The proportionality of Pay-Per-Crawl systems within the framework of article 3(3) of the DSM Directive

Security or barrier?



Information Law and Policy Lab

Amsterdam, 7 September 2026

This policy brief was written by Emma Arends, Simone Roelofsen and Barber Siegert of the Glushko & Samuelson Information Law and Policy Lab (ILP Lab) at the Institute for Information Law (IViR), University of Amsterdam, in collaboration with the Clinic Law Incubator.

The ILP Lab is a student-run, IViR-led initiative that develops and promotes research-based policy solutions to protect fundamental rights and freedoms in European information law.

This report was produced in the context of a project commissioned by COMMUNIA. The analysis, findings, recommendations and conclusions are those of the ILP Lab authors and do not necessarily reflect the views of COMMUNIA.

We are grateful to Paul Keller (Director of Policy, Open Future) and Kristofer Erickson (Deputy Director of CREATe at the University of Glasgow) for agreeing to be interviewed as experts for this research.

Table of contents

1. Introduction 3
2. Legal framework 5
2.1 The TDM-exception 5
2.2 Article 3(3) DSM Directive 5
3. Pay-per-crawl systems 7
3.1 What is Pay-per-Crawl? 7
3.2 Technical measures 7
3.2.1 Rate limiting 8
3.2.2 API-based paid access 8
3.2.3 Paywall access 9
4. Assessing Pay-per-Crawl measures under article 3(3) DSM Directive 10
4.1 Rate limiting 10
4.2 API-based paid access 12
4.3 Paywall access 13
5. Conclusions and Recommendations 16
5.1 Conclusions 16
5.2 Recommendations 17
Bibliography 20
Articles 20
Appendix 1 Interview Paul Keller summary 21
Appendix 2 Interview Kristofer Erickson summary 23

1. Introduction

The 2019 Directive of the European Parliament and the Council on Copyright and related rights in the Digital Single Market (‘’**DSM Directive**’’) introduces an updated regulatory framework aimed at harmonising the rules governing copyright and related rights across the European Union in the digital environment.[1] One of its key innovations is the mandatory text and data mining (”**TDM”**) exception in Article 3(1), which requires Member States to provide an exception to copyright and related rights for reproductions and extractions made by research organisations and cultural heritage institutions for the purposes of scientific research.

At the same time, Article 3(3) DSM Directive permits rightholders to apply measures that are necessary to ensure the security and integrity of the networks and databases on which their works or other subject matter are hosted. While this provision serves a legitimate objective, concerns have been raised that the implementation of such measures may impede the practical exercise of the TDM exception guaranteed under Article 3(1).

This policy paper examines the relationship between Articles 3(1) and 3(3) DSM Directive. In particular, it analyses whether the safeguards available to rightholders under Article 3(3) risk undermining the effectiveness of the mandatory TDM exception.

Against this background, the question arises whether pay-per-crawl practices can qualify as measures necessary to protect the security and integrity of networks and databases within the meaning of Article 3(3) DSM Directive. “Pay-per-crawl” refers to a technical and contractual access model in which operators of websites, databases, or digital repositories charge automated agents (typically web crawlers used for search indexing or AI training) a fee for accessing and retrieving content.

This paper explores whether three different types of pay-per-crawl mechanisms can be justified as measures necessary to ensure the security and integrity of networks and databases within the meaning of Article 3(3) DSM Directive. The analysis

[1] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC OJ L 130/92.

combines doctrinal legal research with empirical research, consisting of two semi-structured expert interviews. The findings are subsequently translated into policy recommendations.

Following this introduction, Section 2 sets out the legal framework of Article 3 DSM Directive and examines the scope and requirements of Article 3(3). Section 3 explains the concept of pay-per-crawl and identifies three types of pay-per-crawl mechanisms relevant to this analysis. Section 4 then assesses whether these mechanisms can be justified as measures aimed at protecting the security and integrity of networks and databases within the meaning of Article 3(3) DSM Directive. Finally, Section 5 concludes and formulates policy recommendations.

2. Legal framework

2.1 The TDM-exception

TDM refers to the automated computational analysis of large quantities of digital data, including text, images and other content, for the purpose of extracting information, identifying patterns and generating new knowledge.[2] The DSM Directive introduced two specific exceptions for TDM in Articles 3 and 4. Article 3 establishes a mandatory exception for scientific research conducted by research organisations and cultural heritage institutions.[3]

Under Article 3(1), Member States must provide an exception permitting reproductions and extractions of lawfully accessible works and other subject matter for the purposes of TDM. The exception applies only where the user has lawful access to the relevant materials. Pursuant to Article 7 DSM Directive, contractual provisions contrary to Article 3 are unenforceable. Furthermore, the exception is not subject to remuneration. The objective of Article 3 is to facilitate scientific research by reducing legal uncertainty surrounding the use of automated analytical techniques on copyright-protected materials.[4]

2.2 Article 3(3) DSM Directive

While Article 3(1) grants research organisations and cultural heritage institutions the possibility to carry out TDM activities, Article 3(3) recognises the interests of rightholders in protecting the technical infrastructure through which copyrighted content is made available. The provision states that rightholders may apply measures necessary to ensure the security and integrity of networks and databases hosting the relevant works or other subject matter.

The first requirement embedded in Article 3(3) is necessity. The use of the term “necessary” suggests that a measure must be objectively capable of addressing a genuine risk to the security or integrity of a network or database. Measures adopted merely to generate revenue, control market access, or restrict research activities are

---

2 Article 2(2) DSM Directive.
3 Rosati 2021, p. 41 en 42.
4 Hugenholtz, *Kluwer Copyright Blog* 2019. See also: Hugenholtz *AMI* 2019/5, p. 167-168.

more difficult to justify under this criterion. Moreover, where less restrictive alternatives are available that provide an equivalent level of protection, the necessity of the measure may be called into question.[5]

Secondly, the measure must pursue the specific objective of protecting the security and integrity of networks and databases. This requirement indicates that Article 3(3) is primarily concerned with technical and operational risks, such as system overload, excessive automated requests, unauthorised access, cybersecurity threats, or disruptions affecting the functioning of digital infrastructure. The provision does not expressly refer to the protection of economic interests or the monetisation of access to content.[6]

Thirdly, although not expressly mentioned in the text of Article 3(3), the principle of proportionality should inform its interpretation. As a general principle of EU law, proportionality requires that measures adopted by rightholders do not go beyond what is necessary to achieve their legitimate objective. Measures that effectively prevent or severely hinder the exercise of the mandatory TDM exception may therefore require particular scrutiny.[7] This is also the case if such measures hinder the effectiveness of the exception.

Finally, Article 3(3) must be interpreted in light of the overall objectives of the DSM Directive. Recitals 8 and 10 emphasise the importance of facilitating scientific research and fostering innovation within the Digital Single Market. Accordingly, Article 3(3) should not be interpreted in a manner that undermines the effectiveness of the mandatory TDM exception established in Article 3(1). The provision seeks to strike a balance between the protection of technical infrastructure and the promotion of scientific research, and its interpretation should reflect both objectives.[8]

---

5 Hugenholtz, *AMI* 2019/5, p. 170.
6 Rosati 2021, p. 58. See also: Senftleben 2026, p. 95-96.
7 Rosati 2021, p. 58.
8 Rosati 2021, p. 58.

3. Pay-per-crawl systems

3.1 What is Pay-per-Crawl?

A pay-per-crawl system refers to a model in which access to online content for automated data collection is conditioned on payment by the party conducting the crawling. Rather than allowing unrestricted automated access to websites or databases, such systems require entities that systematically collect data, such as search engines, aggregators, or developers of artificial intelligence (AI) models or systems, to compensate the content provider for each request or for a defined volume of crawling activity.[9]

The aim of a pay-per-crawl model is to create an economic framework in which large-scale data extraction can take place under controlled conditions while ensuring that right holders receive remuneration for the use of their content. In practice, the implementation of such a system typically relies on various technical measures that regulate and monitor automated access to online content. These measures may include mechanisms that control the frequency of requests, restrict access to certain content, or condition access on authentication or payment.

3.2 Technical measures

This section examines three technical measures that can be implemented within a pay-per-crawl system. It should be noted that rate limiting is not, strictly speaking, a pay-per-crawl mechanism, as it does not necessarily require payment for access. Nevertheless, because it is commonly used alongside monetised access models and similarly regulates automated crawling activity, it is treated as a pay-per-crawl mechanism for the purposes of this analysis and will be referred to as such throughout the remainder of this paper. Similarly, traditional paywalls do not specifically regulate crawling or TDM but access to content more generally. They are included here as a contrasting access-control mechanism in order to assess the limits of Article 3(3).

---

9 Archer, Ghili & Haghpanah 2026, p. 2.

3.2.1 Rate limiting

Rate limiting is a technical measure used by online services to restrict the number of requests that a user, device, or IP address can make to a server within a specified period of time. By imposing such limits, platforms can regulate traffic to their systems and prevent excessive or automated requests that could overload servers or facilitate large-scale data extraction. In practice, rate limiting is commonly implemented in application programming interfaces (”**APIs**’’) and websites by setting thresholds, for example allowing only a certain number of requests per minute. When these thresholds are exceeded, additional requests may be temporarily blocked, delayed, or rejected with an error response. Although primarily designed as a tool for maintaining system stability and preventing abuse, rate limiting can also function as a mechanism to control or discourage automated access to online content and databases.[10] It thus operates as a protective mechanism to mitigate Denial-of-Service (DoS) attacks and control automated web scraping activities.

3.2.2 API-based paid access

The technical base of an API is a software interface that enables different digital systems, such as devices, applications and data servers, to communicate and exchange information with one another. Many cloud services use an API gateway that sits in front of the API within the network architecture. This gateway acts as a single point of entry, where users are authenticated and identified. It also serves a security function by monitoring traffic and centralizing identity management.[11]

A pay-per-crawl system can be implemented by using an API as the access layer and an API gateway to authenticate users, monitor requests and apply usage-based pricing for each data retrieval. These API mechanisms allow right holders to restrict unauthorized crawling while offering controlled, paid access to content. In this way, technical protection measures and monetization strategies operate together to regulate and license data mining activities. APIs, particularly when combined with API gateways, may qualify as ‘appropriate technical measures’ under Article 3(3) DSM,

10 Todd, Morton, Kirby e.a. 2025, p. 1-2.
11 Qazi 2023, p. 2.

insofar as they restrict and control access to protected content and prevent unauthorized TDM. We return to this qualification in section 4.2 below.

3.2.3 Paywall access

A paywall is a system used by online content providers to restrict access to digital content to users who have fulfilled certain conditions, like a payment of a subscription fee. In this sense, a paywall functions as a form of access control, on whether a user is authorised to access protected content.

In its simplest form you have the hard paywall, where access to content is entirely blocked unless the user has entered into a contractual relationship with the provider. These paywalls do not primarily serve to protect the security or integrity of networks and databases, but rather to control access to content and facilitate its monetisation. Because paywalls primarily regulate access to content and commonly serve commercial objectives, their relationship with Article 3(3) raises distinct questions from rate limiting and API-based access. Their compatibility with that provision therefore depends on the purpose, design and effects of the particular measure, as discussed below.

4. Assessing Pay-per-Crawl measures under article 3(3) DSM Directive

4.1 Rate limiting

In principle, rate limiting may fall within the scope of Article 3(3) DSM Directive, as it is a technical mechanism that directly controls the frequency and volume of automated requests sent to a server. It operates by setting predefined thresholds on incoming traffic from a user, device, or IP address, and temporarily delaying, blocking, or rejecting requests once those thresholds are exceeded.[12] Because of this function, rate limiting can be understood as a tool aimed at protecting the operational stability of digital infrastructure and ensuring the availability of networks and databases.

The requirement of *necessity* must therefore be assessed in light of this technical function. Rate limiting is necessary within the meaning of Article 3(3) only where limiting request frequency is objectively required to prevent identifiable technical risks arising from automated or excessive traffic. Such risks include server overload caused by high-frequency crawling, DoS attacks, or degradation of system performance due to uncontrolled API usage. Where rate limiting is calibrated to address these risks based on system capacity and observed traffic patterns, it can be considered compatible with the necessity requirement. However, where thresholds are set without reference to actual system constraints or are deliberately restrictive beyond what is technically required, the necessity of the measure becomes questionable.

In terms of the *objective*, rate limiting is structurally linked to the protection of security and integrity because it functions as a traffic control mechanism designed to preserve system performance and prevent disruption. Its technical design is not primarily aimed at content regulation but at managing the load placed on infrastructure. On this basis, rate limiting aligns with Article 3(3) insofar as it is used to regulate machine-generated request intensity rather than to control the economic value of data access.[13]

---

12 Cloudflare on rate limiting: https://www.cloudflare.com/learning/bots/what-is-rate-limiting/.

13 Akande, Ako, Iyaniwura e.a. 2025, p. 1794.

However, the classification of rate limiting under Article 3(3) depends on how the threshold logic is implemented in practice. Rate limiting systems vary significantly: some apply static limits per time unit, while others use dynamic or adaptive algorithms that adjust thresholds based on behaviour, user type, or system load.[14] Where static limits are applied in a way that does not distinguish between ordinary use and large-scale TDM, they may disproportionately restrict automated research activities. In such cases, the technical design of rate limiting may extend beyond what is necessary for infrastructure protection and effectively interfere with lawful TDM under Article 3(1).

The principle of *proportionality* further requires that rate limiting be technically aligned with actual system needs. This means that thresholds must reflect realistic infrastructure constraints and should not be used as a proxy mechanism to limit access intensity beyond what is required for stability. Rate limiting that technically allows general access but functionally prevents scalable data extraction may therefore be disproportionate in its effect, even if it is formally presented as a security measure.

Rate limiting may also be integrated into commercial API structures where higher request volumes are offered through paid subscription tiers.[15] In such configurations, the same technical mechanism used for infrastructure protection is also used to differentiate access levels based on payment. This dual function transforms rate limiting from a purely protective control system into a pricing mechanism for data access. In such cases, it becomes difficult to justify the measure under Article 3(3), which is limited to technical safeguards ensuring security and integrity rather than mechanisms that structure access according to commercial value.

Overall, rate limiting is technically a request-throttling system that regulates automated access intensity at the infrastructure level. It can therefore fall within Article 3(3) DSM Directive only where its threshold design, implementation, and purpose remain strictly tied to genuine technical constraints of system security. Where it functions as a structural limitation on scalable data extraction or as a monetised access control

---

14 Lyu, Wang, Cheng e.a. 2025, p. 862.

15 Cloudflare on rate limiting: https://www.cloudflare.com/learning/bots/what-is-rate-limiting/.

mechanism, it risks exceeding the scope of Article 3(3) and undermining the effectiveness of the Article 3(1) TDM exception.[16]

4.2 API-based paid access

API-based paid access utilizes an API gateway as a centralized entry point for authentication, traffic monitoring and data delivery. Under Art. 3(3) DSM Directive, such technical measures can be justified as necessary when they serve to protect the integrity and security of the system from the significant resource costs of AI crawling.

The technical necessity of an API gateway often stems from the fundamental difference between human browsing and AI crawling. While human traffic is largely managed through global caching, where popular content is temporarily stored closer to the user, AI crawlers typically perform bulk data collection.[17] AI-scrapers often target bulk data from the long tail of a database, rarely accessed content that is not stored in global caches and thus requires substantial server energy and database operations to retrieve.[18] Both experts interviewed for this project identified high-volume automated access as a potential source of significant infrastructure demands. Consequently, an API can be justified as a necessary traffic management tool to prevent system overload.

The proportionality of charging for this access is frequently demonstrated through the ‘wiki search engine’ model, specifically ‘Wikimedia Enterprise’, frequently described as a specialized search engine for commercial AI developers.[19] This model enhances proportionality by offering additional value-adds, such as higher quality data and access to hidden metadata that is not visible to public crawlers, thereby framing the fee as a payment for premium service that sustains the underlying knowledge infrastructure.[20] By positioning the fee as a payment for a premium, high-efficiency service rather than a toll for the content itself, the model aims to sustain the knowledge infrastructure without violating the TDM exception.

16 Interview Paul Keller.
17 Interview Kristofer Erickson.
18 Interview Kristofer Erickson.
19 Interview Paul Keller.
20 Interview Kristofer Erickson.

However, the measure remains legally fragile if it is used for ‘rent-seeking’ purposes or purely commercial reasons, which Art. 3(3) explicitly prohibits.[21] Rightholders may be tempted to use cybersecurity as an excuse to implement aggressive technical protection measures simply to strengthen their contractual bargaining position. For research organizations, who benefit from a mandatory TDM exception under Art. 3(1), an API based pay-per-crawl system becomes disproportional if it creates a financial or technical barrier that makes research effectively impossible. Furthermore, API-based access functions as a platform-controlled affordance that often imposes strict rate limits and excludes critical metadata or historical content, potentially creating significant data gaps that undermine the ability of researchers to conduct comprehensive auditing or transparency research.[22]

Thus, while an API can proportionately manage the real resource costs of large-scale data extraction, it must not transition from a technical safeguard into a commercial tollbooth that prevents the exercise of mandatory research exceptions.

4.3 Paywall access

Paywall access operates fundamentally different from the other two pay-per-crawl mechanisms examined in this paper. A hard paywall does not regulate the intensity of access as rate limiting does, nor does it provide a structural technical interface for controlled data retrieval, as an API-based system does. Paywall access mechanisms condition any access on prior payment or contractual authorization, making it a binary access-control mechanism rather than a traffic management tool.

Article 3(3) DSM Directive requires that a measure is objectively required to address genuine technical risks to the security or integrity of a network or database. A hard paywall does not respond to identifiable technical threats such as server overload, DoS attacks or system degradation. As Paul Keller observed in his interview, paywalls primarily serve to control access to content and facilitate its monetization.[23] Their design is oriented towards enforcing a commercial access regime, not protecting infrastructure. Article 3(3) concerns technical and operational risks, not the protection

21 Interview Paul Keller.
22 Stalla-Bourdillon 2025, p. 16.
23 Interview Paul Keller.

of revenue streams.[24]. A conventional hard paywall imposed primarily to monetise access will generally be difficult to justify as ‘necessary’ under Article 3(3), insofar as it does not respond to an identifiable risk to the security or integrity of the relevant network or database. This does not exclude the possibility that particular access-control measures may also serve legitimate technical or operational purposes that require assessment in their specific context.

The objective requirement further supports this conclusion. Article 3(3) is limited to measures aimed at protecting the security and integrity of the networks and databases on which works are hosted. A paywall does not pursue this objective. As Kristofer Erickson noted in his interview, a paywall may raise the question of whether they are ‘at all even equivalent to a TPM’ [technical protection measure under Article 6 InfoSoc Directive], or simply ‘a legitimate business model that the content owner can apply’.[25] Measures that function primarily as commercial access controls, rather than as technical safeguards, fall outside the scope of Article 3(3).

The fact that a measure also pursues a commercial objective does not, by itself, make it illegitimate. Rightholders and service providers have legitimate economic interests, including interests connected to the freedom to conduct a business. The relevant question for Article 3(3), however, is whether the particular restriction can be justified as necessary and proportionate to the protection of the security and integrity of the network or database, while preserving the effectiveness of Article 3(1).

Lastly, there is a proportionality assessment. A hard paywall fails this test. Unlike rate limiting, a hard paywall blocks crawler access entirely. This means that research organizations with lawful access under Article 3(1) are completely prevented from exercising the TDM exception, not because of any genuine technical constraint, but because of a condition imposed by the rightholder. Such a measure goes well beyond what is necessary to protect the security and integrity of a network or database. This conclusion is reinforced by Article 7(2) DSM Directive, which explicitly removes the supremacy of online contracts over the scientific TDM exception, reflecting the

---

24 Rosati 2021, p. 58.

25 Interview Kristofer Erickson.

legislature's recognition that commercial access conditions must not override the mandatory research privilege.[26] This is directly contrary to the objectives of the DSM Directive, where the importance of facilitating scientific research is emphasized.[27] There is a risk that rightholders use security measures, not for genuine protection but to strengthen their commercial bargaining position by charging more for access that research organizations are legally entitled to.[28]

Overall, conventional paywalls whose primary function is to monetise access, rather than address identifiable risks to network security or database integrity, are unlikely to be justified solely on the basis of Article 3(3). The assessment should nevertheless remain context-specific and take account of the measure's purpose, technical function, effects on lawful TDM and the interests at stake.

26 Senftleben 2025, p. 12-13.
27 Rec. 8 and 10 DSM.
28 Interview Kristofer Erickson.

5. Conclusions and Recommendations

5.1 Conclusions

This paper examined whether three forms of pay-per-crawl mechanisms, such as rate limiting, API-based paid access, and paywall access, can be justified as measures necessary to ensure the security and integrity of networks and databases under Article 3(3) DSM Directive. The analysis demonstrates that the legality of such measures depends primarily on their purpose, necessity, and proportionality.

The findings indicate that rate limiting may fall within the scope of Article 3(3) where it is genuinely aimed at preventing technical risks such as server overload, denial-of-service attacks, or excessive automated traffic. However, where rate limits are calibrated in a way that effectively prevents lawful TDM activities, they risk undermining the mandatory exception established in Article 3(1).

Similarly, API-based paid access can in certain circumstances be justified as a legitimate traffic-management tool. The expert interviews provided examples and perspectives suggesting that large-scale AI scraping may impose substantial resource costs on content providers and knowledge infrastructures. Models such as Wikimedia Enterprise demonstrate that paid APIs can be used to manage these costs while providing additional services, such as higher-quality data and metadata. Nevertheless, API-based access becomes problematic where fees function primarily as a commercial barrier to lawful research rather than as a proportionate response to genuine infrastructure concerns.

In contrast, conventional paywall access is more difficult to justify under Article 3(3), particularly where its principal purpose is the monetisation of access rather than the protection of network security or database integrity. Unlike rate limiting and API-based access, paywalls are primarily designed to monetise content rather than protect network security or database integrity. Their effect is often to condition or prevent the exercise of the mandatory TDM exception, making them incompatible with the objectives of Article 3 DSM.

Overall, the research demonstrates that Article 3(3) should be interpreted narrowly. While rightholders may adopt measures to protect their technical infrastructure, these

measures must not be used to circumvent or undermine the effectiveness of the mandatory TDM exception. The core challenge lies in distinguishing legitimate cybersecurity measures from restrictions that primarily serve economic or contractual interests.

This distinction has broader implications for scientific research, innovation, and access to knowledge. If measures such as rate limiting and API-based paid access are interpreted too broadly as legitimate safeguards, rightholders may be able to impose technical or financial barriers that significantly hinder the exercise of the mandatory TDM exception under Article 3(1). This could reduce researchers' ability to conduct large-scale data analysis and undermine the DSM Directive's objective of promoting scientific research and innovation. Conversely, if Article 3(3) is interpreted too narrowly, content providers may face increasing infrastructure costs and cybersecurity risks associated with large-scale AI scraping. The challenge for policymakers is therefore to strike an appropriate balance between protecting digital infrastructures and ensuring that lawful TDM activities remain practically effective. As AI technologies continue to develop, maintaining this balance will become increasingly important.

5.2 Recommendations

The following recommendations are addressed primarily to the European Commission, EU and national regulators, policymakers, rightholders, research organisations, and cultural heritage institutions. They aim to clarify the conditions under which technical measures may be adopted under Article 3(3), to safeguard the effective exercise of the mandatory TDM exception in Article 3(1), and to ensure that legitimate security or infrastructure concerns are not used to undermine lawful research activities.

- **Adopt a narrow interpretation of Article 3(3)**

The Court of Justice of the European Union and national courts should interpret Article 3(3) narrowly, taking into account the mandatory nature and purpose of the TDM exception under Article 3(1). Measures should only qualify as "necessary" where they genuinely protect the security and integrity of networks and databases. Technical measures that primarily serve commercial, contractual, or access-control purposes should not fall within the scope of Article 3(3).

- **Clarify the meaning of "necessary" under Article 3(3)**

The European Commission should provide guidance on the interpretation of necessity and proportionality under Article 3(3). Such guidance should clarify how courts, rightholders, and users should distinguish legitimate cybersecurity measures from technical restrictions that effectively undermine lawful TDM activities.

- **Establish safeguards for research organisations**

Research organisations and cultural heritage institutions should be guaranteed practical access to content covered by Article 3(1). Technical measures should not make the exercise of the TDM exception impossible or burdensome. Where access restrictions are imposed, alternative access mechanisms should be easily available.

- **Encourage proportionate API-based solutions**

Where large-scale AI scraping creates substantial infrastructure costs, API-based access models may offer a balanced solution. However, fees should reflect actual resource costs and additional services provided, rather than functioning as a de facto licence fee for exercising a mandatory copyright exception.

- **Prevent the use of cybersecurity justifications for commercial purposes**

Regulators should closely scrutinise situations in which rightholders invoke cybersecurity concerns while primarily pursuing commercial objectives. Technical measures that serve as disguised access-control or revenue-generation mechanisms should not qualify under Article 3(3).

- **Monitor the impact of AI-driven scraping on knowledge infrastructures**

As AI training increasingly relies on large-scale data extraction, policymakers should continue evaluating whether existing legal frameworks adequately balance access to knowledge, scientific research, and the sustainability of public knowledge infrastructures. Future reforms should ensure that both innovation and the long-term viability of research and cultural institutions are protected.

Incorporate recommendations into best practices under Article 3(4) DSM Directive

Article 3(4) DSM Directive requires Member States to encourage rightholders, research organisations and cultural heritage institutions to agree on commonly accepted best practices concerning the application of the measures referred to in Article 3(3). The recommendations set out above should, to the extent possible, be incorporated into such best practices. In particular, best practices could specify the technical and procedural conditions under which measures such as rate limiting and API-based access are considered proportionate, clarify what constitutes a genuine security or integrity concern as distinct from a commercial access restriction, and establish minimum guarantees ensuring that research organisations can effectively exercise their rights under Article 3(1). Such best practices could also address the desirability of providing research organisations with dedicated access channels — such as whitelisting or API access at cost-reflective pricing — so that the exercise of the mandatory TDM exception is not impeded in practice. Developing best practices in this multi-stakeholder format offers a flexible and responsive mechanism to address the rapidly evolving technical landscape of automated data collection without requiring legislative intervention at each stage.

Bibliography

Articles

R. Archer, S. Ghili & N. Haghpanah, *Pay-Per-Crawl Pricing for AI: The LM-Tree Agent*, arXiv (2026).

J.O. Akande, A.M. Ako, A.A. Iyaniwura, S.L. Soleye & N. Ezra, *Reinforcement Learning for Dynamic Rate Limiting in API Security* (2025).

B. Hugenholtz, 'Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining', *AMI: Tijdschrift voor Auteurs-, Media- & Informatierecht* 43(5) (2019).

N. Lyu, Y. Wang, Z. Cheng, Q. Zhang & F. Chen, *Multi-Objective Adaptive Rate Limiting in Microservices Using Deep Reinforcement Learning* (AIIIP – 4th International Conference on Artificial Intelligence and Intelligent Information Processing 2025, ACM 2025).

F.A. Qazi, *Application Programming Interface (API) Security in Cloud Applications* (2023).

E. Rosati, *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790* (Oxford University Press 2021), chapter on Article 3 DSM Directive.

M. Senftleben, *Text and Data Mining, Generative AI, and the Copyright Three-Step Test*, *International Review of Intellectual Property and Competition Law* (2026).

M. Senftleben, K. Szkalej, C. Sganga & T. Margoni, *Towards a European Research Freedom Act: A Reform Agenda for Research Exceptions in the EU Copyright Acquis* (January 2025).

S. Stalla-Bourdillon, *Data Accessibility as a Platform Affordance: The Shaping of Research Possibilities on Online Platforms with Prospective Legal Methods* (December 2025).

P. Todd, M. Morton, H. Kirby & A. James, *Rate Limiting and Threat Detection in Intelligent API Gateways*, ResearchGate (2026).

Appendix 1 Interview Paul Keller summary

Paul Keller is the President of COMMUNIA and Director of Policy at Open Future. While Open Future focuses on research and policy development related to digitalisation, copyright, and access to knowledge, COMMUNIA is primarily engaged in advocacy work in Brussels. Its mission is to protect the public domain and promote free access to knowledge and culture.

According to Keller, copyright debates have changed significantly since COMMUNIA was founded. In the early years, discussions mainly focused on piracy, YouTube, and the transition from physical media to digital access. During the negotiations of the DSM Copyright Directive, most public attention was directed towards Article 13 (now Article 17) concerning upload filters. Looking back, however, Keller argues that Articles 3 and 4 on TDM have proven to be far more consequential, particularly due to the rise of artificial intelligence.

Keller explains that Article 3 was originally designed to enable scientific researchers to conduct automated analyses of works to which they already had lawful access. This was based on the principle that “the right to read is the right to mine.” Researchers frequently encountered contractual restrictions and technical barriers that prevented them from using computational methods to analyse large collections of texts and data.

With regard to Article 3(3) of the DSM Directive, Keller emphasises that the provision was not intended to grant rights holders broad control over research activities. Instead, it allows rights holders to implement measures necessary to protect the security and integrity of networks and databases, for example to prevent server overload or ensure system stability. However, such measures must not go beyond what is necessary and should not be used to effectively prevent lawful text and data mining activities.

Keller also discusses the distinction between Articles 3 and 4. Article 3 provides a strong and mandatory exception for research organisations and cultural heritage institutions, which cannot be overridden by contract. Article 4 has a broader scope and applies to other users as well, but it includes an opt-out mechanism that allows rights holders to reserve their rights through machine-readable means. As a result, Article 3 offers stronger protection for scientific research than Article 4.

Finally, Keller notes that the current debates surrounding generative AI demonstrate how forward-looking Articles 3 and 4 were. Although lawmakers could not foresee the precise technological developments that emerged years later, the provisions were intentionally drafted in technologically neutral language to accommodate future forms of automated analysis. At the same time, disagreements remain about whether AI training falls fully within the original intent of these exceptions.

Key Conclusion

Keller considers Articles 3 and 4 of the DSM Directive to be among the most significant provisions of European copyright law. In his view, Article 3(3) serves as a balancing mechanism: rights holders may protect their technical infrastructure, but they should not use security measures to obstruct lawful text and data mining or scientific research.

Appendix 2 Interview Kristofer Erickson summary

The interview focused on the relationship between copyright law, TDM, cybersecurity measures, and AI-driven web scraping. Kristofer Erickson highlighted that technological protection measures are no longer limited to traditional access controls but increasingly take the form of server-side security measures, such as rate limiting, IP restrictions, automated access blocks, and API-based controls. These measures are often implemented to manage server traffic and prevent abuse, but they can also restrict legitimate research activities and access to knowledge.

A central issue discussed was the difficulty of distinguishing between legitimate cybersecurity measures and restrictions that effectively undermine copyright exceptions. Erickson noted that Article 3(3) of the DSM Directive appears to acknowledge this tension by allowing measures necessary to protect the security and integrity of networks and databases, while at the same time seeking to preserve the effectiveness of the text and data mining exception. However, determining when a security measure is genuinely necessary and when it becomes excessive remains a significant challenge.

Drawing on research conducted with libraries, Erickson explained that institutions frequently encounter restrictions such as automatic access suspensions when too many users access the same resource simultaneously. While these measures may be justified as anti-scraping or anti-denial-of-service protections, they can also prevent legitimate users, including students and researchers, from accessing materials they are entitled to use. These restrictions are often reinforced through licensing agreements, creating an additional barrier to access.

The discussion also addressed API-based paid access, particularly through the example of Wikimedia Enterprise. Erickson explained that Wikimedia has experienced a substantial increase in AI-related crawling and data collection. In response, it created a commercial API service that offers AI developers access to higher-quality data, including metadata that is not publicly available through ordinary web crawling. According to Erickson, this model reflects an attempt to balance the costs of maintaining an open knowledge infrastructure with the growing demand for AI training data.

Regarding paywalls, Erickson questioned whether they should be regarded as TPMs at all. He suggested that paywalls often function as legitimate business models rather than copyright protection measures. Nevertheless, they can create practical problems when combined with cybersecurity restrictions such as geoblocking, IP-range limitations, or restrictions on remote access. As a result, users who have lawfully paid for access may still face significant obstacles when attempting to use content.

A recurring theme throughout the interview was the need to strike a balance between maintaining access to knowledge and ensuring the sustainability of knowledge infrastructures. Erickson argued that AI developers increasingly rely on large-scale scraping of content created by researchers, libraries, and public knowledge institutions. While AI systems generate value from this content, there is a risk that the underlying knowledge infrastructure receives insufficient compensation. He therefore suggested that mechanisms such as paid APIs, licensing arrangements, or other forms of remuneration may be necessary to ensure that creators and knowledge providers continue to be rewarded and that public knowledge resources remain sustainable.

Key Conclusion

Erickson concludes that the growing use of AI and large-scale scraping intensifies the tension between copyright exceptions, cybersecurity measures, and access to knowledge. While security measures may be necessary to protect infrastructure and manage costs, they should not be used to undermine legitimate uses permitted under copyright law. The main policy challenge is therefore to distinguish between proportionate security measures and restrictions that primarily serve commercial interests at the expense of researchers, libraries, and other lawful users.