6.6 Big Data et problèmes liés à l'anonymisation des données

La notion de « Big Data » recouvre les grandes quantités de données provenant de diverses sources qui sont saisies avec une grande rapidité de traitement, enregistrées et mises à disposition à des fins d'évaluation et d'analyse pour une durée et des usages indéterminés. Si la législation fédérale et cantonale sur la protection des données règle le traitement de données personnelles, le Big Data confronte les systèmes actuels de protection des données à des défis de taille, dans la mesure où ils sont fondés sur un contrôle individuel et une souveraineté des données, c'est-à-dire le contrôle que chaque individu exerce sur les données qu'il souhaite faire connaitre. Ce d'autant plus que les séries de données d'origines diverses sont reliées les unes aux autres et peuvent être traitées rapidement par des logiciels automatisés capables d'apprendre (*machine learning* , *deep learning* , intelligence artificielle).

Jusqu'alors, l'utilisation de données anonymisées était considérée, tout particulièrement dans le domaine de la santé, comme la solution pour éviter des conflits juridiques, car elles ne sont plus considérées comme des données personnelles et ne tombent pas donc pas dans le champ d'application du droit de la protection des données. Jusqu'alors, l'utilisation de données anonymisées était considérée, tout particulièrement dans le domaine de la santé, comme la solution pour éviter des conflits juridiques, car les données anonymisées ne tombent pas sous le coup du droit de la protection des données. Or l' anonymisation constitue en soi un traitement de données au sens de la loi sur la protection des données. Anonymiser une série de données signifie qu'il n'est plus possible d'associer une donnée individuelle à une personne concrète ou seulement avec des moyens disproportionnés. Pour ce faire, les données sont modifiées, p.ex. par la suppression complète des caractéristiques dans une série de données, qui permettraient de remonter à la personne, ou par leur généralisation, p.ex. en indiquant des plages de valeurs. Pouvoir remonter à une personne dépend des caractéristiques présentes dans les données. Ce sont donc elles qui déterminent les moyens nécessaires pour pouvoir procéder à une désanonymisation (ou une ré-identification). Anonymiser des données génétiques, qui revêtent un caractère éminemment personnel, n'est possible que dans de rares cas, notamment en l'absence de données liées à la personne ou si celles-ci ne sont pas accessibles. La règle générale suivante s'applique : plus il existe de sources disponibles dont il est possible de tirer des informations contextuelles sur des personnes et qui sont susceptibles d'être utilisées pour procéder à un rapprochement, plus la probabilité d'une ré-identification et d'une désanonymisation est grande. Par ailleurs, il ne faut pas sous-estimer le risque de désanonymisation au vu des techniques d'analyse toujours plus performantes et des possibilités croissantes d'échanges de données. Compte tenu du Big Data Analytics, et en particulier du développement des algorithmes dotés de facultés d'auto-apprentissage, il existe une probabilité élevée et croissante que l'anonymisation irréversible de données ne soit, à terme, fondamentalement plus possible sur le plan technique. En conséquence, les experts et les professionnels avertissent expressément qu'une anonymisation absolument sûre du point de vue technique n'est pas possible.

Dernière mise à jour le 12.06.2026