

[Chronique] Le « data scraping » légalisé aux Etats-Unis, la France va-t-elle suivre ?

par Eric Barbry

Peut-on extraire des données d'applications ou sites publics pour une utilisation tierce (et souvent, le business qui va avec) ? Les Etats-Unis ont tranché, contrairement à la France, rapporte notre chroniqueur.



Le « scraping » : voici un nouveau sujet d'attention et, pour une fois je vais rester neutre. Je me garderais bien d'un coup de chapeau ou d'un coup de gueule. Un vrai Suisse du droit !

C'est que je ne veux me fâcher avec personne... Ni mes clients qui défendent à juste titre le droit sur leurs contenus ; ni mes clients start-up « scrapeur » qui proposent des services toujours plus innovants.

Mais de quoi parle-t-on exactement ? Le « data scraping » est une pratique tout à fait familière des entreprises du web. Elle consiste à extraire des données provenant de sites internet publics ou applications mobiles en vue d'une tierce utilisation.

LinkedIn vs hiQ Labs : la justice a tranché

Très utilisée dans le monde du Web notamment pour des opérations de benchmark ou tout simplement pour proposer de nouvelles offres de service, la légalité du scraping questionne notamment au regard de la propriété de la data scrapée et/ou de la protection de certaines de ces données (données personnelles par exemple).

La justice américaine a finalement tranché clairement cette question à l'occasion d'un litige opposant depuis plusieurs années le réseau social LinkedIn de Microsoft, et la société hiQ Labs, start up américaine dont le modèle économique consiste à utiliser des données publiques pour analyser la fidélité des salariés au sein de leur entreprise.

LinkedIn avait en effet exigé de hiQ qu'elle cesse de scraper les données publiques de sa plateforme tout en commençant à lui bloquer l'accès à celle-ci.



HiQ a alors intenté une action en justice qui a été tranché une première fois par une cour d'appel fédérale en faveur de la société demanderesse. En effet, la 9ème chambre avait, dès 2019, déclaré que LinkedIn ne pouvait pas empêcher hiQ de faire du data scraping sur sa plateforme.

Quelque peu déçu de cette décision ne servant pas ses intérêts, LinkedIn a souhaité porter cette affaire devant la Cour Suprême des Etats-Unis qui a renvoyé l'affaire devant une autre Cour d'appel pour examiner à nouveau le litige notamment au regard de l'affaire Van Buren v. United States de juin 2021 selon laquelle la loi fédérale sur la fraude et les abus informatiques (CFAA) n'empêchait pas le scraping de données accessibles au public.

En effet, dans l'intérêt de sa défense, la société LinkedIn argumentait que la pratique du scraping violait notamment les principes de la loi sur la fraude et les abus informatiques (CFAA) ainsi que ses conditions d'utilisation.

Le 18 avril 2022 les juges américains ont sans doute mis un point final à cette querelle en confirmant la première décision et en confirmant le caractère licite du scraping.

Quelle est la posture de la Cnil en France ?

On peut s'interroger sur la portée d'une telle décision en France et plus généralement au regard du droit européen qui protège le droit d'auteur, les bases de données, les STAD et autres données personnelles.

Mais elle pose une question cruciale : les données publiquement accessibles sur un site web ou une application mobile peuvent-elle être extraites via des scripts pour en faire un autre usage ?

A cette question et pour ce qui est de la France, la Cnil a apporté une réponse formelle à travers une directive du 30 avril 2020 : « Ces données, bien que publiquement accessibles, sont des données personnelles. Dès lors, elles ne sont pas librement réutilisables par tout responsable de traitement et ne peuvent être réexploitées à l'insu de la personne concernée ».

Mais toutes les données scrapées ne sont pas nécessairement des données personnelles et même pour les données personnelles la question se pose dans la mesure où le scraping, enseigné dans toutes les bonnes formations web utilisent des outils qui ne sont traduisent pas nécessairement une collecte illicite ou déloyale qui seules sont sanctionnées pénalement.

Les « pro- base de données et droit des STAD » crieront au scandale ; les scrapers affirmeront que les données étant « publiques » et n'appartenant très souvent qu'à l'internaute lui-même, elles sont libres d'exploitation.

Suisse du droit je suis, Suisse du droit je resterai... Mais à vous de juger, donc.