.
Par son origine multi-personnelle, son impossibilité à le séparer, et son utilité collective, le RDo est donc un bien commun - *res communis* - : un bien qui appartient à tous mais ne peut appartenir à personne en particulier. Son statut est défini en droit français par l'article 714 du Code civil.
C'est aussi un bien où chacun dispose de droits spécifiques (retrait, opposition, oubli) sur son propre apport et ce, dès lors qu'il n'engage pas les droits d'autrui.
Le RDo répond donc de droits collectifs et de droits individuels. La gestion et l'exercice de ces droits doit revenir à un organisme public, garant du contrôle démocratique et souverain et seul à même d'en permettre l'accès et l'usage.
Une telle institution, structurante et référente, créé les procédures, les instances ainsi que les concertations nécessaires. Elle devra donc, tout à la fois, gérer le bien commun et les droits individuels afférents. Sa capacité à ester en justice sera, de ce point de vue, essentielle.
Une agence des données pourrait ainsi être établie. La meilleure base ne serait-elle pas l'actuelle Commission nationale de l'informatique et des libertés (CNIL) ?

Quels droits individuels ?

La faculté technologique nouvelle de captation, de conservation et de traitement informatique des actes de chacun, tandis qu'en parallèle une part croissante de nos vies se déroule sur les réseaux et systèmes numériques, amène à définir - dans ce contexte - la nature et les droits en regard de la personne humaine.

Un être humain est à considérer comme un devenir permanent. C'est cette faculté et cette liberté de devenir qui le caractérise et doit donc être préservée voire accrue.
Le processus de ce devenir, parce qu'il est multiforme, contradictoire et sans limite, parce qu'il n'a de sens que dans un contexte profond et secret, se dénature s'il est observé par autrui et donc jugé et normé. L'alchimie intime et solitaire n'appartient qu'à soi. Un des fondements de la personne humaine est donc le droit au mystère.
De ce processus personnel ressort un personnage que l'on s'est choisi, cette représentation est une variante sociale de soi qui nous définit vis-à-vis des autres. L'intégrité de cette personne sociale doit être préservée. Ainsi, les informations individuelles accessibles, notre histoire personnelle, doivent par principe, et sauf exception motivée, répondre de la volonté individuelle de la personne concernée. C'est le droit au choix de soi.
La personne humaine, pour son accomplissement et la liberté de son évolution, doit se retrouver dans un environnement qui maximise ses choix. Toute réduction du champ des possibles, liée à sa nature réelle ou supposée, ne peut-être qu'exceptionnelle, connue et motivée. Avec chaque réduction de choix éventuelle doit être proposé une alternative commune. C'est le droit à la neutralité du monde.
Par exemple : un site de commerce adapte sans avertissement son offre de produits en fonction de sa supposition du pouvoir d'achat de son client en ligne. Ce faisant, ce site limite la liberté de choix de son client potentiel pour orienter ses décisions et donc le conduire à un choix particulier qui ne serait pas forcément le sien s'il avait accès à la totalité de l'offre. Cette restriction de choix est une atteinte à la liberté individuelle.
Les données personnelles sont une extension de la personne et donc doivent être sous sa maîtrise. Sous réserve des prérogatives judiciaires, la souveraineté individuelle de chacun sur ses données personnelles est garantie.
Par exemple : une personne, dans le passé, a commis une infraction au Code de la route. Cette information disponible pour d'éventuels employeurs compromet bien des possibilités d'embauche. La personne doit avoir la faculté de réduire l'accès à cette information. Le passé ne doit pas être une prison, sauf dérogation temporaire et justifiée.
Enfin, l'accès aux données est un moyen formidable de développement de soi et des autres, équivalent à l'accès à la connaissance. Cet accès, s'il est conditionné par les droits précédents, doit être libre et ouvert à tous. C'estle droit d'accès aux données.
Par exemple : une personne souffre d'une affection peu répandue. Afin qu'elle exerce son jugement et détermine ses choix, l'accès aux données anonymisées de santé des autres malades pareillement atteints serait de la plus grande utilité.
Il faut noter que ces droits individuels sont d'utilité sociale. Qu'en serait-il de la création, de l'innovation, de l'entreprise, de l'imagination et donc du progrès collectif sans la garantie pour chacun de son intégrité informationnelle et la protection de sa liberté de pensée ?
Qu'en serait-il *in fine* de la démocratie sans ces droits qui sont à Internet ce que l'isoloir est à la République ?

Quels droits collectifs ?

Les données - dont font partie les données personnelles - lorsqu'elles sont utilisées par les programmes informatiques adaptés constituent le meilleur moyen de réduire les gaspillages, les dysfonctionnements, les accidents, les pertes de la plupart des systèmes et organisations humaines. Les données sont au cœur de la résolution de nos difficultés actuelles, du progrès positif de nos sociétés, de l'épanouissement des individus, du redémarrage de notre économie, de l'emploi, de la santé et de l'environnement. En ce sens, à la manière du savoir scientifique,

elles constituent un bien commun, non seulement par leur origine, lorsqu'il s'agit de données personnelles, mais par leur destination, ce qui en fait une cause d'utilité publique.
Par exemple : La moitié de la nourriture est gaspillée, notamment par le manque d'informations permettant le réajustement rapide des circuits de distribution. Un tiers de l'essence consommée est perdue en recherche de place pour se garer et donc par l'absence d'information à jour sur les emplacements disponibles. Et plus gravement, selon *IBM*, l'emploi des données permettrait de réduire la mortalité des patients hospitalisés de 20 pour cent.
Les privatisations rivales de données qui sont en cours nuisent au progrès général au sens où elles altèrent définitivement la concurrence. D'une part, par l'effet réseau : le premier acteur ne fera que se renforcer au détriment des autres et, d'autre part, soumettra au seul intérêt privé une ressource d'intérêt général.
C'est pourquoi, les partisans de la non-réglementation des données « pour favoriser l'innovation et la compétitivité » accomplissent, sciemment ou non, un contre-sens. Leur logique aboutit à l'éteignoir du monopole.
De même, l'individualisation juridique des données conduit à atomiser un droit collectif potentiel en une somme de droits privés plus facilement solubles : clic d'acceptation par clic d'acceptation.
Il n'est d'ailleurs pas étonnant que les entreprises du réseau les plus dataphages défendent séparément ou conjointement ces deux thèses : elles leur ouvrent grand les portes de la domination absolue. La première thèse est une extension brute du règne mercantile. La seconde, plus subtile, en phase avec notre tradition juridique, se donne habilement l'allure d'un progrès.
En réalité, la compétition doit se faire non pas sur l'appropriation des données mais sur leur usage.
À chaque entreprise de concevoir les meilleurs programmes informatiques - les algorithmes les plus efficaces - pour en tirer le sens et la valeur. La vraie compétition équitable et productive est là. Ainsi doit être actéel'obligation de mutualisation des données, sous l'égide et la gestion de l'Agence des données, afin d'en permettre l'accès réglementé mais ouvert à tous.
Par exemple : les données recueillies par les thermostats intelligents à domicile peuvent servir aux pouvoirs publics, à l'industrie du bâtiment, aux artisans de l'isolation, aux architectes, aux fournisseurs d'énergie, aux particuliers et à l'ensemble des prestataires informatiques qui concevront les logiciels d'exploitation de ces données pour leurs clients. Laisser ces données aux mains d'un seul acteur, ou de quelques-uns, dévitalise des filières entières.
La reconnaissance de la nature en réseau des données personnelles fait qu'un individu n'a plus la faculté de consentir seul à la cession ou à l'accès à ses données personnelles. Toute captation ou traitement de données personnelles doit passer par un agrément de l'Agence de données, préalable à tout accord individuel.
Par exemple : une personne veut rejoindre un réseau social et lui confier l'accès à ses données personnelles. Elle ne pourra le faire que si ce réseau social est préalablement agréé par l'Agence des données, ce qui garantira tout à la fois ses données et celles d'autrui qu'elle engage forcément.
Par comparaison, un citoyen achète de sa seule intention un produit alimentaire ou un jouet du commerce, cependant la mise sur le marché de ces derniers répond d'une autorisation administrative antérieure.
Nous avons pris d'ailleurs l'habitude de cette sécurité pour la plupart de nos achats et nous l'étendons naturellement aux services en réseau qui pourtant n'en bénéficient pas.
Le transfert de données personnelles vers un service non agréé sera constitutif d'une

infraction y compris à l'égard des personnes dont les données personnelles seraient impliquées.
Par exemple : une personne concède l'accès à son carnet d'adresses à un service de cartographie non agréé. Ce faisant, il livre sans autorisation les coordonnées de tiers qui pourront se retourner contre lui.
C'est donc cette Agence qui agréera toute captation de données sur le territoire national. Son statut public fonde une relation symétrique avec les grandes entreprises du réseau, plus équilibrée que les contrats d'adhésion souscrits d'un clic de souris par des particuliers pressés.
C'est l'Agence également qui agréera les dispositifs et logiciels permettant aux particuliers de recueillir les données personnelles d'autrui. Elle procédera également aux médiations et arbitrages pratiques entre citoyens, tout à la fois capteurs et captés.

Quelles sont les conditions de l'agrément de l'Agence des données ?

- la donnée personnelle doit être captée, conservée, traitée et transférée selon les protocoles et modalités fixées par l'Agence des données ;
- la captation, conservation, traitement ou transfert de données personnelles d'un citoyen européen répond des seules juridictions européennes, ce qui implique, de fait ou de droit, la localisation communautaire des serveursinformatiques ;
- l'exportation des données personnelles de citoyens européens hors du territoire de la communauté est limitée et taxée ;
- la régularisation de la situation fiscale du capteur au regard de l'activité réelle générée par l'usage des données personnelles captées sur le territoire national ;
- l'acceptation de la mutualisation des données sous le contrôle de l'Agence des données.

Quelles modalités pratiques pour le traitement des données personnelles ?

L'information sur un individu est source de valeur pour la collectivité. L'information sur un individu identifié est un risque privatif de liberté pour ce dernier. Il faut donc dissocier la personne - son identité -, de son profil - les informations recueillies - .
Pour atteindre cet objectif, il faut cesser de capter, traiter et transférer les données personnelles sans les protéger. Car toute information numérisée doit être considérée comme publique dès lors qu'elle n'est pas chiffrée. Cette aliénation de fait est une violation des droits individuels susmentionnés.
Ce chiffrement cryptographique doit donc garantir les droits individuels et collectifs afférents aux données sans pour autant en compromettre le meilleur usage.
Le chiffrement proposé est à triple clé. Chacune des clés ne permet le dévoilement que d'une partie seulement des éléments de la donnée.
Ainsi, une donnée est divisée en trois parties :
- *l'identifiant* : ce qui définit l'individu de manière unique : comme son nom, son visage, toute signature biologique (rétine, ADN, voix, empreinte digitale, etc.) ;
- *le profil utilisateur* : ensemble des données relatives à un utilisateur ; le profil est propre à chaque service ou réseau de services ;
- *l'information* : renseignements impliquant au moins une personne ou un profil.

Ce qui se présente de la manière suivante :
Exemple de données personnelles captées et conservées par un musée :
- niveau I : *XXX-XXX-ACTION* : *une visite est comptabilisée par le Musée.*
- niveau II : *XXX-PROFIL-ACTION* : *DR589 a revisité le Musée*.

- niveau III : *IDENTIFIANT-PROFIL-ACTION* : *Karima Dubois a revisité le Musée.*

Le premier niveau est accessible en données publiques.
Les conditions de recherche et de traitement des données personnelles non identifiées sont restreintes par des seuils de granularité et de combinatoire évitant une précision révélatrice d'identité. Il s'agit de garantir l'incertitude sur les personnes identifiées par des tailles d'échantillon en maintenant ainsi un niveau de flou.
Par exemple : « Combien de personnes ont-elles un chien dans tel quartier ? » maintient l'incertitude, tandis que : « Combien de personnes ont-elle un chien dans tel immeuble ? » peut être divulgateur.
Par ailleurs, le rapprochement non autorisé entre une action, un profil et une identité devient un délit.
Le second niveau donne accès à l'historique du profil créé par le collecteur. Les conditions de cet accès sont déterminées par l'Agence des données, de telle manière à préserver le secret de l'identité des profils.
Pour le musée, son travail de statistique et de relations client est fait pour l'essentiel au niveau I et II.
Le troisième niveau n'est accessible que sur une décision judiciaire donnant accès à la clé cryptographique correspondante. Le Ministre de Justice deviendra ainsi le *Garde des sceaux et des clés*.
Pour le musée, les informations de niveau III qu'il génère - comme par exemple les données de paiement recueillies - sont exclusivement réservées aux seuls usages internes éphémères agréés par l'Agence et ne peuvent faire l'objet d'aucun traitement externe ou être partie prenante d'une quelconque transaction avec des tiers.
Cette donnée personnelle encapsulée dans un chiffrage à triple niveau sera associée à des données additionnelles, ou métadonnées, de deux ordres :

- les premières sont d'accès libre et indiquent les conditions d'utilisation des données personnelles ainsi que les droits et restrictions spécifiques qui les accompagnent, instructions dont le respect est obligatoire ;
- les secondes constituent un historique de la capsule de données depuis son origine : c'est-à-dire la succession de toutes les opérations dont elle a été l'objet. Cette mémoire associée fonde, par exemple, l'authenticité de la monnaie virtuelle *Bitcoin*, en l'espèce, par l'historique des transactions rattachée à chaque unité de compte. Cette partie est cryptée et sous contrôle de la clé judiciaire.

En fait, l'encapsulation et les métadonnées font ainsi désormais de la donnée personnelle, jadis inerte, une donnée intelligente.
La capsule peut être elle-même un agent logiciel -un petit programme informatique- qui se peut se comporter et réagir de façon autonome en fonction de contraintes et de conventions spécifiques.
Un exemple nous est donné par le système logiciel *Ethereum* : chaque donnée porte avec elle-même de manière décentralisée les conditions de son usage.
Par exemple : une donnée personnelle encapsulée dans un agent logiciel contenant une position de circulation d'un véhicule est accessible en niveau II aux services de gestion de trafic qui en ont besoin. L'agent logiciel qui reconnaît l'origine autorisée de la demande, l'authentifie, la valide, donne accès à la donnée, puis inscrit en métadonnées, la consultation de l'information.
L'agent logiciel pourra d'ailleurs être partiellement programmé par l'utilisateur à l'origine des données pour déterminer une relation spécifique d'accès limité avec des services agréés - sous

réserve des droits de tiers - et ce à la manière de la licence *Creative Commons*.
Les progrès fulgurants de la capacité des processeurs, du stockage, de la bande passante et de l'efficacité des algorithmes font que le surpoids issu de la protection des données ainsi que des opérations informatiques associées seront vite compensés.
Enfin, les machines devront prohiber par elles-mêmes les usages non autorisés des capsules de données personnelles comme, par exemple, des duplications ou des tentatives d'accès. Et ce, à la manière des photocopieurs ou des imprimantes qui interdisent la copie dès qu'elles reconnaissent que l'image à reproduire est un billet de banque.
L'Agence des données supervise et coordonne cette gestion globale des données. Aucune donnée n'est conservée par l'Agence.
Le code n'est pas ouvert pour garantir l'unicité et la sécurité des versions, éviter les malveillances et les abus et garantir l'immédiateté des mises-à-jour et la stabilité. En revanche, l'ensemble des procédés, logiciels et méthodes de l'Agence est soumise à l'examen contradictoire et publié d'une *Cour des codes*.
Certaines parties des codes source peuvent cependant être examinées sur demande et ainsi ouverte tant à l'inspection qu'à l'amélioration par le public à la manière du logiciel libre. Tous les codes, méthodes et protocoles, sauf exception limitée et motivée, sont à disposition de la Justice.
Le reconnaissance de l'indivision en réseau des données personnelles et de leur statut de bien commun, la reconnaissance des droits individuels et collectifs sur cette ressource ; la création d'une Agence des donnéespour les gérer, la mise en pratique de ce dispositif légal par un triple chiffrement des données personnelles associé à des métadonnées de traitement, un contrôle judiciaire et un contrôle contributif de la société civile sur les méthodes et procédures, voilà qui apportent les garanties civiques nécessaires tout en accélérant le progrès et l'innovation par la mutualisation maîtrisée des données.
Ces principes et pratiques des données en réseau sont applicables d'abord en France mais ont pour objectif ladimension européenne puis mondiale.