# Protection de la vie privée : droit, modèles et outils

Julien Le Clainche, Daniel Le Métayer, Guillaume Piolle, Romuald Thion

INRIA- Grenoble Rhône-Alpes
655 avenue de l'Europe, Montbonnot
38334 Saint-Ismier Cedex - FRANCE

## 1 Introduction

Nous résumons dans le présent document les travaux passés des membres de LICIT sur le thème de la protection de la vie privée ainsi que les perspectives de recherche du groupe dans ce domaine. Les activités de recherche de LICIT concernent de manière plus générale les interactions entre le droit et les technologies de l'information. La démarche promue par LICIT, qui s'applique en particulier à la problématique de la vie privée, met en avant la modélisation formelle comme pivot entre les aspects juridiques et informatiques. Cette démarche se justifie de manière théorique par la nécessité, commune aux deux disciplines, de traiter avec rigueur des problèmes complexes. Elle recouvre également des besoins pratiques, les outils informatiques étant de plus en plus souvent utilisés dans des contextes où ils comportent des implications juridiques (vie privée, contrats électroniques, preuve électronique, protection des biens numériques, etc.). D'un point de vue concret, toutes les activités de LICIT sont menées en étroite interaction avec des partenaires juristes. Les deux parties de ce document (travaux passés et perspectives) couvrent donc un large spectre qui va du droit aux applications pratiques en passant par la modélisation formelle.

## 2 Travaux passés

### 2.1 Vie privée, données personnelles et non discrimination : des droits complémentaires

La complexité et le caractère diffus des nouvelles formes de traitement de l'information conduisent à s'interroger non seulement sur leur impact en matière de protection de la vie privée mais également dans le contexte général de la prise de décision, en particulier de celles qui emportent des conséquences pour les personnes physiques. En effet, les opérations de plus en plus sophistiquées de recoupement, d'analyse, d'inférence sur les données personnelles permettent de profiler les personnes et de leur appliquer

des traitements individualisés. Au regard de ces évolutions, il nous a semblé utile d'examiner les conditions d'une meilleure articulation entre le régime juridique des discriminations prohibées par la loi avec le droit des données à caractère personnel et celui au respect de la vie privée.

On peut distinguer deux types de situations très différentes en matière de collecte de données personnelles :

1. Le recueil de données par des voies traditionnelles dans le cadre de procédures relativement formelles ou à l'occasion d'événements bien marqués et identifiables par les sujets.
2. Les collectes diffuses, imperceptibles ou d'apparence anodines et quasi-permanentes dans le monde numérique. Ces collectes peuvent concerner des informations fournies délibérément (mais dont le sujet n'a pas forcément conscience de la diffusion ou des usages ultérieurs), des informations invisibles pour le sujet (traces, enregistrements vidéo, etc.) ou même des informations inférées automatiquement (fouilles de données, recoupements, etc.), et que le sujet peut ignorer lui-même (par exemple sur les constantes de ses propres comportements).

Les lois de protection des données personnelles ont été conçues essentiellement pour répondre aux problèmes liés au premier type de collecte. Elles s'efforcent de s'adapter aux situations complexes posées par le second type de collecte mais se révèlent de plus en plus ineffectives dans ces cas.

La cause majeure de cette ineffectivité tient dans la philosophie de contrôle "a priori" qui inspire fortement ces lois. Le monde numérique repose sur la circulation et le traitement des données : la collecte d'information n'est plus un événement exceptionnel mais un phénomène permanent. De plus, les frontières s'estompent entre données personnelles et non personnelles, espace public et espace privé, collecte et traitement de données, ... Dans ce cadre, les contrôles préalables sont trop rigides ou simplement impossibles à mettre en oeuvre. Ainsi, des exigences qui peuvent représenter de véritables protections quand le recueil de données se fait par des voies traditionnelles deviennent des obligations de pure forme laissant en réalité les sujets complètement démunis dans le monde numérique. L'exigence du consentement préalable par exemple, se traduit souvent sur Internet par une acceptation machinale des internautes pressés d'accéder à un site ou un service.

Le régime de protection contre les discriminations présente certains avantages par rapport au régime de protection des données personnelles :

- Il est plus flexible car il ne repose pas sur des contrôles a priori (absence de déclaration ou consentement).
- D'origine civiliste, il privilégie les réparations en dommages et intérêts.
- Il prévoit des formes d'actions collectives, ce qui peut, dans une certaine mesure, favoriser un rééquilibrage des rapports de forces.

Les protections contre les discriminations définissent des règles précises aussi bien sur les types d'informations concernés (sexe, âge, religion, etc.) que les finalités prohibées (embauche, fourniture de service, etc.). Cette caractérisation précise constitue également la limite de ce régime : pour lui faire jouer le rôle suggéré ici - de protection contre les traitements illégitimes de données - il serait nécessaire de lever les restrictions actuelles sur ces règles. Un tel élargissement devrait bien entendu être effectué

avec précaution pour maintenir l'effectivité de la protection tout en l'étendant à toutes les discriminations "illégitimes" reposant sur l'usage et le traitement d'informations.

Une autre mesure nécessaire pour rendre véritablement réaliste la voie esquissée ici est le renforcement des moyens d'audit et d'investigation de l'autorité publique pour assurer que l'allègement des contrôles a priori est bien compensé par des contrôles a posteriori suffisamment dissuasifs. Ces moyens sont bien entendu d'ordre financier mais peuvent également reposer sur des solutions techniques permettant de rendre ces contrôles plus sûrs et efficaces.

Pour conclure, il va de soi que nos propositions de "déplacement du curseur" des contrôles (de l'a priori vers l'a posteriori) s'appliquent essentiellement aux situations de la deuxième catégorie (cf introduction) et ne doivent pas remettre en cause les obligations de déclarations, de demandes d'autorisation, ou de recueil du consentement pour la première catégorie où elles demeurent appropriées. Il est clair également que le droit des données personnelles qui, au-delà des droits individuels, vise à protéger un modèle de société démocratique [2], ne doit pas s'en trouver affaibli.

## 2.2 Outils logiques au service du droit : formalisation des normes

Le langage DLP (pour *Deontic Logic for Privacy*), conçu dans le cadre d'une thèse préparée au Laboratoire d'Informatique de Grenoble [8], fournit un cadre formel pour exprimer les différentes normes régulant les traitements sur les données personnelles. DLP est une logique déontique et temporelle s'appuyant sur un ensemble de prédicats orientés suivant six axes réglementaires : l'information de l'utilisateur, son consentement, sa capacité à accéder aux données et à les modifier, la justification de la collecte et du traitement, la conservation des données et enfin leur transmission à des tiers. Les aspects déontiques du langage permettent de construire des normes complexes, via l'expression d'obligations, d'interdictions ou éventuellement de permissions, mêlées à des concepts de nature temporelle. En particulier, des opérateurs dédiés sont proposés pour répondre les problèmes posés par les notions d'obligations avec échéance et d'interdictions maintenues sur une période [9]. Enfin, ce langage autorise la représentation de normes issues de sources multiples (lois, réglements, contrats, préférences...) et possiblement incohérentes entre elles. Il permet de détecter d'éventuels conflits entre ces sources et de les arbitrer en privilégiant les sources identifiées comme les plus importantes, afin de construire une poligique de sécurité cohérente et directement applicable [10].

## 2.3 Outils informatiques au service du droit : les agents de privacy

Une question cruciale en matière d'effectivité du droit est la réconciliation du principe juridique de "consentement explicite" du sujet (avant toute divulgation de données personnelles) avec le caractère de plus en plus continu et spontané (voire invisible) des communications électroniques. A cet effet, nous avons proposé une architecture reposant sur des agents logiciels attachés aux sujets et capables de décider, en leur nom, d'accepter ou de refuser la divulgation de données personnelles. Cette décision peut dépendre de nombreux facteurs, comme les engagements pris par le collecteur des données (finalité de la collecte, durée de rétention, conditions éventuelles de transferts

à des tiers, etc.), le type de données, le contexte (géographique, temporel, etc.). Le sujet peut exprimer ses desideratas dans un langage naturel restreint baptisé SIMPL. Le comportement attendu des agents est spécifié à l'aide de propriétés de conformité sur leurs traces d'exécution. Ces propriétés fournissent indirectement la sémantique du langage SIMPL et permettent ainsi d'éviter toute ambiguïté quant aux souhaits des sujets. Des agents logiciels sont également associés aux collecteurs de données et on a pu dériver, à partir des propriétés locales de conformité, des propriétés globales de correction du système [3]. La solution proposée a été conçue en tenant compte des résultats d'une analyse juridique du consentement conduite dans le cadre de l'action incitative PRIAM [7] qui nous a permis d'identifier les exigences techniques à respecter [6]. Cette solution présente des avantages significatifs par rapport à des propositions antérieures du style de P3P qui permettent d'exprimer de simples déclarations d'intention et comportent différentes sources d'ambiguïté.

Un simulateur expérimental de SIMPL a été réalisé en collaboration avec le projet AMAZONES (CITI de Lyon). Il permet de réaliser des échanges de données entre des agents autonomes et produit les traces résultant de leurs exécutions. Les agents ont été déployés sur des calculateurs légers et à bas coût (faible quantité de RAM, processeur RISC à basse fréquence). Ces calculateurs pourraient par exemple être déployés dans les guichets de services publics pour servir de "bornes d'échange de données personnelles".

Dans le même objectif de renforcer la maîtrise par le sujet de ses données personnelles, nous avons également collaboré avec le projet SMIS (INRIA Rocquencourt) pour définir des politiques de protection de données médicales [1]. Le modèle, qui repose sur des notions d'événements et d'épisodes (ensembles d'événements), permet au patient de spécifier les acteurs autorisés à agir sur des épisodes ou à observer les actions des autres acteurs. Il a été mis en oeuvre par l'équipe SMIS dans le contexte du projet pilote DMSP (Dossier Medico-Social Partagé) des Yvelines.

# 3 Perspectives

Les travaux passés, notamment sur l'effectivité du cadre juridique existant, ont montré la nécessité de compléter les contrôles a priori par des possibilités renforcées de contrôle a posteriori de l'usage des données personnelles. D'un point de vue logique, la notion centrale est celle d'obligation qui, même si elle a déjà été abordée dans certains travaux en sécurité informatique et en logique déontique, n'a pas jusqu'à présent reçu la même attention que celle de droit, classique en sécurité. De fait, on ne dispose pas actuellement d'un cadre formel satisfaisant permettant de décrire les obligations des responsables de traitement de données personnelles. Un tel cadre pourrait servir de base à des vérifications a priori (qu'un système permet d'assurer le respect de certaines contraintes), dynamiques (monitoring) ou a posteriori (audit). Il devrait également permettre de spécifier la répartition des obligations à l'intérieur d'une organisation et d'effectuer des vérifications sur cette répartition (cohérence, complétude, niveau de risque acceptable, etc.). D'un point de vue pratique, nous visons deux objectifs complémentaires : fournir des moyens de conception prenant en compte les exigences de protection des données personnelles et de leur contrôle ("privacy by design",

"accountability by design") et fournir des moyens d'évaluation des systèmes en regard de ces exigences.

## 3.1 Outils logiques pour la protection de la vie privée

Le contrôle a priori peut être abordé sous deux perspectives. Tout d'abord, un travail formel sur les normes elles-mêmes, sur les obligations règlementaires ou contractuelles et sur la conception des politiques de privacy est nécessaire pour faciliter la conception de systèmes respectueux des données personnelles. Les formalismes permettant d'inférer des responsabilités juridiques à partir de systèmes d'obligations complexes sont encore embryonnaires et nécessitent l'introduction des notions de sanction, mais également de réparation des dommages (concept encore peu ou pas traité en informatique). Un travail de fond semble également nécessaire sur les liens entre les obligations et sur leurs aspects dynamiques. En l'état actuel, les outils formels ne permettent pas forcément de modéliser les diverses formes de modification ou d'extinction des contrats. Le raisonnement qui s'ensuivrait devrait permettre de tisser des liens entre les aspects dynamiques des obligations, des responsabilités et des relations organisationnelles elles-mêmes (permettant de modéliser de manière réaliste la gestion des obligations contractuelles par les organisations).

Le contrôle a posteriori consiste en l'analyse des violations des politiques de protection des données personnelles, que ce soit suite à un défaut de conception du système, à une défaillance ou à une action malveillante. Le contrôle a posteriori est particulièrement pertinent en protection des données personnelles, à cause de la complexité spécifique du contrôle a priori et du caractère difficilement observable des violations (pour le sujet). Dans ces situations, il conviendrait de disposer des informations permettant de retracer et de caractériser les manquements, de manière à pouvoir partager des faisceaux de preuves. Il peut être encore plus intéressant de mettre en oeuvre de telles preuves de violation ne nécessitant pas la divulgation de nouvelles informations personnelles.

## 3.2 Méthodes formelles pour la conception de systèmes

On sait que, au-delà de la conception de techniques spécifiques de protection des données personnelles (les "Privacy Enhancing Technologies"), la démarche idéale consiste à intégrer les exigences de protection dans les phases de conception des systèmes eux-mêmes. L'expression "privacy by design" est devenue un slogan en vogue et certains systèmes ont effectivement été conçus dans cet esprit [4] mais cette démarche n'a pas encore été véritablement conceptualisée, systématisée et outillée. En effet, si un consensus s'est formé sur quelques principes généraux, aucune caractérisation précise n'en a été encore proposée. Certains de ces principes, comme la minimisation des données, la transparence, ou la responsabilisation sont pourtant susceptibles d'être définis de manière rigoureuse, voire formelle. D'un point de vue opérationnel, de telles définitions pourraient être utilisées pour élaborer une méthode systématique de conception "privacy friendly" ainsi que des critères d'évaluation de privacy. En d'autres termes, il reste à développer pour la privacy un corpus d'outils et de méthodes comparable à ce qui a été proposé dans le domaine de la sécurité, notamment pour la conception et la certification. Notons que ces méthodes doivent intégrer la complémentarité

entre ”contrôles a priori” et ”contrôles a posteriori” et couvrir aussi bien le ”privacy by design” stricto sensu que ce qu'on appelle parfois l'”accountability by design”. Cette dernière dimension, qui dépasse le cadre de la protection des données personnelles, commence aussi à être étudiée dans le cadre du génie logiciel, notamment pour établir les responsabilités en cas de défaillances [5].

### 3.3 Outils et méthodes au service du droit

Les outils logiques et les méthodes formelles évoqués ci-dessus fournissent un cadre conceptuel qui peut être mis à profit pour analyser la notion juridique de donnée personnelle et les protections offertes par la loi. La définition de donnée à caractère personnel au sens de la loi du 6 janvier 1978 modifiée en 2004 repose sur les possibilités d'identification des personnes, directement ou indirectement et par tous les moyens possibles. Il nous semble que le caractère très général de cette définition est d'autant plus justifié que les évolutions des technologies du numérique et de leurs usages tendent elles-mêmes à élargir constamment le périmètre des données ”personnelles ”. Cependant, vu le volume de plus en plus considérable de données concernées, la très grande généralité de la définition de la loi du 6 janvier 1978 rend également impossible son application stricte. Pour restaurer son effectivité, nous pensons qu'il convient d'adopter une démarche pragmatique reposant sur une analyse des risques et des bénéfices liés à la collecte et au traitement de données. En d'autres termes, il s'agit de troquer des règles rigides devenues impraticables contre une évaluation pragmatique des conséquences pour la personne.

Pour être applicable à grande échelle, une telle démarche doit reposer sur des principes généraux, une sorte de grille d'analyse. Cette grille doit prendre en compte un certain nombre de critères (Cf. partie précédente) qu'il s'agira ensuite d'apprécier au cas par cas. Il conviendra notamment de considérer deux critères fondamentaux :

- Les différents types d'utilisation possibles des données, y compris la probabilité d'une utilisation ” non conforme ”, c'est-à-dire sortant du cadre des finalités de la collecte. Cette probabilité dépend de facteurs comme le degré de confiance accordé au responsable de traitement et à la sécurité de son système informatique , le volume et le type d'informations déjà disponibles sur la personne auxquelles celui-ci pourrait avoir accès, le degré de difficulté d'un tel accès, l'intérêt économique des utilisations non conformes, etc.
- Les conséquences potentielles pour la personne. Ces conséquences peuvent être positives (service rendu ) ou négatives (notamment dans le cas d'utilisations non conformes). Les dommages potentiels pour la personne dépendent d'abord du type d'information concerné (directement ou indirectement, à travers des recoupements, inférences ou ” fouilles de données ”), de leur précision , du type et du nombre d'acteurs qui peuvent les utiliser, des relations et du pouvoir d'action de ces acteurs vis-à-vis de la personne concernée, etc.

On peut imaginer une grille générale et des grilles spécifiques (par type d'information ou secteur d'activité) qui serviraient de référentiels et permettraient l'adoption d'une politique à la fois cohérente et souple (adaptable aux cas particuliers). A chaque combinaison de critères serait associé un ensemble de mesures à prendre par le responsable de traitement (déclaration ou demande d'autorisation à la CNIL, mesures de

sécurité, de recueil du consentement, d'information, facilités offertes pour l'exercice par les personnes des droits prévus par la loi, etc.). Ces grilles d'évaluation contribueraient également à une meilleure appréciation de la situation par les citoyens et une meilleure connaissance des moyens d'exercer leurs droits.

# Références


[1] T. Allard, N. Anciaux, L. Bouganim, P. Pucheral, and R. Thion. Seamless access to healthcare folders with strong privacy guarantees. *Journal of Healthcare Delivery Reform Initiatives*, 2010. À paraître.

[2] J. Le Clainche. *L'adaptation du droit des données personnelles aux communications électroniques*. thèse de doctorat, Université Montpellier 1, Décembre 2008.

[3] D. Le Métayer. A formal privacy management framework. In *Formal Aspects in Security and Trust : 5th International Workshop, FAST 2008 Malaga*, pages 162–176, Berlin, Heidelberg, 2009. Springer-Verlag.

[4] D. Le Métayer. Privacy by design : a matter of choice. In *Conference on Computers, Privacy and Data Protection (CPDP)*. Springer Verlag, 2009. to appear.

[5] D. Le Métayer, M. Maarek, E. Mazza, and M.-L. P. et al. Liability in software engineering : overview of the LISE approach and application on a case study. In *International Conference on Software Engineering, ICSE'2010 (conference version of the research report with same title)*. ACM/IEEE, 2009. to appear.

[6] D. Le Métayer and S. Monteleone. Automated consent through privacy agents : Legal requirements and technical architecture. *Computer Law & Security Review*, 25(2) :136 – 144, 2009.

[7] D. Le Métayer, S. Monteleone, and J. Moret-Bailly. Les ressources du droit alliées aux moyens de la technologie : application à la protection des données personnelles. *Revue Lamy Droit de l'Immatériel (RLDI)*, 51, 2009.

[8] G. Piolle. *Agents utilisateurs pour la protection des données personnelles : modélisation logique et outils informatiques*. thèse de doctorat, Université Joseph Fourier - Grenoble I, Grenoble, France, Juin 2009.

[9] G. Piolle and Y. Demazeau. Obligations with deadlines and maintained interdictions in privacy regulation frameworks. In *8th IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT'08)*, pages 162–168, Sidney, Australia, Décembre 2008. IEEE Computer Society.

[10] G. Piolle and Y. Demazeau. Une logique pour raisonner sur la protection des données personnelles. In *16e congrès francophone AFRIF-AFIA sur la Reconnaissance de Formes et l'Intelligence Artificielle (RFIA'08)*, Amiens, France, January 2008. AFRIF-AFIA.