Bien maîtriser le cadre juridique

Exploiter le potentiel du Big Data tout en respectant la loi nécessite une politique particulièrement rigoureuse de gestion des données. Explications.

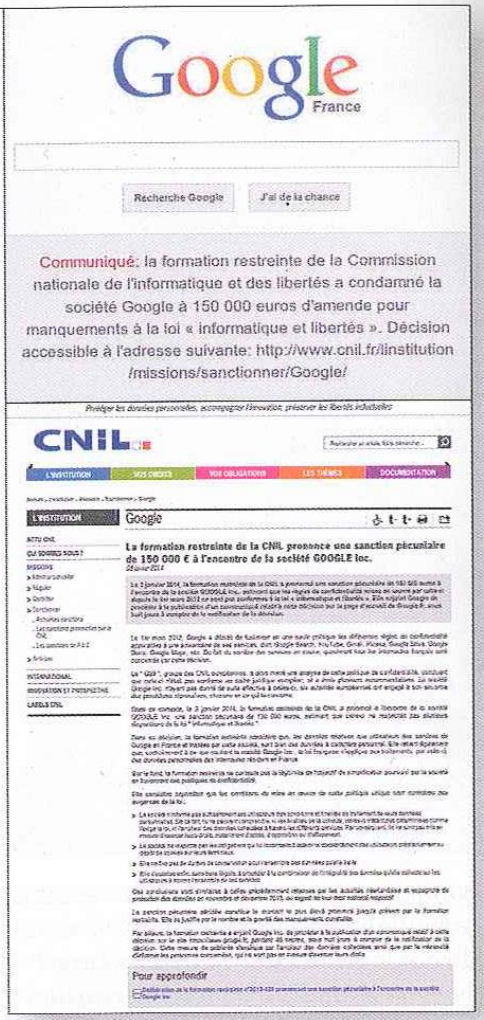
Conformément à la loi Informatique et Libertés du 6 janvier 1978, les données dont le caractère personnel est clairement établi doivent respecter trois grands principes : la finalité du traitement de données, le consentement de la personne et la limitation de durée. Cette loi prend une nouvelle dimension avec le Big Data. En atteste la récente sanction par la formation restreinte de la Cnil à l'encontre de Google, la condamnation mentionnant notamment un fait auquel toutes les entreprises risquent d'être confrontées avec le Big Data : « *elle* [NDLR : Google] *s'autorise enfin, sans base légale, à procéder à la combinaison de l'intégralité des données qu'elle collecte sur les utilisateurs à travers l'ensemble de ses services* ».

LA DÉFINITION DU BIG DATA REMET EN CAUSE LES PRINCIPES DE LA LOI

Des combinaisons dangereuses. Dans les faits, la définition même du Big Data remet en cause les principes de la loi, puisque l'idée est de stocker tout ce qui peut être capté sans présumer de l'usage ultérieur des données et sans limite de durée. Pire encore, elle induit des traitements qui, même lorsque la loi est respectée sur les données initiales, peuvent conduire à des interconnexions de fichier remettant en cause le droit à l'anonymat, à la transparence et à l'oubli. À titre d'exemple, une étude a prouvé que l'identité d'un utilisateur peut être reconstituée à partir de son code postal, de sa date de naissance et de son sexe, démontrant ainsi que la combinaison de quelques caractéristiques peut transformer des données anonymes en données personnelles.

Établir un cadre précis. Partant de ces principes, les entreprises doivent acquérir de nouveaux réflexes et renforcer leur politique de gestion des données personnelles. Dans cette perspective, tout projet de Big Data doit inclure un volet définissant clairement les règles d'utilisation des données et la finalité du projet de façon aussi large que possible afin de réduire les risques opérationnels. Parallèlement, les fichiers assujettis à la loi doivent être identifiés et nettoyés de toutes les données qui ne sont pas utiles à l'analyse. Par exemple, si l'on considère que la date de naissance n'est pas une donnée pertinente, autant l'enlever afin de réduire le risque de réidentification.

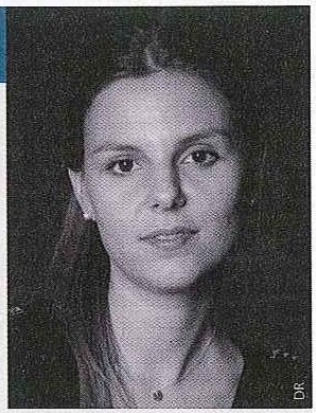
Enfin, les traitements de suppression automatiques des données personnelles doivent être renforcés. Objectif : éviter toute violation de la loi relative à la conservation des données en tenant compte des spécificités technologiques du Big Data. Une solution comme Hadoop favorise en effet la redondance, les données étant recopiées et dispersées sur des myriades de disques. En théorie, ces copies « techniques » échappent aux obligations de déclaration. Mais cette redondance peut provoquer des dérives, difficiles à anticiper aujourd'hui faute de recul.



Début février, Google a dû afficher pendant 48 heures sur sa page d'accueil sa condamnation par la Cnil.

ABIGAIL GOURION, AVOCAT À LA COUR

« Il convient d'encadrer les projets Big Data par des contrats avec les prestataires, les clients, les utilisateurs et les partenaires dans lesquels des clauses stipulent clairement les obligations et garanties de chacun ainsi que les obligations de surveillance et de confidentialité des données stockées. Par ailleurs, il faut anticiper les risques et prévoir la responsabilité applicable en cas de faille de sécurité, de vol, voire de divulgation des données. Les entreprises ont également tout intérêt à établir des règles encadrant le contrôle et l'exploitation des données et à mettre en place des outils de sécurité, notamment dans le cadre de la protection contre les attaques extérieures. Enfin, la démarche doit être complétée par des mesures de contrôle de la conformité des traitements, telles que des audits.


DR.