

Kluwer Copyright Blog

Third European Court Decision on the General Purpose TDM Exception Is Out

Peter Mezei (University of Szeged, Hungary) · Thursday, May 8th, 2025

Regular readers of the Kluwer Copyright Blog may already be familiar with the excellent reviews of the first two rulings on the European Union's new text and data mining (TDM) exception – one from Germany (see the *Kneschke v. LAION* ruling [here](#), [here](#) and [here](#)) and one from the Netherlands (see the *DPG Media v. HowardsHome* ruling [here](#)). The third TDM ruling originates from Hungary. The judgment, issued on 3 December 2024, deserves close attention, as the Municipal Court of Appeals of Hungary had to determine – among other things – whether the scraping of the plaintiff's website by the leading global search engine, for the purposes of indexing relevant content and providing snippet views, falls within the general-purpose TDM exception under [Article 4 of the CDSM Directive](#).

Importantly, the case was not solely about TDM. The defendant presented five distinct arguments to exempt its activities from liability. First, it argued that the European Court of Justice's 'hyperlinking' case law – especially the 'new public theory' – applies to the indexing of press publications. Second, the defendant claimed its use fell within the exception to the press publishers' ancillary right [Article 15(1), last sentence, [CDSM Directive](#), transposed



Photo by [Tim Mossholder](#) on [Unsplash](#)

into Article 82/C point (b) of the Hungarian Copyright Act (HCA)]. Third, it contended that the plaintiff's reproduction right is merely ancillary and cannot be infringed without an act of making content available to the public – something that was claimed not to occur in this case, as the defendant's actions remained within the limits of the exceptions under Article 82/C HCA. Fourth, the defendant argued that either the temporary act of reproduction exception or the TDM exception should cover its scraping and indexing practices. Finally, it claimed that the plaintiff implicitly consented to its practices by not excluding the defendant's bots via the robots.txt file on the plaintiff's website.

The court of first instance was divided on the relevance of these five defenses. However, the Municipal Court of Appeals ultimately accepted the most significant one: the second argument. It concluded that the displayed snippets fell within the exception to the press publishers' right. More importantly for the purposes of this post, the Court of Appeals also overturned the trial court's rejection of the TDM-based defense and its ruling on scraping. (To be clear, the Court of Appeals did not accept the defendant's arguments regarding the new public theory, the ancillary nature of the reproduction right, or the defense based on temporary acts of reproduction.)

In paragraph 51 of the ruling, the court concluded:

“[i]t was not disputed that the defendant had lawful access to the plaintiff's press publications and did not circumvent the technical measures taken by the plaintiff to access them. Moreover, the robots exclusion protocol assigned to the plaintiff's website allowed all search robots, including the defendant's search robot, to crawl and index, and the plaintiff did not object to the indexing in the form required by law. The robots exclusion protocol is not of significance for consent, but primarily for the fact that it is—in accordance with the provisions of the Copyright Act—a machine-readable form by which the right holder could object to the text- and data-mining by the search engine. Since the defendant does not save the download of the page required for crawling, as it does not keep a copy of the page, and the data and information contained in the index are not copies, the statutory condition in Article 35/A(1)(c) [of the Copyright Act] is also satisfied.” (*Municipal Court of Appeals, Case 9.Pf.20.353/2024/6-II, 3 December 2024, para. [51]*). The Hungarian-language decision may be found on the [judiciary's website](#) using the keyword 20.353/2024.)

In short, the ruling declared that web scraping and search engine indexing constitute ‘a form’ of TDM. This conflation of scraping with TDM is not unprecedented. [Measures I.2.2, I.2.3. and I.2.4.](#) of the [Third Draft of the General Purpose AI Code of Practice](#) (see more [here](#)) take a similar position.

This interpretation appears to reflect the interests of the platform industry, however, and it has also sparked significant criticism (see [here](#)). This post takes the position that such a conclusion does not align with the overarching purpose and substance of the [CDSM Directive](#). Accordingly, I feel compelled to critique both the Hungarian judgment and the [Third Draft of the Code of Practice](#).

First, although the [CDSM Directive](#) defines TDM in a broad, technical sense (Article 2(6)), its *telos* was never to exempt all forms of automated data analysis from liability. The recitals of the directive make this clear: the “processing of large amounts of information with a view to gaining new knowledge and discovering new trends possible” may be carried out for research purposes under Article 3 of the [CDSM Directive](#), or for “government services, complex business decisions and the development of new applications or technologies” under Article 4 (see recitals 8 and 18). None of these objectives appear to encompass the decades-old practice of web scraping by search engines.

Recital 9 of the [CDSM Directive](#) also explicitly states that “there can also be instances of text and data mining that do not involve acts of reproduction or where the reproductions made fall under the mandatory exception for temporary acts of reproduction provided for in Article 5(1) of Directive 2001/29/EC,” in which case Article 4 of the [CDSM Directive](#) is not applicable. Moreover, Article 4(2) specifies that any reproduced or extracted information may only be retained for as long as necessary to carry out TDM—where the emphasis is clearly on the *mining*, not on any subsequent services the user might provide.

The ruling of the Municipal Court of Appeals, however, appears to legitimize the permanent storage of collected information for purposes beyond mere mining – namely, for indexing. This interpretation extends the scope of the general-purpose TDM exception in a way that seems unjustified.

Second, including web scraping under Article 4 risks violating the three-step test. The court’s ruling conflates distinct technological processes and subjects them to a single legal provision. Rather than clarifying the law, this approach introduces legal uncertainty and increases the complexity of interpreting and applying the [CDSM Directive](#).

The general-purpose TDM exception – outlined in Recital 18 – was crafted to comply with the three-step test. This is evident from the built-in safeguards of the directive: the lawful access requirement; the purpose-limited nature of the reproduction or extraction; the restriction on retaining collected data only as long as necessary for mining; and the right reservation under Article 4(3). (Of course, differences in implementation may exist across Member States. For example, the Czech transposition of Article 4 into Article 87b of the Czech Copyright Act does not explicitly include the ‘lawful access’ condition.)

However, applying Article 4 to web scraping – arguably the most essential underlying technology of internet browsing – extends the exception far beyond “certain special cases” as required by the first prong of the three-step test. Such an interpretation effectively *transforms the exception into a general rule*.

Furthermore, the third prong of the test – prohibiting “unreasonable prejudice to the legitimate interests of the rightsholders” – also weighs against the court’s interpretation, for at least two reasons.

First, while rightsholders may allow indexing of their content to facilitate discoverability, this does

not imply consent for broader commercial uses of their content. Although robots.txt can (albeit crudely) distinguish between basic scraping and more advanced forms of mining (see [Hanjo Hamann's paper](#) on this point), the defendant challenged only the plaintiff's failure to block scraping—not TDM—via robots.txt. The refusal by both the defendant and the court to differentiate between indexing and automated content analysis undermines the ability of rightsholders to make informed, nuanced decisions about the use of their content by different technologies. If scraping for purposes other than automated analysis were subsumed under the general-purpose TDM exception as well, rightsholders would face the unfair situation, where – either due to the lack of technical expertise or because of the ‘fear of missing out’ from indexing – they would tend not to exclude scraping of their websites. Also, as a consequence to it, they could effectively lose their expressly granted statutory right under Article 4(3) of the CDSM Directive to opt out of TDM – and, with it, their ability to authorize such uses.

Second, indexing has long been recognized as a core function of search engines and is already protected under the E-commerce Directive's safe harbour provisions (currently under the Digital Services Act). It is, therefore, not a practice generally subject to legal challenge. But this exemption should not be extended to the domain of TDM.

The author's manuscript on rights reservation, the AI Act, and the evolving TDM case law is available via SSRN [here](#). See further Martin Senftleben's comments on the right reservation prong of the TDM exception on the Kluwer Copyright Blog [here](#).

To make sure you do not miss out on regular updates from the Kluwer Copyright Blog, please subscribe [here](#).



2024 Future Ready Lawyer Survey Report

Legal innovation:
Seizing the
future or
falling behind?

[Download your free copy →](#)

 Wolters Kluwer

 Future Ready

LAWYER

This entry was posted on Thursday, May 8th, 2025 at 12:51 pm and is filed under [CDSM Directive](#), [Digital Single Market](#), [European Union](#), [Exceptions and Limitations](#), [Hungary](#), [Press Publishers ' Right](#), [Text and Data Mining \(TDM\)](#)

You can follow any responses to this entry through the [Comments \(RSS\)](#) feed. You can leave a response, or [trackback](#) from your own site.