Polish Implementation of TDM Exceptions – General Characteristics

Konrad Gliściński

ABSTRACT

The aim of this article is to analyse the implementation of Directive (EU) 2019/790 on copyright and related rights in the context of Text and Data Mining exceptions within Polish law. It highlights interpretative challenges and uncertainties arising from the regulations, potentially leading to legal disputes. The article begins with an overview of the Directive and then examines the specific provisions in Polish law that implement it, focusing on the general and research exceptions. It discusses the lack of clarity in definitions, the scope of exceptions, and the implications for potential beneficiaries. Additionally, it identifies uncertainties regarding the storage of copies, access conditions, and protections against technical measures. Ultimately, the article concludes with a summary of the main challenges presented by the implementation and their potential impact on the practical use of Text and Data Mining exceptions.

1. INTRODUCTION

The purpose of this article is to provide a general overview of how Polish law has implemented the exceptions related to text and data mining (TDM) as outlined in the CDSM Directive.[1] Two exceptions enabling TDM have been incorporated into Polish law: a general one, based on Art. 4 of the CDSM Directive, and a specific one for scientific research purposes, based on Art. 3 of the directive. Both exceptions are independent of each other. This means that beneficiaries of the research-specific exception will also be able to base their activities on the general exception, and vice versa, as long as the conditions set out in each exception are met.[2] In both instances, the legislator opted not to introduce compensation for the use of works for TDM purposes. The Polish legislator delayed the adoption of the relevant provisions, which only came into effect on 20 September 2024.[3] Although there was ample time for public consultations and adjustments, the current provisions raise concerns and may lead to interpretative disputes. Interestingly, numerous entities participated in these consultations,[4] but in many cases, their input was not reflected in the final version of the law. One significant exception in this regard was the proposal to exclude the use of both exceptions for the purpose of "creating generative artificial intelligence models." After criticism of this solution as potentially inconsistent with EU law, this exclusion was not included in the final text of the law.[5] Additionally, there may be aspects of the Polish regulations that conflict with EU law.

2. TDM – GENERAL INFORMATION

For the purposes of further analysis, it is worth explaining in simple terms what TDM (text and data mining) involves. It seems possible to outline three typical—though not always essential—steps in TDM processes: (1) accessing content, (2) extracting or copying content, and finally, (3) analysing the text or data to uncover knowledge. In the execution of Step 3, we can further distinguish, among others, Stage A (preliminary), which involves cleaning and normalising the texts, and Stage B, which involves the direct analysis of the data.[6] From the perspective of copyright law, we can identify that steps two and three may involve the right to reproduce

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 (Text with EEA relevance.) OJ L 130, 17.5.2019, p. 92–125.

2 See: E. Rosati, *Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790* (OUP:2021), p. 41.

3 Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254)

4 See: https://legislacja.rcl.gov.pl/projekt/12382002/katalog/13037394#13037394.

5 K. Gliściński, 'The Good, the Bad and the Missing – the new proposal for the implementation of the CDSM Directive into Polish law', (*Communia Association*, 1 March 2024) https://communia-association.org/2024/03/01/the-good-the-bad-and-the-missing/ accessed 10 August 2024.

6 E. Rosati, *Copyright in the Digital Single Market. Article-by-Article Commentary to the Provisions of Directive 2019/790* (OUP:2021), p. 68–71.

works. Before the introduction of exceptions for TDM, it was already clear that the copying of works as part of the preparatory activities for TDM constituted reproduction (Step 2). Such reproduction could be carried out with the rights holder's permission (a licence) or under permitted uses provisions.[7] The open question was the status of the analysis itself conducted within the TDM processes (Step 3). Specifically, the question was whether such activities constitute a form of reproduction of works or whether, e.g. due to the lack of human involvement and only machine use, these activities do not constitute reproduction within the meaning of copyright law.

This brings up an important issue. Before the introduction of the discussed exception into Polish law, did the copyright monopoly also cover the activities performed within the scope of Step 3? The Polish structure of economic rights is based on a (dynamic) construction of fields of exploitation.[8] According to Polish copyright law,[9] "...the author shall have an exclusive right to use the work and to dispose of its use throughout all the fields of exploitation and to receive remuneration for the use of the work." However, the Polish law does not introduce a definition of a field of exploitation. Art. 50 only provides examples of such fields.[10] These example fields of exploitation cover copyright provisions defined in EU directives but also include forms of use that have not been harmonised at the EU level.[11] The essence of the dynamic construction of fields of exploitation lies in the fact that, with technological development, new fields of exploitation may emerge,[12] which will automatically fall under the copyright monopoly. However, these fields of exploitation must first be distinguished in contractual practice.[13] However, it seems that (although this statement is not supported by empirical analyses) before the introduction of this exception, there was no widespread practice of distinguishing the activities that make up Step 3 as a separate field of exploitation in copyright agreements. Nor was it common to licence works in this area. Neither the current construction of new exceptions nor the content of Art. 50 directly answers the question of whether the activities carried out in the context of Step 3,[14] in themselves constitute a new field of exploitation. This situation supports the claim that, under Polish law, the activities previously carried out under Step 3 were not covered by the copyright monopoly.

This approach is also supported by the principle of public domain. According to this principle, the existence of exclusive rights imposing obligations on others to refrain from using works in a specific manner should not be presumed if such rights are not explicitly provided for by law.[15] Therefore, since it was not common practice to reserve certain types of activities for rights holders, those performing these activities should not be unexpectedly informed that they were infringing on copyright. This approach is particularly justified in light of the possibility of infringing copyright without fault. Such strict liability, although common in copyright law, should only apply to activities that are objectively defined in the law as falling within the scope of the monopoly but have been infringed without fault. However, if certain activities were not previously specified in the law, the possibility of judicially extending the copyright monopoly to those activities should not be allowed.

Of course, in practice, the issue of assessing the execution of activities involved in Step 3 will only be clarified through jurisprudence. The problem is unlikely to arise in situations where reproduction under Step 2 was carried out under the previously applicable provisions on permitted use, as these provisions could only serve as a basis for reproducing works in a limited range of situations. The issues in this area may particularly concern situations where the other party to the contract obtained, either through a transfer of rights or a licensing agreement, the right to use works in the field of digital reproduction.[16] If, according to the interpretation proposed here, we consider that the analytical activities within Step 3 do not constitute an act of exploitation, this means that such activities fall outside the scope of copyright monopoly. Consequently, performing these activities is not reserved for the rights holder, and simply acquiring or licensing the right to reproduction would be sufficient for conducting TDM. On the other hand, if we determine that the activities carried out within Step 3 are also covered by copyright, simply obtaining a licence or acquiring rights for digital reproduction would not be considered sufficient. Consequently, it would have to be recognised that such a person infringed the copyright of the work.

7 In the Polish legal system, exceptions and limitations to copyright are referred to as "dozwolony użytek," which in English translates to "permitted uses".

8 K. Gliściński, Komentarz do art 17, [w] A. Michalak, Ustawa o prawie autorskim i prawach pokrewnych. Komentarz, Warszawa 2019, p. 147–150.

9 Art. 17 Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

10 Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

11 According to it: "The separate fields of exploitation shall be, in particular: (1) within the scope of fixing and reproduction of works – the production of copies of a work using specific technologies, including printing, reprographics, magnetic fixing, and digital technology; (2) within the scope of trading the original or the copies on which the work was fixed – the introduction to trade, lending for use, or rental of the original or copies; (3) within the scope of dissemination of works in a manner different from that defined in subparagraph 2 – public performance, exhibition, screening, presentation, and broadcasting, as well as retransmission, and making the work publicly available in such a manner that anyone could access it at a place and time selected by them."

12 For example, through the mass identification of a specific method of using works in contracts as a separate source of economic benefits.

13 K. Gliściński, Wyodrębnianie się nowych pól eksploatacji i ich wpływ na obrót prawami do utworów, ZNUJ. PPWI 2010, nr 3, s. 45–60.

14 Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

15 K. Gliściński, Komentarz do art 17, [w] A. Michalak, Ustawa o prawie autorskim i prawach pokrewnych. Komentarz, Warszawa 2019, p. 145–147.

16 Another issue is to what extent and based on what form of permitted uses, before the introduction of the analyzed exception, it was possible to "reproduce" works for the purpose of performing step 2.

3. TDM – DEFINITIONS

According to the CDSM Directive TDM "means any automated analytical technique aimed at analysing text and data in digital form in order to generate information which includes but is not limited to patterns, trends and correlations."[17] The Polish implementation enabling TDM is based on a definition that essentially resembles the definition contained in the CDSM directive. According to it: "The exploration of texts and data involves their analysis solely through the use of an automated technique designed for analysing texts and data in digital form, with the goal of generating specific information, including, in particular, patterns, trends, and correlations."[18] In the Polish translation of the Directive, the term "mining" has been translated as "*eksploracja*" ("exploration" in English). In the literature, certain doubts have been raised regarding the wording of this provision. It states that exploration must occur "solely through the use of an automated technique".[19] According to some, this wording may lead to uncertainties about whether preparatory activities such as pre-processing, data cleaning, or normalisation are covered by the provision Step 2. These activities are performed by humans and are not automated. The issue in this context concerns the use of the word "solely", which does not appear in the text of the CDSM Directive.[20] However, it appears that comparing this definition with the content of the relevant provision of the CDSM Directive introducing the TDM exception for scientific research provides grounds to assert that all reproduction activities are permitted as long as they serve the purpose of text and data mining (see below).

4. TEXT AND DATA MINING FOR THE PURPOSES OF SCIENTIFIC RESEARCH

a. Beneficiaries

The scope of beneficiaries indicated in the CDSM Directive refers to *research organizations* and cultural heritage institutions. As indicated in the literature, the approach taken in the Directive is based on a dual limitation: on one hand, the exception defined in Art. 3 applies only to "scientific research," and on the other hand, it must be carried out by *research organizations*. This means that independent researchers and other entities conducting "scientific research" (e.g., journalists or companies operating research centres) are outside the scope of this exception.[21] The exception in Polish law has three limitations: a formal list of beneficiaries, the purpose of TDM, and a prohibition on obtaining economic benefits (see below). Although it has not been definitively established, it seems that beneficiaries of this exception, according to Recital 11 of the CDSM Directive, can "rely on their private partners for carrying out text and data mining, including by using their technological tools".[22]

The Polish Act defines cultural heritage institutions similarly to how the CDSM Directive does. Consequently, such institutions are defined as: "a library, museum, archive, or a cultural institution whose statutory mission is to collect, protect, and promote collections of film or phonographic heritage."[23] A different legislative technique was used with respect to the second group of beneficiaries. The Polish Copyright Act, referring to the Act on Higher Education and Science, specifies a closed category of entities that are beneficiaries of this exception. They are (i) universities (both public and non-public); (ii) federations of higher education and science entities, scientific institutes of the Polish Academy of Sciences, research institutes;[24] (iii) International scientific institutes established under separate laws operating on the territory of the Republic of Poland; (iv) Łukasiewicz Center; (v) Institutes operating within the Łukasiewicz Research Network, hereinafter referred to as "Łukasiewicz Network institutes"; (vi) The Polish Academy of Arts and Sciences and other entities primarily engaged in scientific activities in an independent and continuous manner.[25] The same scope of beneficiaries has been provided for with respect to related rights[26] and databases protected by sui generis rights.[27]

Thus, this represents a narrower scope of beneficiaries compared to the broader category of *research organisa-*

17 Art. 2(2) 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 (Text with EEA relevance.) PE/51/2019/REV/1 OJ L 130, 17.5.2019, p. 92–125.

18 Art. 6(1)(22) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

19 Art. 6(1)(22) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

20 A. Matlak, M. Wyrwiński, B. Widła, Konsultacje publiczne projektu wdrożenia dyrektyw CDSM i SATCAP II [2024], https://ipwi.uj.edu.pl/documents/122195199/151128292/Konsultacje+publiczne+dotycz%C4%85ce+projektu+wdro%C5%BCenia+dyrektyw+CDSM+i+SATCAB+II+%5B2024%5D/ccbf017d-9501-46b6-94df-c5e04891f792 (10.08.2024), p. 7.

21 Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2021) 71(8) GRUR International 2022, 685–701, Available at SSRN: https://ssrn.com/abstract=3886695 or http://dx.doi.org/10.2139/ssrn.3886695 accessed 04 November 2024.

22 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, rec. 11 art. 77(1) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie.

23 Art. 6(1)(21) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254)

24 Ustawa o instytutach badawczych (Dz.U. z 2024 r. poz. 534).

25 Art. 7 Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2024 r. poz. 1571).

26 Art. 100 Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

27 Art. 8b Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

tions as defined in the Directive. The Directive allows that the beneficiaries of this exception may also include entities whose "primary goal is to [...] carry out educational activities involving also the conduct of scientific research."[28] However, the aforementioned Polish catalogue does not include such educational institutions but only "entities primarily engaged in scientific activities in an independent and continuous manner".[29] Moreover, according to Recital 12 of the CDSM Directive, this definition should be interpreted broadly and include, among others, "hospitals that carry out research".[30] The current wording of the Polish implementation raises doubts as to whether it also covers such hospitals. This does not refer to hospitals run by universities (which should be considered as covered by this exception under Polish law) but rather to other hospitals that also engage in scientific research. Due to the mandatory nature of the exception outlined in Art. 3 of the CDSM Directive, such a narrow scope of beneficiaries is under the threat of being considered incompatible with EU law.

b. Permitted uses and subject matter

Polish law, similar to the CDSM Directive, permits reproduction for the purposes of TDM. This applies—*lege non distinguente*—to reproduction occurring as part of preparatory activities – Step 2 – (such as pre-processing, data cleaning, or normalisation), as well as directly within the TDM process itself (Step 3).

The Polish Copyright Act regulates the reproduction of works and objects of related rights for TDM purposes. Under the Polish Act, the term *works* also encompasses creative databases (protected under Chapter II of the Database Directive) and computer programs. Reproduction of such databases is thus covered by the exception in accordance with the CDSM Directive. At the same time, the Polish legislator, similar to the CDSM Directive, chose not to extend this exception to computer programs.[31] Polish copyright law includes the following under related rights: rights to performances, rights to phonograms and videograms (film fixations), rights to programme broadcasts, rights to first publications and scientific and critical publications, and rights to press publications within the framework of providing services by electronic means. This exception, with respect to all related rights, has been uniformly introduced and covers all related rights existing under Polish law, including those rights that have not been harmonised at the EU level.[32] An analogous exception—contained in the Database Act—allows for the reproduction (extraction) of data without restriction under *sui generis* rights.[33]

c. Direct and indirect economic benefits and TDM for scientific research purposes

The CDSM Directive generally does not prohibit TDM used for scientific research from providing economic benefits to the beneficiaries. It merely specifies that such beneficiaries must: (1) have as their primary goal the conduct of scientific research or carry out educational activities that also involve scientific research, and (2) operate on a non-profit basis[34] or reinvest all profits into scientific research[35] or pursuant to a public interest mission recognised by a Member State.[36] The provision allows

28 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, art. 2(1).

29 Art. 7(1)(8) Prawo o szkolnictwie wyższym i nauce (Dz.U. z 2024 r. poz. 1571).

30 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, rec. 12.

31 Art. 77(1) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

32 Art. 100 Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

33 Art. 8b Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

34 It is worth noting that the English wording of Art. 2(1)(a) of the CDSM Directive raises certain interpretative concerns. According to this definition, such an organization is one that operates "on a not-for-profit basis or by reinvesting all the profits in its scientific research." The issue with this phrasing lies in the fact that the Directive does not define what constitutes a "not-for-profit" organization. To my knowledge, European law also does not provide a clear definition of this term. In relation to certain types of activities, a distinction is often made between "not-for-profit" and "non-profit" organizations. Such distinctions often arise from the specific tax regulations adopted in different countries. In the US, it is noted that "A not-for-profit (NFPO) is an organization that, like a nonprofit, doesn't seek to turn a profit. However, unlike a nonprofit, a not-for-profit doesn't have to exist for the sole purpose of improving society." https://givebutter.com/blog/non-profit-vs-not-for-profit (04.09.2024). The European Commission's proposal includes a definition of organizations operating for "non-profit purposes." According to this definition, "non-profit purposes" means that, "regardless of whether the association's activities are of an economic nature or not, any profits generated are used solely to further the objectives of the organization as defined in its statutes, and are not distributed among its members." Art. 2(c) Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European cross-border associations (Text with EEA relevance) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2023%3A516%3AFIN&qid=1693910621013 (04.09.2024). In the literature, one might encounter statements such as: "The essence of 'not-for-profit' activity is that, alongside its primary mission, it engages in ancillary commercial activities, which both foundations and educational institutions, including public ones, are permitted to undertake. This type of activity differs from 'non-profit' operations typical of administrative entities, which are never considered commercial activities and cannot generate profits." A. Bednarczyk-Płachta, Zysk założyciela szkoły wyższej niepublicznej jako inwestora w odniesieniu do zmian w prawie o szkolnictwie wyższym, PPP 2017, nr 3, s. 10–38. If we consider that a "not-for-profit" organization is one that can generate profit but must reinvest it into its activities, then the wording of Art. 2(1)(a) of the CDSM Directive may be superfluous. This assessment arises from the fact that the provision designates, in addition to "not-for-profit" organizations, another type of organization that can also generate profit but must reinvest it specifically in scientific research.

35 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, art. 2(1)(a).

36 Art. 2(1)(b). "Such a public-interest mission could, for example, be reflected through public funding or through provisions in national laws or public contracts." (recital 12). 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

research organisations, in principle, to generate profits. This further indicates that such organisations may also charge access fees for their analysis results as long as these fees only cover the costs of their activities (e.g., conducting analyses on behalf of external parties, including commercial entities). In this regard, the CDSM Directive only requires that: "access to the results generated by such scientific research cannot be enjoyed on a preferential basis by an undertaking that exercises a decisive influence over such an organisation."[37] This means that such results may be available to these entities, provided that other entities also have the opportunity to access these results under the same conditions, including the same financial terms. Furthermore, the directive directly provides that "research organisations should also benefit from such an exception when their research activities are carried out in the framework of public-private partnerships."[38]

In contrast, the Polish framework introduces a significant restriction. According to it, TDM for research purposes cannot be conducted "for the purpose of obtaining direct or indirect economic benefits."[39] In Polish law, this term appears in many provisions of copyright law. Generally, it is indicated that a financial benefit can be understood "as achieving profit or as reducing incurred costs."[40] This wording indicates that beneficiaries, contrary to the provisions of the CDSM Directive, will not be able to, for example, derive profits from using TDM for scientific purposes. An open question also remains as to whether they will be able to impose fees to cover the costs of providing access to such results and whether they will enter into public-private partnerships. Additionally, the Polish implementation completely overlooks the possibility of recognizing an entity conducting scientific research as a research organisation "pursuant to a public interest mission recognized by a Member State." Such a situation may occur, among other instances, when the research activity is funded by the public sector or is based on relevant provisions in national law or public contracts.[41] In summary, while the CDSM Directive allows for the possibility of deriving financial benefits under Article 3, outlining which entities and purposes are permitted, the Polish law implementing this exception outright prohibits obtaining any economic benefits. It seems that such a restrictive construction is inconsistent with the (already narrowly defined)[42] framework established in the CDSM Directive.

d. Storage and retention of copies created for TDM (Text and Data Mining) for the purpose of scientific research

The CDSM Directive specifies that the storage of copies of works and other subject matters must be done with "an appropriate level of security."[43] The Directive left the Member States the freedom to define the detailed rules for the storage of such copies.[44] The Polish law in this regard has detailed the general security requirement by specifying that: "The storage of works is conducted with a level of security that ensures access to these works is limited exclusively to authorised persons, taking into account authentication procedures."[45] The law itself does not specify who should be considered authorised persons. It seems that this term primarily refers to individuals involved in conducting scientific research on behalf of eligible beneficiaries. The decision of who qualifies as an authorised entity in the context of a particular study should be made by the beneficiary based on their internal procedures. Importantly, access to such copies is not limited solely to researchers directly participating in the study; it may also extend to other individuals (e.g., technicians, IT staff, librarians) who assist in conducting the research on behalf of the institution. Furthermore, according to Recital 11 of the CDSM Directive, beneficiaries of this exception "should also be able to rely on their private partners for carrying out text and data mining, including by using their technological tools". In this context, it can be understood that beneficiaries may designate authorised persons not only among their internal staff but also among private partners they engage for conducting text and data mining on the data copies of works. Given the requirement for "authentication procedures"[46] introduced in the Polish implementation, it seems that

96/9/EC and 2001/29/EC (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125.

37 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, art. 2(1).

38 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 (Text with EEA relevance.) PE/51/2019/REV/1 OJ L 130, 17.5.2019, p. 92–125, rec. 11.

39 Art. 26^2(1) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

40 J. Marcinkowska [w:] Komentarz do ustawy o prawie autorskim i prawach pokrewnych [w:] Ustawy autorskie. Komentarze. Tom I, red. R. Markiewicz, Warszawa 2021, art. 31.

41 Recital 12 of the 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 (Text with EEA relevance.) PE/51/2019/REV/1 OJ L 130, 17.5.2019, p. 92–125.

42 See: Thomas Margoni and Martin Kretschmer, 'A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology' (2021) 71(8) GRUR International 2022, 685–701, Available at SSRN: https://ssrn.com/abstract=3886695 or http://dx.doi.org/10.2139/ssrn.3886695 accessed 04 November 2024.

43 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, art. 3(1).

44 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, rec. 15.

45 Art. 26^2(2) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

46 Art. 26^2(2) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

access to such copies should be granted individually to specific persons.

Similarly to the CDSM Directive, the Polish implementation specifies that works reproduced under this exception "may be stored for scientific research purposes, including the verification of research results."[47] Polish law does not impose any time limits on the storage of copies of works reproduced under this exception.[48] Such a solution should be considered desirable, both from the perspective of the specific nature of conducting scientific research in general and sustainability goals. It is not possible to determine in advance from which point in time duplicated works will no longer be needed. Given the ongoing nature of scientific research, access to such copies may be necessary and desirable at any future time. Therefore, rather than deleting such copies, they should be preserved for future scientific research needs.

The CDSM Directive distinguishes between the "verification" of scientific research and its "review."[49] In the context of Polish law, this distinction can lead to problematic situations. While the TDM exception for scientific research allows entities involved in the verification of research results to access all copies of works used in the TDM process, the situation will be different if a researcher is interested in reviewing those results. In this case, access to these data will not be possible under the TDM exception for scientific research but rather under the general exception for research purposes. This second exception has been recently amended and now allows for the reproduction of "published small works or excerpts from larger works not exceeding 25% of the work's volume."[50] This means that the researcher will be able to physically view the data in its entirety (as long as it does not require the reproduction of data) but will not be permitted to make a complete copy of the data for the purpose of conducting the review. Certainly, such a situation is undesirable from the standpoint of research integrity and transparency. At the same time, this example highlights that the distinction introduced by the CDSM Directive seems unjustified. If the entity conducting TDM research is interested in verifying the results, it will be able to involve third parties to whom it can provide the collected copies. However, if a researcher not affiliated with the original entity conducting the research wishes to review the results, they will only be able to do so based on a limited excerpt of the collected copies.

Certainly, in practice, it will be challenging to distinguish whether a given activity constitutes the verification of research results or their review. Should the determination of whether an activity is one or the other be decided solely by the entity that originally conducted the research (e.g., by specifying a verification stage in the research protocol)? Can a scientist not affiliated with the original entity claim to independently verify the results, and how would such verification differ from a rigorous review of scientific results? Additionally, beyond the scope of this exception's regulations remains the issue of access to such data. Exceptions to the right of reproduction may only grant beneficiaries the right to make copies of certain data but do not impose an obligation on any entities to create such copies. In other words, if a scientist wishes to verify results, but the entity that created the data is unwilling to provide access, the verification cannot be enforced.[51]

e. Measures to ensure the security and integrity of the networks and databases

Following the CDSM Directive, the Polish implementation stipulates that: "Rightholders, in order to ensure the security and integrity of networks and databases in which works are stored, may use only the measures necessary to achieve this goal."[52] The Polish legislation does not specify exactly which measures can be employed by authorised entities, nor does it indicate which measures are considered impermissible. According to Recital 16 of the CDSM Directive, such measures could, for example, "be used to ensure that only persons having lawful access to their data can access them, including through IP address validation or user authentication". These issues are expected to be resolved in practice by judicial rulings at the level of the Court of Justice of the European Union (CJEU). However, it appears that impermissible measures would include those that either prevent or significantly hinder the extraction of data from databases for the purpose of TDM used in scientific research. Currently, there are no publicly known actions by the Polish authorities aimed at fulfilling the obligations arising from Art. 3(4) of the CDSM Directive, including those specified in Art. 3(2) and 3(3) thereof.

f. Protection against contractual override

Art. 7(1) of the CDSM Directive provides that any contractual provision contrary to the exceptions for TDM for scientific research "shall be unenforceable." Consequently,

47 Art. 26²(2) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

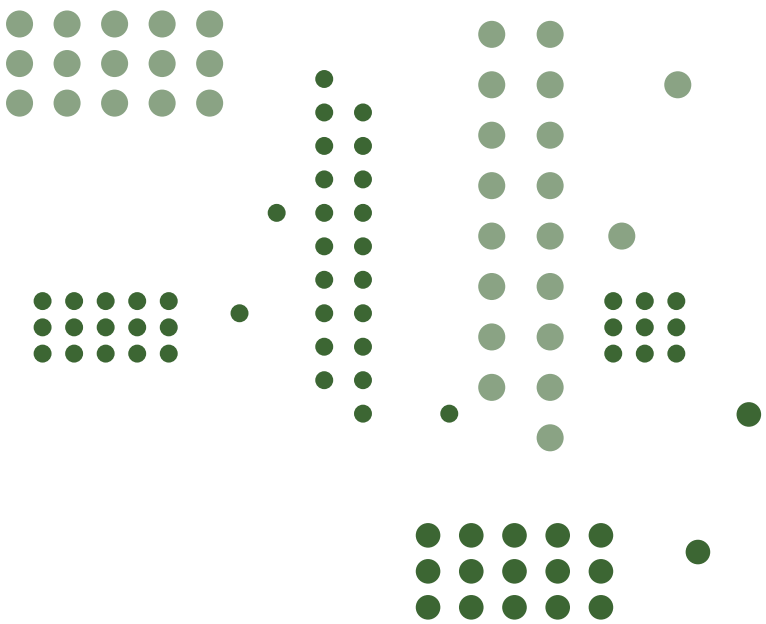
48 A time limit for storing such copies has been introduced in German law, for example, in Section 60d(5) Copyright Act of 9 September 1965 (Federal Law Gazette I, p. 1273) https://www.gesetze-im-internet.de/englisch_urhg/englisch_urhg.html (18.08.2024).

49 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, rec. 15.

50 Art. 27(1) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

51 This issue highlights that copyright law—while it affects scientific activity—does not resolve all the problems associated with it. In this context, it seems important to explore other legal instruments aimed at comprehensively regulating scientific activities in the digital context.

52 Art. 26²(3) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).



Member States are obligated to safeguard this exception against contractual override. This is especially important in the context of licensing agreements entered into by beneficiaries of this exception with database providers. Polish law does not contain a specific provision implementing such protection. In the course of preparing the legislation, it was indicated that: "provisions of the Copyright Act concerning permitted use (Art. 23–35) leave no doubt that they apply regardless of the will of the rights holders, and thus also regardless of any contractual provisions between the rights holder and the beneficiary of the permitted use."[53] The approach adopted by the Polish legislator is difficult to consider correct. First, it is important to highlight that there is a divergence of views in the doctrine on this issue. Some legal scholars argue that the provisions on permitted use are indeed imperative (or semi-imperative), while others believe that it is possible to contractually exclude their application. The lack of consistency in the doctrine in this area, coupled with the absence of case law addressing this issue, means that the position adopted by the legislator lacks strong justification and is not, in itself, a source of law.[54]

Second, even if one assumes that contractual provisions cannot effectively limit the scope of permitted use, using a work in violation of such a provision may still result in liability for breach of contract. This situation creates legal uncertainty and may have a chilling effect. It is crucial to directly regulate this issue, as users often lack knowledge about the legal nature of exceptions and base their decisions on the wording of the provisions.

This problem affects both individual users, such as ordinary citizens who typically accept the terms of agreements automatically, and public institutions that enter into contracts with clauses limiting the scope of permitted use. For such institutions, the legal uncertainty as to whether violating a contractual provision leads to an infringement of copyright law (assuming the non-imperative nature of the provisions) or merely to contractual liability is not so important. In both cases, it may lead public institutions to refrain from using works within the scope of permitted use.

5. GENERAL EXCEPTION OR LIMITATION FOR TEXT AND DATA MINING

a. Beneficiaries, permitted uses and subject matter

The TDM exception for scientific research is based on an open formula indicating that, in the absence of a specific reservation, "it is permissible to reproduce disseminated works for the purpose of text and data mining".[55] This construction means that any entity can benefit from this exception. Such an entity can, therefore, reproduce works of any type (textual, musical, graphic, video, etc.) and in any form and format (particularly in digital formats) for the purpose of TDM. However, the use of computer programs for TDM purposes may be problematic. While the exception allows for the reproduction of such programs, the exclusive rights also cover "translations, adaptations, rearrangements, or any other modifications of the computer program."[56] In many cases, utilising computer programs in this context will require stepping into rights beyond just the right to reproduce.[57]

b. Lawful access v. disseminated work

The only limitation introduced by the Polish legislator is that these works must have been previously disseminated. According to Polish copyright law,[58] a *disseminated work* is that "which, with the permission of its author, has been made available in any manner to the public". However, the Polish concept of a *disseminated work* is not equivalent to the condition of a "lawfully accessible work" as used in the CDSM Directive. The dissemination of a work pertains to the status of the work itself rather than the status of individual copies of it. A work could, therefore, be considered disseminated under Polish law while simultaneously not being a "lawfully accessible work" by the beneficiary. The condition specified in the directive will, therefore, be met first when the rights holder grants the beneficiary appropriate permission to access the work

53 Tabela zgodności, https://legislacja.rcl.gov.pl/docs//2/12382002/13037388/13037389/dokument656773.pdf, p. 14.

54 See: K. Gliściński, Komentarz do art 17, [w] A. Michalak, Ustawa o prawie autorskim i prawach pokrewnych. Komentarz, Warszawa 2019, p. 205–206.

55 Art. $26^2(1)$ Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

56 Art. 74(4)(2) Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

57 B. Widła, Programy komputerowe jako przedmiot eksploracji tekstów i danych w kontekście dyrektywy 2019/790, Europejski Przegląd Sądowy nr 3(210)/2023, p. 13–14.

58 Art. 6(1)(3) Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

(e.g., through a licensing agreement or an open access policy) or second when the work is available without any legal restrictions (e.g., placed on the internet by the rights holder). On the other hand, if a work has been disseminated with the rights holder's permission (e.g., in digital form), but the beneficiary accesses an electronic version of the book from an illegal source, the condition specified in the directive is not met (even though the work is considered disseminated under Polish law). From this perspective, it can be stated that the condition of disseminating a work protects the creator from situations where works are used under permitted uses before their first public release. It is, therefore, related to the moral right of the author "to decide about making the work available to the public for the first time."[59] On the other hand, the condition specified in the Directive pertains to the protection of economic interests related to lawful access to individual copies of the work. As a consequence, the introduction of the requirement for "dissemination of works" in place of "lawful access" may be regarded as incompatible with EU law. In this context, it was pointed out that the absence of this requirement is not necessarily an issue, as Polish law includes a clause referring to the three-step test. Thus, under permitted use, one cannot use works that have been made available illegally.[60] However, such an approach may raise certain doubts.

c. Opt-out mechanism

Art. 4(3) of the CDSM Directive stipulates that the general exception for TDM applies unless it has been expressly reserved by the right holder in an appropriate manner. The Polish legislator, when implementing this solution, specified that such reservations must be made "explicitly and in a manner appropriate to the way in which the work was made available. In the case of works made publicly available in such a way that anyone can access them at a time and place of their choosing, the reservation must be made in a machine-readable format as defined in Art. 2(7) of the Act of 11 August 2021 on open data and the re-use of public sector information),[61] along with metadata".[62] According to this latter provision, a machine-readable format means "a file format structured in such a way that computer programs can identify, recognize, and retrieve specific data and their internal structure."[63] This article, in turn, implements the definition of "machine-readable format" as outlined in Art. 2(13) of Directive 2019/1024. Examples of such formats include XML, JSON, RDF, and CSV.[64]

The legislator has not specified how such a reservation should be made when making works available through other means. Essentially, according to Recital 18 of the CDSM Directive, this can occur through, among other means, "contractual agreements or a unilateral declaration".[65] While in the case of access to works in electronic format, a contractual reservation seems conceivable (e.g., in licensing terms), it is less likely to occur with works available in analogue formats (e.g., printed books). In this latter case, unilateral reservations become significant. It seems that such a reservation should be made on every copy of the work in question. A general reservation, for instance, on the publisher's website or in accompanying materials, may prove to be insufficient. From a practical standpoint, such a reservation can be made alongside the traditional copyright notice typically found in books.

At the same time, in both cases, the legislator did not determine the specific wording of such a reservation. He merely indicated that it should be *explicit*. This means that the content of the reservation should clearly state the prohibition against reproducing the works for text and data mining purposes. On the one hand, it can be argued that using the traditional phrases *all rights reserved* or *no copying allowed*, without explicitly linking them to a prohibition on using the work for TDM purposes, would not meet the requirement for an explicit reservation. On the other hand, it does not seem necessary to cite specific articles from the law or directive to fulfil this requirement. For works distributed digitally but not made publicly available in a manner where anyone can access them at any time and place of their choosing (e.g., music on CDs). However, it seems that the requirement for an *explicit* reservation supports the view that such a reservation should also be made in natural language (e.g., on the packaging of a CD) so that it can be reviewed before purchase. In cases where the reservation does not meet the aforementioned requirements, it should be considered ineffective against individuals conducting TDM activities based on improperly marked or unmarked copies of the work. It is difficult to assert that a purchaser of a work is obliged to seek such a reservation beyond the copy being acquired. Of course, issues related to the effective manner of making reservations have already been raised at the level of the directive itself. However, the Polish legislator did not decide to introduce any specific regulations in this regard.

At the same time, it should be emphasised that opting out does not preclude conducting TDM for scientific

59 Art. 16(4) Ustawa o prawie autorskim i prawach pokrewnych z dnia 4 lutego 1994 r. (Dz.U. z 2022 r. poz. 2509).

60 Raport z konsultacji publicznych projektu ustawy o zmianie ustawy o prawie autorskim i prawach pokrewnych oraz niektórych innych ustaw – załącznik do Oceny Skutków Regulacji https://legislacja.rcl.gov.pl/docs/2/12360954/12887995/12887998/dokument587349.pdf (19.07.2024), p. 15.

61 Ustawa o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego (Dz.U. z 2023 r. poz. 1524).

62 Art. 26^3(2) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

63 Art. 2(7) Ustawa o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego (Dz.U. z 2023 r. poz. 1524).

64 Art. 2 OtwDaneU red. Sibiga/Sybilski 2022, wyd. 1/Garstka/Gos/Sibiga/Sybilski/Szelenbaum, G. Sibiga, D. Sybilski (red.), Ustawa o otwartych danych i ponownym wykorzystywaniu informacji sektora publicznego. Komentarz, Warszawa 2022.

65 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 (Text with EEA relevance.), OJ L 130, 17.5.2019, p. 92–125, rec. 18.

research, nor does it limit activities that are not covered by exclusive rights or those performed with unprotected elements of works.

d. The retention period for copies of reproduced works

Following the text of the directive, the Polish legislator stated that works reproduced under the discussed exception "may be stored solely for the purpose of text and data mining, and only for as long as is necessary to achieve that purpose."[66] This construction, however, leaves some uncertainty regarding the duration for which such copies may be stored. On the one hand, a narrow interpretation of this purpose suggests that once the TDM process is completed, these copies should be deleted. On the other hand, the TDM process can be understood more broadly, encompassing not only the preparation phase and the TDM itself but also subsequent verification activities. These verification activities may be carried out shortly after the TDM or much later.

6. PROTECTION OF BENEFICIARIES FROM TPMS

According to Art. 7(2) of the CDSM Directive, the discussed exception is subject to Art. 6(4) of the InfoSoc Directive. The Polish Copyright Act does not explicitly regulate mechanisms for protecting the beneficiaries of exceptions from Technological Protection Measures (TPMs). In the justification for the draft implementing the CDSM Directive, it was indicated that there is no need to implement Art. 7(2).[67] This approach is based on the assumption that Polish copyright law provisions regarding liability for the removal or circumvention of TPMs (Art. 79(6)) allow for "the removal and circumvention of technical protections if it is intended for the lawful use of works (e.g., within the scope of exceptions for public use)".[68] While it may be agreed that such behaviour is permissible under the current Polish legal framework, the question remains whether this solution complies with Art. 6 of the InfoSoc Directive.[69]

7. CONCLUSIONS

As I have explained throughout this article, the Polish implementation of both TDM exceptions may raise certain concerns. Given that the implementation has only just come into effect, there is a lack of extensive commentary in the legal doctrine on this matter. The chosen method of implementation, largely based on a copy-paste approach, also fails to address many of the issues that were raised concerning the text of the directive. In particular, it does not resolve the issues related to the process of opting out. It remains an open question as to how these provisions will be applied in practice. Will the concerns outlined in this presentation actually translate into practical difficulties in their use? Specifically, will they give rise to legal disputes? All these questions will expectedly find their answers in time.

Konrad Gliściński

PhD in Law. Researcher at the Department of Intellectual Property Law at the Jagiellonian University. Legal advisor to the Board of the Jagiellonian Center of Innovation. Intellectual property expert at Centrum Cyfrowe. Lecturer at the H. Grotius Center for Intellectual Property Rights. He collaborates with the Kalecki Foundation. He completed doctoral studies at the Faculty of Law and Administration of the Jagiellonian University in Cracow and post-graduate studies in company law at the Warsaw School of Economics. He holds an LL. M. of the University of Turin in the field of intellectual property law. A graduate of the Top500 Innovators program at Stanford University on management and commercialization of scientific research. Laureate of the 2010 Minister of Science and Higher Education award for best master's thesis organized by the Patent Office of the Republic of Poland. Author of "All rights reserved. The history of disputes over copyright. 1469 – 1928". Co-author of the commentary on the Industrial Property Law Act (2016) and the Copyright and Related Law Act (2019).

66 Art. 26³(2) Ustawa o zmianie ustawy o prawie autorskim i prawach pokrewnych, ustawy o ochronie baz danych oraz ustawy o zbiorowym zarządzaniu prawami autorskimi i prawami pokrewnymi z dnia 26 lipca 2024 r. (Dz.U. z 2024 r. poz. 1254).

67 Tabela zgodności, https://legislacja.rcl.gov.pl/docs//2/12382002/13037388/13037389/dokument656773.pdf, p. 14.

68 A. Matlak, T. Targosz, E. Traple [w:] Komentarz do ustawy o prawie autorskim i prawach pokrewnych [w:] Ustawy autorskie. Komentarze. Tom II, red. R. Markiewicz, Warszawa 2021, art. 79, s. 1188.

69 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. OJ L 167, 22/06/2001, p. 0010 – 0019.