COUNCIL OF EUROPE CONSEIL DE L'EUROPE

Strasbourg, 11 janvier 2008 T-PD-BUR (2008) 01

COMITE CONSULTATIF DE LA CONVENTION POUR LA PROTECTION DES PERSONNES A L'EGARD DU TRAITEMENT AUTOMATISE DES DONNEES A CARACTERE PERSONNEL

(T-PD)

24e réunion
13-14 mars 2008
Strasbourg, salle G01

L'application de la Convention 108 au mécanisme de profilage.

Eléments de réflexion destinés au travail futur du Comité consultatif (T-PD)

Par Jean-Marc Dinant, Christophe Lazaro, Yves Poullet,
Nathalie Lefever , Antoinette Rouvroy

Version finale

Les experts signataires de ce rapport
expriment ici leur opinion personnelle
qui n'engage pas le Conseil de l'Europe

CENTRE DE RECHERCHES INFORMATIQUE ET DROIT

Document du Secrétariat préparé par
la Direction Générale des affaires juridiques et des droits de l'Homme

Table des matières

1. Tentative de définition du profilage

Le présent rapport fait référence au *profilage abstrait*, reposant sur la mise en évidence d'informations, l'établissement de prédictions et finalement l'inférence.

Or, un profil est parfois généré uniquement sur la base de la collecte et l'analyse de données relatives à la personne concernée, sans inférence ou prédiction issues de sources externes. Il est souvent fait référence à ce second type de profilage par les locutions « *personal profiling* [1]» ou « specific profiling[2]».

En pratique, les profils abstraits sont souvent générés partiellement sur la base de profils spécifiques/personnels et vice versa. Le profilage spécifique tombe pour nous dans le cadre de l'application de la Convention 108 et le profilage spécifique ou individuel nous apparaît comme une donnée à caractère personnel sur laquelle l'individu possède les droits prévus dans ladite Convention.

Notre travail se focalise sur le profilage abstrait parce que certaines opérations cruciales du processus de profilage utilisent comme matière première des données relatives à des individus non identifiables et non identifiés

Dans une opération de profilage (s/e abstrait), nous distinguerons donc trois étapes.

1. La première étape, l'étape d' « *observation* », est une étape de regroupement de données à caractère personnel ou de données anonymes. Si les données se rapportent à une personne identifiable ou identifiée, ces données seront en général anonymisées durant cette étape. Dans ce qui suit nous supposerons que le résultat de cette première étape est un ensemble de données anonymes décrivant certains aspects de la personnalité d'un individu non identifiable. Nous nommerons cette étape « entreposage de données » (**datawarehousing**). L'origine des données peut aussi être interne ou externe. Par exemple, une banque peut établir une liste anonyme de ses clients mauvais payeurs avec leurs caractéristiques. Une entreprise de marketing peut acquérir la liste des « paniers d'achats » des grandes chaînes de supermarché, sans indentification des personnes ayant effectué ces achats.

2. A cette première étape succède un deuxième ensemble d'opérations qui s'effectuent par des méthodes statistiques et qui ont pour objet d'établir, avec une certaine marge d'erreur, des *corrélations* entre certaines variables observables. Ainsi une banque pourra-t-elle établir un lien statistique entre un long séjour à l'étranger et un ou plusieurs défaut(s) de paiement d'un crédit. Nous appellerons cette étape le « **datamining** ». Le résultat concret de cette étape consiste en un **mécanisme** de catégorisation des individus sur base de certaines de leurs caractéristiques observables afin d'en déduire, avec une certaine marge d'erreur, d'autres qui ne le sont pas.

---

1 Clarke Roger, Customer profiling and privacy implications for the finance industry, mai 1997; Future of Identity in the Information Society (FIDIS) Deliverable 7.2: Descriptive analysis and inventory of profiling practices, n° 2.3; 2.6 and 3.3

2 Bygrave Lee Andrew, Data protection law : Approaching its rationale, logic and limits, 2002, p. 303

3. La troisième et dernière étape, que l'on appelle d' « *inférence* », consiste à appliquer le mécanisme décrit ci-dessus afin de pouvoir, à partir des données relatives à une personne identifiée ou identifiable, en déduire des données nouvelles qui sont en fait celles de la catégorie à laquelle il appartient. Bien souvent, seule cette dernière opération est désignée par le terme « profilage ». Toutefois, il nous apparaît essentiel de recentrer cette étape ultime au sein d'un processus. Même si c'est à ce moment que le profilage fait sentir ses effets par rapport à une personne déterminée, le mécanisme de profilage démarre dès l'engrangement des données dans des entrepôts de données, et même, en amont, dès l'observation de l'individu par le recours à des technologies de l'information et de la communication.

A des fins pédagogiques, nous utiliserons tout au long de ce rapport des exemples de profilage :

1. L'ATS (Automatic Target System) qui a été développé aux Etats-Unis pour évaluer la probabilité qu'un individu donné puisse être un terroriste ;
2. La télévision numérique par câble. Il s'agit d'une révolution devant permettre au distributeur de programme de connaître avec précision la chaîne sélectionnée et le zapping effectué par le téléspectateur, alors que ce dernier reçoit une chaîne de télévision via le câble téléphonique selon la technologie DSL ;
3. Le profilage des contribuables effectué par l'Etat afin de détecter les fraudeurs. Ce système en cours de déploiement devrait permettre à l'Etat d'identifier les fraudeurs potentiels afin de mieux cibler les contrôles et d'adapter les dispositifs légaux en vue de lutter contre cette fraude. Plus positivement, le système peut être utilisé afin de signaler des aides fiscales aux personnes concernées ;
4. Le système de publicité en ligne mis en place par Google. Sur la plupart des sites en ligne à grosse fréquentation, apparaît un encart nommé « Ad by Google ». Ce que beaucoup de consommateurs ignorent c'est que les liens commerciaux apparaissant dans cette fenêtre sont générés au cas par cas et en temps réel par Google sur base de la page référante communiquée par leur navigateur. Google peut donc suivre, pas à pas, la navigation de chaque internaute sur chacune des pages des sites à grande fréquentation (eBay, journaux en ligne, moteurs de recherches, sites boursiers, sites de vente immobilière, etc…)
5. Dans un magasin, des tubes de rouges à lèvre sont librement proposés à l'essai devant un miroir. Chaque personne peut donc se maquiller les lèvres devant ce miroir pour en observer les effets. Chaque tube de rouge à lèvre est doté d'une puce RFID qui permet à un système équipé d'une caméra de filmer les lèvres des clients et de connaître les caractéristiques de chaque rouge à lèvre employé. Il y a conservation et analyse des images, mais celles-ci concernent uniquement les lèvres et la personne filmée n'est absolument pas reconnaissable.

1.1.1. Historique et actualité du profilage

Historiquement, le terme profilage a trouvé ses premières lettres de noblesse dans la formation de « crime profiler » aux Etats-Unis. Ces personnes sont réputées être théoriquement capables d'établir le « profil type » d'un criminel en analysant les traces laissées sur la scène du crime. Un des plus fameux profiler fut le psychiatre James A.

Brussel qui parvient à établir de manière précise le profil de « mad bomber » -dans les années cinquante. Grâce à ce profil, le FBI a pu retrouver l'auteur des faits.

Selon Wikipedia, dans notre société moderne, le terme profilage est aussi utilisé comme synonyme pour l'analyse comportementale.

« *L'analyse comportementale d'une personne* *consiste à observer le comportement d'une personne en fonction des stimuli (comportementalisme), plutôt que d'investiguer ses pensées, dans le but :*

- *normalement de l'aider par des conseils mais aussi en agissant sur les stimulus (ex : récompense / sanction) à éviter des comportements nocifs.*
- *parfois aussi de le conditionner à agir dans un sens ne correspondant pas à ses intérêts,*
- *ou encore de le mettre hors d'état de nuire (profilage criminel).*

Cette méthode est parfois critiquée comme correspondant plus à du « dressage » qu'à la résolution de problèmes psychologiques profonds. »

Dans notre propos, nous désignerons comme profilage une méthode informatisée ayant recours à des procédés de data mining sur des entrepôts de données permettant ou devant permettre de classer avec une certaine probabilité et donc avec un certain taux d'erreur induit un individu dans une catégorie particulière afin de prendre des décisions individuelles à son égard.

Cette notion du profilage se distingue du profilage criminel où il s'agit de comprendre, de se mettre dans la peau du criminel, mais se rapproche de l'analyse comportementale dans la mesure où il ne s'agit pas de comprendre les motivations qui poussent ou pourraient pousser un individu à adopter un comportement donné mais d'établir une corrélation mathématique forte entre certaines caractéristiques que l'individu partage avec d'autres individus « semblables », et un comportement donné, que l'on veut prédire ou influencer. Cette approche ne nécessitant pas d'intelligence humaine, mais l'analyse statistique de quantités énormes de données chiffrées relatives à des observations numérisées, elle est praticable à l'aide d'un ordinateur et avec un minimum d'intervention humaine.

1.1.2. Brèves considérations étymologiques sur la notion de profil

Le terme « profil » tire notamment son origine de la sphère artistique. Il désigne les « contours, traits d'un visage vu par un de ses côtés » (1621) ou, plus largement, la « représentation d'un objet vu d'un de ses côtés seulement »3.

Par extension, ce terme a fini par désigner au figuré l' « ensemble des traits caractéristiques d'une chose, d'une situation, d'une catégorie de personnes »[4]. S'agissant des personnes, ce terme renvoie dès lors à l' « ensemble des traits caractéristiques que présente une personne ou une catégorie de personnes » (1925).

---

[3] Voy. *Le Trésor de la langue française informatisé* (TLFI), v° « profil », http://atilf.atilf.fr/dendien/scripts/tlfiv5/advanced.exe?8;s=3843709350.

[4] *Centre National de Ressources Textuelles et Lexicales* (CNRTL), Lexicographie, v° « profil », http://www.cnrtl.fr/lexicographie/profil.

On peut à cet égard faire référence à la notion de *profil psychologique* ou encore plus spécialement à l' « ensemble des caractéristiques que doit présenter une personne pour occuper un poste, être recrutée quelque part » (1967)[5].

Dans le cadre de notre rapport, l'étymologie initiale de cette notion renvoyant au domaine des beaux-arts doit retenir l'attention. En effet, le profil ne constitue qu'une représentation de la personne, construite à partir de différents traits. Esquisse de traits pour le profil du peintre, corrélation de données dans le cadre des activités de profilage. Dans ces deux cas, le profil en tant que représentation ou construction d'un savoir sur la personne ne peut être assimilé à la personne elle-même. Or, comme nous le verrons, un des dangers les plus vifs du profilage réside dans le fait qu'il tend à réduire la personne au seul profil généré par des procédés automatisés, lesquels sont susceptibles de servir de base à la prise de décision.

2. Aperçu de la problématique (→ quelques réflexions en vrac)

Si l'on met de côté la phase préliminaire consistant à collecter et conserver des données (pré-profilage), on peut proposer une définition de la notion de profilage qui s'articule autour d'une double dimension : la constitution d'un profil et son application à une situation déterminée. Pour L. A. Bygrave, le profilage peut être défini de la manière suivante : « *Generally speaking, profiling is the process of inferring a set of characteristics (typically behavioural) about an individual person or collective entity and then treating that person/entity (or other persons/entities) in the light of these characteristics. As such, the profiling process has two main components : (i) profile generation – the process of inferring a profile ; (ii) profile application – the process of treating persons/entities in light of this profile* »[6]. L. A. Bygrave ne distingue pas le data mining et l'entreposage de données. Nous jugeons utile de distinguer ces deux étapes distinctes.

Comme suggéré plus haut, trois étapes sont donc nécessaires pour effectuer ce que nous appelerons un profilage :

- des **observations** numérisées et massives des comportements et caractéristiques des individus
- l'établissement de liens de probabilités (**corrélations**) entre certains comportements/caractéristiques et d'autres comportements ou caractéristiques
- **l'inférence**, à partir de certaines variables comportementales ou caractéristiques observables propres à un individu généralement identifié de nouvelles caractéristiques ou variables comportementale présentes, passées ou futures.

Nous détaillons ci-après, chacune de ces trois phases :

---

5 E. CLAPARÈDE, *Arch. de psychol.*, t. 19, p. 267 ds QUEM. *DDL* t. 29), cité in *Centre National de Ressources Textuelles et Lexicales* (CNRTL), Etymologie, v° « profil », http://www.cnrtl.fr/etymologie/profil.

6 L. A. BYGRAVE, « Minding the machine : Article 15 of the EC Data Protection Directive and Automated Profiling », *Computer Law & Security Report*, 2001, vol. 17, pp. 17-24, disponible en ligne à l'adresse http://www.austlii.edu.au/au/journals/PLPR/2000/40.html.

2. L'entreposage des données d'observation

Les entreprises collectent, stockent pour une durée souvent longue des informations de plus en plus précises relatives au comportement de leurs clients ou de leur personnel. Sur le terrain, on se rend compte que les données sont rarement détruites, nonobstant les déclarations de durée de conservation faites auprès de l'autorité de protection des données. A la convention 108 qui conditionne la licéité de la conservation (qui est un traitement) à la démonstration d'une finalité explicite et déterminée, s'oppose la tentation opérationnelle de ne pas détruire les données « qui peuvent toujours servir ». Toutes ces données, typiquement des transactions financières, des données géographiques, des données d'achat ou de vente, mais aussi, dans le chef du gouvernement, des données relatives aux remboursements et soins médicaux, à l'emploi, au mariage, à la propriété privée, à la fortune, aux biens mobiliers et immobiliers, au revenus, à l'épargne, … peuvent être regroupées dans d'immenses entrepôts de données sous trois formes techniquement distinctes

1. Des données nominatives ou identifiées par un numéro personnel (p.e un numéro de client, le numéro de sécurité social, le numéro national, etc) ou par un pseudonyme identifiant propre au maître du fichier.
2. Des données codées : un tiers en possession d'une « clé » de décodage est capable d'accéder à l'identité de la personne.
3. Des données anonymes : elles ne comprennent aucun d'identifiant. En examinant deux ensembles de données relatives à la même personne, il n'est pas raisonnablement possible, en utilisant les ressources disponibles conformément à l'état de l'art, d'avoir une certitude raisonnable que les deux ensembles de données concernent la même personne.

Afin d'illustrer notre propos, nous pouvons prendre l'exemple d'un ticket de caisse établi dans un grand magasin. Si le ticket comprend un numéro de client ou celui d'une carte de fidélité nominative, on se trouve en présence de données à caractère personnel.

Si le ticket ne comporte pas de numéro de client mais le numéro d'une carte bancaire ayant servi à l'achat, on se trouve en présence d'une donnée codée. Un tiers (ici en l'occurrence la banque) est capable d'identifier la personne ayant effectué un achat.

Si le ticket ne comprend aucune autre information que la liste des produits vendus, la date et une fourchette horaire assez large, nous nous trouvons dans le troisième cas.

Par ailleurs, il convient de préciser que les données choisies peuvent provenir d'une ou de plusieurs sources. Ainsi, dans le cas du profilage, dans le contexte de la gestion des risques d'un Ministère comme celui des Finances, il s'agit de données provenant de différentes sources au sein de l'Administration et de sources externes (autres ministères). Par sources distinctes, on entend bases de données collectées pour des finalités distinctes.

Dans le cas du profilage mené par Google, ce dernier a collecté les données venant de tous les visiteurs des grands sites en ligne sur lesquelles il affiche « Ad by Google ». Si sa finalité est d'assurer un service de publicité sur Internet, Google peut néanmoins suivre le clickstream de chaque individu en particulier, au moins dans les sites à haute fréquentation où il est présent via des hyperliens invisibles. Il lui est

ainsi possible de mesurer et ou de doser l'exposition publicitaire de chaque individu en particulier, de manière générale. Dans ce but, Google traite l'adresse IP qui permet au fournisseur d'accès à Internet, en vertu des règles obligatoires de conservation des données de trafic, d'identifier l'abonné détenteur de l'adresse IP à un moment donné. Il s'agit donc d'une donnée codée. Par ailleurs, si Google, à l'instar de DoubleClick, injecte un cookie identifiant et rémanent sur le disque dur du terminal de l'utilisateur, il y a lieu de considérer qu'il s'agit d'une identification par un numéro personnel, ce numéro unique au monde étant associé au clickstream particulier d'un individu, potentiellement révélateur de son identité sociale, économique ou culturelle.

Dans le cas du rouge à lèvre, il est possible d'imaginer que les données soient purement anonymes dans la mesure où le système ne filme que les lèvres d'une personne et qu'il est raisonnablement impossible, en règle générale, à partir d l'image des lèvres d'une personne de l'identifier avec certitude. Toutefois, il serait possible de manière automatique, éventuellement avec une intervention humaine d'apprendre au système à distinguer un homme d'une femme, un homme à la peau blanche d'un homme à la peau noire. Dans la mesure où le système de surveillance visé est couplé avec un lecteur de puce RFID et où cette puce RFID contient, ainsi que la norme l'exige, un identifiant global unique (numéro de série), il serait possible, lors du passage aux caisses du bâton de rouge à lèvre flanqué de sa puce RFID, de lier un client particulier à une séquence vidéo particulière et d'ainsi pouvoir en déduire son appartenance ethnique, son genre et, voir éventuellement, ses préférences sexuelles si l'essayeur du rouge à lèvre est identifié comme un homme.

Dans le cas de la télévision numérique, la communication du programme est effectuée techniquement par le fournisseur d'accès à Internet. Le décodeur sert à capter le flux des paquets IP venant du central téléphonique via une connexion DSL, à ré assembler ces paquets et à les encoder pour fournir un signal vidéo classique (p.e. PAL, SECAM, VGA, DVI, etc.). Lors d'un changement de chaîne de télévision, le décodeur transmettra, à la demande de l'utilisateur, des paquets IP au central via la connexion DSL (qui est bidirectionnelle) afin de demander l'envoi des paquets IP d'une autre chaîne. Dans le contexte des méthodes traditionnelles de télédiffusion par voie hertzienne (antenne terrestre ou satellite&parabole) qui sont unidirectionnelles, l'émetteur ne peut pas connaître le programme qui est en cours d'écoute. Il ne peut même pas savoir si la télévision est ou non allumée. Dans le cas de la télévision numérique, le fournisseur du service peut savoir si le décodeur est allumé et quel est le programme sélectionné. Il peut suivre et mémoriser le moindre zapping. Il peut ainsi conserver un profil d'écoute parfaitement précis de chaque abonné. Il devient alors techniquement possible de modifier les publicités en fonction du profil de l'utilisateur. Nonobstant le fait que le fournisseur d'accès connaît l'identité (au sens des données comme les nom, prénom et adresse de des abonnés) de son abonné, l'adresse IP du poste de télévision constitue une donnée à caractère personnel. L'adresse IP du poste permet en effet de le distinguer de tous les autres postes de télévisions (voir infra les travaux de Pfitzmann). De manière plus générale, il nous semble qu'il ne saurait y avoir d'anonymat (et donc de donnée non personnelle) lorsque un appareil de télécommunication possède une adresse unique, dans la mesure où c'est cette adresse unique qui va permettre de transmettre une information individuelle à un poste unique, différencié de tous les autres postes de télécommunication.

Dans le cadre d'un entrepôt de données, les données choisies peuvent faire l'objet de diverses opérations soit d'anonymisation, soit de codage, éventuellement par un tiers, préalablement à leur entrée dans l'entrepôt de données qui permettra la ou les opérations de data mining.

On pourrait imaginer que l'opérateur de télévision numérique vende à un tiers les profils d'écoute de « ses » téléspectateurs, identifiés par leur adresse IP, mais sans aucune donnée nominative. Ce tiers pourrait à son tour acheter les données « anonymes » de clickstream auprès de cybermarketeurs (comme DoubleClick ou Google[7]). Il serait alors possible de croiser, sur base de leur adresse IP et de la date et l'heure de la connexion les comportements de téléspectateurs et des surfeurs afin de dresser un profil particulièrement pointu de tout ce qu'un ménage consulte sur le web et regarde à la télévision,

Il convient par ailleurs de souligner que les données choisies peuvent faire l'objet de diverses opérations soit d'anonymisation, soit de codage par un tiers ou non préalablement à leur entrée dans l'entrepôt de données qui permettra la ou les opérations de data mining.

3. Le « datamining »

Dans le cadre de cette étude consacrée au profilage, on ne peut faire l'économie de l'évocation des développement des techniques informatiques qui aujourd'hui rendent les activités de profilage de plus en plus aisées et sophistiquées, notamment en améliorant considérablement les approches traditionnelles d'analyse des données et de statistiques. Ces différentes techniques, basées sur les innovations récentes en matière d'intelligence artificielle (arbres de décision, règles d'associations, réseaux de neurones, grilles de score, etc.), ont permis l'émergence du *data mining.*

3.1. Notions

Le *data mining* (littéralement : « forage de données », plus significativement « extraction de la connaissance » ou « exploitation stratégique des données ») peut se définir comme « l'application des techniques de statistiques, d'analyse de données et d'intelligence artificielle à l'exploration et à l'analyse sans *a priori* de (souvent grandes) bases de données informatiques, en vue d'extraire des informations nouvelles et utiles pour le détenteur de ces données »[8]. Cette notion recouvre donc l'ensemble des nouvelles techniques et méthodes qui ont pour but d'explorer et d'amener à la surface de manière exhaustive des relations à partir d'une masse importante de données pouvant relever de sources et de bases de données diverses[9]. En d'autres termes, l'intérêt du data mining réside dans le fait qu'il s'agit d'un outil informatique capable de « faire parler » les données collectées.

---

7 Google a racheté DoubleClick en mai 2007…

8 S. TUFFÉRY, *Data mining et statistique décisionnelle. L'intelligence dans les bases de données*, Ed. Technip, Paris, 2005, p. VII.

9 B. MOXTON, « Defining Data Mining », DBMS Data Warehouse Supplement, 1996, cité par S. OMARJEE, *Le data mining: aspects juridiques de l'intelligence artificielle au regard de la protection des données personnelles*, Univeristé de Montpellier I, ERCIM, 2001-2002, disponible en ligne à l'adresse http//www.droit-ntic.com.

De manière générale, les méthodes sur lesquelles repose le data mining se répartissent en deux catégories : les unes *descriptives* et les autres *prédictives*, selon qu'il existe ou non une variable « cible » que l'on cherche à expliquer ou à prédire.

Ainsi, les méthodes descriptives visent à mettre en évidence des informations présentes mais cachées au sein de la masse des données. Les méthodes prédictives permettent, quant à elles, « d'exploiter un ensemble d'événements observés et historiés pour tenter de prévoir l'évolution d'une activité en dessinant des courbes de projection. Cette méthode peut s'appliquer à la gestion de la relation client pour prédire le comportement d'un client. L'objectif est par exemple de déterminer les profils d'individus présentant une probabilité importante d'achat ou encore de prévoir à partir de quel moment un client deviendra infidèle »[10].

3.2. Les applications du data mining

Les possibilités offertes par le data mining en termes de profilage sont nombreuses et couvrent différents domaines d'application.

3.2.1. La gestion de la relation client et le marketing

De manière générale, le data mining s'avère un outil extrêmement précieux en matière de marketing et de gestion de la clientèle. Il est un des moyens permettant le passage d'un marketing de masse à un marketing véritablement personnalisé. A ce titre, le data mining participe du souci de personnalisation qui caractérise les tendances contemporaines comme le *marketing one to one* et le *customer relationship managment*[11]. En effet, les entreprises souhaitent connaître de manière de plus en plus approfondie les habitudes, les goûts et les comportements d'achat de leurs clients afin de personnaliser leur offre sous la forme de sollicitations ciblées. Dans cette perspective, elles ont donc progressivement enrichi leur base de données commerciales, voire ont développé des « entrepôts de données », des *data warehouses*. Or, l'exploitation efficace de ces bases de données exige des outils spécifiques pour en tirer des informations pertinentes.

En matière de *marketing stratégique*, le data mining peut notamment servir pour l'aide à la création de « packages » et de promotions ; l'aide à la conception de nouveaux produits ; la politique de fidélisation de la clientèle ; l'adaptation de la communication marketing, voir du prix des produits et services proposés (dans une optique de « dynamic pricing »), à chaque segment de clientèle, etc. Ainsi, par exemple, la grande distribution a recours aux méthodes de data mining afin de gérer des bases de données au volume considérable qui ont été alimentées par le développement des cartes de fidélité privatives et enrichies par les informations comportementales provenant des tickets de caisses. L'analyse des associations de produits sur les tickets de caisse permet à l'entreprise d'établir le profil de clients et,

---

[10] H. BENALI, « Analyses décisionnelles et data mining », *SUPINFO Projects*, Ecole Supérieure d'Informatique, Paris, 20 août 2006, disponible en ligne à l'adresse http://www.supinfo-projects.com/fr/2006/decisionnel%5Fdatamining/.

[11] PH. LEMOINE, "Commerce électronique, marketing et liberté", in Groupe d'études Société d'information et vie privée (P. Tabatoni sous dir. de), *La protection de la vie privée dans la société d'information*, t. II, 2000, disponible en ligne à l'adresse http://www.asmp.fr/travaux/gpw/internetvieprivee/rapport2/chapitr7.pdf.

ce faisant, de mieux sélectionner les produits et d'adapter de manière plus subtile sa stratégie de « merchandising ».

La *gestion de la relation client* s'impose aussi comme une des applications privilégiées du data mining. La relation client recouvre une grande diversité d'activités qui toutes peuvent bénéficier des apports du data mining. A titre d'exemple, citons les avantages suivants[12] : l'identification des « prospects » les plus susceptibles de devenir client, le meilleur taux de réponse lors des campagnes de marketing; la personnalisation des pages du site web de l'entreprise, en fonction du profil de chaque internaute; le choix du meilleur canal de distribution ou la détermination des meilleures implantations pour les agences d'une banque ou les établissements d'une chaîne de grand magasins, l'analyse des lettres de réclamation de la clientèle[13, etc.]

Certains chercheurs du MIT s'emploient à développer des logiciels sophistiqués (*recommendations systems*) qui ne se limitent pas à un seul domaine d'application et à certaines données spécifiques, comme ceux utilisés dans le cadre de sites de commerce électronique tels qu'Amazon ou eBay. L'objectif visé par ces recherches est de ne pas se restreindre aux données propres à l'*utilisateur*, envisagé dans le contexte d'une application unique et déterminée mais de privilégier le profilage de la *personne* dans son ensemble : « we must start modeling the person, rather than the user »[14]. Qualifiée de « *social data mining* », la technique est destinée à construire de profils plus riches, notamment en analysant les données collectées sur des sites de « réseautage virtuel » (*web-based social networks*) qui fleurissent ces derniers temps sur Internet tels que Friendster, MySpace ou Facebook. Ces plates-formes numériques en vogue se révèlent des sources précieuses d'informations sur les individus et les communautés socioculturelles dans lesquelles ils s'inscrivent. En effet, les individus ne font pas seulement qu'y mentionner leurs amis et connaissances, ils s'y décrivent eux-mêmes et tiennent également à jour un inventaire détaillé de leurs activités, de leurs intérêts et de leurs passions (littérature, musique, télévision, spectacles, films, sports, nourriture, etc.)[15].

3.2.2. La gestion du risque

C'est dans le domaine de la *gestion du risque*, qui tend à revêtir une importance considérable dans nos sociétés contemporaines, que le data mining tend à s'imposer comme un instrument incontournable. Dans ce domaine, il peut notamment être utile pour la recherche des caractéristiques des clients à risque afin d'adapter la tarification en matière d'assurance, la prévention des impayés, l'aide à la décision de paiement concernant les comptes courants dont le découvert dépasse la limite autorisée dans le secteur bancaire ; l'utilisation d'un « score » de risque pour

---

[12] Les exemples cités ont été partiellement repris de l'ouvrage de S. TUFFÉRY, *Data mining et statistique décisionnelle. L'intelligence dans les bases de données*, Ed. Technip, Paris, 2005.

[13] Une telle opération se base sur l'analyse de données textuelles, relevant de ce que l'on appelle le *text mining*.

[14] H. LIU & P. MAES, "InterestMap: Harvesting Social Network Profiles for Recommendations", *Workshop: Beyond Personalization 2005 IUI'05*, 9 January, 2005, San Diego, disponible en ligne à l'adresse http://ambient.media.mit.edu/assets/_pubs/BP2005-hugo-interestmap.pdf.

[15] Ces « présentations de soi » s'apparentent en quelque sorte à des profils pré-constitués par les individus eux-mêmes.

proposer le montant de crédit le plus adapté à chaque client, ou refuser l'octroi de crédit, en fonction de la probabilité qu'il a de rembourser aux échéances et conditions prévues par le contrat, etc.

Une application récente et inhabituelle du data mining a été développé en matière de risque judiciaire en Grande-Bretagne et au Pays de Galles. Baptisée, projet OASys (Offender Assessment System). Cette application est destinée à évaluer systématiquement les délinquants et à cerner leur profil afin, notamment, d'estimer le risque de récidives et de passage à l'acte[16]. Les variables sur lesquelles repose le système couvrent différents facteurs associés de près ou de loin à la délinquance comme : le niveau d'instruction et de formation, la gestion des revenus, le domicile, le mode de vie, le cercle de relations, l'analyse des délits, etc. Globalement, un tel système est censé offrir aux praticiens et aux responsables un outil permettant d'assurer la qualité et l'efficacité de la pratique en matière d'évaluation, de planification de la prise en charge et de personnalisation des interventions individuelles.

4. Application des règles de profilage à un individu en particulier

Cette étape permet de distinguer le profilage des traitements statistiques. Le traitement statistique a pour but de générer des résultats qui doivent conduire à décrire et comprendre une situation voire à une prise décision privées ou publiques abstraites qui, une fois prises, auront des effets sur des personnes. En d'autres termes, ce qui caractérise le traitement statistique, c'est que son but est l'aide à la prise de décision non individuelle mais globale. Si le profilage inclut donc des opérations statistiques, il nous paraît cependant poursuivre une finalité distincte de celle des opérations statistiques dans la mesure où le profilage a pour but de nourrir une décision modifiant une ligne d'action, et entend appliquer automatiquement et de manière plus effective les options déjà prises. Ainsi, lorsque Google définit les profils appropriés en fonction des caractéristiques propres à tel produit ou service dont il souhaite assurer une publicité « one to one » il ne s'agit pas de connaître l'état du marché ni le cas échéant prendre une décision stratégique quand à l'évolution de ses propres produits (ce que des opérations statistiques peuvent l'aider à faire), mais, en aval d'une décision déjà prise, le profilage permet simplement d'assurer à celle-ci une effectivité maximale lors de son application individuelle par la recherche des critères les plus appropriés et des corrélations les plus riches.

Il est clair que le profilage permet une application immédiate de son résultat. Ainsi, lorsqu'une personne réagit de telle ou telle manière devant son poste de télévision interactive, et se caractérise par son choix de programmes,la correspondance constatée en temps réel entre ses choix et le profil X, permettra l'envoi immédiat de telle ou telle bannière ou spots publicitaires.

---

[16] R. MOORE, "Le système OASys d'évaluation des délinquants en Angleterre et au Pays de Galles", *Probation en Europe*, Bulletin de la Conférence Permanente Européenne de la Probation, Juin 2006, pp. 12-13, disponible en ligne à l'adresse http://www.cep-probation.org/bulletin/june06-F.pdf. Ce système a également pour objectifs d'identifier et de classer les besoins liées aux délits, d'évaluer le risque de passage à un acte grave, de lier l'évaluation au programme d'exécution des peines, de statuer sur le besoin de réaliser des évaluations spécialisées complémentaires, de mesurer le changement pendant l'exécution de la peine, etc.

Le profilage dans d'autres cas donne lieu à une application individuelle qui peut être postposée dans le temps. Dans l'exemple du ministère des finances, on peut parfaitement imaginer que les indices qui constituent le profil d'un fraudeur soient, avant d'être appliquées à des cas individuels, donnent lieu à un repérage préalable des citoyens correspondant à ce profil.

4.1. Les décisions sur base d'un traitement automatisé

La Convention 108, étrangement, ne contient pas de disposition interdisant qu'une décision administrative ou privée impliquant une appréciation sur un comportement humain puisse avoir pour seul fondement un traitement automatisé d'informations donnant une définition du profil ou de la personnalité de l'intéressé.

En cela, elle se différencie de la directive 95/46/CE qui en son article 15 traite explicitement des « décisions individuelles automatisées ». L'article 15 constitue sans aucun doute une disposition originale dans le corps de la directive car son objet porte sur un type de décision et non, comme tel, sur un traitement de données. Plus encore, son originalité réside aussi dans le fait qu'elle est la seule disposition de la directive à porter sur certains aspects des activités de profilage[17].

Dans le cadre de cette étude, il n'est pas inutile de s'attarder quelque peu sur les fondements et les limites d'une telle disposition. L'article 15 de la directive 95/46/CE dispose que :

> *1. Les Etats membres reconnaissent à toute personne le droit de ne pas être soumise à une décision produisant des effets juridiques à son égard ou l'affectant de manière significative, prise sur le seul fondement d'un traitement automatisé de données destiné à évaluer certains aspects de sa personnalité, tels que son rendement professionnel, son crédit, sa fiabilité, son comportement, etc.*
>
> *2. Les Etats membres prévoient, sous réserves des autres dispositions de la présente directive, qu'une personne peut être soumise à une décision telle que celle visée au paragraphe 1 si une telle décision :*
>
> *a) est prise dans le cadre de la conclusion ou de l'exécution d'un contrat, à condition que la demande de conclusion ou d'exécution du contrat, introduite par la personne concernée, ait été satisfaite ou que des mesures appropriées, telles que la possibilité de faire valoir son point de vue, garantissent la sauvegarde de son intérêt légitime*
>
> *ou*
>
> *b) est autorisé par une loi qui précise les mesures garantissant la sauvegarde de l'intérêt légitime de la personne concernée.*

Il convient de souligner ici que, dans le cas précis où l'individu fait l'objet d'une décision prise sur le seul fondement d'un traitement automatisé, le droit d'accès aux

---

[17] Il faut aussi tenir compte de l'article 12, a), 3ème tiret, de la directive qui, en matière de droit d'accès, précise que la personne concernée a le droit d'obtenir du responsable de traitement « *la connaissance de la logique qui sous-tend tout traitement automatisé des données la concernant, au moins dans le cas de décisions automatisées visées à l'article 15 paragraphe 1* ».

données inclut « *la connaissance de la logique qui sous-tend tout traitement des données le concernant* » (art 12 de la Directive 95/46).

Dans son étude consacrée à l'article 15 de la Directive, L. A. Bygrave montre que cette disposition dérive de plusieurs préoccupations dans le chef du législateur européen[18].

La préoccupation majeure a trait à l'automatisation croissante des processus décisionnels à l'égard des individus. Comme le révèlent les travaux préparatoires, le législateur européen en est venu à s'inquiéter d'une telle automatisation tant elle diminue le rôle joué par les personnes dans les processus de décision: « *This provision is designed to protect interest of the data subject in participating in the making of decisions which are of importance to him. The use of extensive data profiles of individuals by powerful public and private institution deprives the individual the capacity to influence decision-making processes within those institutions, should decisions be taken on the sole basis of his 'data shadow'* »[19].

Une autre préoccupation concerne le fait que l'automatisation galopante des processus de décision engendre une acceptation quasi-automatique de la validité et de la pertinence de ces décisions et, corrélativement, un désinvestissement et une déresponsabilisation de décideurs « humains ». A cet égard, la Commission relève que *« the results produced by the machine, using more and more sophisticated software, and even expert system, has an apparently objective and incontrovertible character to which a human decision-maker may attach too much weight, thus abdicating his own responsibilities* »[20].

Enfin, les travaux préparatoires de la directive évoquent aussi le risque que ces « images informationnelles » des individus – ces *data shadows* – finissent par usurper l'autorité constitutive de la personne physique elle-même, et ce, malgré leur caractère souvent abstrait et réducteur. Cette préoccupation, sans doute de nature plus générale, concerne ici le risque d'aliénation de la personne et d'atteinte à la dignité humaine[21].

Malgré sa vocation à couvrir les risques en matière de décisions individuelles automatisées, l'article 15 recèle plusieurs ambiguïtés susceptibles de rendre son application délicate. Sans entrer dans une analyse exhaustive des conditions d'application de cette disposition, relevons tout de même certaines difficultés.

L'article 15 concerne une décision affectant de « *manière significative* » la personne concernée[22]. Que faut-il entendre par significatif ? Cette expression revêt-elle un quelconque sens objectif, indépendant des propres perceptions de la personne

---

[18] L. A. BYGRAVE, « Minding the machine : Article 15 of the EC Data Protection Directive and Automated Profiling », *Computer Law & Security Report*, 2001, vol. 17, pp. 17-24, disponible en ligne à l'adresse http://www.austlii.edu.au/au/journals/PLPR/2000/40.html.

[19] COM(90) 314 final – SYN 287, 13 September 1990, p. 29.

[20] COM(92) 422 final – SYN 287, 15 October 1992, p. 26. Voy. à cet égard le propos édifiant de S. TUFFÉRY, (*op. cit.*, p. 1) : « Le data mining permet de *limiter la subjectivité humaine* dans les processus de décision, et aussi, grâce à la puissance grandissante des outils informatiques, de traiter de plus en plus rapidement de grands nombres de dossiers ».

[21] L. A. BYGRAVE, *op. cit.*, p. 4. Voy. le considérant n° 2 de la directive : « considérant que les systèmes de traitement de données sont au service de l'homme ; qu'ils doivent, quelle que soit la nationalité ou la résidence des personnes physiques, respecter les libertés et les droits fondamentaux de ces personnes, notamment la vie privée, et contribuer (…) au bien-être des individus ».

[22] L'article 1 de la recommandation n° R (97) 18 fait référence, dans un style plus succinct, à « *des décisions ou des mesures relatives à une personne déterminée* ».

concernée ? L'effet de la décision doit-il être de nature purement matérielle ou économique ? Faut-il que la décision ait nécessairement un effet contraire aux intérêts de la personne concernée ? Sans doute, cette dernière condition est-elle nécessaire mais pas suffisante pour considérer que la décision a un effet significatif. En effet, l'envoi d'une brochure commerciale à une liste de personnes sélectionnées sur base d'un traitement automatisé ne semble pas pouvoir être considérée comme affectant la personne de manière significative au sens de l'article 15. Par contre, d'autres types de publicité en cybermarketing apparaissent plus problématiques, notamment lorsqu'elles impliquent une discrimination abusive basée sur l'analyse du clickstream (par exemple, la personne visitant un site web se voyant offrir des produits ou des services à un prix plus élevé que les autres ou la personne se voyant refuser l'opportunité d'acheter des produits ou des services par ailleurs disponibles pour les autres).

Une autre difficulté à la nature du traitement visé par l'article 15. Que faut-il entendre par « *traitement automatisé de données destiné à évaluer certains aspects de sa personnalité* » ? Un tel traitement vise-t-il précisément l'élaboration et l'utilisation de *profils* à des fins d'aide à la décision ? On notera à cet égard que dans cette disposition le traitement est censé porter sur des données, celles-ci n'étant pas explicitement qualifiées de « personnelles ». Dans ce cas, le traitement visé par cette disposition pourrait concerner l'élaboration d'un profil dérivé de données qui ne revêtent pas nécessairement et directement un caractère personnel au sens de la législation en la matière, ce qui est fréquent dans les diverses activités de profilage auxquelles nous avons déjà fait allusion. Cela dit, la référence faite dans cette disposition à « certains » aspects de la personnalité ''un individu est également ambiguë. Est-ce à dire que tous les aspects personnels ne sont pas pertinents pour l'application de cette disposition ? Mais alors où et comment placer la limite à cet égard, en dépit des quelques exemples cités à l'article 15 comme le rendement professionnel, le crédit, la fiabilité ou le comportement ?

Si l'article 15 de la directive semble de prime abord s'imposer comme un contrepoids précieux face aux risques liés aux traitements automatisés propres aux activités de profilage, il faut cependant constater que son application est rendue délicate en raison des ambiguïtés de sa formulation. A cela, s'ajoute le fait que son applicabilité est tributaire de la réunion de plusieurs conditions cumulatives :

- une décision doit avoir été prise ;
- cette décision doit avoir des effets juridiques à l'égard d'une personne ou l'affecter de manière significative ;
- cette décision doit être prise sur le seul fondement d'un traitement automatisé de données ;
- les données traitées doivent être destinées à évaluer certains aspects de la personnalité de l'individu concerné par la décision.

Si l'une de ces conditions vient à manquer, le droit consacré par l'article 15 ne sera pas reconnu.

5. Analyse de la loi suisse

La loi suisse est intéressante dans la mesure où il s'agit, à notre connaissance, de la seule norme nationale qui traite explicitement du profilage.

La loi même appréhende le phénomène du « profilage » à travers une définition, l'article 3 d. définit le « profil de la personnalité » comme « *un assemblage de données qui permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique* ». Cette définition soulève quelques interrogations quant à son champ d'application. Le profil est un assemblage de données. Sa caractéristique est de rassembler des données diverses non naturellement corrélées dans la mesure où l'assemblage de telles données est statistiquement liée à une caractéristique de l'individu ou d'un groupe d'individus (comme une famille, les étudiants de telle classe, les membres d'un quartier) qui permet à celui qui produit ou utilise le profil de prendre une action vis-à-vis des personnes ou groupe de personnes.

Ainsi, il pourrait être acquis, simple exemple, que 80% des personnes de sexe mâle qui paient avec une carte de crédit, font leurs courses le samedi soir à partir de 16h00, et ont dans leur panier, des produits diététiques, des bouteilles de vin de tel prix, sont des cibles idéales pour des voyages de courte durée dans des hôtels de luxe dans des îles paradisiaques et lorsqu'ils sont de professions indépendantes, constituent à 85% des fraudeurs à l'impôt direct.

Cet exemple nous amène à la seconde partie de la définition : l'assemblage de données permet d'apprécier les caractéristiques essentielles de la personnalité d'une personne physique ».

La notion de « caractéristiques essentielles » soulève difficulté. Considère-t-on que les goûts particuliers en matière de voyage d'un groupe ainsi profilé ou la fraude à l'impôt sont des caractéristiques essentielles de la personnalité des membres de ce groupe ?

Sans doute, faut-il relativiser cette notion de « caractéristiques essentielles » et l'envisager par rapport à l'opération que celui qui utilise le profil entend mener. Ainsi, pour celui qui contrôle le respect de la réglementation fiscale, ce qui importe, c'est de découvrir les divers assemblages de données qui caractérisent le fraudeur, de même, pour l'organisateur de voyages, ce qui importe c'est d'élucider les données qui lui permettront d'identifier avec un aléa minima les preneurs des produits qu'ils proposent.

La loi suivie donne à la donnée « profil » un statut proche de celle des données sensibles. Ainsi, on retrouve cette assimilation aux articles 13, 17, 18 ».

Analysons ces trois dispositions :

- l'article 13 évoque les « motifs justificatifs » qui permettent de justifier une atteinte à la personnalité. En dehors du consentement de la personne concernée, il est affirmé que l'atteinte à la personnalité est justifiable par un intérêt prépondérant du maître du fichier.

Il est ainsi noté que l'évaluation du crédit d'une personne est un motif justificatif valable à condition qu'elle ne s'appuie pas sur des données sensibles ni ne soient constitutives d'un profil de personnalité. Cette disposition étonne dans la mesure où elle interdirait la constitution de systèmes de crédit rating dont le fonctionnement repose sur le profilage des différents types de clientèle.

Les articles 17, 17a, 18 concernent des traitements publics. L'article 17 exige que le traitement de « profils » ne soit possible que par une loi au sens formel ou exceptionnellement si la personne concernée y consent ou « a rendu ses données

accessibles à tout un chacun », un cas d'autorisation du Conseil fédéral considérant l'absence de menace des droits des personnes concernées, et enfin si le traitement du profil est absolument exigé par l'accomplissement d'une tâche clairement définie par la loi. Ce dernier cas exclurait la possibilité pour le fisc d'utiliser des profils pour faciliter le contrôle de la perception de l'impôt car la réalisation de cette tâche légale est possible par d'autres modes de contrôle que le profilage. Par contre, si la loi prévoit à des fins de sécurité des aéroports l'utilisation des méthodes modernes d'identification des personnes suspectées de terrorisme.

L'article 17a introduit depuis la loi du 24 mars 2006 permet, au Conseil fédéral sur base de l'avis préalable de l'autorité de protection des données, d'autoriser le traitement de profils dans le cadre d'essais pilotes précédant l'adoption d'une loi au sens formel. Ainsi, on peut imaginer que soient expérimentées des actions administratives, pressenties comme objet de futures lois, auprès de personnes recrutées sur base de profils définis.

Mentionnons enfin l'article 18 qui note à propos des traitements publics que la collecte de données essentielles ou de profils de la personnalité doit être effectuée de façon reconnaissable pour cette dernière. La disposition accentue ainsi l'obligation d'information du responsable de traitement lorsqu'il s'agit de telles données.

6. Des données anonymes.

Il existe une polémique par rapport au concept de données à caractère personnel et à l'anonymat sur Internet. Il convient de relever le concept de Globally Unique Identifier qui a été largement utilisé, notamment pour la rédaction de Privacy Policies via P3P. Par exemple, la « privacy policy[23] » de Microsoft spécifie que *« Le système d'exploitation Windows génère un identificateur global unique (GUID, Globally Unique IDentifier) qui est stocké sur votre ordinateur afin d'identifier celui-ci de façon unique. Le GUID ne contient aucune information permettant de vous identifier personnellement et ne peut pas être utilisé pour vous identifier.* ».. DoubleClick, société de Cybermarketing récemment rachetée par Google déclarait pour sa part : *" DoubleClick does not collect any personally-identifiable information about you, such as your name, address, phone number or e-mail address."*[24].

Il semble donc que des acteurs majeurs de l'industrie des NTIC n'admettent pas que des données comme un numéro associé à un terminal de télécommunication (qu'il d'un numéro de série associé à un composant software ou hardware (voire composé), voire injecté par un acteur particulier (p.e. cookie rémanent). Ainsi en va-t-il de l'adresse IP, dont la classification comme donnée à caractère personnel a été remise en question, malgré la prise de position extrêmement claire du groupe 29, se basant sur le considérant 26 de la Directive 95/46[25,26].

---

[23] Vu sur http://v4.windowsupdate.microsoft.com/fr/default.asp en mai 2004.

[24] Vu sur http://www.doubleclick.net/company_info/about_doubleclick/privacy

[25] http://www.cnil.fr/index.php?id=2244&news[uid]=484&cHash=f2a66a27ee

[26] Voir également à ce sujet les conclusions de l'avocat général Juliane Kokott dans l'affaire C-275/06 Productores de Música de España (Promusicae) contre Telefónica de España SAU en ligne sur http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:62006C0275:FR:HTML

Pour permettre de voir plus clair dans le débat, il s'avère intéressant d'examiner la position des techniciens sur les concepts fonctionnels qui émergent derrière le concept de données à caractère personnel et en dessinent les enjeux. Sur ce terrain, il est indispensable de citer la norme ISO 15408 « Common Criteria for Information Technology Security Information ». « *Ces critères communs permettent l'évaluation des fonctions de sécurité à travers un jeu de onze classes fonctionnelles et d'exigences de garantie : audit de sécurité, communication, support cryptographique, protection des données utilisateur, identification et authentification, gestion des fonctions de sécurité, vie privée, protection des fonctions de sécurité, utilisation des ressources, accès aux composants, et voies de confiance. Ces onze classes fonctionnelles sont divisées chacune en soixante-six familles, chacune comprenant des critères de composants* » [27,28,]. Il est intéressant de constater que sur base de cette norme, la vie privée apparaît comme un aspect de la sécurité des NTIC. Un apport essentiel de la norme ISO 15408 est qu'elle distingue, dans son volet « vie privée » quatre niveaux de protection[29] que nous résumons ci-dessous :

- La non observabilité (unobersvability) : à ce niveau, il est impossible de distinguer un comportement humain du bruit qui l'entoure.
- L'anonymat (anonymity) : ce niveau est atteint si, dans un contexte donné, un ensemble de données concerne plusieurs individus et n'est pas rattachable à un individu en particulier
- Le pseudonymat (pseudonimity) : à ce niveau, l'individu peut-être identifié, dans un contexte donné, par un tiers. Dans un contexte donné, un individu peut posséder un ou plusieurs pseudonymes. C'est ce que nous appelons par ailleurs les données codées.
- La châinabilité (chainability) : c'est un niveau de protection où il est possible de faire un lien entre un même individu, ou, plus précisément, entre deux ensembles de données le concernant.

Dans le même contexte, il convient ici de faire état des travaux d'Andreas Pfitzmann de l'Université de Dresde en Allemagne. L'intérêt des travaux de Pfitzmann et de son équipe est qu'il entend définir une terminologie qui pourrait à terme être approuvée par le W3C et servir de base de vocabulaire dans les régulations techniques (normes) issue du World Wide Web Consortium. D'autre part, il est question de créer une communauté composée de consommateurs, de l'industrie et d'expert afin d'évaluer le niveau de « privacité » ou de « privacidité » des technologies ambiantes.

Au niveau juridique, considérant qu'une « exigence essentielle » des équipements terminaux de télécommunication consiste en *l'incorporation de mesures de sauvegarde pour assurer que les données à caractère personnel et la vie privée des utilisateurs et abonnés soient protégés* (selon l'article 3.3c de la directive 99/5), il est techniquement possible que la Commission Européenne fasse usage de la faculté qui lui est donnée dans son article 5.2[30] de la directive 99/5 afin d'imposer des règles

---

27 http://www.cases.public.lu/documentation/normalisation/ISO_15408/index.html

28 http://www.commoncriteriaportal.org/

29 Voir à ce sujet : Andreas Pfitzmann : Anonymity, Unobservability and Pseudonymity – A Proposal for Terminology, disponible en ligne sur http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

30 Where a Member State or the Commission considers that conformity with a harmonised standard does not ensure compliance with the essential requirements referred to in Article 3 which the said

techniques de préservation de la vie privée et non, sensu stricto, des données à caractère personnel.

Il est par ailleurs à relever que la convention 108, à l'instar de la Directive 95/46, même dûment complétée par la Directive 2002/58 ne sauraient revendiquer à elles seules le titre de défenseur exclusif du droit au respect de la vie privée[31]. D'autres normes juridiques internationales, transposées dans les droits nationaux des pays membres protègent la vie privée. Citons par exemple la convention du Conseil de l'Europe sur le cybercriminalité qui punit en ses articles 2 et 3 l'accès illégal ou l'interception de données informatiques, quelles soient ou non des données à caractère personnel. Dans le cadre européen la Directive 2002/21[32] en son article 8.4.c impose aux autorités réglementaires nationales de « *contribuer à assurer un niveau élevé de protection des données à caractère personnel et de la vie privée* »

Il est par ailleurs piquant de constater que l'adresse email a toujours été considérée, tant par l'industrie que par les autorités de protection des données comme une donnée à caractère personnel. Or, force est de constater qu'il existe de nombreuses adresses email anonymes (p.e. toto234321@yahoo.com) qui ne permettent pas d'identifier une personne mais juste de la contacter.

En conclusion de ce point sur les données à caractère personnel,

- La vie privée des individus doit être protégée, indépendamment ou non de l'utilisation de données les identifiant ou permettant de les identifier. Ce principe transparaît d'ailleurs au travers d'autres textes normatifs de l'Union Européenne ou du Conseil de l'Europe. L'observation et l'enregistrement habituel des comportements d'individus anonymes demeure certainement une intrusion dans leur vie privée.
- La convention 108 du Conseil de l'Europe et les directives européennes de protection des données ne constituent pas des instruments juridiques qui suffisent à protéger la vie privée des individus dans le monde des NTIC.
- Les travaux de normalisation technique menés au sein de l'ISO détaillent des niveaux de protection de la sécurité des données (unobservabilité, anonymat, pseudonymat, non-traçabilité) s'appliquant à des informations relatives aux individus, que ceux-ci puissent être identifiés ou non. Techniquement, il n'y a pas d'anonymat lorsqu'un élément d'un jeu de données relatif à un individu permet de relier celles-ci à un autre jeu de données relatives au même

---

standard is intended to cover, the Commission or the Member State concerned shall bring the matter before the committee » or the article 15 of the directive 2002/58 that state « 2. Where provisions of this Directive can be implemented only by requiring specific technical features in electronic communications networks, Member States shall inform the Commission …3. Where required, measures may be adopted to ensure that terminal equipment is constructed in a way that is compatible with the right of users to protect and control the use of their personal data, in accordance with Directive 1999/5/EC …"

[31] Il convient d'ailleurs de rappeler ici que, de par son titre même, la convention 108 prétend à la protection des personnes à l'égard du traitement automatisé des données à caractère personnel tandis que la directive 95/46 revendique la protection des personnes physiques à l'égard du traitement des données à caractère personnel…

[32] Directive 2002/21 du Parlement et du Conseil relative à un cadre réglementaire commun pour les réseaux et services de télécommunications électroniques (directive « cadre »), J.O.C.E du 24 avril 2002, L108/33

individu. L'anonymat suppose la non-traçabilité, ou, plus concrètement l'impossibilité de procéder de manière automatique à un appariement de deux ensembles de données, ou, de manière plus précise encore, l'absence d'un identifiant codé ou non, raisonnablement rémanent dans le temps qui serait régulièrement présent dans différents jeux de données concernant le même individu, voir sa famille.

- Il ne nous semble pas sérieusement contestable que le profilage d'un individu, dans sa troisième étape, dans la mesure où il s'agit d'une personne identifiable ou identifiée, est clairement un traitement de donnée à caractère personnel, même si ce mécanisme a été originellement conçu en se basant sur des données entièrement anonymes.

7. De la finalité « statistique » et « profilage »

7.1. Considérations générales

La recommandation (97) 18 a pour champ d'application la collecte et le traitement automatisé de données à caractère personnel à des fins statistiques[33]. Ce champ d'application couvre non seulement les activités à caractère strictement statistique, mais aussi les activités fondées sur des procédés statistiques et comportant des opérations de collecte et de traitement des données à caractère personnel, comme les sondages d'opinion ou les études de marché[34].

Pour qualifier les activités de traitement des données à des fins statistiques, la recommandation propose les définitions suivantes (art. 1) :

- "*L'expression « **à des fins statistiques** » se réfère à toutes opérations de collecte et de traitement de données à caractère personnel nécessaires aux enquêtes statistiques ou à la production de résultats statistiques. De telles opérations excluent toute utilisation de l'information obtenue pour des décisions ou des mesures relatives à une personne déterminée*".

- "*L'expression « **résultats statistiques** » désigne une information obtenue par le traitement de données à caractère personnel en vue de caractériser un phénomène collectif dans une population considérée*".

De manière générale, la statistique est une discipline scientifique et une activité ayant pour objectif de mettre en évidence les caractéristiques collectives d'une population ou d'un groupe déterminé et de le faire par synthèse d'informations individuelles.

Comme le précise l'exposé des motifs de la recommandation, « *la statistique a pour objet l'analyse des phénomènes de masse. Elle permet, grâce à un processus de condensation, de tirer une affirmation générale d'une série d'observations individuelles systématiques. Les résultats de ce processus se présentent le plus*

---

[33] Recommandation N° R (97) 18 du Conseil de l'Europe concernant la protection des données à caractère personnel collectées et traitées à des fins statistiques, adoptée par le Comité des Ministres le 30 septembre 1997.

[34] Voy. l'exposé des motifs de la recommandation, n° 61 (ci-après « l'exposé des motifs »).

souvent sous la forme d'informations chiffrées sur le phénomène ou la population considérés. Ainsi, alors même que la statistique repose sur des observations individuelles, son objectif n'est pas la connaissance des individus en tant que tels, mais la production d'informations synthétiques et représentatives de l'état d'une population ou d'un phénomène de masse. L'activité statistique se distingue donc d'autres activités notamment du fait qu'elle ne vise pas des décisions ou des mesures individualisées, mais bien plutôt la connaissance de grands ensembles – tels que les cycles économiques, les conditions de vie d'un groupe social ou la structure d'un marché commercial – ainsi que l'analyse de phénomènes tels que les épidémies, les tendances d'opinion, la fertilité ou les comportements de consommation des ménages – et donc des jugements ou décisions de portée collective »[35].

C'est en sens que doit être comprise la "finalité statistique" telle qu'exprimée à l'article 1 de la recommandation (97) 18. On retiendra donc que toute activité statistique vise à caractériser un phénomène de masse[36]. Si la statistique s'appuie au départ sur les données relatives aux divers individus composant la population ou le groupe étudié, **le résultat statistique détache en dernière instance l'information de la personne**.

Toutefois, à ne s'en tenir qu'à la finalité de la statistique et au fait qu'elle ne produit que des résultats anonymisés, on manquerait le lien étroit que cette discipline entretient avec la problématique de la protection de la vie privée et des données à caractère personnel. En effet, la statistique repose intrinsèquement sur la possibilité de collecter et de traiter un ensemble de micro données, dont certaines peuvent revêtir un caractère personnel.

En matière de statistiques, deux types de risques peuvent être pointés :

- le détournement de finalité : « *On ne peut donc pas exclure le risque que les données en question puissent être détournées de la finalité pour laquelle elles ont été collectées et qu'elles soient alors utilisées à des fins personnalisées. Ce pourrait être le cas, lorsque la statistique côtoie l'administration et la police, où l'on serait tenté d'utiliser pour des jugements et des décisions individualisées des données rassemblées à des fins statistiques* »[37].

- le recoupement des données : « *Par ailleurs, malgré leur caractère anonyme et agrégé, les résultats statistiques peuvent parfois être susceptibles d'analyses ou de recoupements permettant l'identification des personnes dont relèvent les données de base* »[38]. C'est pourquoi, le principe 2.1. inclut les résultats statistiques dans le champ d'application de la recommandation.

Pour prendre la mesure de l'impact des activités statistiques sur la vie privée, il convient dès lors de prendre en considération l'entièreté du processus de production et de diffusion de l'information statistique.

---

35 *Ibidem*, n° 2.
36 *Ibidem*, n° 8.
37 *Ibidem*, n° 3.
38 *Ibidem* n^{os} 3 et 27d.

De manière générale, il importe en outre de souligner le fait que la connaissance statistique n'est pas une fin en soi. Ainsi, peut-elle servir à des fins diverses qu'il est d'usage de classer en trois catégories :

- des fins d'information générale ;
- des fins d'aide à la planification et à la décision ;
- ou encore des fins scientifiques.

Il faut donc tenir compte des « finalités médiates » sous-jacentes à ce type d'activité, tout en ne perdant pas de vue que l'information statistique fournie pour ces finalités médiates porte toujours sur des phénomènes de masse et ne peut dès lors entraîner des conséquences directes ou individualisées pour les personnes[39]. Ainsi, en ce qui concerne la finalité d'aide à la planification et à la décision, les données à caractère personnel qui sont collectées et traitées à des fins statistiques ne peuvent en aucun cas servir de base à la prise de décisions individuelles. Sont visées ici notamment les décisions de nature administrative, judiciaire, fiscale ou financière, ainsi que de décisions d'autorité et de mesures affectant les personnes dans leur cadre de travail ou dans leurs activités associatives ou corporatives[40] (comme celles qui seraient relatives à une admission ou une radiation, une imposition, une allocation, un traitement médical, etc.[41]).

Les activités statistiques couvertes par la recommandation ont donc pour vocation d'interférer le moins possible avec les personnes qui fournissent l'information de base et tout au plus peuvent-elles aboutir à des décisions de portée générale (lois, barèmes, campagnes de vaccination, organisation des transports, conception de modèles, mises en production, etc.) qui, en dépit de leur impact favorable ou défavorable sur certaines personnes, n'impliquent pas à un retour personnalisé sur les individus concernés.

7.2. Les principes

En matière de statistiques, il est d'usage de distinguer deux types de collecte. La collecte primaire effectuée directement auprès des personnes et la collecte secondaire effectuée auprès d'organismes privés ou publics disposant de données sur les personnes.

D'un point de vue méthodologique, le processus de collecte primaire se déroule généralement de la manière suivante : la collecte est précédée d'une phase destinée à fixer le champ de la collecte, c'est-à-dire l'ensemble de personnes a priori pertinent dans le cadre de la statistique projetée. Cette phase permettra ensuite de procéder à la collecte proprement dite à partir de différentes techniques ; la plus connue étant l'enquête réalisée sur la base de questionnaires. Ensuite, la collecte est suivie d'une phase de contrôle destinée à s'assurer de la qualité et de la pertinence des données. Enfin, peut alors avoir lieu le traitement statistique au sens strict qui aboutit aux résultats statistiques. On notera cependant que ce processus méthodologique ne se déroule pas toujours de manière aussi linéaire. En effet, il peut, par exemple, arriver que certaines anomalies apparaissent dans les résultats statistiques et obligent à

---

39 *Ibidem* n° 12.
40 *Ibidem*, n° 68a.
41 *Ibidem*, n° 13.

reprendre les contrôles en amont ou encore que la collecte soit étalée dans le temps, de sorte que l'établissement de certains résultats se fasse avant qu'elle ne soit achevée.

La méthode statistique ne se limite cependant pas au traitement de données collectées au moyen d'enquêtes. En effet, il est fréquent de recourir à des procédés dits de collecte secondaire. Ce deuxième type d'acquisition des données présente assurément de nombreux avantages que ce soit en termes de temps, de coût ou de fiabilité des données, lesquelles, en principe, ont été déjà contrôlées par le premier collecteur.

C'est en ayant égard à un tel cadre méthodologique qu'il convient d'analyser les principes assurant la protection des données à caractère personnel collectées et traitées à des fins statistiques.

7.3. La finalité statistique

Comme nous l'avons souligné, bien que les données ne soient utilisées que pour produire des résultats statistiques et que ceux-ci soient impersonnels, l'activité statistique peut entraîner certains risques à l'égard de la vie privée et des données à caractère personnel. Aussi, pour assurer une certaine protection, deux conditions doivent être réunies : une utilisation exclusivement statistique des données et une production de résultats impersonnels[42].

La finalité statistique étant strictement définie par la recommandation (97) 18, celle-ci impose que les données ne doivent pas être utilisées de manière incompatible avec la finalité de leur collecte[43]. Le principe 4.1 de la recommandation précise à cet égard que *"Les données à caractère personnel collectées et traitées à des fins statistiques doivent servir uniquement à ces fins. Elles ne doivent pas être utilisées pour prendre une décision ou mesure relative à la personne concernée ou pour compléter ou corriger des fichiers dont les données à caractère personnel sont traitées pour des finalités non statistiques".*

Ce principe est une traduction spécifique de l'article 5 de la Convention 108 en vertu duquel les données collectées pour une finalité spécifique ne doivent pas être utilisées pour d'autres finalités incompatibles avec celle-ci[44].

Comme le relève l'exposé des motifs de la recommandation, cette disposition entraîne trois conséquences[45] :

---

42 *Ibidem*, n° 27.

43 On notera que cette disposition fait écho à la définition de l'expression « à des fins statistiques » contenue sous le principe 1 de la recommandation : *L'expression « **à des fins statistiques** » se réfère à toutes opérations de collecte et de traitement de données à caractère personnel nécessaires aux enquêtes statistiques ou à la production de résultats statistiques. De telles opérations excluent toute utilisation de l'information obtenue pour des décisions ou des mesures à une personne déterminée*"(nous soulignons). La recommandation, comme pour en souligner l'importance fondamentale, évoque donc à deux reprises la règle selon laquelle un résultat statistique ne peut servir à la prise de décision à l'égard d'une personne déterminée.

44 En particulier, l'article 5, b), de la Convention dispose que : "*Les données à caractère personnel faisant l'objet d'un traitement automatisé sont enregistrées pour des finalités déterminées et légitimes et ne sont pas utilisées de manière incompatible avec ces finalités*".

1. lorsque des données à caractère personnel ont été collectées et traitées à des fins statistiques, elles ne doivent en aucun cas être utilisées pour prendre des décisions ou des mesures à l'égard des personnes concernées ;
2. les données collectées et traitées à des fins statistiques ne doivent pas être utilisées pour compléter ou corriger des fichiers de données servant à des finalités non statistiques. En effet, si des données étaient ainsi détournées de leur finalité statistique initiale, rien ne permettrait de garantir qu'elles ne puissent être utilisées pour prendre des décisions ou des mesures à l'égard des personnes concernées ;
3. un responsable de traitement peut éventuellement procéder à des collectes ayant une pluralité de finalités, statistiques d'une part et non statistiques d'autre part. Dans le cadre d'une collecte à finalités multiples, le principe 4.1 postule le principe d'une séparation fonctionnelle des traitements en vertu de leurs finalités respectives.

L'article 4.2. envisage le cas de la collecte secondaire. Plus particulièrement, il autorise le traitement à des fins statistiques de données à caractère personnel collectées à des fins non statistiques, cette opération n'étant pas incompatible avec la/les finalités pour lesquelles les données ont été initialement collectées. A l'inverse, le traitement à des fins non statistiques de données collectées à des fins statistiques est considéré comme totalement illicite.

7.4. L'anonymisation

Pour prévenir les risques d'atteinte à la vie privée, le principe 3.3. de la recommandation n° R (97) 18 prescrit l'anonymisation des données : *Les données à caractère personnel collectées et traitées à des fins statistiques doivent être rendues anonymes dès qu'elles ne sont plus nécessaires sous une forme identifiable*.[46]

La question est celle du risque induit par la détention de données identifiables et, en particulier, de données dites d'identification.

Au moment de la collecte peuvent en effet être réunies ce qu'on appelle des données d'identification. La notion de « données d'identification » est définie par la recommandation et recouvre « *les données à caractère personnel qui permettent l'identification directe de la personne concernée et qui sont nécessaires à la collecte, au contrôle et à l'appariement des données, mais qui ne sont pas utilisées par la suite pour établir des résultats statistiques* »[47]. Plus particulièrement, il s'agit d'informations qui identifient les personnes pour les besoins de la collecte et du contrôle (par exemple, la date de naissance ou le lieu de résidence) ; celles-ci peuvent notamment être collectées afin de mener des enquêtes répétitives, d'assurer le contrôle du travail des enquêteurs ou encore de vérifier certaines données douteuses auprès des enquêtés.

---

[45] Exposé des motifs, n° 68.

[46] Des dispositions spécifiques en ce sens sont développées aux principes 8.1 et 10.1 de la recommandation.

[47] Voy. le principe 1 de la recommandation consacré aux définitions.

Dans ce contexte, l'anonymisation s'impose comme une mesure élémentaire de protection. On notera au passage que la recommandation ne définit pas en quoi consiste cette opération ; toutefois, l'exposé des motifs apporte certaines précisions. Ainsi, « l'anonymisation consiste à supprimer les données d'identification afin que les données individuelles ne puissent être attribuées nommément aux diverses personnes concernées »[48]. Plus particulièrement, en matière de statistiques, l'anonymisation sera la plupart du temps réalisée par la séparation des données d'identification, et ce, dès que celles-ci ne sont plus nécessaires aux opérations de traitement.

S'agissant des données d'identification, la recommandation dispose explicitement en son principe 10. 1 que « *lorsque des données d'identification sont collectées et traitées à des fins statistiques, elles doivent être séparées et conservées séparément des autres données à caractère personnel, sauf si cela est manifestement déraisonnable ou infaisable* ». [49]

Une mesure de protection comme l'anonymisation n'est cependant pas infaillible dans la mesure où il subsiste un risque que les données, en principe anonymes, soient réidentifiées[50]. Comme le précise l'exposé des motifs, les données seront cependant considérées comme anonymes si « l'identification exige des activités déraisonnables, c'est-à-dire des opérations excessivement compliquées, longues et coûteuses. Les conditions d'anonymat sont relatives notamment aux moyens techniques dont on peut disposer pour identifier les données et percer ainsi leur anonymat. De la sorte, compte tenu des progrès rapides des techniques et des méthodes informatiques, les délais et activités pour identifier une personne qui seraient aujourd'hui considérés comme 'déraisonnables', pourraient ne plus l'être à l'avenir ». Toutefois, selon l'exposé des motifs, la formulation actuelle est suffisamment flexible pour englober de tels développements[51].

7.5. La licéité

La recommandation (97)18 impose des conditions de licéité à l'égard de la collecte à des fins statistiques de données à caractère personnel. Il s'agit d'une application spécifique du principe de licéité tel qu'il est consacré à l'article 5 a), de la Convention 108 exigeant que les données à caractère personnel soient obtenues loyalement et licitement.

---

48 Exposé des motifs, n° 53b.

49 On lira en parallèle le principe 11.1, al. 2, de la recommandation relatif à la conservation des données : « *En particulier, les données d'identification doivent être détruites ou effacées dès qu'elles ne sont plus nécessaires : aux opérations de collecte, de contrôle et d'appariement ; ou pour assurer la représentativité de l'enquête ; ou pour répéter une enquête avec les mêmes personnes* ».

50 En outre, bien que les résultats statistiques ne puisent être considérés comme des données personnelles, ils permettent, moyennant certaines opérations de recoupement, de retrouver, au moins approximativement, les données propres à certaines personnes et d'établir leur identité.

51 Exposé de motifs, n° 52d. Voy. le principe 1 de la recommandation consacré aux définitions. « *L'expression 'données à caractère personnel' signifie toute information concernant une personne physique identifiée ou identifiable (personne concernée). Une personne physique n'est pas considérée comme 'identifiable' si cette identification nécessite des délais coûteux et des activités déraisonnables. Lorsqu'une personne physique n'est pas identifiable, les données sont dites anonymes* ».

Le principe 4.3 de la recommandation envisage différentes hypothèses. Ainsi, la collecte et le traitement seront considérés comme licites :

- dans les cas où la collecte et le traitement des données à caractère personnel à des fins statistiques sont prévus par la *loi*. Cette hypothèse vise notamment les activités statistiques effectuées dans le cadre d'une mission d'intérêt public impliquant une obligation de renseignement de la part des citoyens.
- dans les cas où la loi l'autorise[52] :
  - si la personne concernée ou son représentant légal a donné son *consentement*. Cette hypothèse vise les situations où la personne concernée est directement interrogée lors de l'enquête statistique.
  - si la personne a été informée de la collecte ou du traitement de ses données et ne s'y est *pas opposée*. Cette dernière hypothèse vise les cas particuliers où la personne concernée a la faculté de s'opposer au traitement de données dans le cadre d'une collecte secondaire[53].

    Notons que dans le cas d'une collecte de ce type –dont nous avons déjà souligné qu'elle constitue un instrument méthodologique précieux– l'information ne peut être fournie aux personnes concernées de la même façon et avec les mêmes moyens que lors d'une collecte primaire, effectuées directement auprès des intéressés. Comme le souligne l'exposé des motifs, « *il serait en effet très coûteux, voire impossible, de retrouver toutes les personnes concernées. De plus, elles pourraient être surprises et même s'inquiéter sans motif de recevoir cette information* »[54].

    Dans ce contexte, l'obligation d'information incombant au responsable de traitement doit être garantie, en vertu du principe 5. 4 de la recommandation, au moyen d'une « publicité appropriée ». Dans le cas des statistiques d'entreprise, la publicité appropriée peut par exemple consister dans des communications d'office adressées aux clients et aux fournisseurs ou figurant d'emblée dans les factures et les ordres de commande. Dans le cas des statistiques publiques, il peut s'agir d'une publication officielle du texte légal ou de tout acte qui autorise la collecte secondaire[55].

Le principe 4.4 de la recommandation formule deux conditions additionnelles de licéité à l'égard de la collecte secondaire de données initialement obtenues à des fins non statistiques. L'objectif de cette disposition est, comme le principe 4.4 l'évoque explicitement, d'éviter que les mêmes données ne soient collectées une nouvelle fois. Le traitement de pareilles données à des fins statistiques sera considéré comme licite si cela est nécessaire[56] :

---

52 La collecte et le traitement sont considérés comme « autorisés » dans le sens où rien ne s'y oppose légalement.

53 La recommandation prévoit alors trois conditions spécifiques : la personne concernée doit être informée de façon appropriée ; elle ne doit manifester aucune opposition au traitement de ses données et ce traitement ne peut porter sur des données sensibles.

54 Exposé des motifs, n° 81.

55 *Ibidem*, n° 81a.

56 Dans les mêmes conditions, les données collectées pour une finalité statistique peuvent également être traitées pour d'autres finalités statistiques.

- à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique ;
- à la réalisation de l'intérêt légitime poursuivi par le responsable de traitement à condition que ne prévalent pas les droits et libertés fondamentales de la personne concernée.

Comme le souligne l'exposé des motifs, « cet intérêt peut relever de considérations de différents ordres : scientifique (la collecte secondaire peut parfois fournir des données d'une meilleure qualité que celles obtenues au moyen d'une collecte directe), technique (l'organisation est plus simple), économique (les coûts sont moins élevés), ou de simple courtoisie à l'égard des personnes concernées (celles-ci ne sont pas dérangées par un enquêteur qui solliciterait des renseignements qu'elles ont déjà fournis dans le passé). Un ou plusieurs de ces critères peuvent donc amener à la décision de procéder à une collecte secondaire »[57].

7.6. La proportionnalité

Ici encore la recommandation se fonde sur une des exigences fondamentales de la Convention 108 : les données collectées doivent être adéquates, pertinentes et non excessives par rapport aux finalités poursuivies par le responsable du traitement (article 5c). Dans le domaine de la statistique, le principe 4.7 de la recommandation dispose que « *la collecte et le traitement des données à caractère personnel doivent être limités aux seules données nécessaires aux finalités statistiques poursuivies* ». En outre, la recommandation porte une attention particulière aux données dites d'identification, lesquelles ne doivent être collectées et traitées que si cela est nécessaire. Par exemple, pour effectuer des contrôles de nature technique ou pour constituer des fichiers statistiques[58].

La notion de « données nécessaires aux finalités statistiques poursuivies » doit cependant être comprise dans le cadre particulier de l'activité statistique. Cela implique d'avoir égard à la méthodologie qui préside à l'organisation et aux modalités de collecte et de traitement des données à des fins statistiques. Dans cette perspective, la recommandation n'impose pas de critères de pertinence ou de nécessité car cela conduirait à battre en brèche l'autonomie scientifique des statisticiens relative à la définition et à la combinaison des variables nécessaires à la caractérisation d'un phénomène collectif. En effet, la pratique révèle que le travail du statisticien porte d'abord sur la constitution d'un jeu de variables et, dans un deuxième temps seulement, sur les données à collecter. Or, dans le cas de certaines enquêtes, il peut s'avérer complexe de définir précisément quels sont les données nécessaires pour parvenir à établir le jeu de variables souhaité[59].

---

[57] Exposé des motifs, n° 72b.

[58] Voy. *supra*, p.

[59] Exposé des motifs, n° 75b : « Ainsi, si on peut aisément définir les données 'nécessaires' pour établir des taux de réussite scolaire ou des statistiques sur le revenu et les dépenses des ménages, il n'en va pas de même lorsqu'il s'agit d'établir des statistiques et des indicateurs sur les inégalités entre les hommes et les femmes ou sur les conditions de vie d'un groupe social. De surcroît, certaines collectes de données à caractère personnel, telles que les recensements de la population, ne visent pas uniquement à produire un ensemble de résultats statistiques prédéfinis, mais aussi à constituer une base d'information fondamentale susceptible d'être exploitée à de multiples fins statistiques particulières dans une longue période - par exemple, au cours de dix ans ».

Comme le souligne l'exposé des motifs, « les rédacteurs ont donc reconnu que la finalité statistique d'une collecte ne peut pas toujours être définie par rapport à un résultat statistique attendu ou à un nombre défini de résultats statistiques. Ainsi, ils sont convenus que le principe de proportionnalité doit certes demeurer un critère général de référence lors de l'établissement des questionnaires et des plans d'enquête, mais que son application doit, d'une part, respecter l'autonomie scientifique des approches méthodologiques du statisticien et, d'autre part, tenir compte du degré de détermination des objectifs de chaque collecte »[60].

8. La finalité du profilage

Dans le cadre de notre étude, il importe toutefois d'identifier et de s'interroger sur les différents types de constitution et d'utilisation de profils, lesquels ne recèlent pas tous le même potentiel privacide. En effet, l'utilisation d'un profil peut servir des objectifs très divers. A cet égard, nous proposons la typologie suivante[61] :

Utilisation du profil afin de :

a) établir une règle générale :
   i. en science
   ii. comme ressource commerciale et marketing
   iii. utilisée par l'organisation pour prendre des décisions organisationnelles d'ordre général
   iv. utilisée par l'organisation pour prendre des décisions organisationnelles d'ordre spécifique

b) appliquer à un cas individuel et concret
   i. décider d'offrir ou non quelque chose à un groupe
      1. les individus n'étant pas encore réellement intéressés
      2. les individus souhaitant l'avoir
   ii. décider d'offrir quelque chose à un individu
      1. en tant que tel
      2. avec une remise
      3. avec un surcoût
   iii. décider ou non d'accéder à la requête d'un individu
      1. qui n'est pas vitale pour l'individu
      2. qui est vitale pour l'individu

Au regard de cette typologie, on peut par exemple constater que la ligne de démarcation entre statistiques traditionnelles[62], techniques de data mining et de profilage n'est pas toujours facile à tracer.

---

60 *Ibidem*.

61 Cette typologie s'inspire de la réponse formulée par B.-J. KOOPS (replier), in M. HILDEBRANDT & S. GURWITZ (eds.), *Implications of profiling practices on democracy and rule of law*, Future of Identity in the Information Society (FIDIS), Report D7.4, 5 September 2005, pp. 66-67.

62 On notera cependant que, contrairement à la statistique traditionnelle, la mise en œuvre des techniques de data mining ne repose pas sur les mêmes bases méthodologiques. Ainsi, on peut observer que le data mining vise à exploiter des données plus volumineuses, fluides et souvent mal échantillonnées, parfois lacunaires, initialement collectées pour d'autres fins que l'analyse statistique. La question de la qualité des données se pose à cet égard avec une acuité particulière.

En effet, les techniques descriptives de data mining semblent beaucoup s'apparenter à la statistique traditionnelle. Ainsi en est-il de l'analyse du panier de la ménagère dans la grande distribution qui permet de déterminer les produits souvent achetés simultanément et, en conséquence, d'agencer les rayons et d'organiser les promotions. Parmi les techniques descriptives, on peut aussi citer les opérations de classification automatique connue en marketing sous le nom de segmentation de la clientèle.

La modélisation permise par le data mining est ici très proche des finalités visées par la recommandation (97) 18 puisqu'elle contribue à la connaissance de grands ensembles, de groupes sur base de laquelle des jugements ou décisions de portée collective peuvent être pris.

Plus particulièrement, dans le cadre de la statistique traditionnelle, les résultats statistiques visent à mettre en évidence ce qu'on appelle des « agrégats », c'est-à-dire des totaux, des moyennes, des pourcentages, des dénombrements (par exemple, la répartition de la population suivant diverses caractéristiques individuelles comme l'âge, la profession, le lieu de résidence). Quand ils ne sont pas purement numériques, les résultats statistiques visent à établir des « typologies », « c'est-à-dire la mise en évidence de groupes caractérisés par la coïncidence de certains états ou comportements, un zonage sur un territoire ou des regroupements dans une nomenclature de professions, etc. »[63].

Cependant, si l'on ne peut reprocher à une entreprise de chercher à caractériser sa clientèle et de procéder à des sélections en fonction de variables pertinentes pour orienter sa stratégie et son action commerciales, il faut être attentif au fait que les techniques de segmentation, par exemple, autorisent la constitution de profils, c'est-à-dire des classes homogènes de clients en fonction des comportements observés à partir de diverses variables. La segmentation, dite « comportementale » permet en effet, à partir d'informations déduites de l'observation des comportements, d'établir le profil socio-économique, voire psychologique d'une personne, laquelle sera ensuite classée dans un « segment ». Par exemple, dans une visée de prospection commerciale, les entreprises ont développé des méthodes et des outils permettant de segmenter la clientèle afin de sélectionner les clients susceptibles d'être intéressés par les produits ou services proposés.

Si le segment lui-même peut difficilement être considéré comme une donnée à caractère personnel[64], son utilisation peut devenir problématique lorsqu'il est associé à une personne identifiée ou indirectement identifiable et figure dans un traitement automatisé. A cet égard, la CNIL considère que « le segment, dans la mesure où il résulte de traitements statistiques, n'appartient pas aux informations de base recueillies auprès des personnes intéressées et, en conséquence, ne constitue pas à lui seul une information nominative : qu'il le devient toutefois dès lors qu'il est associé

---

63 Exposé des motifs de la recommandation N° R (97) 18, n° 9.

64 De manière identique, les résultats statistiques ne constituent pas des données à caractère personnel puisque, comme tels, ils ne sont pas liés à une personne physique identifiée ou identifiable. Toutefois, par le biais de recoupements, il peut arriver qu'un résultat statistique permette d'identifier des données individuelles et d'établir un lien entre celles-ci et l'identité des personnes concernées. Pour assurer la protection de la vie privée à l'égard de ces résultats, le principe 14 de la recommandation (97)18 formule certaines exigences en matière de communication et de publication de ce type de résultats.

à une personne identifiée ou indirectement identifiable et figure dans un traitement automatisé »[65].

Un autre danger associé à la technique de segmentation est qu'elle s'appuie sur des grilles comportant un nombre très conséquent de variables. Dans le secteur bancaire, par exemple, on peut se demander d'une telle pratique si elle n'entre pas en contradiction avec le principe de légitimité et de proportionnalité de l'information prévu par la Convention 108. Ainsi en est-il lorsque le banquier s'appuie sur des données comme le nombre et le montant des prélèvements vers des organismes de crédit concurrents, le solde net du foyer et non du titulaire du compte, la détention d'une assurance extérieure à la banque, le nombre de paiements par carte à l'étranger, le montant de l'assurance-vie[66].

S'agissant de la constitution de profils, les techniques prédictives de data mining semblent encore plus délicates à appréhender. Une attention particulière doit en effet être accordée aux processus de *scoring* qui se révèlent être une des applications les plus usitées dans les secteurs comme la banque, l'assurance ou la téléphonie ; le scoring étant précisément rendu possible par les techniques de segmentation comportementale susmentionnées.

Le scoring est basé sur un ciblage plus pointu des individus. Il peut s'agir de prédire :

- le *risque*, par exemple le risque de défaillance d'un demandeur de crédit.
- l'*appétence*, c'est-à-dire la probabilité d'achat plus ou moins forte d'un client ; ce qui, par exemple, est particulièrement utile pour concentrer les mailings sur les clients les plus susceptibles d'y répondre favorablement.
- l'*attrition*, c'est-à-dire la perte ou le départ d'un client chez un prestataire concurrent. Par extension, cela vise la capacité de l'entreprise à retenir ses clients, à les fidéliser ; il permet de définir le taux de défection.

Le système de scoring de la clientèle permet donc d'identifier et d'élaborer une typologie des clients en attribuant à chacun d'eux une note, accessible au réseau commercial, utilisée pour apprécier l'opportunité d'octroi d'un produit ou d'un service ainsi que le risque de défaut[67]. En matière de crédit, par exemple, le mécanisme est le suivant : les logiciels de *credit scoring* associent à des informations personnelles[68]

---

[65] Commission nationale de l'informatique et des libertés, Délibération 93-032 du 6 avril 1993 relative au contrôle effectué le 2 octobre 1992 à la Caisse Régionale de Crédit Agricole de la Dordogne : « Considérant en conséquence que dans ce cas, cette information doit être adéquate, pertinente et non excessive par rapport aux finalités qui ont conduit à son enregistrement ; qu'en particulier, les finalités qui ont présidé à la collecte des informations de base ne doivent pas être méconnues ; que la segmentation ne doit pas reposer sur des informations dont la collecte est interdite ou qui seraient complètement étrangères aux activités de l'entreprise, notamment dans la mesure où elles concerneraient des éléments de la vie privée dont elle n'a pas à connaître ».

[66] Commission nationale de l'informatique et des libertés, « Note d'information sur les conséquences au regard de la loi 'Informatique et libertés' modifiée de la notation de la clientèle bancaire (ratio prudentiel « Mac Donough » - Bâle II) », 3 mars 2005, p. 5.

[67] *Ibidem*, p. 4.

[68] La CNIL identifie trois catégories d'informations de nature personnelle : les informations « administratives » sur le clients telles que l'âge, la catégorie socio-professionnelle, l'ancienneté du client dans la banque, etc.), les informations de comportement telles que le nombre et le type de produits bancaires détenus, le solde moyen du compte bancaire sur les douze derniers mois,

relatives aux demandeurs de crédit des pondérations particulières issues de données statistiques et de probabilités, de sorte qu'au-dessus d'un montant de points – une note – le crédit est automatiquement accordé. Au contraire, si une mauvaise note est attribué au client par le système informatique, la banque aura fortement tendance à refuser la demande.

Dans ce contexte, la constitution de profils doit être pratiquée dans le respect de la législation en matière de protection des données à caractère personnel dès lors que le profil est associé à une personne identifiée et identifiable et figure dans un traitement automatisé.

8.1. Finalités

Les techniques de data mining ont ceci de particulier qu'elles permettent une utilisation diversifiée des données collectées. Ainsi, les données collectées pour une finalité précise peuvent être ultérieurement traitées pour d'autres objectifs. Par exemple, des informations transactionnelles collectées dans le but d'effectuer un paiement par carte de crédit peuvent être utilisées par la suite pour d'autres finalités, notamment pour des opérations de data mining. Par essence, le data mining s'apparente toujours à un traitement ultérieur qui peut modifier la finalité des traitements et conduire à un risque pour la vie privée.

Au regard de la législation en matière de protection des données à caractère personnel, une des difficultés principales serait liée, pour certains, au fait qu'un bon programme de data mining ne permettrait pas, à l'avance, de délimiter quel serait la finalité déterminée – médiate – de sa mise en œuvre, son but étant d'extraire des informations inconnues d'un vaste ensemble de données disponibles. Comme l'affirme Ann Cavoukian, "the data miner does not know, cannot know, at the outset, what personal data will be value or what relationships will emerge. Therefore, identifying a primary purpose at the beginning of the process, and then restricting one's use of the data to that purpose are antithesis of a data mining exercise"[69].

8.2. Interconnexion

La question de la finalité du traitement devrait en outre être croisée avec celle de l'interconnexion (et des notions apparentées comme la corrélation et l'appariement).

- Voy. le principe 4.6 de la recommandation (97) 18 : « *Des données à caractère personnel ou des ensembles de données à caractère personnel peuvent être appariés ou mis en relation à des fins statistiques si le droit interne ménage des garanties appropriées pour empêcher leur traitement et leur communication à des fins non statistiques* ».

---

les revenus réguliers et la surface financière, etc. et enfin les « événements » bancaires comme la présence ou non d'oppositions.

[69] A. CAVOUKIAN, *Data Mining : Staking a Claim on Your Privacy*, Information and Privacy Commissioner Ontario, January 1998, p. 13, disponible en ligne à l'adresse http://www.ipc.on.ca/images/Resources/up-datamine.pdf

- Voy. la définition proposée à l'article 2 de la loi luxembourgeoise du 13 août 2002 relative à la protection des personnes à l'égard du traitement des données à caractère personnel[70]

(j)"interconnexion" : toute forme de traitement qui consiste en la corrélation de données traitées pour une finalité avec des données traitées pour des finalités identiques ou liées par un ou d'autres responsables du traitement

- Voy. Article 25 I de la loi française n° 78-17 du 6janvier 1978 relative à l'informatique, aux fichiers et aux libertés :

Sont mis en œuvre après autorisation de la Commission nationale de l'informatique et des libertés, à l'exclusion de ceux qui sont mentionnés aux articles 26 et 27 :

Les traitements automatisés ayant pour objet :

- l'interconnexion de fichiers relevant d'une ou de plusieurs personnes morales gérant un service public et dont les finalités correspondent à des intérêts publics différents ;

- l'interconnexion de fichiers relevant d'autres personnes et dont les finalités principales sont différentes.

La CNIL semble se diriger vers une interprétation large de l'interconnexion[71] : « Il s'agit de tout traitement automatisé mis en œuvre par un ou plusieurs responsables qui consiste à mettre en relation (à corréler) des données ayant une finalité avec d'autres données ayant une finalité identique ou différente. Cette mise en relation (ou corrélation) peut consister à transférer un fichier pour alimenter un autre fichier ou pour réaliser la fusion de ces fichiers, à mettre ponctuellement en relation plusieurs fichiers normalement gérés séparément, par exemple en constituant un fichier d'appel à partir de l'un de ces fichiers qui servira à interroger les autres fichiers et sera enrichi par les résultats de cette interrogation. Il peut également s'agir d'assembler des informations provenant de plusieurs fichiers au sein d'une même base de données (exemple des bases dénommées «entrepôts de données» alimentés par des informations provenant de différents fichiers) avec un éventuel recours à des techniques logicielles de mises en relations ponctuelles (outils dits de datamining) ou de créer un lien technique entre plusieurs bases de données nominatives qui permettra, par exemple, de les consulter simultanément (par exemple, des sites portails permettant par des « liens hypertextes » d'assurer une mise en relation avec d'autres bases) ».

La Commission s'oriente également vers une interprétation large en ce qui concerne « l'interconnexion de fichiers relevant d'autres personnes et dont les finalités principales sont différentes ». Il s'agirait aussi bien : d'une interconnexion de fichiers de plusieurs personnes relevant du secteur privé ; d'une interconnexion entre fichiers d'une personne du secteur public et fichiers d'une personne du secteur privé ; d'une interconnexion réalisée au sein d'une même personne morale de droit privé entre de fichiers présentant des finalités principales différentes.

---

70 *Mémorial*, 13 août 2002, A – N° 91, p. 1836.

71 http://www.cnil.fr/index.php?id=1735.

8.3. Autorégulation

Il faudrait aborder la question de l'autorégulation. Par exemple, en ce qui concerne la détermination de normes relatives :

- aux procédures internes : des procédures devraient au sein des entreprises être définies en ce qui concerne l'utilisation du système de notation par la clientèle ;
- à la notation ou l'appartenance au segment : la notation ou l'appartenance au segment ne devrait pas être accessible en permanence sur la fiche informatique du client afin d'éviter toute stigmatisation ;
- à l'élaboration des segments : les segments ne devraient pas comporter des qualificatifs péjoratifs, défavorables ou subjectifs sur des individus regroupés (tels que « tempérament de jouisseur »).

9. Conclusions et recommandation

9.1. Le profilage est-il un traitement de données à caractère personnel ?

Dans le rapport explicatif de la Convention 108, il est rappelé, que la présente Convention a pour objet de renforcer la protection des données, c'est-à-dire la protection juridique des individus vis-à-vis du traitement automatisé des données à caractère personnel les concernant.

On peut se demander si le profilage constitue un traitement de « données à caractère personnel ? De manière traditionnelle, tant la convention 108 du Conseil de l'Europe que la directive européenne 95/46 protège les atteintes aux libertés individuelles et à la vie privée mais uniquement dans le contexte d'un traitement inadéquat des données à caractère personnel, c-à-d des données se rapportant aux personnes identifiées ou identifiables. Une particularité du profilage est qu'il procède à la fois du traitement de données qui seront en règle générale soit anonymes, anonymisées ou codées dans les deux premières étapes et du traitement de données à caractère personnel, sensu stricto, dans l'application des règles du profilage à une personne individuelle identifiable ou identifiée.

Encore faut-il s'entendre sur ce que signifient des données parfaitement anonymisées. Nous considérons qu'il convient de se référer à la définition fonctionnelle de Pfitzmann relative à l'anonymat[72] : « ***Anonymity* of a subject means that the subject is not identifiable**[73] **within a set of subjects, the *anonymity set*.**[74] En d'autre termes, des données qui contiennent un identifiant propre à un individu ne sont pas anonymes si

---

[72] Anon Terminiology disponible en ligne sur http://dud.inf.tu-dresden.de/Anon_Terminology.shtml

[73] "not identifiable within" means "not uniquely characterized within".

[74] From [ISO99]: "[Anonymity] ensures that a user may use a resource or service without disclosing the user's identity. The requirements for anonymity provide protection of the user identity. Anonymity is not intended to protect the subject identity. [...] Anonymity requires that other users or subjects are unable to determine the identity of a user bound to a subject or operation." Compared with this explanation, our definition is more general as it is not restricted to identifying users, but any subjects.

d'autres données personnelles contiennent ou peuvent contenir le même identifiant relatif au même individu.

Le traitement de données purement anonymes dès leur genèse (p.e. images d'individus non reconnaissables issues de la vidéo surveillance, panier de la ménagère payé en liquide, etc) sort du ratio materiae de la convention 108. Cela ne signifie pas pour autant que de telles opérations sont licites et légitimes au regard d'autres normes juridiques, en particulier au regard de l'article 8 CEDH.

Il convient d'insister lourdement sur ce dernier point qui peut paraître évident aux professionnels de la protection des données. Dans l'esprit du grand public et dans l'industrie des NTIC , semble émerger la conviction que les traitements de données personnelles d'individus non identifiés et non indentifiables ne sont soumis à aucune règle de droit. Il n'en est évidemment rien, d'une part parce que de plus en plus de normes juridiques (notamment des directives européennes) utilisent le vocable bicéphale « vie privée et protection des données personnelles » et d'autre part parce que le droit au respect de la vie privée demeure un droit fondamental garanti par de nombreuses constitutions nationales et par l'article 8 CEDH. L'individu a droit à une intimité qui ne s'accommode pas ou mal d'une surveillance anonymisée et d'observations systématiques de ses moindres faits et gestes, fussent-ils routiniers.

9.2. Finalité statistique et finalité de profilage

Le « profiling » né des techniques actuelles des « datawarehouse » (entrepôt de données) et de « data mining » constitue un traitement qui, même s'il utilise les méthodes statistiques, ne correspond pas aux traitements à « finalité statistique » pour lesquels des règles spécifiques comme celles prévues par la Recommandation R(97) 18 ont été élaborées.

Les traitements de finalité de « profiling » présentent a priori un risque supérieur en matière de protection des données, en particulier des risques d'application discriminatoire dans la mesure où ils ont pour finalité non d'aboutir non à une simple constatation de faits vue ou non de modifier une ligne d'action mais à une application médiate ou non vis-à-vis de situations individuelles, le profilage constituant une méthode permettant une efficacité plus grande au service de finalités comme celle publicitaire, de contrôle du respect de la réglementation ou d'octroi de crédit.

Cette seconde réflexion a bien évidemment une conséquence, il va de soi que les limites posées aux traitements statistiques doivent toute la mesure du possible s'impose également aux traitements dits de profilage, étant donné que d'autres mesures supplémentaires s'imposent en ce qui concerne ces dernières et que par ailleurs certaines règles qui visaient à encourager les traitements statistiques soient a priori déclarés non applicables.

Ainsi, le principe selon lequel « le traitement à des fins statistiques de données à caractère personnel collectées à des fins non statistiques n'est pas incompatible avec la/les finalités pour lesquelles les données ont été initialement collectées, dans la mesure où des garanties appropriées sont prévues notamment pour empêcher l'utilisation des données à l'appui de décisions ou de mesures relatives à la personne concernée (Principe 4.1. de la Recommandation, repris également par l'article 6 de la directive 95/46/CE).

Le refus d'application de ce principe entraîne nécessairement le besoin de trouver dans les autres fondements de légitimité traditionnellement retenus le fondement de l'utilisation des données collectées pour une finalité donnée dans le contexte de la finalité de profilage. Il paraît difficile de permettre la collecte secondaire à des fins de profilage lorsque « celle-ci est nécessaire à l'exécution d'une mission d'intérêt public ou relevant de l'exercice de l'autorité publique » comme l'indique l'article 4.4. de la recommandation à propos des traitements statistiques de l'autorité publique. Sans doute, de tels traitements doivent-ils relever d'une autorisation directe de la loi au sens matériel du terme et, dans le secteur privé, peut-on admettre le profilage simplement parce que « nécessaire à la réalisation de l'intérêt légitime poursuivi par le responsable du traitement », comme indiqué par le même article. Certes, la Recommandation ajoute « à condition que ne prévalent pas les droits et libertés fondamentales de la personne concernée » mais cela suffit-il et ne faut-il pas limiter les traitements de profilage au moins par des conditions procédurales qui permettront de juger de cette prépondérance évidente des intérêts du responsable du traitement vis-à-vis des intérêts de la personne concernée.

9.3. Le profilage comme finalité et les finalités du profilage

Comme nous l'avons écrit supra, le profilage n'est pas une finalité en soi mais une modalité technique d'aboutir à un résultat. En pratique, il s'agira toujours (cfr notre taxinomie décrite supra)

- de refuser des biens et services ou de les proposer avec des modifications de prix (à la hausse ou à la baisse) ;
- de contacter un individu pour une finalité de marketing direct ;
- de prendre d'autres décisions individuelles susceptibles d'affecter plus ou moins gravement un individu.

On comprend aisément que le profilage peut être bien plus lourd de conséquences qu'un simple traitement statistique dans la mesure où il peut exclure un individu de l'accès à un travail (p.e. il semblerait que toutes les administrations publiques américaines effectuent un scoring ATS avant d'engager un fonctionnaire fédéral), au crédit (sur base d'un mauvais crédit scoring) , au logement, etc. Dans d'autres cas, les conséquences sont moins dramatiques (marketing, décision de procéder à un contrôle fiscal,…). Si les conséquences affectent donc plus ou moins l'individu, le mode de décision lié au profilage peut aggraver celles-ci. Le profilage comporte fatalement un certain taux d'erreur que l'on peut supposer minime. Il n'en demeure pas moins que certains individus se retrouveront, à tort, touchés par une décision prise, dans leur cas particulier, sans fondement. Ceci est particulièrement gênant dans la mesure ou l'individu se trouve confronté à une machine, par exemple via Internet, qui est incapable de faire preuve de bon sens. A l'instar de la directive 95/46 en son article 15, il nous semble important que la personne concernée puisse avoir un droit de recours non automatisé, lorsqu'il est l'objet de décisions de profilage automatisées, en particulier lorsque ces décisions affectent l'exercice d'un droit fondamental.

9.4. Licéité, transparence et proportionnalité

Ces trois piliers d'un traitement de données à caractère personnel devraient impérativement intervenir de manière précoce dans une opération de profilage, même si des données humaines non personnelles font l'objet d'un traitement. Il nous apparaît que le risque d'un traitement illicite, peu transparent ou disproportionné naît, de manière précoce, dès les opérations d'entreposage et de datamining. Une politique de prévention des risques (principe de précaution) devrait donc naître en amont de l'application d'un profil à un individu déterminé. Les mesures de sécurité prévues à l'article 7 de la Convention peuvent concerner les opérations de collecte de données même parfaitement anonymisées.. Ainsi, par exemple, il nous semblerait relever d'une bonne politique de sécurité l'interdiction de l'entreposage de données sensibles et du data mining sur des corrélations entre des données sensibles comme la race ou les déviances sexuelles et le profil de consommation. Un tel datamining pourrait, in fine, mettre en évidence des corrélations logiques entre certains paniers d'achats et certaines déviances sexuelles ou certaines races. Il serait alors possible d'induire ou de déduire la valeurs de ces deux variables sur base du panier d'achat, ce qui semble a priori illicite.

9.5. De la nécessité d'un mode de protection particulier contre les opérations de profilage

Nous avons souligné le régime particulier de protection proposé par les articles 12 et 15 de la Directive 95/46. qui constituent un fondement juridique permettant à un individu de ne pas être soumis aux décisions produisant des effets juridiques ou l'affectant de manière significative prises sur le seul fondement d'un traitement automatisé destiné à évaluer certains aspects de sa personnalité.

L'individu, toujours sur la base de cette Directive, peut connaître la « logique » qui sous-tend le traitement automatisé de données, au moins dans le cas cité ci-dessus.

La convention 108 ne propose pas de régime particulier d'information, d'accès ou d'opposition aux traitements automatisés destinées à évaluer certains aspects de la personnalité d'un individu.

Techniquement un tel regime de protection pourrait être ajouté à la convention 108 sous la forme d'un protocole additionnel.

La motivation de ce régime de protection trouve sa base sur la nature particulière des risques nouveaux et croissants liés à la multiplication des opérations de profilage. Ces risques sont liés à la triple montée en puissance des opérations de data mining (multiplication des types de données stockés avec stockage automatique et systématique d'opérations routinières, augmentation de la taille des entrepôts de données et des possibilités d'interconnexion, montée en puissance des techniques de data mining, de la puissance de calcul, de son opacité et de sa complexité).

De manière courante, de plus en plus d'opérateurs commerciaux ont et auront recours à ces techniques pour accorder ou refuser l'accès à leurs biens ou services ou pour en moduler le prix. Le risque d'exclusion

De manière plus générale, pris dans le filet d'une opération de profilage dont la complexité, voire la logique ou même l'existence lui échappe, l'individu risque de se retrouver malgré lui dans un schéma kafkaïen. Il devient alors incapable d'avoir une quelconque maîtrise sur son image informationnelle ni même de comprendre les mécanismes présidant à la création de cette image.

Sans ce régime particulier de protection, il existe un risque important que les opérateurs économiques recourent de plus en plus souvent et de manière systématique à un profilage rapide et peu coûteux de leur clientèle. Ce profilage génèrera, fatalement, l'exclusion de certains individus à certains biens ou services ou l'augmentation du prix de leur accès. On risque ainsi d'opposer à un individu la prédiction d'un futur qui n'est pas le sien mais le fruit d'une prévision basée sur les comportements antérieurs d'autres individus qu'il ne connaît et auquel il est, formellement étranger.

Nous pensons, dans le contexte économique et technologique ambiant et raisonnablement prévisible, que ce protocole additionnel doit aller au delà du mode de protection prévu dans la Directive 95/46 et, qu'en particulier, il n'y a pas lieu de limiter la protection de l'individu aux décisions automatisées produisant des effets juridiques ou l'affectant de manière significative. Le simple fait d'avoir été l'objet d'un profilage automatisé destiné à évaluer certains aspects de sa personnalité doit ouvrir à l'individu le droit à en être informé, à pouvoir accéder à la logique du profilage et à s'opposer, au moins dans certains cas à un traitement automatisé censé pouvoir évaluer certains aspects de sa personnalité.

9.6. La recommandation d'une recommandation

En conclusion de notre étude nous pensons qu'il serait opportun pour le Conseil de l'Europe de produire une recommandation tendant à encadrer les activités de profilage. Plusieurs éléments nous semblent de nature à justifier semblable recommandation :

a) Le Conseil de l'Europe a déjà, très judicieusement, produit une recommandation relative aux statistiques. Or, le profilage constitue pour nous une finalité (ou une modalité technique) distincte de celle des statistiques, et ce malgré la confusion qui peut régner dans l'esprit du grand public. L'utilisation grandissante de profils risque d'affecter de manière grave un nombre croissant d'individus en leur attribuant un profil qui ne correspond pas nécessairement à leurs particularités ou se révèle trop invasif par rapport à leur vie privée. *A fortiori*, une recommandation sur le profilage s'impose-t-elle donc.

b) Le profilage constitue une activité qui a connu un essor prodigieux ces dernières années. Les nouvelles technologies de l'information et de la communication se sont aujourd'hui largement affranchies des problèmes d'encombrement, de coût, de lenteur ou fiabilité qui les caractérisaient voici une dizaine d'année. A l'aube du 21ème siècle, il n'aura jamais été aussi facile, aussi rapide, aussi peu visible et bon marché d'observer et de stocker la plupart des micro actions anodines et quotidiennes des êtres humains (faire un achat ou une vente, effectuer une recherche, se déplacer, lire un journal, consulter un livre, envoyer et recevoir du courrier, changer de chaîne de télévision, tester un tube de rouge à lèvre dans un magasin, …)

c) Il règne actuellement une ambiguïté importante au sujet de la qualification des données comme données à caractère personnel et de nombreux acteurs risquent de considérer la Convention 108 et les Directives européennes 95/46 et 2002/58 comme les seules normes juridiques de protection de la vie privée. Dans le cas où ces normes ne trouveraient pas à s'appliquer ratio materiae, faute de la qualification des données individualisées comme des données à caractère personnel, certains maîtres du fichier pourraient croire, à tort, que leur traitement est entièrement compatible avec le respect de la vie privée, tel que décrit à l'article 8 de la Convention Européenne des Droits de l'Homme. Une Recommandation sur le profilage pourrait avoir pour effet de rappeler à ces acteurs les exigences de dignité et d'intimité qui constituent les fondements de l'article 8 CEDH, exigences qui, bien sûr, s'appliquent à tous les individus, identifiables ou non.