Information Law and Policy Lab

Mine your own business

Reviewing the practical implications of the CDSM Directive's Text-and-Datamining-exception for Research Organisations and Cultural Heritage Institutions

Amsterdam, November 2025

This policy brief was written by Elinor de Bruijn, Mathijs Borchardt, Rifka Roos and Isabel Teeuwissen of the Glushko & Samuelson Information Law and Policy Lab (ILP Lab) at the Institute for Information Law (IViR), University of Amsterdam. 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.

The report was produced in the context of a project in collaboration with Stichting Internet Archive. It reflects the recommendations and conclusions of the ILP Lab authors. We are grateful for the input received from our supervisor, João Pedro Quintais (IViR).

Questionnaires were completed by, and expert interviews conducted with, the following individuals and representatives of various institutions: representatives of the Koninklijke Bibliotheek (National Library of the Netherlands), TU Delft, the Vrije Universiteit Amsterdam, Rijksuniversiteit Groningen, Universiteit van Amsterdam, and TU Eindhoven. Further contributions came from the University of Maastricht and Elsevier. In addition to institutional representatives, interviews were also held with Matthijs van Otegem, Paul Keller, and Martijn Davids of the Mediafederatie.

The research for this brief was concluded in July 2025.

Glossary

- **CDSM Directive –** 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

- **InfoSoc Directive –** Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society

- **EU –** European Union

- **TDM –** Text and data mining

- **ROs –** Research organisations

- **CHIs –** Cultural heritage institutions

The **Samuelson & Glushko Information Law and Policy Lab** develops and promotes research-based policy solutions that protect fundamental rights in the field of European information law.

Table of Contents

1. Introduction

The 2019 Directive of the European Parliament and the Council on Copyright and related rights in the Digital Single Market (CDSM Directive) creates an updated framework that harmonises the EU's rules on copyright and related rights in the digital age.

Its aim is to support the effective functioning of the internal market and ensure fair competition. It seeks to protect rightholders and enable fair use of works through harmonised rules at EU level. This is expected to promote innovation, creativity, investment, and the production of new content, including in digital contexts.

This policy paper presents an analysis of Article 3 CDSM Directive in light of the upcoming revision of the directive scheduled for 2026. The goal of this paper is to explore the positive and negative effects of the implementation of this exception, identifying its challenges and limitations as well as opportunities for improvement. Our analysis combines doctrinal research with empirical research, including perspectives from stakeholders, obtained through quantitative interviews. Our findings are then translated into policy recommendations.

After this introduction, section 2 sets out to define key concepts of our analysis. Section 3 then explains the regime and requirements of the exception in Article 3 CDSM Directive.

Building on the legal analysis, section 4 provides an overview of the outcome of the qualitative research that was conducted by interviewing relevant stakeholders. First, the stakeholders and their main objectives are identified, after which the main points of consensus and discord are discussed.

Due to their impact on stakeholders, two of the main problems identified in connection with the CDSM Directive's regime are given special attention: contractual obligations and the impact of AI on copyright. These problems are analysed in respectively section 5 and 6.

Section 7 concludes by providing recommendations for the revision of the CDSM Directive in this area. Section 8 concludes with our final remarks.

2. Key concepts

The benefits of Article 3 CDSM Directive are only applicable to a select few parties: research organisations (ROs) and cultural heritage institutions (CHIs). In addition, TDM under this provision may solely occur for the purpose of scientific research. It is therefore crucial to understand the meaning of these concepts. Before going into the specific actors, however, it is important to first define TDM-activities.

2.1 TDM-Activities

2.1.1 TDM and Copyright

TDM refers to automated techniques used to analyse digital content—such as text, sounds, images, or data—in order to extract useful information like trends and correlations. These technologies enable large-scale computational analysis, allowing for the discovery of new knowledge and patterns across vast datasets. TDM technologies are prevalent across the digital economy and can particularly benefit the research community and support innovation. TDM techniques are widely used by both private and public entities for various purposes, including government services, business decisions, and the development of new applications or technologies, in addition to their significance in scientific research. However, TDM also comes with copyright implications, because TDM-activities are often and to a large extent based on data that is copyright protected.

As discussed below, EU law provides exceptions and limitations for TDM to support scientific research and prevent internal market fragmentation. The main exception in this regard is found in Article 3 of the CDSM Directive. With TDM becoming more common, clear and harmonised rules are essential. Researchers must meet certain conditions, such as having lawful access to content. Legal uncertainty could weaken the EU's research competitiveness, highlighting the need for a unified legal framework.

2.1.2 Types of TDM

To understand the impact of TDM regulation, it is important to understand the scope of TDM-activities. After all, Article 3 of the CDSM Directive only applies if an activity falls within the scope of TDM.

Courts across Europe have affirmed the broad and growing relevance of TDM-activities, consistently recognising that many automated processes qualify as TDM under copyright law. The following section is a short overview of the first three rulings by different European national courts on the TDM-exceptions from the Directive. These decisions help illustrate the wide scope of copyright relevant TDM-activities.

In the LAION case, the Hamburg regional court found that scraping billions of image-text pairs from websites to train AI models constituted TDM, since it involved analysing material for "*patterns, relations, correlations and probabilities*". Crucially the court emphasised that even technical preprocessing steps in AI development, such as dataset curation, fall within the legal scope of TDM.[1]

Similarly, the Amsterdam District Court ruled in the DPG Media v HowardsHome case that the automated analysis of news content for recurring topics or trends, using publicly available data, involved extracting insights through automated means and thus constitutes a TDM-activity.[2]

Lastly, web scraping for the purpose of search engine indexing has been recognised to fall under the scope of copyright relevant TDM. A Hungarian Municipal Court of Appeals ruled that Google's web crawling and content analysis to generate indexed search results constitutes TDM as it needs the automated extraction and interpretation of textual data to uncover relevant connections. However, this decision is not without controversy. It has been argued that this interpretation of TDM-activities exceeds the CDSM Directive's intent and risks violating the three-step test.[3]

At least two conclusions can be drawn from these cases. First, that national courts consider web scraping activities to be a type of TDM. Second, despite some lack of clarity about how far this concept may reach, TDM-activities are increasingly relevant for a wide array of activities in today's digital economy.

[1] Goldstein, Stuetzle & Bischoff, 2024
[2] Valk & Toepoel, 2025,
[3] Mezei, 2025

2.2 Scientific Research

Both ROs and CHIs may rely on the TDM-exception of Article 3 CDSM Directive if if the purpose of such uses is scientific research. Recital 12 CDSM clarifies that this includes both human sciences as well as the natural sciences. Rosati defines scientific research within the boundaries of the CDSM as: *"any activity aimed at generating information that allows the uncovering of new knowledge or insights that are based on or characterized by the methods and principles of science"*[4], a formulation now widely accepted as the definition of scientific research for TDM use.

Crucially, the non-commercial nature of an institution affects only its entitlement to the exception, not the definition of the research itself. In other words, irrespective of whether the ultimate use of the findings is commercial, any endeavour which can be defined as scientific research will likely fall within Article 3's scope. However, the institution invoking the exception must not pursue those TDM-activities for commercial gain.[5] That means that all research, regardless of content or possible future uses may qualify, but those conducting the research must fulfil the other criteria listed in Article 3 CDSM Directive. The Hamburg court's decision in Kneschke v LAION further illustrates the scope of "scientific research". The Hamburg court held that preparatory steps for TDM research are covered by the exception, focusing on the non-commercial nature of the activity, regardless of who uses the data or the research outcome.[6]

2.3 Research Organisations (ROs)

Article 2(1) defines "research organisations" as typically non-profit or publicly motivated entities whose results are not preferentially accessed by influential undertakings. Recital 7 highlights their diversity and shared non-profit or public-interest purpose.

[4] Rosati, 2021, p. 43
[5] Stamatoudi & Torremans, 2021.
[6] Goldstein, 2024

2.3.1 Non-commercial and public character

Recital 12 defines research organisations broadly as entities whose main objective is to conduct scientific research, either by themselves or in together with educational institutions. These entities may have a different legal forms and structures, but all research must be undertaken on a non-commercial basis, either as no profit is to be made or as all revenue is to be reinvested into the scientific research or based on a public interest mission which is recognised by a Member State.[7]
It also clarifies that a public interest mission may, for example, be indicated by the presence of public funding, national legislation, or the existence of public service contracts. The recital explicitly states that institutions that are controlled by, or upon which commercial undertakings have significant influence, cannot be seen as a research organisation. What exactly constitutes significant influence is not specified. Thus, only those institutions which conduct scientific research without commercial interest can rely on Article 3 CDSM Directive.

2.3.2 Collaborations

Recital 11 encourages research organisations to form public-private partnerships, in line with EU research policy.[8] When an RO works with a third party that is not an RO, both can still rely on the exception in Article 3. The RO would still be the beneficiary. This broader scope reflects how research organisations often collaborate with others, relying on private partners' tools and resources. To support these partnerships, the Directive allows third parties to use the exception as long as the TDM-activities are necessary for non-profit research. However, the research must remain non-commercial, so anyone using the exception cannot be motivated by commercial interests. Also, the exception only applies if the partnership is led by the research organisation, meaning the non-commercial RO must keep decisive control.[9]

While these partnerships aim to boost collaboration and research, they create uncertainty about their scope. A research organisation may develop

[7] Quintais, 2025
[8] Ducato & Strowel, 2021
[9] Ducato & Strowel, 2021

a tool with a commercial partner and rely on the TDM-exception since the research is non-commercial. However, it is unclear how much the commercial partner can use the results, as the non-commercial RO enables the TDM-activities while the commercial party provides funds and tools. This mutual reliance makes it unclear who has the right to use the research outcomes.[10]

2.3.3 Limits of allowed activities

There are several limitations to the scope of the exception. These are detailed in the following section.

Firstly, TDM-activities must always stay within the scope of the scientific purposes and cannot be conducted outside non-commercial research. The focus is on the purpose of the research, not the researcher's conduct. Organisations allowed to conduct TDM under Article 3 cannot use it for other interests, such as administrative tasks. Researchers working independently or informally cannot rely on the exception, which also excludes activities such as scholarly reviews.[11]

Secondly, although both the European Commission and the Court of Justice of the European Union (CJEU) consistently stress the indispensable role of journalists in a healthy democracy, journalists remain excluded from the scope of Article 3. This exclusion could seriously hamper their work, as access to TDM would allow them to draw more accurate and nuanced conclusions from data sets.[12] The EU thus forfeits potential impacts that could be achieved through TDM-activities.[13]

Thirdly, by changing the wording of the Directive from "non-commercial" (as in InfoSoc Article 5(1)) to specifically "research organisations" and "cultural heritage institutions", the TDM-exception can never apply to small and medium-sized enterprises (SMEs), even if they carry out their activities for non-commercial purposes, unless they are collaborating with either of the approved organisations. They therefore lose out on a

[10] Chappell, 2015
[11] Ducato & Strowel, 2021
[12] EC Impact Assessment SWD, 2016
[13] Margoni & Kretschmer, 2022

competitive advantage other companies in other parts of the world have regarding TDM opportunities.

2.4 Cultural Heritage Institutions (CHIs)

Cultural Heritage Institutions (CHIs) are referred to in Recital 13 of the CDSM as *"publicly accessible libraries and museums regardless of the type of works or other subject matter that they hold in their permanent collections, as well as archives, film or audio heritage institutions. They should also be understood to include, inter alia, national libraries and national archives, and, as far as their archives and publicly accessible libraries are concerned, educational establishments, research organisations and public sector broadcasting organisations."*

CHIs have the authority to establish access conditions for their collections, as they own the physical premises and infrastructure. However, the scope of access they can grant is often constrained by the copyright restrictions imposed by the original copyright holders when works were added to the collection.[14] These limitations can include prohibitions on reproduction, distribution, and public display, which CHIs must adhere to, even when they possess the physical items. As a result, CHIs may face challenges in providing broader access or enabling certain uses of their collections without obtaining additional permissions or licenses from copyright holders.

Although CHIs primarily preserve and collect cultural artifacts, collection alone does not justify TDM-activities. All TDM-activities by CHIs must be part of non-commercial research. Practically, a CHI cannot perform TDM without securing proper licenses first.

Lastly, special attention must be drawn to the collaboration between CHIs and commercial entities. When institutions digitise their collections, they often collaborate with commercial entities to supply the technology and funding needed to turn physical items into digital files. However, this may

[14] Janssens, 2020

interfere with the licensing for publication, distribution, and exceptions to these publication and distribution rights.[15] To illustrate this problem, the British Library—Google Library Project can be seen as an example. Under their agreement, Google was granted permission to scan large portions of the Library's public-domain holdings. Although the underlying works had entered the public domain (as their copyright had expired), the digitised reproductions themselves became subject to licensing by Google. [16] Since Google now controls and provides access to these files, it's unclear whether either Google or the British Library (if it were still under EU law) could use the EU's non-commercial TDM-exception.

3. The scientific TDM-exception and the concept of lawful access

The previous sections laid down some criteria for applicability of Article 3. This section explains the basic regime of the exception and explores its most crucial requirement, i.e "lawful access".

3.1. The basic regime

3.1.1 Introducing Article 3

Article 3 CDSM Directive provides an exception for ROs and CHIs to carry out TDM-activities if these activities are administered solely for the purpose of scientific research. Within this scope, such institutions are allowed to extract, copy, and represent information, works or data that is copyright protected for the sake of conducting scientific research, as long as they have obtained lawful access to the material they will be utilising.

Article 3(2) explains that copies made under the TDM-exception must be stored securely and may be retained for scientific purposes including to verify research findings. This means that the exception does not extend to the publication of the original protected content. While the exception

[15] Arends, et.al., 2024
[16] Cruiz, 2011

regards the copying and the analysing of the works, it does not allow the publication of copyrighted works.[17]

Article 3(3) allows rightholders to ensure the "*security and integrity of the networks and databases where the works or other subject matter are hosted*". These security and integrity measures must be limited to what is necessary to achieve their purpose, while still allowing rightholders to restrict lawful access, provided they do not hinder the effective use of the exception.

Article 3(4) encourages stakeholders to develop best practices for storage and security, but Article 3 remains effective even in the absence us such measures.

3.1.2. National Implementation

Article 3 of the Directive is mandatory and had to be transposed by Member States by June 7, 2021. While it sets a minimum standard, Article 25 of the Directive allows Member States to maintain or adopt broader TDM exceptions, as long as they comply with the InfoSoc and Database Directives. For instance, national exceptions based on Article 5(3)(a) of the InfoSoc Directive (for teaching or scientific research) are permitted, though they are not fully harmonised at EU level and may not cover all rights. Additionally, Recital 15 CDSM Directive permits national arrangements, developed with stakeholder input, provided they are proportionate and ensure secure, authorised use only.

3.2 The “lawful access” requirement

3.2.1 Lawful access

The concept of “lawful access” as referenced in Article 3(1) CDSM Directive plays a crucial role in defining the scope of the TDM exception. However, the term itself is not explicitly defined within the Directive and lacks consistent usage across EU copyright law, where related expressions such as “lawful user,” “legal access” and “lawful acquirer” are employed with

[17] Article 3(2) CDSM Directive

varying meanings. This ambiguity creates legal uncertainty regarding what precisely constitutes lawful access. The Directive's recitals offer guidance on how lawful access should be understood in this context. This section examines these interpretative routes.

Recital 14 describes lawful access as "*covering access to content based on an open access policy or through contractual arrangements between rightholders and research organisations or cultural heritage institutions, such as subscriptions, or through other lawful means. (…) Lawful access should also cover access to content that is freely available online*". This implies several routes to fulfil the requirement: through contractual arrangements with rightholders, open access policies, content that is freely available online and any other lawful means of access.

3.2.2 Access based on open access policy

The EU's 2016 'Background note on open access to scientific publications and open research data' defines open access as *"the practice of providing on-line access to scientific information that is free of charge to the user and that is re-usable"*. Between 2010 and 2022, the share of open access publications in the EU rose from 56.8% in 2010 to 77.7%, reflecting the growing influence of open access policy.[18]

By eliminating cost and licensing barriers, open access policies expand the range of materials that can be mined without prior authorisation, enhancing the practical utility of the exception. This also contributes to a more equitable research environment, particularly for institutions or researchers with limited access to subscription-based databases.

3.2.3 Freely available content

The concept of "freely available online" content has been relevant for copyright related discussions for some time resulting in extensive jurisprudence of the CJEU when interpreting the exceptions to the right of reproduction and the scope of the right of communication to the public.[19] Content is generally considered freely available online if it was

[18] Belderbos, 2025

[19] E.g. in cases C-435/12, *ACI Adam and others v Stichting Thuiskopie*, ECLI:EU:C:2014:254, C-466/12 *Svensson and others v Retriever Sverige AB*, ECLI:EU:C:2014:76, C-160/15 *GS Media bv v*

uploaded with the rightholder's consent, is accessible to anyone on the internet, and does not require bypassing technical restrictions.

The main challenge in this case is determining whether the uploading of content was authorised by the rightholder. Since verifying this individually is impractical, the general rule is that content not clearly uploaded illegally can be accessed in good faith, unless the user is a commercial entity. For example, using articles from a site that appears affiliated with rightholders and presents itself as a scientific platform is generally acceptable, while accessing content from known piracy sites is not. Many cases fall into a grey area where authorisation is unclear. However, it would be unpracticable to require ROs and CHIs to employ lawyers to check all their sources before engaging in TDM-activities, especially considering the potentially automated and large-scale nature of the access (e.g. web scraping the open web for scientific research purposes).

For these reasons the most workable approach to assessing the legality of freely accessible online content, for the purpose of research TDM-activities, would be to consider that as long as the content is not obviously or manifestly illegal, ROs and CHIs should be able to perform TDM-activities.

3.2.4 Other lawful means

Recital 14 leaves a lot open to interpretation by including "*or through other lawful means*". It is unclear what this catch-all category entails concretely – works in the public domain are not in copyright (anymore), so any use of them is by definition lawful. [20] Beyond this, as this category is not clarified further, one must consider the general rule applicable to exceptions on copyright in EU-law based on the three-step test as seen in Article 5 InfoSoc.

Sanoma Media Netherlands BV, ECLI:EU:C:2016:644 and C-527/15 *Stichting Brein v Jack Frederik Wullens*, ECLI:EU:C:2017:300.

[20] Article 1 of the Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights (consolidated version 31/10/2011).

3.2.5 Limitation of the scope of lawful access through security and integrity measures

Article 3(3) lets rightholders take necessary security and integrity measures, as long as these do not obstruct the effective use of the exception or go beyond what is needed to meet their purpose.[21]

Only a select type of measures suitable to this aim are legally protected from circumvention[22] and thus have the power to limit this scope.[23] These are the so-called "technological protection measures" (TPMs) defined in Article 6(3) InfoSoc Directive as: "*any technology, tool, or component intended to stop or limit the use of copyrighted works or other protected material without the permission of the rightholder as long as they allow rightholders to control access or use of the material, successfully protecting the content*". For example, a paywall to access an online service would likely fall under this definition but a machine readable rights reservation expressed via robots.txt would not, as it does not allow the rightholder to effectively control the access of the content. It is also the case that the rights reservation mechanism under Article 4(3) cannot "*affect the application of Article 3 of this Directive*".[24]

These TPMs are applicable in all the aforementioned scenarios of access. For example, they can be used to enforce contractual terms of access if the contract would stipulate that access is confined to a specific country by implementing geo-blocking measures. On the other hand, a paywall is an example of a TPM that affects and shapes the conditions of access to content that would otherwise be freely available online. Circumventing any of these measures would automatically result in the deterioration of the legality of the access as circumvention of measures that qualify as TPMs under Article 6(3) InfoSoc and would be in violation of the legal protection afforded under Article 6(1) InfoSoc.

In conclusion, the requirement of lawful access if fulfilled through a myriad of means but can be halted through some protective measures.

[21] Recital 16 CDSM Directive

[22] Based on Article 7 CDSM Directive no contractual circumvention of Article 3 CDSM Directive is allowed.

[23] Szkalej, 2025, p. 315

[24] Article 4(4) CDSM Directive

3.3. Contractual Obligations

3.3.1 Contractual Arrangements

The recitals mention two concrete examples of contractual arrangements that give ROs and CHIs lawful access to content: subscriptions and licenses. Furthermore, ROs and CHIs are protected by the prohibition to exclude TDM-activities from contractual arrangements through Article 7 CDSM Directive.

While such arrangements give rightholders room to shape access conditions, Article 7 offers strong protection against contracts that attempt to exclude TDM for research purposes. Still, the Directive leaves unaddressed the possible economic consequences. If, for example, TDM rights are factored into subscription pricing, it could create disparities between well-resourced and smaller institutions. For publishers, increasing prices to reflect TDM's added value might seem reasonable, but this risks undermining the Directive's goal of promoting cross-border research collaboration. Member States should, therefore, not provide for compensation for rightholders as regards uses under the text and data mining exceptions introduced by this Directive.

This creates a significant problem for ROs and CHIs who display a tendency toward overcompliance due to risk aversion. Faced with complex contractual language or ambiguous legal safeguards, these users may choose to[25] challenge potentially unlawful restrictions, thereby nullifying the utility of Article 7(1). Moreover, rightholders can circumvent Article 7(1) by simply omitting explicit references to TDM in their contracts and instead defining access conditions that render the exception ineffective, as discussed in section 3.2.5.

In conclusion, while Article 7 CDSM Directive certainly has use in safeguarding the ability of ROs and CHIs to make use of Article 3 CDSM Directive, it is not a perfect solution.

[25] Szkalej, 2025

Part two: Empirical study

4. Stakeholder interviews

Building on the previous legal analysis, this section presents the empirical research that we conducted for the purpose of this policy brief. Firstly, the method and parameters of the research will be explained. Secondly, our findings will be presented in a concretised overview, grouping the main points of consensus and discord across various stakeholder groups. Lastly, these findings will be summarised and translated into policy recommendations from which we will draw our conclusions.

4.1 Research method

The research conducted consists of qualitative interviews. Guided by a diversity criterion we interviewed different types of participants that relate in various ways to the TDM-exception, offering different perspectives on how this provision shapes practice. To get a general overview of the TDM-exception in practice, the following questions were asked in the context of semi-structured interviews:

1. In what ways are you working with text and data mining (TDM)?
2. Have you noticed any effects from Article 3 CDSM Directive?
3. Do you have strong opinions about possible exceptions to copyright for research organisations (ROs) or cultural heritage institutions (CHIs)?
4. How do you interpret Article 3 CDSM Directive?
5. What are some challenges or issues you encounter regarding TDM-practices?
6. What improvements would you like to see?
7. In an ideal, utopian world, what would the playing field look like for you regarding TDM's legal landscape?

The main goal of these questions was to identify the successes and shortcomings of Article 3 in practice. Moreover, each interviewee was asked additional questions based on their individual areas of expertise.

4.2 Stakeholders involved

The landscape of TDM is populated by a diverse array of stakeholders, each with distinct interests and varying levels of knowledge. Navigating these differing perspectives is crucial for understanding the ongoing policy debates surrounding TDM. The following stakeholder groups can be identified and classified according to the interviews: research organisations and university libraries; cultural heritage institutions; publishers; and policy experts. The following subsection elaborates on the various stakeholders and their general interests in TDM more broadly, as well as the application of the Article 3 exception in practice.

4.2.1 ROs and ULs

The research organisations and university libraries interviewed include the Dutch Royal Library and Matthijs van Otegemen from the University of Utrecht. A questionnaire was also sent to various ROs, with responses from several Dutch universities.

These institutions are actively involved in TDM for scientific research. Their main interest is to secure easy, cost-free access to copyrighted content for publicly funded research, to foster scientific progress and ensure reproducibility. This includes the ability to archive downloaded content, specifically to ensure academic integrity, which often clashes with publisher terms.

They are aware that the exception of Article 3 CDSM Directive requires lawful access but are unclear on its limits. According to them, publishers often introduce restrictions or extra costs for TDM, such as download limits or service blocks, even where access is legally granted. There is little clarity on what “lawful access” means in practice, leaving these institutions and their researchers unsure of their legal boundaries. This causes ROs and researchers to remain hesitant to use TDM due to legal and technical barriers. Based on the stakeholder views, there is limited insight into the figures on how many researchers conduct TDM and use the TDM exception to its full potential. Some institutions, like TU Delft, note that their infrastructure is not yet suited for large-scale TDM.

Their key priority is to see clearer legal protections and rules, possibly supported by universities retaining publishing rights. Above all, they seek the removal of obstacles that prevent TDM for legitimate scientific research.

4.2.2 CHIs

Among the relevant parties that were interviewed for CHIs' perspectives are the National Library of the Netherlands (Koninklijke Bibliotheek, hereafter KB) and Matthijs van Otegem. Although Matthijs van Otegem is listed in the previous chapter on ROs in his current role as director of the university library, his perspectives are relevant to the CHIs chapter, based on his previous experiences in libraries and as a member of the OCLC Global Council.

These parties state that they do not conduct TDM activities themselves but instead play a facilitative role by collecting, preserving, digitising and providing access to existing materials for others. For example, the KB said to have recently developed a dataset of digitised newspapers freely available without IP restrictions, thanks to agreements with rightholders. Generally, though, such databases are readily copyright protected and therefore pose implications for the TDM exception.[26]

Additionally, CHIs recognise that privacy concerns justify the need for contracts, even when data is provided. For example, collecting newspaper articles may raise privacy concerns because the texts contain private information of the subjects of the news reports.

The KB also noted that it was clear to them that researchers who want to use their datasets for research through TDM are generally unfamiliar with the exception, despite it being specifically designed for them. It experiences ambiguities regarding the notion of “lawful access” and notes that it is uncertain whether or not ROs can invoke the TDM-exception against intermediaries (like CHIs) who provide access to content on the basis of a licence, given that the TDM-exception clearly is geared to content platforms exploited by right holders themselves. This distinction

[26] The implications of contractual agreements on TDM in practice, through the experience of stakeholders, will be discussed in section 5.1.

is important because it would result in different parameters for the definition of "lawful access".
This suggests that CHIs experience legal uncertainty as to the requirements for Article 3 CDSM Directive, and that tackling this uncertainty should be a priority when revising the CDSM Directive.
Therefore, CHIs, in this qualitative research also have legal certainty of the requirements for Article 3 CDSM Directive as a key feature for the revision of 2026.

4.2.3 Publishers

The parties interviewed that fall under the category of publishers include Elsevier and the Media Federation *(Mediafederatie*). For publishers, their involvement in TDM and general interests primarily focus on safeguarding the economic value of copyrighted content, ensuring the sustainability of their business models in the digital age, and defending the rights and interests of the creators whose works they publish.
During the interviews it was noted that there is an increase in TDM activities since the implementation of the CDSM Directive. They emphasise the importance of the "lawful access" requirement for their revenue model. They advocate that authors do not want their work diluted by automatic programs used for commercial ends, such as AI-model development. They argue that it is practically impossible to trace whether a specific protected work has been used in AI-model training, which complicates enforcement and licensing efforts.

Additionally, they experience significant unclarity around the classification of ROs and CHIs. They want more certainty that the party relying on the TDM exception is indeed an RO or CHI, so the exception applies to the requesting party. Additionally, there is uncertainty about the difference between non-commercial and commercial entities. This includes uncertainty about the scope of the exceptions in Article 3 and 4 and their relationship, namely how to draw a line between research activities and commercial TDM activities that might be subject to a rights reservation or opt-out by rightholders, namely in the context of web scraping activities.

Lastly, the Media Federation, as a collective management organisation (CMO), acknowledges the significant challenges posed by generative AI,

particularly concerning the use of protected works for LLM training and the competitive impact of AI-generated outputs.[27]

Overall, publishers are not against the notion of a research exception but need more clarity about its scope, and assurances that the scope can be monitored and enforced.

4.2.4 Experts

An expert on EU copyright policy, Paul Keller, has observed that the exception has so far not enabled research that was previously impossible. Rather its main benefit is providing universities with greater protection against unfair licensing practices. He points out that although the Directive aims to support innovation, including AI, LLMs were not a core focus during its drafting, creating a gap between policy goals and actual implementation. He warned that the restrictive nature of Articles 3 and 4, especially Article 4's opt-out mechanism, could limit the full potential of TDM and weaken the EU's competitiveness in AI research.
Furthermore, some experts at TIB - Leibniz Information Center for Science and Technology and University Library, point to a limited understanding in practice of what TDM entails, for all actors involved. They stress the importance of educating both rightholders and researchers on the relevant legal framework.

4.3 Main findings

The following sections detail what the main consensus as well as discords are for the parties involved. Through this discussion, the major uncertainties with the current regime are showcased.

4.3.1 Consensus

Across stakeholders, several points of consensus can be identified. Firstly, there is a general consensus that the exception for research purposes (as in the general concept of Article 3 CDSM Directive) is justified and important for scientific research and innovation. Even CMOs like the Media

[27] These challenges will be discussed in more detail in section 5.2.

Federation express no fundamental objections to its application, provided it does not jeopardise regular commercial exploitation. Moreover, TDM is broadly recognised as a powerful tool for generating new knowledge, discovering trends, and supporting innovation across various sectors. All parties, from researchers to publishers, agree on the necessity for clearer rules and better implementation to ensure the effective application of TDM exceptions, particularly concerning publisher restrictions and the evolving landscape of AI.

4.3.2 Discord

While there are areas of agreement, significant divergences arise primarily from the fact that different parties have differing interests. The primary discrepancies are set out below.

Remuneration for TDM

With regards to the remuneration for TDM, ROs, ULs and Paul Keller advocate that TDM for non-commercial scientific research should be free of extra costs or restrictions, as the access has already been lawfully paid for. They suggest publishers should be compelled to cooperate for free, with TDM-enabling APIs included in the license cost. Keller specifically argues that scientific research under Article 3 does not cause damage justifying remuneration.[28]

Publishers, on the other hand, insist on compensated access to protected content for commercial purposes. They express significant financial objections to the lack of remuneration for the use of their works in AI-training and for the economic harm caused by AI-generated content competing with original works.

Control and opt-out mechanisms for Article 3

The Media Federation calls for a right for rightholders to set conditions on research involving their works and the introduction of an opt-out

[28] When discussing the Knescke v. LAION case, Keller noted that a major underlying concern of advocates regarding LAION is companies shielding themselves behind the scientific exception (Article 3) when they are actually engaged in commercial activity. He argues against this by asking where the real justified claim for remuneration lies, arguing that scientific research under Article 3 doesn't cause damage.

possibility for Article 3, backed by enforceable sanctions, to allow publishers to assess and control commercial impacts.

ROs and ULs generally oppose an opt-out for Article 3, arguing that it would undermine the mandatory nature of the exception.

The Media Federation also proposes adding a transparency clause to give rightholders clearer insight into TDM-activities. This would help them assess potential commercial risks, especially when protected works are used to train AI models that could compete with original content and disrupt existing revenue streams (this will be detailed more specifically in section 6). Keller, instead, suggests that such a transparency clause for TDM could be important for reproducibility. However, he continues, such a clause should never be used as a controlling tool, as that would be detrimental to academic freedom.

"Copyright trumps licenses" and digital ownership

Keller advocates for the principle of "digital ownership", where a library, once it lawfully buys content, should own it and have the right to conduct TDM. He argues against "double protection" where publishers hold both copyright and license, limiting access. He believes Article 3 could be Europe's "redemption" on a digital cultural level by allowing cultural institutions to act as a counterweight to concentrations of power. Keller suggests CHIs should have a "making available" right of works under Article 3.[29]

For publishers, their business model is founded largely on licensing and paywalled access. While they allow TDM for lawfully accessed works, this still generates revenue from the initial access. Therefore, their business model depends on the very licensing system Keller opposes.

Role and activism of CHIs

It has been observed that CHIs are generally risk-averse and not heavily engaged in TDM activities, seeing their primary role as a trusted information source rather than a processor. It has been suggested they

[29] All TDM-activities done before making the results available to the market should be covered by an unconditional exception for ROs, CHIs and commercial entities, as long as they are doing research. Publishing would then require some form of remuneration.
Including a making available right for CHIs , like libraries, to utilize their right under Article 3.

might take on a greater role in protecting information if granted a more meaningful “making available” right.
The KB confirms their facilitative role, focusing more on preserving, digitising, and making material available. It does, however, carry out TDM-activities on a limited scale, and says they act responsible and compliant with regards to the rules involved. It also actively engages in dialogue with CMOs and publishers to facilitate TDM for scientific research by external researcher on digitised and born-digital works in the KB collection. By doing so, it has achieved significant innovation through constructive dialogue with CBO’s and beneficiaries.
they have achieved significant innovation through constructive dialogue with CBO’s and beneficiaries.

However, they do actively engage in dialogue with CMOs and publishers to facilitate TDM for scientific research by external researchers on digitised and born-digital works in the KB collection.

The KB – who does carry out TDM-activities, but on a limited scale – says it rather acts responsible and compliant with regards to their activism, and says they were able to achieve a significant innovation through constructive dialogue with CBO’s and beneficiaries.
The KB confirms their facilitative role and limited TDM activities, focusing more on preserving and making material available. However, they do actively engage in dialogue with CMOs and publishers to facilitate TDM for scientific research by external researchers on digitised and born-digital works in the KB collection.

4.3.3 Uncertainties

From the points described above, the following uncertainties of stakeholders can be distilled and categorised into legal and practical or operational uncertainties.

Legal uncertainties

ROs and CHIs face legal uncertainty due to the unclear extent to which publishers can impose restrictions on TDM activities. Specifically, it is unclear whether publishers can legally add limitations or costs to TDM when institutions already have lawful access to the material. Clarifying these boundaries is essential to ensure Article 3 remains effective.
Moreover, the scopes of Articles 3 and 4 are unclear. Mainly the boundaries between the exception for TDM for scientific research and the broader TDM exception with opt-out function are easily crossed.
Lastly, there is a lack of international harmonisation of TDM related laws which complicates cross-border application. Nor is there a clear framework for web archiving or collecting digital material.

Lawful access

Additionally, the ROs and CHIs have collectively expressed uncertainty with regards to the term "lawful access". For CHIs, this especially concerns digitally obtained content versus physical access. The KB noted that their position is different from that of ROs. ROs and ULs have a focus on how they can use the exception to connect researchers and should therefore be treated as such. The KB clarified that a problem they ran into is that the Directive permits ROs to perform TDM on lawfully accessed materials, but it remains unclear whether such access encompasses digitised content obtained through external parties or if legal access within physical premises also implies TDM rights.

Commercial vs non-commercial use

Additionally, the line between commercial and non-commercial use is increasingly blurred. Uncertainty exists about what constitutes commercial use of TDM datasets. Additionally, the use of the results of the research might become commercialised, or partly belong to a commercial entity, increasing the confusion about the boundaries of 'commercial'. Policy expert Paul Keller, however, dismisses legal uncertainty regarding the qualification of ROs and CHIs as an immediate problem. He views this as a "distraction" tactic, noting that most are non-profit and well-defined. He acknowledges that there are some edge cases, such as private universities, but also those readily have relationships with the National ministries. Boundary cases for CHIs exist around the museum sector (the Modern Contemporary Museum Amsterdam (Moco) is mentioned as

potentially more of a commercial undertaking than a heritage institution) and entities with purely digital collections such as SIA and Europeana that can be qualified as CHIs. Nonetheless, ROs are typically non-profit.
Despite this point by Keller, with which we tend to agree, the fact remains that many of the stakeholders we have interviewed mentioned this. As such, this paper flags it has a potential issue.

Practical / operational uncertainties
Various stakeholders have expressed practical and operational uncertainties on how to effectively circumvent or legally challenge technical barriers imposed by publishers, such as download limits or system blackouts that prevent TDM. The varying ways publishers deliver data (APIs, dumps, spreadsheets) make it labour-intensive and slow for researchers to obtain consistent data from multiple sources, for example by a 100-article-per-day download limit. These technical measures have to be manually circumvented through downloading batches of only 50 articles multiple times.
The conflict between publishers' requirement to delete raw data after research and the need to retain it for research reproducibility also poses a problem. Publishers require deletion of the used raw data after the research has been completed. However, there are increasing requirements that a study be reproducible and that the data used is preserved after the study is completed.
Additionally, the impossibility of tracing whether a specific protected work has been used in AI-model training datasets creates significant challenges for enforcement and licensing for rightholders. Rightholders argue that they are not receiving any form of compensation for their content's use in AI-training, nor for the market erosion that occurs when AI-generated outputs substitute their original works.

4.3.4 Interim conclusion

This section outlined the main problems and needs expressed by various stakeholders. Two central concerns will be examined together in the following chapter; *Highlighting Two Core Issues: Contractual Arrangements and Generative AI*. The first relates to beneficiaries and the problems caused by unclear terms in contractual agreements. The second concerns publishers and rightholders, focusing on the use of copyright-

protected materials in the training of generative AI models, such as large language models.

5. Highlighting Two Core Issues: Contractual Arrangements and Generative AI

5.1 Contractual Agreements

This next section will give insight into the main concern that was expressed by multiple beneficiaries of the research exception with regards to contractual agreements.

5.1.1 Beneficiaries concern

The main concern expressed by ROs and CHIs is that reliance on contractual agreements adversely affects their ability to conduct research using TDM technologies. ROs and CHIs predominantly rely on contractual agreements to gain lawful access to the copyrighted content.[30] These agreements, typically in the form of licences, often contain clauses that explicitly or implicitly prohibit or restrict TDM activities. As a result, ROs and CHIs are hesitant to engage in TDM-activities themselves or allow others to do so on licensed data. This hesitation persists even though Article 7 CDSM Directive renders such provisions unenforceable.

5.1.2 Contractual practices

Contractual agreements are typically negotiated on a case-by-case basis, resulting in TDM research being dependent on the parties involved, the nature of the activity and the licensed content. During our interviews, the University of Amsterdam highlighted that the variation in contractual practices negatively impacts TDM research by making the process more complex and time-consuming. Stakeholders identified several contractual practices that pose significant obstacles to TDM, such as the bundling of licensed works in ways that render the content non-downloadable.

[30] Senftleben 2022, p. 43

Additionally, the implementation of technical barriers is identified as an obstacle, as they restrict access to content and prevent large scale-data analysis.

5.1.2.1 Prohibition on contractual override

Lack of Protection

The protection of Article 7(1) CDSM Directive is observed by beneficiaries to have little impact in practice, as restrictive terms and conditions are still part of licensing agreements. These terms, either directly or indirectly prohibit TDM. For example, in a 2024 study authored by Ana Lazarova the following contractual provision was studied:

> *"It is prohibited to:* ***systematically or programmatically download****, whether manually or by using programs such as* ***robots*** *or searchbots, spiders,* ***crawlers or other automated*** *downloading programs,* ***algorithms*** *or devices, to* ***continuously or automatically search, scrape, extract, deep link or index*** *all or a substantial portion of the [Licensed Materials], such as an entire issue of or article from a journal. Additionally: It is* ***prohibited to: conduct text and data mining.".***[31]

This clause not only explicitly prohibits TDM, buts also indirectly restricts TDM-based research through the use of descriptive language that corresponds with activities covered by the definition of TDM in the CDSM Directive.[32] The same study found that only 9 out of 100 analysed contracts permitted TDM, while the other 91 were either unclear or restrictive or explicitly prohibited TDM.[33] The University of Amsterdam similarly observed that when TDM is not explicitly permitted by contract, publishers often respond negatively to requests, thereby rendering TDM research on licensed content practically impossible. Regarding a similar vein, the University of Groningen noted a lack of clarity on the extent to which restrictions may be imposed or additional payments may be requested.

A lack of knowledge

[31] Lazarova, 2024, this anonymised clause can be found in the second figure in the blogpost, as the full study is yet unpublished,
[32] See Part One, paragraph 2.1 of this paper.
[33] Lazarova, 2024

During a interview, some experts, noted that restrictive contractual terms often result from a lack of legal understanding of the research exception by both the beneficiaries and rightsholders. While Articles 3 and 7 of the CDSM Directive offer a clear legal framework in theory, in practice rightholders often misunderstand what TDM involves and respond with overly restrictive terms. At the same time, beneficiaries like libraries may not have the legal knowledge to assert their rights.

Moreover, agreements with non-European rightholders are especially challenging, as these parties are usually unfamiliar with EU law and negotiate standard terms that overlook the research exception. Other beneficiaries report similar experiences.

Still, experts at TIB – Leibniz Information Center recognise a general willingness to comply with EU law once rightholders have a clear understanding of TDM and the implications of the research exception. They stress the importance of educating rightholders on the relevant legal framework. The sentiment that a harmonised understanding of TDM is needed is shared by stakeholders.

5.1.2.2 Inequality of bargaining power

Several stakeholders attribute the contractual obstacles for TDM to an underlying inequality of bargaining power between rightholders and beneficieries.[34] ROs and CHIs typically have less leverage in contractual negotiations, while rightsholders are often in a position to dictate terms.
The underlying issue of power disbalance was also identified in the COMMUNIA Report on Unfair Licensing Practices with regard to the experience of libraries.[35] For example, the report found, that publishers may threaten to withdraw from negotiations or refuse to enter into agreements altogether when the inclusion of TDM allowance was requested.[36]

[34] Among them Matthijs van Otegem and Paul Keller
[35] COMMUNIA report on unfair licensing practices: the library experience, 2025, p. 3.
[36] COMMUNIA report on unfair licensing practices: the library experience, 2025, p. 7.

The KB noted that the transition from print to a digital sales model has deepened this imbalance, as licensed materials often remain under the control of the rightsholder rather than the licensed party.

However, Paul Keller emphasised that this is a structural issue, rather than a legislative one, and thereby cannot be sufficiently addressed by Article 3 and 7 CDSM Directive alone.

Several other stakeholders seem to share this sentiment, as they expressed general satisfaction with the protection offered by the legislative framework but found that barriers arise in practice.

Moreover, Paul Keller noted that beneficiaries have a sense of dependency with regard to rightsholders. He mentioned as example the fact that contractual terms are predominantly bargained for based on the terms proposed by rightsholders rather than propositions of beneficiaries.

5.1.2.3 Conclusions

In general, there seems to be a lack of consensus on what the limits to restrictions of publishers are as well as what they should be. The first problem might be tackled through more education about the rights flowing from Articles 3 and 7 CDSM Directive, while the latter is influenced by the bargaining powers of the respective parties.

5.2. The exception and the rise of generative AI

In the interviews a fear was perceived by publishers and rightsholders that the TDM exception facilitates the training of generative AI. As another recurring issue seems to be that none of the parties involved are sufficiently aware of the workings of the legal regime of this exception, the significance of TDM and the article 3 CDSM Directive exception for the development of generative AI will be discussed below.

5.2.1 The relationship between TDM and the development of GenAI

Breakthroughs in AI-development have led to generative AI models, capable of producing a wide range of content such as music, literary works and visual art. In the first step of AI model-training, massive amounts of data (potentially including copyrighted works) are analyzed through automated systems to abstract patterns that will become the basis of the generative AI model. This process is a textbook example of TDM, qualifying as such under the CDSM Directive.[37] Additionally, the collection of large datasets through web scraping for this purpose could also be qualified as an act of TDM under the broad definition of the Directive, as seems to be supported both by the national court cases mentioned above[38] and references in the AI Act[39] and the draft GPAI Code of Practice.[40]

With the directive using such a broad definition, it is beyond doubt that acts essential for the development of generative AI models fall within its scope. As such, EU copyright law and its handling of TDM-activities will be crucial for the future of European AI-development industry.[41]

5.2.2 Rightsholders' fears

During the conducted interviews, the rise in generative AI-models and the financial ramifications of that to industries that rely on copyright came up several times. Both rightsholders and libraries indicated that the popularity of these models has influenced the discussion regarding the merits of the TDM exception.

Rightsholders seem to feel that they are missing out on financial compensation in several ways. Firstly, some rightsholders believe that the exception allows the use of protected materials to train commercial generative AI models without providing reasonable remuneration to the rightsholders whose work was used. Secondly, rightholders argue that this

[37] Article 2(2) CDSMD
[38] See Chapter 2.1.2 above.
[39] Recital 105, AI Act, Regulation (EU) 2024/1183.
[40] Code of Practice for General-Purpose AI Models – Copyright Chapter. (2024). *Code of Practice*. European Commission. pp. 4–5.
[41] Hugenholtz, 2019

will lead to more structural financial issues, as they fear that generative AI models will compete with creators in the production of their works. Lastly, publishers view it as a missed chance to profit from selling valuable datasets for AI model training. This has led to several parties mentioning that if they had the knowledge they now have about generative AI, they would not have supported the scientific TDM exception in its current form.

5.2.3 The role of the exception in Article 3 for the development of generative AI

For this analysis, a distinction is made between the "direct" and "indirect" impact of the exception. The first refers to the direct beneficiaries of the exception: ROs and CHIs. The latter refers to public-private partnerships where a commercial entity partners up with a direct beneficiary of the exception for the duration of a project.

Direct impact of the exception

The direct impact of the exception is limited as the pool of beneficiaries of the exception is quite restricted.

Conducting TDM practices under the exception to develop a potentially competing generative AI-model would by definition not be covered by the exception.

Additionally, the lawful access requirement offers rightsholders another way to limit TDM. Publishers with contracts can legally exclude TDM use for commercial AI training. Although Article 7 prohibits contracts from blocking the exception, this only applies when the exception itself applies. As noted, TDM by ROs or CHIs for commercial AI development falls outside the exception and thus outside Article 7's protection.

Lastly the exception only covers acts of reproduction and extraction but no other exclusive rights like distribution or communication to the public.[42] This matters because any activity further than the initial TDM itself might be an infringement of copyright if the mined dataset contains protected materials and is "made available" after the primary mining

[42] Contardi, 2025, p. 53.

activity. Some even go as far as arguing that the transferring of mined data to a centralised collection, gathered and organised for use in AI training, could be considered "making available to the public" and thus fall outside the protection of the Article 3 exception.[43]

Indirect impact: Private-Public partnerships

As explained in section 2.3.2, commercial actors might indirectly benefit from the exception through public-private partnerships, if that collaboration is under the control of the RO or CHI. [44]

As mentioned above, the limited scope of copyrighted acts that fall under the exception serves as an additional failsafe. The exception only extends to acts of extraction and reproduction, but not subsequent acts such as the dissemination of the results if those include protected materials, such as an AI-development-relevant data corpus.[45] This represents an essential step for the potential use of protected TDM acts under the exception of the further development of an AI model. Any such downstream use that involves further reproduction and communication to the public of protected materials would require individual authorization from rightsholders.

Mainly because of its limited scope of beneficiaries, both directly and through public-private partnerships, the research exception is not suitable to enable *TDM activities* for developing generative AI models for commercial use that may compete with copyright holders. These commercial actors, if engaging in TDM activities, will fall under the scope of Article 4 of the CDSM Directive. It is therefore essential for all parties to understand that Article 3 does not by itself enable training of competing commercial generative AI models. Of course, it is possible that a model trained with recourse of TDM for scientific purposes is afterwards made available to third parties and fine-tuned for commercial use. But in that case, any subsequent fine-tuning and commercial availability or exploitation are not activities covered by the TDM exception in Article 3 CDSM Directive. Rather, any TDM that takes place in that context would

[43] Contardi, 2025, p. 54.
[44] Senftleben, et. al., 2025, p. 18.
[45] Senftleben, et. al., 2025, p. 21.

have to be assessed under Article 4 of the Directive. In our view, therefore, the research exception should not be weakened or restricted due to concerns it was never meant to address.

6. Conclusions and Recommendations

So far, this (policy) paper has outlined the approach of several key stakeholders regarding the exception. Several problems were underscored, regarding among others, the vagueness of the scope of the provision, the necessity for lawful access, the obstacles arising from contractual agreements, and the uncertainty and inconsistency the development of AI has caused.

6.1 Conclusions

In light of the upcoming revision of the CDSM Directive in 2026, this policy brief has sought to explore the practical implications of the TDM exception for research purposes of Article 3 CDSM Directive. Through doctrinal analysis and empirical research, we have found that the exception plays a critical role in enabling scientific research and innovation, but its effectiveness is hindered by several legal and practical barriers.

Sections 2, 3 and 4 of this paper outlined the legal scope and limitations of Article 3 CDSM Directive by clarifying the definition of TDM classifying the relevant categories of beneficiaries to the exception and the key requirement of "lawful access". Several concerns are highlighted, including the restrictive wording of the exception and its interplay with contractual practices. This legal analysis shows the importance of a clear and harmonised implementation of the exception.

Building on the doctrinal research, Section 5 and 6 present the findings of the empirical research, derived from stakeholder interviews. These sections examine the practical implications of the exception and the extent to which it achieves its intended purpose. It is found that stakeholders are generally in agreement about the positive potential of Article 3 CDSM Directive. However, they remain cautious to conduct TDM due to several uncertainties surrounding the application of the exception, such as the scope and meaning of "lawful access", the legal boundaries of

permitted TDM-activities and the enforceability of AI-related uses under the exception.

Our analysis shows that the main concerns raised by stakeholders largely stem from a lack of awareness and understanding of the legal framework. Both beneficiaries and rightholders demonstrate limited knowledge of the scope and implication of the exception, which contributes to their hesitation to conduct or allow TDM activities. As outlined in Section 5, beneficiaries report that reliance on contractual agreements often undermines their ability to do TDM, particularly given their weaker bargaining position in negotiations with rightsholders. However, the analysis also shows that a greater awareness of the exception and the prohibition on contractual override, can empower beneficiaries to assert their rights more effectively. Section 6 addresses rightholders' concerns about the potential application of Article 3 on the development of generative AI models. It is shown that these concerns are mostly unfounded, as the exception has a narrowly defined scope of beneficiaries, commercial use is excluded, and strict limitations are placed on public-private partnerships. As such, the use of TDM for the purpose of training commercial AI models falls outside the scope of Article 3. Concerns related to such uses are more appropriately addressed under general exception for TDM of Article 4 CDSM Directive, or under potential copyright infringement stemming from the release of Clarification is needed to distinguish Article 3 from Article 4, especially for AI, ensuring Article 3 excludes commercial AI training that substitutes human-created works.

6.2 Recommendations

6.2.1 Stakeholder Analysis

A key focus for the 2026 revision of the Directive must be a thorough, truly inclusive consultation with all affected parties before any text is finalised. Although several key stakeholders were involved in the earlier discussions, several organisations, researchers, universities, and independent researchers felt excluded from the deliberations. As a result, many now struggle to interpret critical terms—such as "lawful access" and "scientific research"—and practical obstacles remain unaddressed.

To remedy this, the revision should:

1. Establish a permanent TDM Advisory Forum with rotating seats for independent researchers, doctoral candidates, university libraries, non-profit research centres and experts in digital humanities and AI ethics.
2. Publish all consultation materials—agendas, position papers, minutes and impact summaries—so contributors can track how their input influences the draft.
3. Mandate regular Member State reporting on the exception's real-world application, drawing on both quantitative usage data and qualitative feedback from diverse research settings.

Embedding a transparent, ongoing consultation process will ensure that Article 3 is not only legally sound but also practically effective. It is important that when conducting these talks the different relationships between players are taken into account. While one interviewee expressed the view that a contract is fair simply because both parties consented to it, this policy paper challenges that assumption. In practice, significant power imbalances, especially between ROs, CHIs, and large publishers, can lead to unfair contractual terms, even when formal consent is given. It is not so much that these institutions are afraid to push back (although that too might be a factor), but rather that they often lack awareness or understanding of their actual legal position, particularly regarding rights under the TDM-exception.

To address this, several solutions are possible. First, awareness campaigns could play a key role in empowering ROs and CHIs, particularly in the lead-up to the possible 2026 Directive revision. Some experts even stated that, ideally, TDM rights should not even appear in contracts at all, as these rights are granted by law. They therefore should not need additional negotiation.

Additionally, structured focus groups and discussion forums should be created to ensure that all stakeholders have the opportunity to participate in shaping a potential revised version of the Directive.

6.2.2 Legal Certainty

The majority of the stakeholders interviewed indicated to struggle with several concepts within the CDSM Directive. Three possible solutions to these uncertainties are proposed:

Lawful access

One of the most pressing sources of legal uncertainty under Article 3 CDSM Directive is the ambiguous notion of "lawful access". Researchers and institutions often lack clarity about whether specific forms of access—such as through interlibrary loan systems, shared institutional repositories, or public databases—meet the threshold required for lawful TDM. Moreover, rightholders frequently exploit this ambiguity by imposing contractual terms or technical restrictions that obstruct TDM-activities even where access is formally given.

To resolve this, EU legislation or official interpretative guidance should at least clarify that "lawful access" covers:

- Access obtained via standard library or university subscriptions;
- Access through publicly accessible repositories;
- Access via national legal deposit systems or institutional agreements;
- Any lawful acquisition not contingent on additional permissions.

Simultaneously, it should be made clear that contractual terms cannot limit TDM-rights once lawful access has been obtained, consistent with Article 7 of the Directive. That is to say, where lawful access is provided, no TDM restrictions may be imposed. Without such clarification, research organisations will continue to operate under uncertainty, and the harmonising goal of the CDSM Directive will remain unmet.

Strengthening enforcement of Article 7

Although Article 7 prohibits contractual provisions from overriding the exceptions and limitations provided for in the Directive (including Article 3), this safeguard has proven ineffective in practice. Researchers report that publishers routinely use indirect mechanisms (e.g. technical measures, usage quotas, or opaque license conditions) to restrict or discourage TDM, despite the legal guarantee that these activities cannot

be contractually excluded. To enhance the practical enforceability of Article 7, Member States should be required to establish robust enforcement frameworks, including:

- A fast-track complaints mechanism allowing researchers and libraries to report violations;
- Administrative penalties or sanctions for rightsholders who impose contractual or technical barriers incompatible with Article 3;
- A clear monitoring mandate for national copyright authorities or independent regulators.

Such mechanisms would empower ROs and CHIs to assert their rights confidently and deter rightsholders from circumventing the law through backdoor restrictions.

Widespread confusion remains regarding the division of scope between the two articles. This legal uncertainty is particularly problematic in the context of AI development, where the boundary between scientific and commercial use is often blurred. Research that involves the training of general-purpose AI models – including of the generative type –for example, often occurs in hybrid public-private partnerships and raises questions about which exception, if any, applies.

To address this ambiguity, the European Commission should issue formal guidance or an interpretative communication that:

- Confirms that Article 3 applies to all TDM conducted for the purposes of scientific research, regardless of the funding structure, institutional affiliation, or public-private collaboration. This would include, for instance, applied machine learning research carried out by universities in cooperation with industry partners.
- Emphasises that Article 4 is intended for general purpose or fully commercial TDM, where rightsholders may lawfully opt-out, and where uses are not linked to a scientific research mandate.

Making this division explicit would not only prevent rightsholders from misapplying Article 4 opt-out provisions to uses that should fall under Article 3, but would also promote legal certainty, foster user confidence, and support compliance across Member States. Moreover, by structurally distinguishing research-driven TDM from commercial data exploitation, the legal framework will be better positioned to adapt to future

technological developments and help realise important scientific discovery. As AI-technologies evolve and new computational methods emerge, a well-defined division of legal bases will provide the necessary regulatory agility, allowing the EU to remain responsive to innovation without compromising the rights of authors or the integrity of research.

Bibliography

Books

Rosati, E. (2021). *Copyright in the Digital Single Market: Article-by-Article Commentary to the Provisions of Directive 2019/790* (First edition). Oxford University Press. https://doi.org/10.1093/oso/9780198858591.001.0001

Articles

Contardi, Magali, 'El rol de la excepción de minerías de datos en la inteligencia artificial (IA) generativa – perspectiva europea' (2025) 24(1) *Revist@ e-Mercatoria* 37–76. DOI: https://doi.org/10.18601/16923960.v24n1.02

Ducato, R., & Strowel, A. (2021). Ensuring text and data mining: Remaining issues with the EU copyright exceptions and possible ways out. *European Intellectual Property Review*, *43*(5), 322–337.

Geiger, C., Frosio, G., & Bulayenko, O. (2018b). Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data? Deleted Journal, 49(7), 814–844. https://doi.org/10.1007/s40319-018-0722-2

Janssens, M. C. (2020, 8 oktober). *inDICEs: empowering IPR for Cultural Heritage Institutions*. inDICEs. https://indices-culture.eu/indices-empowering-ipr-for-cultural-heritage-institutions/?utm_source=chatgpt.com

Margoni, T., & Kretschmer, M. (2022). A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology. *GRUR International (Print)*, *71*(8), 685–701. https://doi.org/10.1093/grurint/ikac054

Nobre, T., Hackett, T. (2025). COMMUNIA report on unfair licensing practices: the library experience https://communia-association.org/wp-content/uploads/2025/05/Unfair-licensing-practices_the-library-experience.pdf

Quintais, J. P. (2025). The concept of "research organisation" and its implications for text and data mining and AI research. Available at SSRN: https://ssrn.com/abstract=5155685 or http://dx.doi.org/10.2139/ssrn.5155685

Quintais, J. P. (2025). What Is a 'Research Organisation' and Why It Matters: From Text and Data Mining to AI Research. *GRUR International (Print)*. https://doi.org/10.1093/grurint/ikaf030

Senftleben, M., Szkalej, K., Sganga, C. and Margoni, T., 2025. Towards a European Research Freedom Act: A Reform Agenda for Research Exceptions in the EU Copyright Acquis. *Available at SSRN*.

Stamatoudi, I., & Torremans, P. (2021). THE DIGITAL SINGLE MARKET DIRECTIVE. In *EU Copyright Law*. Edward Elgar Publishing Limited. https://doi.org/10.4337/9781786437808.00027

Szkalej, K, (2025). The Paradox of Lawful Text and Data Mining? GRUR International 74(4), 2025, 307–319. https://doi.org/10.1093/grurint/ikaf029

Vučković, R. M., Kanceljak, I., & Jurić, M. (2021). Cultural heritage institutions during and after the pandemic: the copyright perspective. EU and comparative law issues and challenges series (ECLIC), 5, 379-397.

Internet sources

Arends, A., De Bont, A., Rosenberg, M., Quintais, J. P., & Van Daalen, O. (2024, 27 november). *Report for OFF: Demonopolizing the European public domain – ILP Lab*. Samuelson & Glushko Information Law And Policy Lab. https://ilplab.nl/2024/11/27/report-for-off-demonopolizing-the-european-public-domain/

Belderbos, H. (2025, April 16). *EU advances in research and innovation, but continued investment remains important*. Open Access Government. https://www.openaccessgovernment.org/eu-advances-in-research-and-innovation-but-continued-investment-remains-important/191674/

Bernt Hugenholtz, P. (2019, July 24). *The New Copyright Directive: Text and Data Mining (Articles 3 and 4)*. Kluwer Copyright Blog. https://copyrightblog.kluweriplaw.com/2019/07/24/the-new-copyright-directive-text-and-data-mining-articles-3-and-4/

Chappell, B. (2015, September 9). *National Geographic Reshapes itself in $725 million deal with 21st Century Fox*. NPR. https://www.npr.org/sections/thetwo-way/2015/09/09/438853832/national-geographic-reshapes-itself-in-725-million-deal-with-21st-century-fox

Cruiz, J. (2011, August 24). *Access to the agreement between Google Books and the British Library*. Open Rights Group. https://www.openrightsgroup.org/blog/access-to-the-agreement-between-google-books-and-the-british-library/

Goldstein, P. (2024, November 13). *Kneschke vs. LAION – Landmark Ruling on TDM-exceptions for AI training data – Part 1*. Kluwer Copyright Blog. https://copyrightblog.kluweriplaw.com/2024/11/13/kneschke-vs-laion-landmark-ruling-on-tdm-exceptions-for-ai-training-data-part-1/

Goldstein, Stuetzle & Bischoff, (2024), "Kneschke v Laion-Landmark ruling on TDM-exceptions for AI training data- Part 2". Kluwer Copyright Blog.

Lazarova, A. (2024). *Libraries, licenses, limitations: an Empirical Insight into the Contractual Conditions Regulating TDM for Research*. https://digrep.bg/en/copyright/libraries-licenses-limitations/

Mezei, P. (2025, May 8). *Third European Court Decision on the General Purpose TDM-exception is out*. Kluwer Copyright Blog. https://copyrightblog.kluweriplaw.com/2025/05/08/third-european-court-decision-on-the-general-purpose-tdm-exception-is-out/

Valk, E., & van Toepoel, I. (2025, January 16). *DPG Media et al vs. HowardsHome – a national ruling on DSM's press publishers' rights and TDM-exceptions*. Kluwer Copyright Blog. https://copyrightblog.kluweriplaw.com/2025/01/16/dpg-media-et-al-vs-howardshome-a-national-ruling-on-dsms-press-publishers-rights-and-tdm-exceptions/

Official documents

COMMISSION STAFF WORKING DOCUMENT IMPACT ASSESSMENT Accompanying the document Proposal for a Directive of the European Parliament and of the Council on common rules for the internal market in electricity (recast) Proposal for a Regulation of the European Parliament and of the Council on the electricity market (recast) Proposal for a Regulation of the European Parliament and of the Council establishing a European Union Agency for the Cooperation of Energy Regulators (recast) Proposal for a Regulation of the European Parliament and of the Council on risk preparedness in the electricity sector SWD/2016/0410 final - 2016/0379 (COD)

European Commission: Directorate-General for Research and Innovation and Senftleben, M. R., *Study on EU copyright and related rights and access to and reuse of data*, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/78973

Peukert, A., & Castets-Renard, C. (2025). *Code of practice for general-purpose AI models: Copyright chapter*. European Commission, Directorate-General for Communications Networks, Content and Technology. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai
Triaille, J.-P., Meeûs d'Argenteuil, J.d., & Francquen, A.d., European Commission: De Wolf & Partners, Directorate-General for the Internal Market and Services. (2014). *Study on the legal framework of text and data mining (TDM)*, Publications Office. https://data.europa.eu/doi/10.2780/1475

Case law

ACI Adam and others v Stichting Thuiskopie, case C-435/12, ECLI:EU:C:2014:254.
British Horseracing Board Ltd and Others v. William Hill Organization Ltd, Case C-203/02, ECLI:EU:C:2004:695.
DPG Media v. HowardsHome, Rechtbank Amsterdam, 30 October 2024, ECLI:NL:RBAMS:2024:6563.
GS Media BV v. Sanoma Media Netherlands BV and Others, Case C-160/15, ECLI:EU:C:2016:644.
Infopaq International A/S v. Danske Dagblades Forening, Case C-5/08, ECLI:EU:C:2009:465.
Kneschke v. LAION, LG Hamburg, Urteil vom 27.09.2024 – 310 O 227/23. https://openjur.de/u/2495651.html
Municipal Court of Appeals (Hungary), Case 9.Pf.20.353/2024/6-II, 3 December 2024.
Nils Svensson and Others v. Retriever Sverige AB, Case C-466/12, ECLI:EU:C:2014:76.
Stichting Regionaal Opleidingen Centrum Noord-Kennemerland (ROC Noord-Holland) v. Stichting Autoriteit Financiële Markten (AFM), Case C-235/05, ECLI:EU:C:2006:778.
Stichting Brein v. Jack Frederik Wullems (Filmspeler), Case C-527/15, ECLI:EU:C:2017:300.