L’extraction et l’indexation de données par les crawlers sur internet – Point juridique

publié le **21/12/2016**, vu **9926 fois**, Auteur : Maître Matthieu Pacaud

La collecte automatisée de données sur internet est-elle légale ? Rapide analyse du statut juridique des robots d'indexation.

L’explosion du data mining et du big data pousse à s’intéresser à la légalité de la collecte automatisée de données.

Les données peuvent être collectées volontairement, au moyen de formulaires remplis par des utilisateurs, ou en obtenant le droit d’utiliser des bases de données. Elles peuvent également être collectées au moyen de robots - dits crawlers web - qui parcourent le web à la recherche de données pour les indexer et en permettre la consultation ultérieure par des tiers.

Qui est propriétaire des données collectées ? Est-il possible de collecter les données sans autorisation ?

Il s’agit d’un risque juridique à analyser pour toute société procédant à ce type de collecte.

La propriété et la protection des données des sites internet

Les données figurant sur un site internet appartiennent au propriétaire du site internet, à la personne l’ayant autorisée à le faire, ou à toute personne les ayant mises en ligne (dans le cas du contenu utilisateur).

Ces personnes sont titulaires de plusieurs droits de propriété intellectuelle sur leurs bases de données :

- Droits d’auteur sur leur propre contenu, sous réserve que celui-ci soit « original » ;
- Droit sui generis des producteurs de bases de données, sous réserve d’avoir fait un « investissement substantiel » pour la créer.

Le droit d’auteur protège l’expression du contenu de la base de données (les textes, photographies, etc) ainsi que sa mise en forme. A l’inverse, le droit sui generis protège les données figurant au sein de la base (leur structure, leur organisation logique, etc).

L’atteinte à ces données peut être qualifiée de contrefaçon de droit d’auteur ou du droit sui generis du producteur, selon le cas.

La mise à disposition de ces données sur internet par leurs propriétaires, en accès libre ou limité, n’a pas pour effet de limiter le degré de protection par le droit de la propriété intellectuelle.

Les sociétés procédant à la collecte des données doivent donc s’assurer de respecter les droits des tiers sur les données dès lors qu’elles pourraient rapidement se trouver dans une situation de

contrefaçon.

L’autorisation de la collecte des données du site internet

Il convient de différencier les données collectées pour indexation, de celles extraites pour réutilisation.

L’indexation des sites internet par un robot

L’indexation consulte les données accessibles sur internet afin de les cataloguer, ce qui peut entraîner des problématiques de droit d’auteur mais également de droit des bases de données.

Les données collectées pour indexation ne sont en réalité pas extraites de la base de données et font toujours référence à leur site d’origine. Il s’agit d’une simple prestation technique d’indexation, qui a un effet favorable sur le site indexé dès lors qu’il va lui permettre d’attirer du trafic. En conséquence, la jurisprudence indique qu’il ne s’agit pas d’une contrefaçon. Cela a été confirmé par la jurisprudence (voir par exemple décision du Tribunal de Grande Instance de Paris - Adenclassified du 1er février 2011).

En outre, chaque webmaster peut, via son fichier robot.txt, contrôler la manière dont les données de son site sont visitées par les crawlers, notamment en interdisant l’accès à certaines d’entre elles. La Cour d’Appel de Paris, dans un arrêt SAIF c/ Google du 26 janvier 2011, indique ainsi que les sites internet disposent de moyens pour s’opposer à l’indexation via ce fichier, ce qui permet de protéger leur propriété intellectuelle.

Il n’est donc, dans ce cas, pas nécessaire de demander une autorisation spécifique pour indexer les données.

Il est toutefois important de tenir également compte des conditions d’utilisation des sites concernés. Ceux-ci peuvent interdire explicitement l’extraction ou l’indexation de leur base de données, de manière contractuelle, au sein de leurs CGU. Le non-respect de ces conditions d’utilisation peut être sanctionné.

Il convient toutefois de différencier les données personnelles affichées sur ces sites (par exemple issues des réseaux sociaux), qui sont elles soumises à un régime d’autorisation spécifique, des autres données. Si les données sont indexées, conformément à la loi LCEN de 1978, il est systématiquement nécessaire d’obtenir l’autorisation de la personne concernée pour cet usage (par exemple via un outil d’opt-in lors de l’inscription sur un site).

L’extraction des données du site internet pour réutilisation

La collecte de données sans autorisation est interdite par les articles L342-1 et L342-2 du Code de la Propriété Intellectuelle, qui excluent l’extraction substantielle des données et la réutilisation de celles-ci.

Si l’extraction ou la réutilisation porte sur les éléments protégés par le droit sui generis du producteur, elle est constitutive d’un acte de contrefaçon.

Il en va de même vis-à-vis du droit d’auteur pour le contenu (textes, photographies, etc) ou la mise en forme de la base de données.

Pour extraire et réutiliser des données issues d’une base de données, il conviendra donc d’obtenir une autorisation préalable du producteur. Elle donne en général lieu à la conclusion d’un contrat

pouvant faire l’objet d’une rémunération, ces données pouvoir avoir une valeur économique non-négligeable.

A contrario, si l’extraction n’est pas substantielle (par exemple la recherche de mots spécifiques sur une page pour créer un corpus) et non répétée, il est possible de réutiliser les données. Il convient toutefois de s’assurer en amont de limiter l’extraction afin de réduire le risque. L’évaluation du caractère substantiel se fait au cas par cas par la jurisprudence.

En cas de doute, n’hésitez pas à nous contacter afin que nous puissions évaluer ensemble votre situation juridique.

Matthieu Pacaud

Avocat au Barreau de Paris

0609318009

contact@pacaud-avocat.fr

www.pacaud-avocat.fr