université de BORDEAUX

Les données de la recherche : Entre RGPD et anonymisation

François Pellegrini
Professeur, Université de Bordeaux
francois.pellegrini@labri.fr

Donnée

- L'information est un bien immatériel
  - N'a pas de « propriétaire » et ne se « vole » pas
- Différents régimes juridiques applicables aux différentes catégories de données
  - Droit d'auteur
    - Pour les œuvres de l'esprit exhibant une originalité
  - Droit des données à caractère personnel
    - Pour les données relatives à une personne physique
  - Secret des affaires
  - Etc.

université de BORDEAUX

Lois « Informatique & Libertés » (1)

- Créées en réaction au mésusage des données à caractère personnel par les États durant la première moitié du XX^{e} siècle
  - En France, loi « Informatique et Libertés » de 1978
- Création d'organes de contrôle indépendants de l'exécutif et des administrations
  - Modèle juridique original d'« Autorités administratives indépendantes »
    - Ne peuvent appartenir aux autres pouvoirs en vertu de la séparation des pouvoirs
  - En France, la CNIL
    - 18 commissaires et ~210 personnels

Lois « Informatique & Libertés » (2)

- Article 1 de la loi « Informatique & Libertés » :

  « L'informatique doit être au service de chaque citoyen. Son développement doit s'opérer dans le cadre de la coopération internationale. Elle ne doit porter atteinte ni à l'identité humaine, ni aux droits de l'homme, ni à la vie privée, ni aux libertés individuelles ou publiques.

  Toute personne dispose du droit de décider et de contrôler les usages qui sont faits des données à caractère personnel la concernant, dans les conditions fixées par la présente loi. »

- Principe d'« autodétermination informationnelle »

Donnée à caractère personnel

- Art. 4 RGPD :
  - « Toute information se rapportant à une personne physique identifiée ou identifiable [...] ; est réputée être une « personne physique identifiable » une personne physique qui peut être identifiée, directement ou indirectement, notamment par référence à un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne, ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale »

Donnée sensible

- Art. 9 RGPD :
  - « Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits. »

Critères de licéité

- Licéité appréciée selon les critères suivants :
  - Finalité : déterminée, explicite et légitime
    - Base légale
  - Responsable du traitement
  - Destinataires des données traitées
  - Durée de conservation
  - Mesures de sécurité de conservation
  - Exercice des droits des personnes : information, droit d'accès, de rectification, d'opposition, etc.
- Contrôles effectués :
  - Proportionnalité des moyens mis en œuvre

Périmètre du RGPD

- Le RGPD n'applique pas aux données :
  - Qui ne concernent pas des personnes physiques
    - P. ex. : données d'entreprises
    - Mais certaines données d'entreprises peuvent être des données personnelles
  - Concernant des personnes mais ne permettant pas de les identifier

Anonymisation et réidentification (1)

- L'anonymisation des données est une tâche délicate
  - Elle ne peut se réduire à la seule pseudonymisation
- Plusieurs types d'attaques permettent de réidentifier les personnes concernées par des données pseudonymisées
  - En particulier les attaques par corrélation

Anonymisation et réidentification (2)

- La portée des attaques par corrélation dépend de la granularité des données collectées
  - Au menu : spaghetti, macaroni ou semoule ?
- Le niveau de granularité détermine un compromis entre :
  - L'utilisabilité des données pour la finalité recherchée ou des finalités ultérieures
  - La protection des personnes contre la réidentification

Protection intégrée de la vie privée

- Nécessité de ne collecter que les données strictement nécessaires aux finalités
  - Principe de minimisation des données
- Prendre en compte la protection de la vie privée dès la conception des dispositifs (« *privacy by design* »)
  - Intégrer les mécanismes d'attaque dans l'analyse préalable à la collecte des données
- Gérer l'archivage et le partage ultérieur