Part 1 : NOTE DATA / OPEN DATA

La Data est aujourd'hui annoncée comme une source de richesses multiples, elle est valorisable en argent mais aussi en connaissance de l'Homme. Tout en globalisant et systématisant les personnes, elle se nourrie cependant de l'individualité de chacune d'entre elles pour exister. Quoi de plus valorisable que la connaissance parfaite et subtile de l'Homme dans ce qu'il est et ce qu'il fait.

Loin cependant de percevoir son réel potentiel, elle demeure à l'heure actuelle une « donne » assez mal connue qui interroge.

Les questions sont aussi variées que légitimes : Que faire de la DATA lorsque j'en possède ? Comment faire pour la valoriser ? Comment l'utiliser au service de la Société ? Comment la façonner pour améliorer mon service à la Société ? … Autant de questions dont le législateur s'est par ailleurs emparé.

La DATA est désormais une nécessité pour le développement économique et l'innovation. Néanmoins, elle n'est pas sans poser un certains nombres de questions fondamentales sur les libertés et les droits qu'elle est susceptible d'enfreindre de par sa récolte et son utilisation. Des questionnements dont s'empare aujourd'hui le juriste.

A- Le contexte : DATA / OPEN DATA

La DATA est une information brute mais précise qui est récoltée à l'occasion de la réalisation d'une action à laquelle participe une personne ou un service auquel elle accède. D'ordre privé, ces données peuvent être ouvertes au public dès lors qu'elles entrent dans le champ des informations publiques communicables au sens de la loi de 1978. On parle alors d'OPEN DATA.

Les conditions d'accès à l'Open DATA sont strictement prévues par la loi de 1978. A juste titre, car elles sont de nature à porter gravement atteintes aux libertés et droits fondamentaux des personnes. Un constat pourtant difficile à démontrer dès lors que l'on sait que l'OPEN DATA ne peut être constitué que de données anonymisées. Certes, cependant l'anonymisation ne peut que rarement être garantie. En effet, testées à l'aune des critères posés par le G29 (avis du groupe européen des autorités de protection du 10 avril 2014 sur les techniques d'anonymisation), à savoir l'individualisation (est-il toujours possible d'isoler l'individu ?), la corrélation (est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?) et l'inférence (peut-on déduire de l'information sur un individu ?), les techniques classiques d'anonymisation présentent, en effet, des risques de ré-identification. Qu'elles aient pour objet de transformer les données pour qu'elles ne se réfèrent plus à une personne réelle (« pseudonymisation » et « masquage ») ou de généraliser les données pour qu'elles ne soient plus spécifiques à une personne, mais communes à un ensemble de personnes (« agrégation »), aucune de ces techniques n'est infaillible. En définitive, tout dépend du niveau de sécurité des outils utilisés. Ainsi, plus le niveau d'agrégation est élevé et plus les outils de codage sont poussés, moins il y a de risque de ré-identification.

Surtout, la possibilité d'établir des liens entre les données anonymisées constitue la principale faiblesse de l'ensemble de ces dispositifs, puisqu'il suffit de croiser les informations pour qu'elles redeviennent identifiantes : à titre d'exemple, il a été prouvé que 89 % des patients d'un hôpital peuvent être identifiés à partir de leur code postal, du mois et de leur année de naissance, de leur sexe et du mois de sortie de l'hôpital en question (P.-L. Bras, A. Loth, Rapport sur la gouvernance et l'utilisation des données de santé, 2013, p. 27). De même, la technique du « carroyage » utilisée par l'INSEE, qui consistait à diviser le territoire en carrés de 200 mètres de côté pour calculer l'imposition moyenne des contribuables, a été abandonnée dès lors que les résultats mis en ligne permettaient d'identifier les contribuables résidant dans un espace faiblement peuplé, ne comptant qu'un seul foyer fiscal (G. Gorce, F. Pillet, L'open data et la protection de la vie privée, rapport d'information, Sénat, n° 469, avr. 2014, p. 49).

Le risque de ré-identification par recoupement de données, dans un contexte d'explosion quantitative – « Big Data » – et de diffusion sans limite, est ainsi avéré. Les fuites peuvent alors être lourdes de conséquences, pour les citoyens, bien sûr, mais aussi pour les administrations. Leur responsabilité pourrait en effet être engagée pour négligence dans l'anonymisation, par

Charles Cardine

Dirigeant marketing de SIH-SOLUTIONS

Lyon Area, France | Hospital & Health Care

EXPÉRIENCE

Directeur marketing de SIH-SOLUTIONS (2005 - Maintenant)
SIH-SOLUTIONS

COMPÉTENCES ET EXPERTISE

Healthcare

un citoyen victime d'une ré-identification rendue possible par une mise en open data mal maîtrisée.

Lina WILLIATTE-PELLITTERI
lwilliatte@williatte-avocats.fr

Avocat au Barreau de Lille – Cabinet WT
(www.williatte-avocats.fr)
Professeur de Droit à la Faculté de Droit de l'Université Catholique de Lille
Directrice du Master 2 Droit de la Santé et de la Responsabilité médicale
Membre du Bureau Exécutif et du Conseil d'Administration de la Société Française deTélémédecine