

A step towards legalising web scraping under French law



Charles Bouffier*

Victoire Danès**

As new technologies develop rapidly, the law has to constantly catch up and change. The relatively new activity of web scraping has the potential to breach copyrights and other protections if not regulated carefully. The French law has now taken a step towards this, with an important exception for text and data mining.

Legal issues of web scraping

Web scraping refers to a set of techniques allowing a user to automatically extract the content of one or more websites, using a script or program, with the aim of transforming them to be used in another context.

This practice of extracting content from websites may constitute a violation of various intellectual property rights (e.g. copyrights, neighbouring rights, rights of database producers). Indeed, the free accessibility of content on the internet does not presume that content is not subject to protection under the law.

Web scraping can also constitute „data theft“ on the basis of Article 323-3 of the French Criminal Code if fraudulent intent is established. This article provides that:

„The fact of fraudulently introducing data into an automated processing system, extracting, holding, reproducing, transmitting, deleting or fraudulently modifying the data it contains is punishable by five years' imprisonment and a fine of €150,000.

When this offence has been committed against a system of automated processing of personal data implemented by the State, the penalty is increased to seven years' imprisonment and a fine of €300,000“.

Finally, web scraping can be punished on the basis of protection against unfair competition and free riding. An entity that uses this technique to feed the content of their website or to enrich their database will not have made the same efforts and/or investments as the owner of the website from which the data are extracted.

However, web scraping is increasingly used in practice, especially in connection with artificial intelligence (AI) and machine learning. Therefore, it has become necessary to adapt the law to this practice.

The new copyright exception

A new exception was created by Directive (EU) 2019/790, transposed into French law by Order No 2021-1518 of 24

November 2021, and by Decree No 2022-928 of 23 June 2022. This exception concerns text and data mining¹ that limits copyrights, neighbouring rights and the right of database producers.

Once the work has been disclosed, „digital copies or reproductions of a work for the purpose of text and data mining“² are therefore made legal.

Text and data mining are defined by the texts as: „the implementation of an automated analysis technique of text and data in digital form in order to extract information, including patterns, trends and correlations“³.

The „Conseil Supérieur de la Propriété Littéraire et Artistique“ (CSPLA) specifies the choice of this wording („text and data mining“) by „its broad meaning, including data extraction“⁴.

The extraction, copy and reproduction of data are thus covered by the exception.

The exception for text and data mining concerns copyright as well as neighbouring rights and the rights of database producers⁵.

The exception covers two distinct cases⁶:

- 1) digital copies and reproductions of works for the purpose of text and data mining for scientific research on the one hand, and
- 2) digital copies and reproductions of works for the purpose of text and data mining by any person or entity, regardless of the purpose, on the other hand.

The first case allows web scraping without authorisation from the authors „for the sole purpose of scientific research by research organisations, libraries accessible to the public, museums, archive services or institutions holding film, audiovisual or sound heritage, or on their behalf and at their request by other persons [or entities], including within the framework of a non-profit partnership with private actors“⁷.

Right holders cannot prevent a copy or digital reproduction of their works in this specific context. An appropriate level of

* Partner in IP/IT & Data Protection, at Racine Law firm.

** Lawyer in IP/IT & Data Protection, at Racine Law firm.

¹ Articles L122-5-10°; L122-5-3; L342-3-6°; R122-27 and R122-28 of the French Code of Intellectual Property (FCIP).

² Article L122-5-10° of the FCIP.

³ Article L122-5-3 para. 1 of the FCIP.

⁴ Mission Report „Transposition of text and data mining exceptions: issues and proposals“ by CSPLA, December 2020, p.14.

⁵ Article L122-5-3-III of the FCIP; Article L211-3-8° of the FCIP; Article L342-3-6° of the FCIP.

⁶ Article L122-5-3 of the FCIP.

⁷ Article L122-5-3-II of the FCIP.