



UvA-DARE (Digital Academic Repository)

Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining

Hugenholtz, B.

Publication date

2019

Document Version

Final published version

Published in

AMI : Tijdschrift voor Auteurs-, Media- & Informatierecht

[Link to publication](#)

Citation for published version (APA):

Hugenholtz, B. (2019). Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining. *AMI : Tijdschrift voor Auteurs-, Media- & Informatierecht*, 43(5), 167-171. <https://www.ami-online.nl/art/3996/artikelen-3-en-4-dsm-richtlijn-tekst-en-datamining>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Artikel in de serie over de DSM-richtlijn

Artikelen 3 en 4 DSM-richtlijn: tekst- en datamining

Bernt Hugenholtz

Prof. mr. P.B. Hugenholtz is hoogleraar IE aan de Universiteit van Amsterdam en redacteur van dit blad.

Tekst- en datamining (TDM) – het geautomatiseerd analyseren van grote hoeveelheden ongestructureerde tekst en data – is niet meer weg te denken uit de wetenschap, en speelt daarnaast een steeds grotere rol in een veelheid van andere domeinen, variërend van journalistiek tot artificiële intelligentie. De nieuwe DSM-richtlijn voorziet daarom in een tweetal beperkingen van het auteursrecht en het databankenrecht ten aanzien van tekst- en datamining: de ene ten behoeve van non-profit wetenschappelijk onderzoek (art. 3), de andere voor andere doeleinden (art. 4). Daarmee ontstaat de indruk dat de richtlijn voor TDM ruim baan heeft gemaakt, maar deze schijn bedriegt. Rechthebbenden kunnen door middel van een ‘opt-out’ tekst- en datamining voor commerciële doeleinden uitsluiten of aan licenties onderwerpen. In dit artikel worden de TDM-bepalingen van de DSM-richtlijn kritisch besproken. Daaraan gaat vooraf een kort exposé over tekst- en datamining en de stand van zaken naar huidig auteursrecht.

I. Tekst- en datamining

‘Tekst- en datamining’ wordt in de DSM-richtlijn gedefinieerd als “geautomatiseerde analysetechniek die gericht is op de ontleding van tekst en gegevens in digitale vorm om informatie te genereren zoals, maar niet uitsluitend, patronen, trends en onderlinge verbanden” (art. 2 lid 2). Eenvoudiger gezegd: tekst- en datamining (TDM) is de computergestuurde analyse van grote hoeveelheden gegevens om kennis te vergaren.¹ Door de toegenomen rekenkracht van computers en de onuitputtelijke bron van ‘mijnbare’ tekst, beelden, geluiden en andere gegevens op het internet is TDM in de afgelopen jaren een belangrijk onderzoeksinstrument geworden in de wetenschap en andere domeinen. In de linguïstiek kan TDM bijvoorbeeld worden gebruikt om teksten te analyseren teneinde syntactische of grammaticale patronen te ontleden.² TDM wordt ook toegepast op tal van andere wetenschappelijke terreinen, variërend van astronomie tot musicologie tot de sociale wetenschappen.

TDM dringt tevens steeds sterker door in de industriële en commerciële wereld: in farmaceutisch onderzoek, journalistiek, information retrieval, zoekmachines, marktonderzoek en consumenteninformatie – om maar een handvol gebieden te noemen. TDM is ook een essentieel instrument bij het ontwikkelen van zelflerende systemen die grote hoeveelheden ruwe tekst en gegevens nodig hebben om complexe taken zoals vertaling of spraakherkenning aan te leren. De ontwikkeling van kunstmatige intelligentie is daarom in belangrijke mate afhankelijk van TDM.³

Het proces van tekst- en datamining omvat verschillende fasen. In de eerste fase worden (doorgaans grote hoeveelheden) ruwe tekst en data onttrokken aan de bron, bijvoorbeeld door het downloaden van bestanden of het ‘crawlen’ en ‘scrapen’ van websites. In de tweede fase worden de verzamelde gegevens voorbewerkt en in ‘genormaliseerde’ bestanden ondergebracht, zodat de gegevens voor automatische data-analyse geschikt zijn. In de derde fase worden de

1 Wikipedia, https://en.wikipedia.org/wiki/Data_mining_mining. Zie J.-P. Triaille a.o., ‘Study on the legal framework of text and data mining (TDM)’, rapport voor de Europese Commissie, maart 2014, p. 15-18.

2 Wikipedia, https://en.wikipedia.org/wiki/Text_mining_mining.

3 Zie <https://www.leru.org/news/better-rules-on-text-and-data-mining-are-needed-to-maximise-the-potential-of-artificial-intelligence-in-europe>.

gegevens geladen en softwarematig geanalyseerd. In alle fasen van dit proces vinden mogelijk auteursrechtelijke relevante handelingen plaats.⁴

2. Auteursrecht en TDM

Hoewel de term tekst- en datamining relatief nieuw is, is het geautomatiseerd analyseren van tekst- of databestanden dat niet. De vraag of tekst- en data-analyse auteursrechtelijke gevolgen heeft, is dan ook al vroeg in de literatuur gesignaleerd, maar bleef lange tijd een hoofdzakelijk academische kwestie.⁵ Door de explosieve groei van TDM van de afgelopen jaren – vooral in het commerciële domein – zijn rechtshabbers zich pas recent gaan realiseren dat hier kansen liggen.

Daarbij heeft de zeer ruime definitie van het reproductierecht in de Auteursrechtlijn een rol gespeeld. Volgens art. 2 van deze richtlijn strekt dat recht zich uit tot elke “directe of indirecte, tijdelijke of duurzame, volledige of gedeeltelijke reproductie [...], met welke middelen en in welke vorm ook “. Daarmee is een conflict tussen auteursrecht en TDM vrijwel onvermijdelijk geworden. In een op verzoek van de Europese Commissie in 2014 geschreven studie concludeert Triaille dan ook: “we consider that data analysis generally involves an act of copying – in accordance with Article 2 of the Infosoc Directive – of whole or part of the data being processed during the data analysis activities”.⁶

De schaarse rechtspraak bevestigt dat TDM met reproductie in auteursrechtelijke zin gepaard kan gaan. In het eerste *Infopaq*-arrest oordeelde het HvJ EU dat een “data-capture-procedure” (in casu: het scannen en automatisch doorzoeken van persartikelen door een attenderingsdienst) als reproductie in de zin van art. 2 Auteursrechtlijn aangemerkt kan worden, en dat de tijdelijke reproductie-uitzondering van art. 5 lid 1 Auteursrechtlijn toepassing mist tenzij de door het verwerkingsproces gegenereerde kopieën na afloop automatisch worden gewist.⁷ Dit oordeel werd in het tweede *Infopaq*-arrest bevestigd.⁸ Art. 5 lid 1 van de Auteursrechtlijn, dat in Nederland is geïmplementeerd in art. 13a Aw, biedt derhalve slechts beperkte ruimte voor TDM, zoals in overweging 9 bij de DSM-richtlijn ook wordt erkend.

Een interessante TDM-gerelateerde uitspraak van eigen bodem betreft de zaak *Anne Frank Fonds/KNAW*, die in 2016 leidde tot een opvallend vonnis van de Amsterdamse rechtbank. Het Fonds maakte bezwaar tegen de digitalisering en conversie tot een XML-TEI-bestand van het bekende dagboek ten behoeve van wetenschappelijk tekstueel onderzoek. De rechtbank oordeelde dat hier sprake was van ongeoorloofde verveelvoudiging die niet door enige bestaande beperking van het auteursrecht kon worden gerechtvaardigd. Maar de handhaving van het auteursrecht door het Fonds stuitte naar het oordeel van de rechtbank wel af op de in art. 13 van het EU-Handvest gewaarborgde vrijheid van wetenschap.⁹

In de afgelopen jaren is gewaarschuwd voor de voor TDM mogelijk desastreuze gevolgen van het zeer ruim geïnterpreteerde reproductierecht,¹⁰ en is onderzocht of bestaande beperkingen wellicht toch soelaas zouden kunnen bieden. Zo is onder meer gewezen op de in de Auteursrechtlijn opgenomen, optionele wetenschapsbeperking,¹¹ die onder het nieuwe DSM-regime onverkort van kracht blijft¹² maar in Nederland tot op heden niet als zodanig is geïmplementeerd. Ook is gepleit voor invoering van een open geformuleerde wettelijke beperking, enigszins naar het voorbeeld van de Amerikaanse fair use-regel.¹³ In dit verband is opmerkelijk dat het meest gezaghebbende federale hof van beroep in de Verenigde Staten, het 2nd Circuit van het Court of Appeals, inmiddels tweemaal tot het oordeel is gekomen dat het digitaliseren, doorzoeken en doorzoekbaar maken van enorme aantallen auteursrechtelijk beschermde boeken *fair use* is.¹⁴ Tekst- en datamining lijkt derhalve naar Amerikaans auteursrecht zonder meer geoorloofd, zelfs als dit geschiedt met winstoogmerk.

De meest fundamentele kritiek uit wetenschappelijke hoek betreft overigens niet zo zeer het tekortschieten van het repertoire aan wettelijke beperkingen, als wel de ruime uitleg van het reproductierecht als zodanig. Zo schreef de European Copyright Society in een opinie naar aanleiding van de DSM-voorstellen van de Commissie:

“Copyright allows a right owner to control the exploitation of a work. Text and data mining of copyright protected works is not ‘exploitation’. Works which are

4 Triaille (noot 1), p. 28 e.v.

5 Zie P.B. Hugenholtz, *Auteursrecht en information retrieval*, Deventer: Kluwer 1982, p. 31-33.

6 Triaille (noot 1), p. 31.

7 HvJ EU 16 juli 2009, C-5/08 (*Infopaq I*), r.o. 64. In deze zin reeds: P.B. Hugenholtz (noot 5), p. 32.

8 HvJ EU 17 januari 2012, C-302/10 (*Infopaq II*).

9 Rb Amsterdam 23 december 2015, IER 2017/5, p. 27 m.nt. C.F.M. de Vries (*Anne Frank Fonds/KNAW*). Zie M. Caspers, <http://copyrightblog.kluweriplaw.com/2016/01/20/the-role-of-anne-franks-diary-and-academic-freedom-for-text-data-mining/>. Art. 13 EU-Handvest luidt: “De kunsten en het wetenschappelijk onderzoek zijn vrij. De academische vrijheid wordt geëerbiedigd.”

10 European Copyright Society, ‘General Opinion on the EU Copyright Reform Package’, 24 januari 2017, <https://europeancopyrightsociety-dotorg.files.wordpress.com/2015/12/ecs-opinion-on-eu-copyright-reform-def.pdf>.

11 Art. 5 lid 3(a) Auteursrechtlijn. Zie H. Osinga, ‘De panacee voor tekst- en datamining’, *AMI* 2018, afl. 4, p. 137. Zie ook C. Geiger e.a., ‘Text and Data Mining in the Proposed Copyright Reform: Making the EU Ready for an Age of Big Data?’, *IIC* 2018, p. 814-844.

12 Zie art. 1 lid 2 DSM-richtlijn.

13 P.B. Hugenholtz & M.R.F. Sentleben, ‘Fair use in Europe. In search of flexibilities’, Amsterdam, november 2011, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1959554.

14 Authors Guild v. HathiTrust, 755 F.3d 87 (2d Cir. 2014); Authors Guild v. Google, Inc. 804 F.3d 202 (2d Cir. 2015).

subject to TDM are not “used as works”. TDM does not affect the market for these works. Yet, due to the currently prevailing, formalistic interpretation of the reproduction right – see e.g. Infopaq cases I (C-5/08) and II (C-302/10) – the current proposal would continue to allow copyright owners to inhibit the purely technical copies made through TDM activities by non-research organisations and companies. The need for a licence that emerges from this situation is inconsistent with copyright logic and creates chilling effects on research activities that are in the general interest. At the same time, it does not produce any incentive effect on new creations or productions. Such an outcome runs counter to the goals of copyright and the functions of economic rights.”¹⁵

Inderdaad kan worden geconstateerd dat de TDM-problematiek een direct uitvloeisel is van een ‘formalistische’ interpretatie die de economische ratio van het reproductierecht uit het oog verliest. Uitgaande van een meer economisch-normatieve uitleg van het reproductierecht zouden louter technische kopieerhandelingen niet als reproductie moeten worden aangemerkt indien het werk daardoor niet in economische zin geëxploiteerd wordt.¹⁶ Hiermee in direct verband staat het argument dat de bescherming van het auteursrecht zich uitstrekt tot de vorm van het werk, en niet tevens de daarin besloten gegevens of andere inhoudelijke elementen omvat. Het onttrekken van gegevens aan een werk zou daarom zonder meer toelaatbaar moeten zijn.

Naast het auteursrecht kan ook het sui generis databankenrecht een obstakel tegen TDM opwerpen, indien het ‘mijnen’ zich (mede) richt op databanken. Bedenk hierbij dat het begrip ‘databank’ in databankenrechtelijke zin zeer ruim is, en bijvoorbeeld ook een enkele website kan omvatten. Hoewel het sui generis recht ruimte laat voor extractie en hergebruik van niet-substantiële delen van een databank, zal het mijnen zich daartoe meestal niet beperken. Daar komt bij dat ook het ‘herhaald en systematisch’ opvragen en hergebruiken van kleine gedeelten van een databank inbreuk oplevert indien de onttrokken gegevens bij elkaar genomen een ‘substantieel deel’ van de inhoud vormen.¹⁷ Dat datamining strijd kan opleveren met het databankenrecht, blijkt impliciet uit het arrest van het Europese Hof in de *Directmedia*-zaak. Volgens het Hof maakt “het zonder toestemming kopiëren van de inhoud van een databank, gevolgd door het bewerken hiervan, deel uit van de verrichtingen waartegen deze richtlijn de fabrikant van een databank via de invoering van een recht sui generis beoogt te beschermen”.¹⁸

3. TDM in de DSM-richtlijn

Ook al wordt de relevantie van bestaande wettelijke beperkingen in de overwegingen bij de DSM-richtlijn onderstreept, en al blijven de regels van het bestaande acquis onverkort van kracht,¹⁹ toch heeft de EU-wetgever de invoering van een uitdrukkelijke TDM-uitzondering noodzakelijk geacht, met name met het oog op de rechtszekerheid.²⁰ Hoewel de noodzaak om een ruime TDM-vrijheid in het wetenschappelijk onderzoek te waarborgen altijd al evident was, is de aanpak van de Uniewetgever ten aanzien van tekst- en datamining in het commerciële domein meer ambivalent. In het oorspronkelijke ontwerp van de DSM-richtlijn van de Europese Commissie werd slechts een verplichte TDM-uitzondering voorgesteld ten behoeve van niet-commerciële onderzoeksorganisaties en culturele erfgoedinstellingen.²¹ Tijdens de besprekingen in de raads werkgroep is op voorstel van de Nederlandse delegatie een optionele uitzondering toegevoegd die onder voorwaarden TDM voor commerciële doeleinden toestaat. Uiteindelijk heeft het Europees Parlement ook aan die uitzondering verplichte status toegekend. De DSM-richtlijn bevat nu dan ook twee verplichte TDM-bepalingen, artikelen 3 en 4. Beide bepalingen gelden in gelijke mate voor het auteursrecht en het sui generis databankenrecht. De twee beperkingen zijn echter niet even robuust.

Art. 3 voorziet in een vrijstelling voor reproductie- en extractiehandelingen door ‘onderzoeksorganisaties en instellingen voor cultureel erfgoed’. Gezien de minimale schade die rechthebbenden door deze beperking naar verwachting zullen lijden, hoeven de lidstaten niet te voorzien in financiële compensatie.²² Een ‘onderzoeksorganisatie’ is een “universiteit, inclusief de bibliotheken ervan, een onderzoeksinstelling of een andere entiteit die hoofdzakelijk tot doel heeft wetenschappelijk onderzoek te verrichten of onderwijsactiviteiten te verrichten die ook de verrichting van wetenschappelijk onderzoek omvatten”, die handelt zonder winstoogmerk dan wel uit hoofde van een erkende publieke taak (art. 2 lid 1). Een ‘instelling voor cultureel erfgoed’ wordt gedefinieerd als “een voor het publiek toegankelijke bibliotheek of een voor het publiek toegankelijk museum, een archief of een instelling voor cinematografisch of audio(visueel) erfgoed” (art. 2 lid 3). Publieke omroeporganisaties en commerciële onderzoeksinstellingen kunnen dus geen beroep doen op art. 3, maar moeten terugvallen op art. 4.

Behalve het mijnen als zodanig is ook de veilige bewaring van kopieën van gemijnde werken en ander materiaal toegestaan

15 ECS Opinion (noot 10). Zie tevens R. Ducato & A. Strowel, ‘Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to “Machine Legibility”’, IIC (2019) 50:649-684.

16 Zie diverse opstellen in: P.B. Hugenholz (red.), *Copyright Reconstructed: Rethinking Copyright’s Economic Rights in a Time of Highly Dynamic Technological and Economic Change*, Alphen aan den Rijn: Wolters Kluwer 2018.

17 Art. 2 lid 1 (b) Databankenwet. Zie HvJ EG 7 november 2004, C-203/02 (*British Horseracing*), r.o. 76.

18 HvJ EU 9 oktober 2008, C-304/07 (*Directmedia*), r.o. 39; zie tevens HvJ EU 19 december 2013, C-202/12 (*Innoweb*).

19 Art. 1 lid 2 DSM-richtlijn.

20 EC Impact Assessment, p. 104-105; zie tevens overweging 8 DSM-richtlijn.

21 M. Caspers, ‘Een auteursrechtelijke uitzondering voor TDM: is het genoeg?’, AMI 2017, afl. 2, p. 80-86.

22 Overweging 17.

“ten behoeve van wetenschappelijk onderzoek, met inbegrip van de verificatie van onderzoeksresultaten” (art. 2 lid 2). Dit is belangrijk omdat voor empirisch wetenschappelijk onderzoek gebruikte onderzoeksgegevens beschikbaar moeten blijven ter verificatie, voor herhalingsexperimenten en vervolgonderzoek.

Art. 3 lid 1 staat TDM enkel toe ten aanzien van werken of ander materiaal (bijv. databanken) waartoe de begunstigde organisaties “op rechtmatige wijze toegang hebben”. Volgens overweging 14 omvat ‘rechtmatige toegang’ de toegang tot inhoud op grond van contractuele regelingen (bijv. abonnementen of open-access licenties), alsook tot “content die vrijelijk online beschikbaar wordt gesteld”. De eis van ‘rechtmatige toegang’ betekent echter niet dat de rechthebbenden tekst- en datamining contractueel mogen uitsluiten; op grond van art. 7 van de DSM-richtlijn zijn contractsbepalingen in strijd met art. 3 niet afdwingbaar. De mogelijkheid van een ‘opt-out’ wordt enkel voorzien in art. 4. Wel staat het de rechthebbenden vrij om “maatregelen [te] nemen om de veiligheid en de integriteit te verzekeren van de netwerken en databanken waar de werken of andere materialen worden gehost. Deze maatregelen gaan niet verder dan wat nodig is om die doelstelling te verwezenlijken” (art. 3 lid 3). Blijkens overweging 16 zijn dergelijke maatregelen alleen te rechtvaardigen om redenen die verband houden met de veiligheid en integriteit van de systemen of de databanken,²³ en dus niet om zuiver commerciële redenen. Mochten rechthebbenden desalniettemin technische maatregelen toepassen waardoor tekst- en datamining door onderzoekinstellingen wordt belemmerd, dan moeten de lidstaten op grond van art. 7 lid 2, tweede zin, DSM-richtlijn jo. art. 6 lid 4 Auteursrechtrichtlijn hiertegen passende maatregelen nemen.

De tweede TDM-uitzondering van de richtlijn heeft betrekking op een veel breder scala van gebruikers, maar is beperkter van omvang. Art. 4 lid 1 staat in het algemeen reproductie- en extractiehandelingen toe “ten behoeve van tekst- en datamining”, en reproducties en extracties mogen voor datzelfde doel worden behouden. De bepaling staat TDM dus toe voor alle denkbeeldige doeleinden, ongeacht onderliggende commerciële motieven, maar biedt de rechthebbenden wel de mogelijkheid om aan deze vrijstelling te ontsnap-

pen. Op grond van art. 4 lid 3 geldt de beperking enkel voor zover de rechthebbenden zich hun rechten niet uitdrukkelijk hebben voorbehouden, “op een passende wijze, zoals machinaleerbare middelen in het geval van inhoud die online openbaar wordt gemaakt”. Volgens overweging 18 “moet het alleen passend worden geacht deze rechten voor te behouden indien hierbij machinaal leesbare middelen worden gebruikt, waaronder metagegevens en de voorwaarden van een website of een dienst. [...] In andere gevallen kan het passend zijn om rechten voor te behouden met behulp van andere middelen, zoals contractuele overeenkomsten of een eenzijdige verklaring”.

Ergo, rechthebbenden kunnen tekst- en datamining voor commerciële doeleinden effectief verbieden door online metagegevens van het robot.txt-type aan hun inhoud toe te voegen. Daardoor creëert en legitimeert de richtlijn in wezen een afgeleide markt voor tekst- en datamining, die rechthebbenden naar believen kunnen controleren, in licentie geven of zelfs volledig verbieden. Hoewel de meeste contenteigenaren niet geneigd zullen zijn om datamining te verbieden of te gelde te maken, zullen sommige rechthebbenden daar anders over denken. Wetenschappelijke uitgeverij, bijvoorbeeld, zijn zich er terdege van bewust dat hun uitgeefportfolio's een informatieve waarde bezitten die verder gaat dan de gepubliceerde artikelen die zij hebben geaggregeerd. Sommige uitgeverij bieden reeds betaalde TDM-diensten aan, en zullen terughoudend zijn om TDM-licenties aan derden te verlenen. Andere uitgeverij zijn nog bezig met het ontwikkelen van licentiestrategieën om te profiteren van deze opkomende markt.²⁴

Geconcludeerd kan worden dat de TDM-bepalingen van de DSM-richtlijn minder vrijheid voor tekst- en datamining waarborgen dan ze op het eerste gezicht lijken te doen. Door de opt-out clause van art. 4 in het bijzonder worden data-mijners met winstoogmerk overgeleverd aan de willekeur van de rechthebbenden. Dit plaatst commerciële onderzoeksinstituten, journalisten, AI-ontwikkelaars en andere innoverende ondernemers in de EU in een nadelige concurrentiepositie ten opzichte van de Verenigde Staten, waar tekst- en datamining zonder meer geoorloofd is, zelfs als dat gebeurt met winstoogmerk.²⁵

23 Overweging 16, eerste zin, luidt: “Aangezien het aantal verzoeken om toegang tot en de hoeveelheid downloads van hun werken of andere materialen zeer hoog kan zijn, moeten rechthebbenden de mogelijkheid krijgen om maatregelen te nemen wanneer het risico bestaat op aantasting van de veiligheid en integriteit van hun systemen of databanken.”

24 M. Caspers, L. Guibault e.a., ‘Baseline report of policies and barriers of TDM in Europe. FutureTDM’, 2016, p. 84-88.

25 Zie Pam Samuelson, [http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-](http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/)

[mandatory-text-data-mining-exception-restrictive/](http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/). Ook diverse andere landen buiten Europa, zoals Japan en China, maken ruim baan voor TDM; zie Future TDM, p. 72 e.v. Vgl. tevens Handke, C., Guibault, L., & Vallbé, J. J. (2015). Is Europe Falling Behind in Data Mining? Copyright's Impact on Data Mining in Academic Research. In B. Schmidt, & M. Dobrevá (Eds.), *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust: proceedings of the 19th International Conference on Electronic Publishing* (pp. 120-130). Amsterdam: IOS Press.

Artikel 3**Tekst- en datamining met het oog op wetenschappelijk onderzoek**

1. De lidstaten voorzien in een uitzondering op de rechten bedoeld in artikel 5, onder a), en artikel 7, lid 1, van Richtlijn 96/9/EG, artikel 2 van Richtlijn 2001/29/EG, en artikel 15, lid 1, van deze richtlijn voor reproducties en opvragingen door onderzoeksorganisaties en instellingen voor cultureel erfgoed om met het oog op wetenschappelijk onderzoek tekst- en datamining te verrichten op werken of andere materialen waartoe zij rechtmatige toegang hebben.
2. Kopieën van werken of andere materialen die worden gemaakt overeenkomstig lid 1, worden opgeslagen met een passend beveiligingsniveau en mogen worden bewaard voor doeleinden in verband met wetenschappelijk onderzoek, inclusief de verificatie van onderzoeksresultaten.
3. Rechthebbenden kunnen maatregelen nemen om de veiligheid en de integriteit te verzekeren van de netwerken en databanken waar de werken of andere materialen worden gehost. Deze maatregelen gaan niet verder dan wat nodig is om die doelstelling te verwezenlijken.
4. De lidstaten moedigen rechthebbenden, onderzoeksorganisaties en instellingen voor cultureel erfgoed aan om algemeen aanvaarde beste praktijken vast te stellen met betrekking tot de toepassing van de in lid 2 bedoelde verplichting en de in lid 3 bedoelde maatregelen.

Artikel 4**Uitzondering of beperking voor tekst- en datamining**

1. De lidstaten voorzien in een uitzondering op of beperking van de rechten als bedoeld in artikel 5, onder a), en artikel 7, lid 1, van Richtlijn 96/9/EG, artikel 2 van Richtlijn 2001/29/EG, artikel 4, lid 1, onder a) en b), van Richtlijn 2009/24/EG en artikel 15, lid 1, van deze richtlijn voor reproducties en opvragingen van rechtmatig toegankelijke werken en andere materialen met het oog op tekst- en datamining.
2. Reproducties en opvragingen op grond van lid 1 mogen worden bewaard zolang dit nodig is voor tekst- en datamining.
3. De uitzondering of beperking als bedoeld in lid 1 is van toepassing op voorwaarde dat het gebruik van de in dat lid bedoelde werken en andere materialen door de rechthebbenden ervan niet op passende wijze uitdrukkelijk is voorbehouden, zoals machinaal leesbare middelen in het geval van online ter beschikking gestelde content.
4. Dit artikel doet geen afbreuk aan de toepassing van artikel 3 van deze richtlijn.