Authentification, identification et tiers de confiance

Michel Arnaud

Michel Arnaud

Université Paris X Nanterre
CRIS, université Paris X Nanterre

AUTHENTIFICATION, IDENTIFICATION ET TIERS DE CONFIANCE

L'étude des structures organisationnelles, des usages des outils numériques, des représentations individuelles et des projets collectifs montre que les technologies de l'information et de la communication (TIC) peuvent être considérées comme des dispositifs de médiation dans la communication des organisations. La médiation en matière de protection des données personnelles est cruciale pour la rendre effective aussi bien que pour en contrôler l'accès. La médiation confiée à un tiers de confiance permet de gérer l'intentionnalité des individus et les jeux d'interactions : elle garantit les procédures d'authentification nécessaires aux parties prenantes pour sécuriser les transactions, sans pour autant révéler l'identité des personnes, sauf dans certaines conditions bien précises. Les procédures de pseudonymisation fournissent des pistes pour rendre inviolable l'accès aux données personnelles confiées aux tiers de confiance, tout en aussi faisant sauter les restrictions de l'anonymisation, qui freinent le développement du commerce électronique.

La protection des données personnelles et ses limites

La définition et le traitement des données personnelles

L'article 2 de la directive européenne 95/46 du 24 octobre 1995 définit les « données à caractère personnel » ainsi : « toute information concernant une personne physique identifiée ou identifiable (personne concernée) ; est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale. » Cette définition plutôt large concerne aussi bien les données directement nominatives (nom, prénom, date de naissance…) que celles qui le sont indirectement comme un matricule, une adresse, un numéro de téléphone, un élément biométrique, une adresse IP Internet, les traces des données de connexion, etc. – tout ce qui permet de remonter indirectement à la personne.

Pour limiter la collecte et les échanges de données personnelles, la CNIL a établi un certain nombre de préconisations qui ont été améliorées au cours du temps et peuvent se résumer comme suit :

- Ne collecter les données qu'au niveau de finesse strictement nécessaire.
- Répartir les données dans des fichiers ou des systèmes informatiques distincts du reste des applications en veillant à ce qu'aucun lien logique ne soit établi entre ces fichiers et les autres applications.
- Ne pas fournir systématiquement un logiciel d'interrogation généraliste, c'est-à-dire permettant de croiser n'importe quels critères. Le logiciel d'interrogation généraliste doit être réservé à un petit nombre d'experts nommément habilités, les autres utilisateurs n'ayant accès aux statistiques que via des requêtes préprogrammées.
- Effacer si possible immédiatement après le traitement qui les a utilisées les données personnelles permettant d'identifier la personne.

L'anonymisation des traces

Dans le considérant 26 de la directive européenne 95-46, il est spécifié que les principes de la protection ne s'appliquent pas aux données rendues anonymes d'une manière telle que la personne concernée n'est plus identifiable. L'anonymat consiste en la possibilité de garder les traces d'une personne (caractéristiques, comportements, etc.) sans avoir la moindre possibilité de connaître sa véritable identité. La CNIL a approuvé les procédés qui permettent de rendre anonymes des données personnelles particulièrement sensibles dans le domaine de la santé publique (études épidémiologiques), des décisions de justice publiée sur Internet, de l'agrégation des données du recensement de la population ou encore des archives électroniques des entreprises privées, sans pour autant en faire une promotion inconditionnelle

dans la mesure où ce principe peut entrer en contradiction avec la défense de l'ordre public ou encourager les dénonciations anonymes et multiplier les mises en cause calomnieuses.

Sur le plan technique, le paradoxe anonymat *et* suivi est résolu grâce à l'utilisation d'une fonction de hachage, opération consistant à calculer une valeur numérique utilisée comme identifiant unique et irréversible à partir des donné directement ou indirectement nominatives d'une personne, cette valeur étant ensuite substituée aux données à partir desquelles elle a été calculée. La CNIL s'appuie sur la quasi-impossibilité mathématique de retrouver à partir du résultat final (la valeur numérique d'anonymisation) les données directement ou indirectement nominatives ayant données lieu au calcul de hachage. La méthode usuelle consiste à multiplier les clefs ou les parties de clef, chacune détenue par des personnes ou organismes différents. Toutefois, la levée de l'anonymat est encore possible pour corriger les anomalies avec des tables destinées à maintenir la correspondance entre un code identifiant de la personne et la valeur d'anonymisation calculée par la fonction de hachage.

Si la meilleure riposte est de demander l'anonymat dès le traitement des données, il est très difficile ensuite de l'imposer *a posteriori* à cause de deux pressions d'origines différentes mais allant dans le même sens d'une conservation sans anonymisation : l'obsession sécuritaire et l'impératif commercial.

La pression sécuritaire pousse les citoyens à vouloir être protégés des terroristes et à placer les méthodes d'identification à cet effet avant la protection des données personnelles. Dans la mesure où les terroristes se cachent souvent sous des apparences banales trompeuses, il est essentiel d'utiliser des moyens de détection qui empiètent sur la vie privée pour tenter de les identifier. Les Etats en réponse à ce besoin s'exonèrent des injonctions de la CNIL, par exemple en France, en étendant la mise en fiche systématique des citoyens avec les fichiers de police (STIC, JUDEX, Ardoise), la vidéosurveillance, la biométrie – avec le fichier des empreintes génétiques (FNAEG) appliquée à tous les degrés de délinquance –, le contrôle des étrangers en situation irrégulière (ELOI), l'autorisation de stockage des données personnelles étant donnée une fois pour toutes par un magistrat après la création de ces fichiers par décrets. L'administration fiscale a des facilités qui lui sont accordées à cause du traitement des recouvrements. Les avantages du « tiers autorisé » font qu'elle a accès à tous les fichiers et peut les interconnecter sans problème. Au niveau européen, le contrôle des passagers aériens arrivant dans l'Union (PNER européen avec 19 données personnelles) s'aligne sur le système américain (qui recueille 34 données différentes sur chaque voyageur débarquant aux Etats-Unis sans aucune garantie ni d'effacement des données personnelles ni d'anonymisation des traces).

D'autre part, les consommateurs manifestent un désir toujours plus impérieux d'avoir à leur disposition des services adaptés à leurs besoins, même si c'est au prix de la divulgation de leurs données personnelles. Il s'agit d'une mutation culturelle encore plus forte et plus durable que celle provoquée par la lutte anti-terroriste. L'ère de la consommation de masse est révolue en ce sens que la personne n'éprouve plus d'intérêt à se conformer à un comportement grégaire. Le communisme et le fascisme ont été des doctrines de masse où la

personne se fondait dans le collectif de manière anonyme. Le capitalisme se targue de savoir individualiser la relation au client. Sans cultiver le paradoxe, l'anonymisation des traces peut être vue comme un retour en arrière. Le service public est confronté à une contradiction : en s'adressant de plus en plus à des clients et non plus seulement des administrés, il doit faire face au risque de ne plus être accessible à tous car le coût de la personnalisation du service est élevé. Cette évolution est pourtant inéluctable car l'administré consommateur demande une combinatoire de services en vue d'une simplification des actes de sa vie quotidienne. Dans le secteur privé, l'époque de la publicité de masse est révolue au profit du marketing direct. L'approche par profils socio-économiques et segmentation des cibles de marché décriée par Packard et Marcuse est à présent remplacée par un contrôle de l'attention, obtenu par un ciblage fin et précis grâce à l'accumulation de données personnelles stockées à partir des traces laissées volontairement ou involontairement par le consommateur et dont l'analyse permet de prévoir ses moindres envies. Les banques ne sont pas en reste pour l'évaluation des risques liés au crédit à la consommation. La CNIL tend à interdire le traitement automatique de la gestion des comptes et des prêts par la technique du *scoring*. Mais les banques continuent à donner accès à leurs fichiers de mauvais payeurs à qui en a besoin et déposent des recours devant le Conseil d'Etat pour continuer leurs pratiques, contestant ainsi l'autorité de la CNIL.

Complémentarité des intérêts des Etats et des multinationales

Les flux d'information croissant sans cesse, savoir gérer l'information sur l'information personnelle est un enjeu du pouvoir, pour les Etats comme pour les entreprises.

Au fur et à mesure que les moteurs de recherche se perfectionnent, les utilisateurs confient de plus en plus d'informations à la Toile, en posant des questions de plus en plus intimes. Les grands fournisseurs d'accès, comme Google, Yahoo!, AOL ou Microsoft, qui regroupent la majorité des internautes américains et multiplient les services en ligne, ne prospèrent qu'en augmentant le chiffre d'affaires généré par la publicité en ligne.

Le modèle économique de la gratuité d'accès aux services sur la Toile repose sur les achats générés par les clics des usagers sur les bannières publicitaires. Ce budget ne progresse que si les annonceurs estiment qu'ils peuvent de mieux en mieux cibler les consommateurs potentiels, réclamant à ces opérateurs des données de moins en moins anonymes sur leurs clients et de plus en plus d'informations personnelles pour prévoir ou influencer leurs achats. Car les opérateurs peuvent collecter à distance les données personnelles de l'internaute stockées par son navigateur et son système informatique. Les fichiers que l'on appelle cookies gardent la trace des sites visités ainsi que des informations sur les profils des utilisateurs.

En août 2006, le fournisseur d'accès à Internet AOL a involontairement rendu public sur le Web le détail de recherches en ligne effectuées par 600.000 de ses abonnés en avril et mai 2006, au total près de 20 millions de requêtes, classées par mots clefs. Avec le numéro de

sécurité sociale facilement disponible, les numéros de téléphone, les achats de produits spécifiques, il a été facile aux journalistes de retrouver l'identité des internautes même si AOL avait pris la peine de l'anonymiser sous un numéro d'identification.

De son côté, Amazon a déposé un brevet pour un système d'interrogation de sa base de données clients – soit plus de 59 millions de personnes –, permettant non seulement d'obtenir les données personnelles habituelles, l'endroit où l'acheteur habite mais aussi son niveau de revenu, son orientation sexuelle, sa religion, sa race.

Quant à Google, la société avait prévu de conserver ses cookies jusqu'en 2038. Sous la pression du G29 représentant les CNIL européennes, la compagnie a décidé de la réduire à 2 ans à partir de mars 2007. En avril 2008, les Etats européens ont jugé cette durée trop longue et ont réussi à forcer Google à la réduire à 6 mois depuis septembre 2008. Mais Google entend différencier « protection des données personnelles » et « rétention sans limite de temps des données anonymisées » pour ne pas tuer la poule aux œufs d'or.

Cette chasse aux données personnelles ne vient pas seulement s'ajouter aux systèmes de surveillance mis au point par les États : elle contribue aussi à les alimenter. C'est ce que montre un rapport publié par la Cour des comptes américaine en mars 2008, dénonçant l'achat en 2005 par des services publics comme la justice, la police ou les impôts des données personnelles à des entreprises privées, sans respecter les principes de protection au nom de l'assurance sociale et de la lutte contre le terrorisme et la fraude fiscale 1. C'est encore ce que révèle le projet LifeLog de la DARPA 2, qui consiste à mettre au point un système permettant d'enregistrer l'intégralité des évènements, états et relations d'une personne, pendant sa vie entière, à partir des informations fournies par les cartes bancaires, les sites web visités, le contenu des conversations téléphoniques, les livres lus, les émissions de télévision ou de radio sélectionnées, les données repérées par GPS, les informations biomédicales. Le but est d'identifier ainsi les préférences, plans, buts, et autres marqueurs d'intention de chaque citoyen. On le voit, le contrôle des individus par les Etats passe désormais par la récupération des données détenues par les firmes gérant l'accès à Internet : surveillance commerciale et politique vont de pair, même si elles ne visent pas les mêmes objectifs.

Vers une redéfinition des données personnelles

La relation à l'identité a changé : les appareils d'accès au numérique deviennent des compagnons de vie, nouvelles prothèses pour lire, écrire, écouter, voir et échanger. Ces

---

[1] United States Government Accountability Office, *PRIVACY Government Use of Data from Information Resellers Could Include Better Protections*, Statement of Linda D. Koontz, Director Information Management Issues, Testimony Before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, March 11, 2008 http://www.gao.gov/htext/d08543t.html

[2] Defense Advanced Research Projects Agency : agence de projets de recherche avancée de defence <http://en.wikipedia.org/wiki/LifeLog>.

matériels accroissent les interactions dans le monde virtuel, avec des conséquences sur la personne réelle qui s'en trouve non seulement façonnée dans sa gestion du temps et de l'espace mais aussi dans la relation à sa propre intimité puisque les réseaux sociaux l'amènent à révéler ce qui reste d'habitude caché. Si le plaisir de se voir dans le miroir numérique tendu aux autres personnes rencontrées sur les réseaux est réel, comment considérer que les données sont encore personnelles quand elles sont ainsi exposées ? On peut se demander si les internautes qui révèlent tout d'eux-mêmes sur les réseaux sociaux sont conscients des dangers ou bien s'ils sont manipulés. Le fait est que leur liberté de choix doit leur être reconnue comme un droit fondamental.

Au lieu de définir comme données personnelles l'ensemble des identifiants et des traces permettant directement ou indirectement de retrouver la personne, ne vaudrait-il pas mieux séparer les données identitaires (nom, prénom, date de naissance, adresse, etc.) des traces laissées sur les réseaux lors des déplacements par exemple, des données liées aux paiements et de leur appliquer des traitements différents ? Un pas dans cette direction est franchi avec la directive européenne 2002-58 qui différencie les données proprement identitaires des « données relatives au trafic », qui sont conservées en tant que telles avec des systèmes d'identification des numéros appelés et des « données de localisation » indiquant la position géographique de l'équipement terminal d'un utilisateur d'un service de communications électroniques. L'examen objectif de la situation oblige cependant à constater que l'injonction d'effacement des données personnelles et d'anonymisation des traces est rarement observée. Nous proposons donc d'élargir cette définition de données non directement personnelles aux données comportementales, à savoir tout ce qui est relatif aux traces laissées sur les réseaux permettant de reconstituer les déplacements, les achats, etc. La loi française sur la confiance numérique du 21 juin 2004 oblige les opérateurs de communications électroniques, les fournisseurs d'accès à Internet (FAI) et les hébergeurs à conserver les données liées aux contenus. Le décret d'application liste toutes les données susceptibles d'identifier tout créateur de contenu en ligne et à conserver durant un an : adresse IP, mot de passe, login (nom de connexion), pseudonyme, terminal utilisé, coordonnées de la personne physique ou morale, identifiants des contenus.

Les nombreuses entorses aux règles de la CNIL et du G29 (groupe intergouvernemental chargé d'appliquer la directive européenne 95-46) prônant l'effacement immédiat des données personnelles, l'anonymisation de toutes les traces et la non interconnexion des fichiers, nous pousse à déplacer le curseur de l'interdiction définitive d'accès aux données personnelles vers un contrôle renforcé des procédures d'accès aux données. La notion de données personnelles est à revoir : elle pourrait être scindée en deux blocs, données identitaires d'un côté, tous les types de traces de l'autre. Cela permettrait à la fois de réduire le champ à couvrir pour rendre la protection plus effective et de mieux contrôler les procédures établissant des liens entre les deux types de données, limitant la reconstitution des comportements à des conditions bien définies.

Les potentiels de la pseudonymisation

L'étape actuelle de la découverte des réseaux sociaux se caractérise par le recours à des modèles de comportement liés au contexte. Pourquoi est-on enclin à révéler sa véritable identité quand on décrit ses goûts intimes sur les réseaux ? Est-ce à cause de l'idéologie libertaire qui a caractérisé les débuts d'Internet ? Celle-ci privilégie en effet la recherche de l'authenticité, l'affirmation de sa propre identité étant proportionnelle à l'intensité du plaisir à communiquer, en opposition à ce qui est connoté comme hypocrite (la double vie, les faux semblants, l'usage des pseudos). La duplicité a également acquis une image péjorative sur les réseaux à cause des adultes qui se font passer pour des adolescents sur les forums de discussion. Mais entre se défouler en disant tout et refuser de s'assumer en se cachant derrière un pseudo, il y a la place pour un sens de la protection de soi qui se développera au fur et à mesure que les risques se concrétiseront. Si la communication par l'intermédiaire des réseaux touche à la gestion des pulsions, nous sommes probablement dans une phase de découverte adolescente. Le pseudo pourrait être un moyen de passer à un comportement plus mature.

Assurer sous certaines conditions le caractère hermétique d'un pseudo peut permettre de rétablir la frontière entre vie publique et vie privée. Les procédures permettant de différencier la vraie personne, connue par ses données identitaires, des personnalités d'emprunt qu'elle peut adopter pourraient être essentielles au maintien des droits fondamentaux. S'il est actuellement facile pour un spécialiste de découvrir l'identité réelle derrière un pseudo, on peut envisager des mesures plus efficaces de protection permettant de séparer l'identification de l'authentification avec l'intervention de tiers de confiance comme c'est le cas avec « mon.service-public.fr » et le dossier médical personnel.

Le concept de pseudonymat est différent de l'anonymat : il s'agit de la possibilité accordée à une personne de disposer d'une autre identité qui ne pourra pas être facilement rattachée à sa véritable identité. En matière de liberté d'expression et de contrôle des traces, cette approche semble appropriée car elle permet aux internautes de garder leurs différentes sphères (privée, associative, professionnelle) séparées tout en évitant d'accorder un moyen de diffuser des contenus illégaux sans en assumer la responsabilité ou de commettre des méfaits sous une autre identité. Le pseudo est lié à un numéro d'identification – par exemple un matricule, un code client, un numéro d'adhérent ou le numéro de sécurité sociale – qui permet d'identifier de façon certaine une personne si l'on dispose de la table de correspondance. Son accès doit être soigneusement réglementé car il est garant de la liberté de ne pas révéler qui on est vraiment. Ces procédures à la fois techniques et administratives demandent à être validées par des juristes et des techniciens avec les certifications correspondantes. Il s'agit de caractériser les circonstances requérant l'acte d'identification qui peut intervenir dans des conditions précises (sur requête de la police, des services de l'impôt, des banques, de l'administration, etc.) et de valider juridiquement et techniquement l'acte d'authentification, de telle sorte qu'il inspire confiance aux interlocuteurs du pseudo. L'importance de ces deux aspects complémentaires mais parfaitement distincts donne la mesure de la place à accorder au tiers

de confiance, institution chargée de garantir la valeur du pseudo qu'elle certifie en termes d'authenticité (ce pseudo correspond bien à une personne réelle), ce qui implique d'assurer d'autres critères liés, tels que sa solvabilité par exemple. Dans ce cas, le pseudo certifié comporte un taux de crédit à préciser, correspondant à celui de la personne qui se cache derrière le pseudo. Le traitement des traces associées à un pseudo ne pose plus problème puisqu'il y a à la fois absence d'identification de la personne réelle et possibilité de bâtir son profil à des fins de suivi commercial et administratif.

L'institution tiers de confiance devient pivot dans la nouvelle économie des flux puisqu'elle permet d'obtenir les garanties nécessaires à l'établissement de transactions instantanées et sécurisées avec l'émission de certificats électroniques correspondant aux pseudos certifiés. Le tiers de confiance regroupe les autorités de certification, d'enregistrement et l'opérateur de certification. L'architecture « Liberty Alliance » propose des solutions dans ce sens tandis qu' « Open ID » ne donne aucune garantie sur le tiers de confiance. Les cartes d'authentification forte viennent compléter le dispositif. Est-ce que les données identitaires doivent être stockées sur une carte à puce détenue par la personne ou bien déposées dans une banque centrale des identités ou un réseau de banques d'identités indépendantes des pouvoirs publics ? L'émission de pseudos certifiés (20 par an et par personne par exemple) pourrait constituer, comme dans le cas de la masse monétaire, un moyen de contrôle de la flexibilité et du volume des identités permises et un dispositif de garantie pour l'exercice de la liberté individuelle. Les pseudos non certifiés, eux, (avatars, etc.) seraient laissés au libre choix du citoyen. La captation de pseudos certifiés et la révélation intempestive d'une identité hors l'action d'un juge seraient sévèrement réprimées.

Il est temps à nos yeux de dépoussiérer la loi française « informatique et libertés » de 2004 et d'adapter la notion de données personnelles aux nécessités du commerce électronique et du suivi administratif des personnes, en instituant la différenciation entre identification et authentification, en instaurant le contrôle sur la préservation de l'identité, en offrant les facilités liées à des pseudos certifiés, véritable refondation de la liberté individuelle à l'ère des réseaux. Au lieu du panoptique de Foucault qui s'installe insidieusement avec la biométrie utilisée pour tous les actes de la vie quotidienne, il convient de retrouver le sens premier du symbole, sensé détourner la pulsion en la sublimant, comme l'ont noté Godelier et Malinowski. Faciliter l'usage de pseudos certifiés sur les réseaux permet à notre sens d'encourager un transfert culturel car ils renforcent le lien social en multipliant les signes de reconnaissance et d'échange entre les personnes. Ils sont les instruments d'une sagesse collective encore à construire dans le cadre de la culture du XXI^{e} siècle.

REFERENCES BIBLIOGRAPHIQUES

Foucault Michel, *Surveiller et punir*, Paris : Gallimard, 1993, 360 p.
Godelier Maurice, *L'idéel et le matériel, Pensée, économies, sociétés*, Paris : LGF, 1992, 348 p.

Manilowski Bronislaw, *Journal d'ethnographe,* Paris : Seuil, 1985, 301 p.

Marcuse Herbert, *One-Dimensional Man*, Boston : Beacon Press, 1964, 260 p.

Packard Vance, *The hidden persuaders*, New York : Ig Publishing, 2007, 200 p.

Pierce John R., *An Introduction to Information Theory, Symbols, Signals and Noise*, Dover Sciences Books, Mineola N.Y., 1980, 295 p.

Rheingold Howard, *Les communautés virtuelles. Autoroutes de l'information : pour le meilleur ou pour le pire ?* Paris : Addison Wesley France, 1995.