PROTECTION DE LA VIE PRIVÉE ET DES DONNÉES PERSONNELLES DANS L'ENVIRONNEMENT NUMÉRIQUE

Jean-Noël Colin *et Cécile* de Terwangne

Le présent ouvrage est consacré à la protection de la vie privée et des données à caractère personnel. Dans l'approche suivie, il ne s'agit pas de proposer au lecteur un précis sur le droit au respect de la vie privée dans son ensemble. La matière est en effet tentaculaire et un tel ouvrage devrait couvrir le droit à l'intégrité physique et morale, au nom, à l'honneur, à l'image, à une vie familiale, à un environnement sain, et la liste est loin d'être close. L'objectif de cet ouvrage consiste plutôt à mettre à disposition des praticiens et de toute personne confrontée à des questionnements liés à la protection des individus face aux développements techniques et sociétaux et à la tournure que prend notre société, un outil apportant les réponses juridiques à de tels questionnements.

Le développement spectaculaire des technologies de l'information et de la communication (TIC) offre de grandes possibilités et de nombreux avantages. Le recours aux ordinateurs et aux réseaux de communication, et en particulier à Internet, a permis le déploiement de services inimaginables, tout en accroissant l'efficacité et l'accessibilité des services classiques. En bien des situations, cette évolution des choses facilite grandement la vie.

L'utilisation de ces technologies présente toutefois aussi de nouveaux dangers pour la vie privée et les libertés de chacun. Données recueillies à l'insu des personnes, données réutilisées pour des finalités inavouées, données conservées des mois, voire des années, données transmises à des tiers, données confidentielles diffusées : la réalité concernant le sort des données à caractère personnel dans l'environnement numérique d'aujourd'hui a bien des faces noires. Les individus faisant usage du réseau et de toute la variété de services en ligne existant désormais, que ce soit par le biais d'un ordinateur, d'un téléphone portable (*smartphone*) ou d'une tablette, perdent, dans une grande mesure, la maîtrise de leurs données. Ils ne savent pas ce qui en est fait, ils ne peuvent contrôler à distance qui y accède. Une série d'acteurs de l'Internet et des nouveaux médias, par contre, connaissent leurs goûts, leurs centres d'intérêt, leurs mouvements, les endroits et les personnes qu'ils fréquentent…

Les technologies offrent par ailleurs des possibilités de contrôle et de surveillance jamais égalées par le passé. Caméras de vidéosurveillance, puces RFID (permettant l'identification par radiofréquence), techniques de géolocalisation, pour ne citer que ces outils-là, dessinent peu à peu les contours d'une société qui a pu déjà être qualifiée de « société de surveillance ». La surveillance technique est omniprésente, que ce soit dans les rues et les parcs, au travail, dans les aéroports et les gares, dans les magasins, dans les entrées d'immeubles, dans les écoles, dans les crèches, et jusque dans les maisons grâce aux paquets-cadeaux permettant l'installation de circuits de caméras privés que l'on peut s'offrir à Noël…

L'ensemble de cette réalité met en cause le droit au respect de la vie privée ainsi que le droit à la protection des données.

Il convient, à l'entame du présent ouvrage, de définir la portée des deux notions clés qui seront déclinées dans tous les contextes différents alimentant les nombreux chapitres de l'ouvrage : la notion de « vie privée » et celle de « protection des données à caractère personnel » (chapitre 1.1).

Il semble aussi opportun d'éclairer le lecteur non versé dans les spécificités de la « société de l'information » ou de la « société de surveillance », en présentant en exergue de l'ouvrage (chapitre 1.2) les principaux développements technologiques qui soulèvent de nouveaux risques et suscitent de nouveaux défis pour la vie privée et la protection des données, dans le sens où ces dernières sont atteintes, mises à mal ou, tout simplement, en jeu en présence de ces développements. Les risques découlant non de la technologie elle-même, mais des usages qui sont apparus en s'appuyant sur les potentialités offertes par la technologie, seront, eux aussi, présentés (chapitre 1.3).

CHAPITRE 1.1. NOTIONS DE « VIE PRIVÉE » ET DE « PROTECTION DES DONNÉES À CARACTÈRE PERSONNEL »

La **vie privée**, dans ce contexte, ne doit pas se comprendre de façon traditionnelle comme une sphère intime à protéger, contenant un ensemble d'informations privées, voire confidentielles, que l'on souhaite garder cachées. Elle est à entendre comme faculté d'autodétermination, d'autonomie, capacité de l'individu à effectuer des choix existentiels[1]. En la matière, il s'agit plus précisément **d'autodétermination informationnelle**[2], c'est-à-dire du droit pour l'individu de « savoir ce qui se sait sur lui », de connaître les données le concernant qui sont détenues, d'en maîtriser les circuits de communication, d'en contrecarrer les utilisations abusives. La vie privée ne se réduit donc pas à une quête de confidentialité, c'est la **maîtrise** par chacun de son image informationnelle.

C'est en ce sens que l'Assemblée parlementaire du Conseil de l'Europe a veillé à compléter sa Résolution 428 (1970). En effet, le droit au respect de la vie privée garanti par l'article 8 de la Convention européenne des droits de l'homme (CEDH) avait été défini par l'Assemblée en janvier 1970 dans sa « Déclaration sur les moyens de communication de masse et les droits de l'homme » contenue dans cette Résolution comme « le droit de mener sa vie comme on l'entend avec un minimum d'ingérence ». Près de trente ans après l'adoption initiale de ce texte, l'Assemblée parlementaire a précisé que, « [p]our tenir compte de l'apparition des nouvelles technologies de la communication permettant de stocker et d'utiliser des données personnelles, il convient d'ajouter à cette définition *le droit de contrôler ses propres données* »[3].

---

1. Pour la reconnaissance explicite d'un droit à l'autodétermination ou l'autonomie personnelle contenu dans le droit au respect de la vie privée de l'article 8 CEDH, voy. Cour eur. D.H., arrêt Evans c. Royaume-Uni, 7 mars 2006, req. n° 6339/05 (confirmé par la Grande Chambre dans son arrêt du 10 avril 2007) ; arrêt Tysiac c. Pologne, 20 mars 2007, req. n° 5410/03 ; arrêt Daroczy c. Hongrie, 1er juillet 2008, req. n° 44378/05.
2. Voy. H. Burkert, « Le jugement du tribunal constitutionnel fédéral allemand sur le recensement démographique », *Droit de l'Informatique et des Télécoms*, 1985, 8-16 ; C. de Terwangne, « Le rapport de la vie privée à l'information », *in* E. Montero (dir.), *Droit des technologies de l'information. Regards prospectifs*, coll. Cahiers du CRID, n° 16, Bruxelles, Bruylant, 1999, p. 144 ; Idem, « Loi relative à la publicité de l'administration et loi relative à la protection des données personnelles : regards croisés sur deux voies d'accès à l'information », *in Transparence et droit à l'information*, coll. Formation permanente CUP, Liège, vol. 55, 2002, p. 90 ; Th. Leonard et Y. Poullet, « Les libertés comme fondement de la protection des données nominatives », *in* F. Rigaux, *La vie privée : une liberté parmi les autres ?*, Travaux de la Faculté de droit de Namur, n° 17, Bruxelles, Larcier, 1992, pp. 231 et s.
3. Résolution 1165 (1998) de l'Assemblée parlementaire du Conseil de l'Europe sur le droit au respect de la vie privée, adoptée le 26 juin 1998 (c'est nous qui soulignons).

La **protection des données à caractère personnel**[1] est une émanation du droit au respect de la vie privée pris dans la dimension de droit à l'autodétermination qui y est lié. C'est le droit pour chacun de contrôler ses propres données, qu'elles soient privées, publiques ou professionnelles.

S'autonomisant du droit au respect de la vie privée, le droit à la protection des données suppose la prise en compte, d'une part, des déséquilibres de pouvoirs entre la personne concernée et celui qui traite les données, déséquilibres engendrés par les capacités de traitement des données à disposition de ce dernier et dramatiquement exacerbés aujourd'hui du fait des développements techniques et, d'autre part, de l'impact que les traitements de données peuvent avoir sur les divers droits et libertés des individus. D'autres droits et libertés que la seule vie privée entrent en effet en ligne de compte, tels que la liberté de se déplacer, celle de s'assurer, celle de se loger, celle de trouver un emploi, celle de s'informer et de s'exprimer en toute transparence, etc.

La Charte des droits fondamentaux de l'Union européenne, texte devenu juridiquement contraignant depuis l'entrée en vigueur du Traité de Lisbonne et qui sera analysé dans les pages du titre 1er de cet ouvrage dédiées au droit européen, est le premier catalogue international de droits de l'homme à distinguer explicitement les concepts de vie privée (art. 7) et de protection des données (art. 8)[2].

---

1. Pour une définition précise de la « donnée à caractère personnel », voy., *infra*, chapitre 1.2.
2. *Art. 7.* – « Toute personne a droit au respect de sa vie privée et familiale, de son domicile et de ses communications. »
   Art. 8. – « 1. Toute personne a droit à la protection des données à caractère personnel la concernant.
   2. Ces données doivent être traitées loyalement, à des fins déterminées et sur la base du consentement de la personne concernée ou en vertu d'un autre fondement légitime prévu par la loi. Toute personne a le droit d'accéder aux données collectées la concernant et d'en obtenir la rectification.
   3. Le respect de ces règles est soumis au contrôle d'une autorité de protection des données. »

CHAPITRE 1.2. DÉFIS POUR LA VIE PRIVÉE ET LA PROTECTION DES DONNÉES POSÉS PAR LA TECHNOLOGIE

La puissance de calcul et de stockage toujours plus importante, la connectivité toujours plus étendue rendent possible le développement de nouvelles technologies et applications qui constituent de véritables défis pour la vie privée et la protection des données. Elles impliquent bien souvent une collecte massive de données personnelles sur les citoyens, acheteurs en ligne, utilisateurs de réseaux sociaux, etc. Cela se réalise fréquemment à l'insu de ceux-ci. On observe, par ailleurs, l'utilisation de plus en plus répandue d'identifiants permettant de lier un utilisateur à ses actions, sa position géographique ou ses données. C'est le cas notamment de l'adresse IP (Internet Protocol)[1], de l'identifiant présent sur un tag RFID (voy. *infra*) ou d'un numéro de session dans un cookie (voy. *infra*). De plus, ces informations peuvent être analysées et corrélées pour en déduire d'autres, à des fins de profilage par exemple. Enfin, le stockage et la diffusion des informations collectées ou inférées échappent de plus en plus souvent au contrôle de la personne concernée, qui se retrouve impuissante devant l'utilisation parfois abusive qui en est faite. Les conséquences en sont des risques accrus de fuites d'information et de traçage des personnes, mettant ainsi à mal la vie privée de celles-ci.

Dans les points qui suivent, nous passerons en revue une série de technologies émergentes ou en mutation, en décrivant d'une manière générale leurs applications et leur fonctionnement, mais aussi les menaces potentielles qu'elles présentent pour la vie privée et la protection des données.

1. Convergence des moyens de communication

L'évolution des moyens de communication et des services de diffusion et de partage d'information conduit à une convergence de plus en plus grande entre ces différents

1. Une adresse IP est un numéro d'identification unique attribué à chaque appareil connecté à un réseau informatique (Internet) utilisant l'Internet Protocol. Elle est attribuée à l'appareil de façon permanente ou provisoire.

systèmes, avec, pour conséquence, un manque de plus en plus important de transparence quant aux véritables outils utilisés et, surtout, une perte de contrôle de la diffusion de l'information, qui circule, est agrégée, remise en forme, réexpédiée...

Ainsi, le téléphone, doté d'une puissance de calcul et de stockage, devient par là « intelligent » (*smartphones*) ; l'ordinateur permet de téléphoner ; la vidéoconférence est disponible sur des baladeurs MP3 ; un numéro de fax est en fait une façade pour un courriel ; les appels vers un GSM peuvent être redirigés vers un poste fixe, avant d'échouer sur la boîte vocale d'un service de type VOIP (*Voice Over IP* – téléphonie sur réseau IP) consultée sur un PC. Ces exemples montrent à quel point il devient très compliqué pour un utilisateur de déterminer le type de moyen de communication utilisé et, surtout, où vont et d'où proviennent les informations envoyées ou reçues.

Mentionnons encore à ce sujet des développements tels que le « Outlook Social Connector » de Microsoft, qui permet aux destinataires d'un courriel d'obtenir le statut Facebook de l'expéditeur. Ceci montre la confusion de plus en plus grande entre des sphères qui, jusqu'ici, étaient clairement distinctes et les risques de diffusion d'information non souhaitée que cela entraîne.

2. Géolocalisation

Des moyens de plus en plus sophistiqués et précis permettent d'établir la position géographique d'un utilisateur, que ce soit directement d'après des informations obtenues par son terminal (au moyen d'une puce GPS, de plus en plus répandue dans les téléphones portables) ou via le réseau auquel il est connecté (par triangulation des bornes GSM ou l'utilisation de bases de données reprenant la localisation des réseaux Wi-Fi – voy., à ce sujet, les informations collectées par la *Google Car*[1]).

La gestion des transports publics de manière électronique permet aussi de suivre les déplacements des usagers, par exemple à partir de la validation de leur titre de transport auprès de bornes.

La position de l'utilisateur est parfois conservée ou communiquée à des tiers sans informer ni obtenir son consentement, avec, pour conséquence, un traçage possible des déplacements, un profilage des absences du domicile... Cela met en cause la liberté de circuler anonymement.

---

1. Voy., à ce sujet, http ://pro.clubic.com/entreprises/google/actualite-343282-google-acces-wi-fi-repertories-grande-bretagne.html ou http ://www.infos-du-net.com/actualite/17071-google-wi-fi-reseaux.html.

De façon encore plus pernicieuse, l'information de géolocalisation lors de la prise de photos (p. ex., avec un téléphone portable), combinée aux technologies de reconnaissance faciale, telles qu'implémentées, entre autres, dans les logiciels Apple iPhoto® ou Google Picasa®, permet de déterminer la localisation d'une personne figurant sur une photo, à son insu.

3. Traçabilité des utilisateurs

Contrairement à ce que l'on pense, la navigation sur Internet laisse bien davantage de traces que déambuler et agir dans la vie réelle. Les actions que l'on effectue sur Internet laissent entre les mains de différentes personnes des traces de ce que l'on a fait (adresse IP, fournisseur d'accès, page d'où l'on vient, historique de la navigation...). Les outils comme l'adressage IPv6 et les cookies (voy. *infra*) permettent d'individualiser un ordinateur et, dès lors, son utilisateur. À l'inverse de ce qui se passe dans le monde physique réel, il n'est pas question de se promener sur les inforoutes, d'entrer dans les magasins virtuels, de lire le journal, d'être intéressé par une annonce commerciale... sans que cela se sache. On ne peut manquer de s'interroger sur cette transparence permanente qui ne serait sans doute pas tolérée dans le monde réel.

4. Adressage IPv6

En raison de la prolifération des systèmes connectés à Internet, la plage d'adresses définie par la norme IPv4[1] est épuisée, ce qui menace l'expansion d'Internet. En réponse à ce problème, la norme IPv6 a été créée, qui supporte un nombre beaucoup plus important d'adresses distinctes[2]. À titre d'illustration, IPv6 permettrait à chaque individu sur terre de disposer de plusieurs dizaines de milliards de milliards de milliards d'adresses pour son usage personnel.

L'assignation d'une adresse IPv6 à un équipement peut être réalisée de différentes manières, dont l'une utilise l'adresse physique (adresse MAC) de l'appareil pour

1. IPv4 définit un format d'adresse sur 4 bytes, représentant chacun une valeur entre 0 et 255, soit $2^{32} \simeq 4.10^{9}$ adresses possibles.
2. IPv6 utilise un format d'adresse sur 16 bytes, représentant chacun une valeur entre 0 et 255, soit $2^{128} \simeq 256.10^{36}$ adresses distinctes.

générer l'adresse IPv6, ce qui permet alors de lier le trafic à une machine, voire de conduire à une personne. D'autres modes permettent d'éviter cette situation, en générant des adresses de manière pseudo-aléatoire ou en recourant à un serveur d'adresses qui les assigne de manière automatique[1].

Le caractère identifiant ou non de l'adresse IPv6 dépendra donc soit des paramètres de configuration par défaut du système utilisé, soit de la compétence de l'utilisateur.

5. Cookies

Le mécanisme des cookies est défini par le protocole de navigation Web (http) et permet à un serveur Web de transmettre au navigateur de l'internaute une série d'informations que celui-ci lui retournera lors des visites ultérieures (vers ce site uniquement). Le cookie a une durée de vie limitée, soit liée à la fermeture du navigateur, soit à une date d'expiration. Les cookies sont donc stockés localement par le navigateur, typiquement sur le disque dur de l'utilisateur.

Les cookies sont utilisés par les serveurs Web à des fins de gestion de session et de personnalisation, mais ils peuvent aussi servir comme un moyen de traçage. De plus, il faut noter que, lors de la visite d'un site, le navigateur peut recevoir des cookies provenant de sites tiers, ceci étant dû à l'inclusion dans le site consulté originellement de contenu provenant de ces sites tiers. Cette technique est fréquemment utilisée pour la mesure d'audience ou le profilage publicitaire.

Bien que les navigateurs les plus répandus permettent aux internautes de gérer, voire de bloquer les cookies, ces fonctions sont rarement utilisées, soit par méconnaissance, soit, plus simplement, parce que le blocage des cookies rendrait la navigation Internet impraticable.

6. Réseaux de distribution intelligents

L'on assiste à une évolution des réseaux de distribution d'énergie vers une forme intelligente (*Smart Grid*) dans laquelle sont incorporées des technologies informati-

1. Ce mécanisme utilise le protocole DHCP.

ques afin d'optimiser la production et la distribution, l'objectif étant d'ajuster au mieux la production et la consommation, conduisant ainsi à des économies d'énergie, l'évitement de pannes... Le *Smart Grid* fonctionne à partir de compteurs intelligents, munis de capteurs et reliés via un réseau à un système qui collecte, intègre et analyse les données de consommation.

Les compteurs intelligents communiquent les informations de consommation en temps réel à l'opérateur, ce qui peut constituer un moyen de profiler les consommateurs : absence ou présence dans le bâtiment, utilisation d'appareils possédant une « signature » énergétique...

De plus, à l'intérieur même du bâtiment, des appareils peuvent aussi être connectés au compteur intelligent, l'informant de la consommation instantanée, mais lui permettant aussi d'agir sur celle-ci, par exemple en adaptant automatiquement la température d'un thermostat ou en désactivant l'air conditionné lors d'un pic de consommation.

Dans ce cas encore, l'on assiste à une collecte massive d'informations pouvant être liées à une personne ou un groupe de personnes, permettant d'en déduire des caractéristiques et des comportements de manière très ciblée. Lorsqu'en plus, cette information est collectée par des tiers, comme c'est le cas pour le système PowerMeter[1] de Google, le risque de diffusion non contrôlée de l'information est encore plus grand.

7. RFID et l'Internet des objets

La technologie RFID (*Radio-Frequency IDentification*) est une technique d'identification qui se base sur trois composants :

- l'étiquette, ou tag, qui est collée ou intégrée à l'entité à identifier ;
- le lecteur, utilisé pour interroger le tag lorsque celui-ci est à sa portée ;
- le système d'information, qui reçoit l'information du lecteur et la traite.

Le tag est composé d'une antenne et d'une puce électronique, qui contient, au minimum, un identifiant. Lorsque le tag est interrogé par un lecteur (par l'utilisation d'ondes magnétiques), il transmet son identifiant au lecteur. La structure du tag est

1. Google PowerMeter est un système permettant à un utilisateur de visualiser sur le Web sa consommation énergétique, ce système étant alimenté à partir du compteur intelligent installé chez l'utilisateur. L'accès à cette information est normalement réservé à l'utilisateur en question.

très simple, de manière à permettre une production de masse à un coût autorisant son utilisation massive, typiquement quelques centimes[1]. La lecture du tag ne nécessite pas de contact entre celui-ci et le lecteur ; en fonction du type de tag, la distance de lecture peut varier entre quelques centimètres ou quelques dizaines de centimètres, voire au-delà.

Les tags RFID sont utilisés dans la gestion des stocks et de l'approvisionnement, pour les péages routiers, dans la grande distribution pour la gestion de l'inventaire, des caisses ou du service après-vente, dans les aéroports pour le suivi des bagages ou comme moyen de marquage des animaux. Dans certains cas, les tags peuvent être implantés chez des êtres humains, par exemple pour assurer la sécurité d'enfants ou de personnes âgées, ou, dans un registre plus léger, pour surveiller l'accès ou gérer les consommations dans une discothèque.

L'identifiant étant spécifique à un tag, la lecture de celui-ci permet donc de suivre ses déplacements, d'après la position du lecteur, et donc ceux de l'objet ou de la personne qui le porte. La lecture se faisant à distance, l'utilisateur n'est pas nécessairement conscient de celle-ci, ce qui peut conduire à des fuites d'information ou un traçage à son insu. L'interrogation simultanée d'un grand nombre de tags permet d'identifier très rapidement les objets ou personnes marqués dans un environnement proche et, donc, là aussi, d'aboutir à un profilage du porteur.

Différentes solutions techniques existent (et d'autres continuent d'être développées) qui permettent de limiter les possibilités d'utilisation malveillante des technologies RFID. Mais, bien souvent, leur mise en œuvre fait augmenter significativement le coût de fabrication, rendant difficile leur utilisation à large échelle.

L'Internet des objets (*Internet of Things*) pousse l'idée de l'Internet et de l'identification un (grand) pas plus loin, en décrivant un monde où tout est interconnecté : les personnes, mais aussi les objets. Internet sort donc du monde strictement virtuel pour intégrer les objets du monde réel, physique, en utilisant des technologies telles que la RFID, les communications sans fil à courte portée (NFC – *Near Field Communication* ou communication en champ proche), la géolocalisation et les réseaux de capteurs. Dans ce scénario, les objets connectés agissent avec un haut degré d'autonomie, capables d'acquérir et de transmettre des informations collectées au travers de capteurs, de les traiter, et d'interagir avec les utilisateurs et leur environnement.

Bien que l'Internet des objets soit encore une discipline récente, dont les utilisations scientifiques et commerciales en sont encore à leurs balbutiements, il est cependant évident qu'il se base sur des collectes et des traitements massifs d'informations,

1. Notons qu'un tag peut être plus élaboré : il peut contenir plus d'informations que l'identifiant et posséder sa propre batterie, pour atteindre des distances de transmission plus élevées ou agir comme capteur, par exemple.

pour la plupart pouvant être liées directement ou indirectement à des individus et, par là même, menacer leur vie privée.

8. Robots d'indexation

Un robot d'indexation (*Webcrawler* ou *Webspider*) est un logiciel écrit pour explorer le Web de manière automatique, afin d'indexer le contenu visité et d'alimenter ainsi les moteurs de recherche pour permettre une recherche plus efficace et, donc, un accès plus aisé à l'information. Il fonctionne par analyse des pages visitées, en suivant récursivement les hyperliens.

Certains robots malveillants analysent les pages pour en extraire les adresses électroniques afin de constituer des listes de diffusion pour l'envoi de spams. D'autres peuvent aussi parcourir des pages, afin d'agréger et de corréler les informations collectées et d'en inférer d'autres.

9. Données biométriques

Des moyens biométriques, c'est-à-dire liés à des caractéristiques physiologiques de l'individu telles que ses empreintes digitales, son empreinte rétinienne, son empreinte vocale ou son ADN, sont de plus en plus utilisés pour authentifier une personne (vérifier son identité), que ce soit dans le domaine des paiements électroniques, du contrôle aux frontières, du contrôle d'accès, de la reconnaissance faciale...

Les données biométriques doivent d'abord être collectées, avant de pouvoir être confrontées à celles fournies lors de l'authentification et ainsi valider celle-ci. Cela implique le stockage d'une grande quantité de données à caractère personnel, dont certaines, telles que l'ADN, percent l'intimité de l'individu, y compris celle de son ascendance et de sa descendance.

10. Privacy by Design

Le terme « Privacy by Design » fait référence à un ensemble de principes élaborés pour être utilisés lors de la conception, du développement et de l'exploitation de systèmes d'information, afin de garantir que les dimensions « vie privée » et « protection des données » ont été correctement prises en compte dès la conception et que, dès lors, ces systèmes sont en conformité avec les exigences légales et réglementaires en la matière.

C'est Ann Cavoukian, commissaire à l'Information et à la Vie privée de la province d'Ontario, au Canada, qui est à l'origine de cette initiative, fondée autour des valeurs de respect de l'utilisateur, de transparence à son égard pour ce qui concerne la collecte et le traitement des données, et de refus de compromis dans lesquels la vie privée serait sacrifiée au profit d'autres objectifs. Les principes de base sont le caractère proactif des mesures de sécurité, le fait que, par défaut, la protection des données est assurée, toute dérogation devant faire l'approbation de la personne concernée, le fait que la protection des données doit être considérée comme partie intégrante des fonctions du système d'information, plutôt qu'une fonctionnalité annexe, et qu'elle doit être maintenue tout au long du cycle de vie de l'information collectée.

Ces principes sont applicables aussi bien au domaine IT qu'à celui des pratiques-métiers et de l'infrastructure physique.

Un portail Internet est consacré à cette approche[1], qui, outre une présentation générale, démontre l'applicabilité de la démarche au travers de nombreux cas d'études, montrant ainsi qu'il est possible de concevoir des systèmes efficaces et répondant aux exigences-métiers sans pour autant sacrifier à la protection des données.

11. Cloud Computing

Le « Cloud Computing » est un paradigme récent d'architecture IT, qui rend complètement transparent pour l'utilisateur l'endroit où les données qu'il manipule et les services qu'il utilise sont effectivement stockés ou mis en œuvre. Le terme fait référence à la fois aux services accédés et délivrés via Internet, et aux systèmes d'information et à l'infrastructure matérielle et logicielle qui fournit ces services.

1. http ://www.privacybydesign.ca.

Le « Cloud » permet une grande flexibilité dans la gestion et l'allocation des ressources, où le modèle d'investissement s'oriente plus vers un modèle de facturation à l'usage, ainsi qu'une grande souplesse dans l'intégration de services, de manière intra- ou interorganisationnelle, indépendamment de l'implantation géographique.

Les services de type « Cloud » peuvent être offerts à divers niveaux ; on distingue généralement trois modèles différents :

- *Infrastructure as a Service* (IaaS) : les services offerts sont de type « infrastructure », soit principalement du matériel et du logiciel de base, ainsi que de la connectivité ; la gestion de cette infrastructure est laissée au client ;
- *Platform as a Service* (PaaS) : les services offerts prennent la forme d'une plateforme opérationnelle, composée de l'infrastructure, mais aussi de l'environnement logiciel permettant au client de développer ou d'exploiter ses propres applications ; la gestion de l'ensemble est donc partagée entre le fournisseur de services et son client ;
- *Software as a Service* (Saas) : le fournisseur offre ici une solution applicative complète à son client, en prenant en charge à la fois l'infrastructure, mais aussi l'application. De tels services sont offerts, par exemple, par salesforce.com, pour la gestion commerciale, ou par Google, au travers de ses services mail, documents, agenda...

Le « Cloud Computing » constitue donc une extension du périmètre de sécurité vers Internet où il est fort compliqué d'effectuer un contrôle efficace. Le stockage de ses données est confié par l'utilisateur à un tiers, le fournisseur de services, qui les héberge et les traite dans des conditions bien souvent inconnues de l'utilisateur. Ceci nécessite une réelle relation de confiance entre l'utilisateur et le fournisseur de services. Cette confiance peut être renforcée par des garanties contractuelles.

Les principaux défis se situent autour de la protection des données confiées au *Cloud*, de la préservation de leur intégrité et du maintien d'un contrôle d'accès approprié.

12. Deep Packet Inspection

L'information circulant sur un réseau est classiquement transmise sous forme de paquets, formés d'un en-tête et d'un corps ; l'en-tête contient l'information nécessaire pour permettre aux équipements réseaux traversés de mener le paquet jusqu'à sa destination.

Le filtrage du trafic réseau, opéré typiquement par des pare-feu (*firewalls*) se base, pour autoriser ou non le transit, sur les informations de routage, présentes dans l'en-tête des paquets, soit principalement l'origine et la destination du message. Le « Deep Packet Inspection » se base en plus sur des critères de contenu, en analysant non seulement l'en-tête, mais aussi le corps du message, soit son contenu.

La technique est évidemment plus coûteuse en temps et en ressources. Elle permet d'améliorer la sécurité des systèmes d'information, en détectant et filtrant le contenu malicieux. Mais elle peut aussi être détournée à des fins de surveillance ou de censure.

CHAPITRE 1.3. DÉFIS POUR LA VIE PRIVÉE ET LA PROTECTION DES DONNÉES LIÉS AUX USAGES

1. Collecte et traitement d'informations recourant aux technologies de l'information et de la communication (TIC)

1.1. Par les autorités étatiques

Le développement de l'administration électronique ou de l'e.gouvernement (*e.government*) à partir de l'utilisation des TIC par les administrations publiques conduit à une organisation en réseau des autorités étatiques. Cette évolution se base essentiellement sur le partage de données entre autorités, la création de fichiers de référence et de vastes entrepôts de données, de même que sur l'interconnexion de bases de données autrefois indépendantes. Ce modèle suscite d'importantes interrogations relatives à la protection de la vie privée. Le modèle antérieur de l'administration « en silos », chaque entité disposant d'informations propres, isolées, destinées à réaliser la mission légale de l'entité, était présenté comme la garantie contre un État omniscient à l'égard duquel le citoyen serait totalement transparent. L'« obscurité pratique » était la clé de l'équilibre dans la relation administration-administrés. Cette garantie a disparu au nom de l'efficacité. On doit aujourd'hui impérativement poser la question de la maîtrise par chacun des informations collectées à son propos, de la transparence des échanges et de la proportionnalité des traitements.

Le recours aux identifiants uniques servant d'instruments d'interconnexion et d'accès transversal aux données d'un individu augmente encore les risques de perte de contrôle et de non-respect de la proportionnalité.

Les inquiétudes face aux traitements de données personnelles par les autorités publiques sont accentuées par le fait que ces traitements servent de base à la prise de décisions tels l'octroi d'une pension, la reconnaissance d'un statut particulier, l'établissement de l'impôt, l'ouverture d'enquêtes pénales...

1.2. Par les entités commerciales

Les données personnelles représentent une valeur économique. Cette valeur est importante à trois niveaux :

- pour les acteurs offrant des services via Internet, car connaître le profil des internautes intéressés par les produits ou services et pouvoir détailler très précisément leur intérêt (pages Web lues, liens cliqués, fréquence des visites...) permettent de configurer l'offre de manière optimale ;
- pour les acteurs exploitant commercialement des bases de données nominatives : récolter des données tous azimuts permet de constituer de très riches bases de données exploitables et revendables pour des activités de marketing et de mailing ;
- pour le fonctionnement même du Web : la gratuité de la plupart des services offerts sur le Web n'est que de façade. L'exposition publicitaire des utilisateurs finance l'offre. Le modèle économique repose sur le marketing. Celui-ci sera d'autant plus rentable que le profil des destinataires est précis et permet de cibler efficacement les messages publicitaires[1].

Dans ces trois schémas, la collecte et le croisement d'informations conduisant à dessiner les profils des utilisateurs deviennent des opérations cruciales. Ces opérations se font toutefois, dans de trop nombreux cas, à l'insu des personnes concernées. Elles impliquent souvent une utilisation des données au-delà des finalités originelles. Et la quantité des données collectées pose inévitablement la question de la proportionnalité. Est-il nécessaire ou tout simplement normal, par exemple, que les moteurs de recherche (comme Google) conservent durant des mois tous les mots introduits par une personne (individualisée grâce à un cookie) ? Cet ensemble de mots est le plus souvent incroyablement révélateur de ses centres d'intérêt, ses activités, ses projets...

1.3. Par les employeurs

Les TIC ont mis entre les mains des employeurs des outils de surveillance inimaginables autrefois. Les cartes magnétiques d'accès aux locaux disent à l'opérateur du réseau qui se trouve où et à quelle heure, alors que les clés classiques étaient muettes à ce sujet. Les réseaux de caméras permettent de surveiller les visiteurs aussi bien que le personnel. La surveillance du personnel s'effectue également par le contrôle de la navigation sur Internet et l'usage du courrier électronique mis à la disposition des travailleurs. Pour ceux qui prestent hors des murs de l'entreprise, les systèmes de localisation et de suivi géographique des travailleurs permettent de

1. Pour Google et Facebook, le profit tiré des activités de marketing opérées sur leurs sites s'élève annuellement à plusieurs milliards de dollars.

gérer à distance une flotte de taxis, de dépanneuses ou de camions et de surveiller leurs pérégrinations en temps réel.

Les TIC représentent aussi des instruments de connaissance. Bon nombre d'employeurs se renseignent à la source du Web sur les candidats employés. Google et Facebook, notamment, jouent ainsi le rôle d'indicateurs et révèlent au futur patron des facettes des candidats qui ne se trouvent pas sur leurs CV...

1.4. Par les individus eux-mêmes

Dans bien des cas, les individus ne prennent pas la pleine mesure de la portée de leurs actions sur le réseau. Le Web 2.0 leur a donné la possibilité d'interagir, d'apporter des commentaires, de diffuser eux-mêmes du contenu, de partager en continu savoirs, photos, vidéos, informations, états d'âme... Toutefois, le rayonnement de l'information sur Internet dépasse parfois largement ce à quoi on s'attend. L'exemple des informations tirées des pages publiques de Facebook et jointes automatiquement, à l'insu de la personne concernée, par un logiciel de courrier électronique aux courriels envoyés a déjà été cité *supra*. La puissance des robots « ratisseurs » qui alimentent les moteurs de recherche permet de faire remonter des informations trouvées à des endroits épars, publiées dans des contextes qu'on croyait particuliers à des personnes qu'on croyait restreintes. Ce qui est émis dans un certain cercle (p. ex., un commentaire déposé sur un forum de discussion) risque donc de réapparaître, sorti de son contexte et juxtaposé à d'autres informations.

Une fois l'information (texte, image, vidéo) diffusée, on ne peut plus contrôler son parcours. L'effacer du site initial n'empêchera pas qu'elle perdure dans les lieux où elle a été copiée ou téléchargée avant son effacement. Et il est illusoire de vouloir contrôler que l'usage qui est fait de l'information (notamment aux antipodes et par des inconnus) respecte la finalité de sa diffusion première.

Cette perte de contrôle est d'autant plus inquiétante qu'elle s'accompagne de l'*eternity effect*. À l'inverse de la mémoire humaine, la mémoire électronique n'efface rien si ce n'est volontaire. Des éléments peuvent remonter éternellement du passé tant qu'on n'a pas pris la décision, le temps et l'énergie de les supprimer (là où on a la maîtrise de la suppression).

Des actes individuels malveillants peuvent aussi susciter des inquiétudes. Diffuser une information diffamatoire ou confidentielle sur Facebook, poster une vidéo intime ou humiliante sur YouTube, ou créer un faux article sur quelqu'un dans Wikipédia peut causer des dommages d'une ampleur sans précédent dans la vie *off line*.

2. Profilage des internautes

Le profilage consiste à appliquer des algorithmes à des quantités d'informations agrégées, pour mettre au jour des corrélations entre les données et faire surgir des profils. Ces derniers sont appliqués à un individu, pour décider du traitement à lui réserver (le considérer ou non comme fraudeur fiscal ou comme cible de marketing de tel produit ou comme voyageur candidat terroriste…). Motivé par un intérêt économique (*cf.*, *supra*, point 1.2), sécuritaire ou autre, le profilage est facilement réalisable à partir des informations disponibles à grande échelle (traces, mots introduits dans les moteurs de recherche, etc.) et du recours aux cookies, notamment.

Le profilage répond à des besoins ou intérêts légitimes de la société : analyse du risque, identification des fraudes, segmentation des marchés, ajustement de l'offre à la demande, etc. Toutefois, il peut amener à priver des individus de manière injustifiée de l'accès à certains services. L'existence de profils conduit à ce que l'information offerte est filtrée, triée, sélectionnée en fonction du destinataire. Cela vaut aujourd'hui massivement pour les informations commerciales. Sera-ce demain le cas pour toute information ? Le profilage risque aussi d'être un instrument de discrimination. Comment contester l'élaboration d'un profil ou son application inappropriée ? La plupart du temps, l'existence des profils échappe à la connaissance des individus concernés et la compréhension de leurs critères d'élaboration échappe à ceux qui les appliquent. Enfin, l'activité de profilage suscite de graves préoccupations concernant la proportionnalité. Les quantités de données collectées et la durée de leur conservation sont, dans bien des cas, totalement excessives.

3. Rétention des données

Les données liées à l'utilisation d'Internet et des nouveaux moyens de communication représentent une mine de renseignements précieux pour les activités de recherche policière et de lutte contre la criminalité.

Depuis les attentats du 11 septembre 2001, des textes ont été votés au niveau européen pour harmoniser les situations dans lesquelles des données relatives au contenu ou des données de trafic ou de localisation sont conservées pour être tenues à la disposition des autorités pénales. Ces données portent sur la durée, la date, les destinataires, le lieu de toutes les communications, le volume des SMS/textos et des courriels…

Il est intéressant de voir la progression de ces textes. La Convention du Conseil de l'Europe sur la cybercriminalité de novembre 2001 prévoit que les États peuvent imposer la conservation rapide de telles données, à la demande d'une autorité, pour des données spécifiées et pour maximum nonante jours. La directive 2006/24 du 15 mars 2006 sur la conservation de données, quant à elle[1], impose aux fournisseurs de services de communication (Internet, téléphones, mobiles, fax) la rétention des données de trafic et de localisation de tout le monde, de façon systématique et pour une durée variant entre six mois et deux ans...

1. Voy. la présentation de cette directive dans le chapitre 2.