

KVAN

Kennisdocument
**Tekst- en Datamining
in de Auteurswet**

April 2026

Managementsamenvatting

De snelle opkomst van kunstmatige intelligentie (AI) dwingt erfgoedinstellingen tot het formuleren van beleid. AI-modellen hebben grote hoeveelheden data nodig om te trainen, en digitale archiefcollecties zijn hier een aantrekkelijke bron voor. Het ontbreekt bij veel instellingen nog aan expliciet beleid, wat juridisch betekent dat het 'minen' van deze data door commerciële partijen in beginsel is toegestaan.

Dit kennisdocument beantwoordt de vraag:

"Hoe moeten archiefinstellingen omgaan met het beschikbaar stellen van collectiemateriaal voor AI?"

Het antwoord vraagt om nuance: een generiek beleid van "alles open" of "alles dicht" is niet wenselijk en juridisch onhoudbaar. Een gedifferentieerde aanpak wordt geadviseerd, gebaseerd op de aard van het materiaal en afhankelijk van wie de rechthebbende is.

1. Beleidsuitgangspunten afhankelijk van het type werk

De kern van het advies is dat de archiefinstelling zijn publieke taak voorop stelt en **geen** drempels opwerpt, tenzij de wet of een rechthebbende dit vereist.

- **Publiek Domein & Openbare Overheidsinformatie:** Maak voor dit type materiaal geen voorbehoud voor tekst- en datamining. Bij dit materiaal – dat niet (meer) auteursrechtelijk beschermd is, of overheidsinformatie betreft dat onder de openbaarheidswetgeving valt – weegt de publieke taak van de sector zwaarder dan de gevolgen van een voorbehoud. De instelling kan hier geen 'immaterieel voorbehoud' op plaatsen. Een voorbehoud uit technische noodzaak (bijv. serveroverbelasting), een 'materieel voorbehoud', blijft mogelijk.
- **Materiaal waarvan de instelling rechthebbende is:** Maak voor dit type materiaal geen voorbehoud voor tekst- en datamining. Voor materiaal waarvan de instelling zelf de rechten bezit (rapporten, metadata, eigen publicaties), gelden de regels van de WHO en WOO, een voorbehoud hierop zou wel eens niet mogelijk kunnen zijn.
- **Materiaal waarvan derden rechthebbenden zijn:** Treedt voor dit type materiaal faciliterend op. Bij materiaal van derden – waaronder fotografen en auteurs – kan de instelling niet zelfstandig een auteursrechtelijk (immaterieel) voorbehoud maken. De instelling faciliteert hier de keuze van de rechthebbende.
 - Zwijgt de rechthebbende? Dan is TDM wettelijk toegestaan (opt-out principe).
 - Maakt de rechthebbende bezwaar? Dan moet de instelling dit honoreren en technisch implementeren.

2. Technische implementatie

Elk beleid voor een voorbehoud op Tekst- en Datamining moet machinaal leesbaar zijn. Omdat één perfecte standaard nog ontbreekt, adviseert dit document een hybride aanpak. Maak gebruik van:

- **Robots.txt:** Voor grofmazige toegangcontrole op siteniveau (de 'poortwachter').
- **TDM Reservation Protocol (TDMRep):** Voor fijnmazige rechtendauiding op objectniveau (de 'boekhouder').

Aangevuld met een expliciete passage in de algemene voorwaarden van de website. Een zin in een disclaimer of footer van de website wordt niet aangeraden.

1.1. 3. Actieve communicatie

Omdat de wet uitgaat van een 'opt-out' systeem, heeft de instelling een zorgplicht om rechthebbenden te informeren. Het document adviseert een transparante communicatiestrategie:

1. **Informeert:** Leg op de website uit dat collecties gebruikt kunnen worden voor AI.
2. **Faciliteert:** Biedt een duidelijk meldpunt waar rechthebbenden een opt-out kunnen aanvragen.
3. **Effectueert:** Koppel deze meldingen in het beheersysteem direct aan de technische standaarden (TDMRep), zodat de wens van de maker automatisch wordt gerespecteerd door AI-bots.

Inhoud

Managementsamenvatting	1
1. Beleidsuitgangspunten afhankelijk van het type werk	1
2. Technische implementatie	1
3. Actieve communicatie	2
Inhoud	3
Introductie	5
Privacy en eigendomsrechten	5
Hoofdvraag	6
Dank aan het Stadsarchief Amsterdam	6
Hoofdstuk 1: AI binnen de erfgoedsector	7
Wat is AI?	7
Technische aspecten van AI	7
Gebruik AI in de erfgoedsector	8
Hoofdstuk 2. Juridisch kader	9
Relevante juridische termen en wetgeving	9
Tekst- en Datamining (TDM)	11
Publieke raadpleging beperkt openbare archiefbescheiden	13
Juridische analyse	13
Rechtszaken	15
LAION v Robert Kneschke	15
DPG Media v. HowardsHome	15
Bologportal v. Redata	15
Aansluitovereenkomst GPT-NL	16
Hoofdstuk 3. Standaarden voor machinaal voorbehoud	17
Robots.txt	17
TDM reservation protocol	17
Hoofdstuk 4. Praktische voorbeelden uit de sector	19
Stadsarchief Amsterdam	19
Maria Austria Instituut	20
ANPFoto	20
Nederlands Fotomuseum	21
Rijksoverheid	21
Pictoright	21
Koninklijke Bibliotheek, nationale bibliotheek	22

Nationaal Archief	23
Hoofdstuk 5. Communicatie	24
Hoofdstuk 6. Conclusies	25
Principes	25

Introductie

Elke erfgoedorganisatie moet zich vandaag de dag verhouden tot de rappe opkomst van kunstmatige intelligentie (AI). Erfgoedinstellingen ontwikkelen beleid en starten projecten waarin actief wordt gekeken naar de inzet van deze technologie, de mogelijke meerwaarde ervan en de grenzen van de huidige techniek.

De archiefinstelling heeft een maatschappelijke taak in het behouden, beheren en beschikbaar stellen van archief- en erfgoedinformatie. Dit verhoudt zich direct tot de snelle ontwikkelingen in AI. Bij brede digitale beschikbaarstelling wordt materiaal vaak ook opgehaald voor de training van AI-modellen, vaak zonder dat de instelling zich hiervan bewust is en zonder dat er expliciet beleid voor is geformuleerd. Zonder een dergelijk beleid – en het bijbehorende juridische voorbehoud – is het 'minen' van dit materiaal in beginsel wettelijk toegestaan.

Het onbeperkt laten minen van auteursrechtelijk beschermd materiaal kan problemen veroorzaken. Enerzijds kan het leiden tot een overmatige belasting van technische systemen.¹ Anderzijds ontstaan er juridische risico's wanneer een rechthebbende niet wil dat diens werk wordt gebruikt voor AI-training, maar het archief dit (onbedoeld) toch toestaat.

Privacy en eigendomsrechten

De regels rondom tekst- en datamining (TDM) vallen onder het intellectueel eigendomsrecht, zoals het auteursrecht, naburige rechten en het databankenrecht. Binnen dit kader heeft de erfgoedinstelling een actieve, sturende rol: zij faciliteert de beschikbaarstelling en communiceert de voorwaarden waaronder het collectiemateriaal mag worden gebruikt. De instelling kan voorbehouden maken en derden te beperken in bepaalde handelingen, zoals commerciële tekst- en datamining.

Dit is een fundamenteel andere positie dan de rol die een instelling heeft ten aanzien van de Algemene Verordening Gegevensbescherming (AVG). Onder de AVG is elke partij die gegevens verwerkt zelf verantwoordelijk voor die verwerking. Dit betekent dat een partij die data verzamelt (minet), zelf een eigen grondslag moet hebben en moet voldoen aan de AVG-voorwaarden. Het is niet de taak of de bevoegdheid van de erfgoedinstelling om via TDM-beleid beperkingen op te leggen aan dit verdere gebruik van persoonsgegevens door derden. De verantwoordelijkheid van de erfgoedinstelling beperkt zich tot de eigen rechtmatige verwerking en de beslissing om de data al dan niet beschikbaar te stellen; de verantwoordelijkheid voor privacy-compliance bij het minen en verwerken van persoonsgegevens verschuift naar de partij die de mining uitvoert.

Dit document gaat dan ook niet in op de privacy-aspecten van het gebruik van persoonsgegevens door derden. Het richt zich op het voorbehoud dat de erfgoedinstelling kan maken ten aanzien van intellectueel eigendom. Voor een breder overzicht van het gebruik van collectiemateriaal voor AI verwijzen wij naar het kennisdocument Erfgoed en AI.²

¹ [Are AI Bots Knocking Cultural Heritage Offline?](#)

² [Kennisdocument 'AI en Erfgoed' gepubliceerd - KVAN](#)

Hoofdvraag

Dit kennisdocument stelt de vraag:

"Welk beleid moeten archiefinstellingen hanteren om tekst- en datamining (TDM) op hun collecties conform het intellectueel eigendomsrecht en technisch uitvoerbaar te faciliteren?"

Het document biedt de benodigde context en geeft praktische handvatten voor de bewegingsruimte van het archief. Eerst wordt een technisch en theoretisch kader geschetst, gevolgd door voorbeelden uit de sector, waarna een advies volgt over hoe een archief zich kan verhouden tot deze theorie en praktijk.

Dank aan het Stadsarchief Amsterdam

Dit kennisdocument komt mede voort uit een samenwerking tussen Maarten Zeinstra (Coördinator Auteursrecht voor KVAN) en het Stadsarchief Amsterdam. KVAN bedankt het Stadsarchief voor het delen van de gezamenlijk ontwikkelde inzichten en perspectieven.

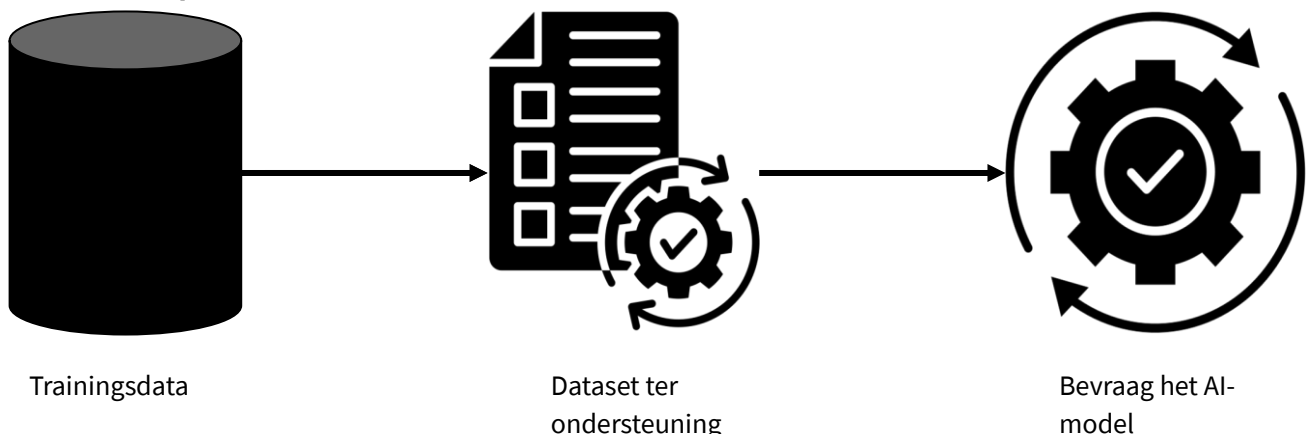
1. AI binnen de erfgoedsector

1.1 Wat is AI?

Iedereen heeft wel een idee van wat kunstmatige intelligentie, artificiële intelligentie of AI is.³ Populaire media hebben ons decennialang dystopische beelden voorgespiegeld van de impact van kunstmatige intelligentie op de samenleving. Kunstmatige intelligentie is zeker (nog) niet menselijk, we hebben het in dit document dan ook niet over algemene artificial intelligence, zoals we die kennen van films als ‘I-Robot’, ‘A.I.’, ‘Her’, etc. Het is ook niet duidelijk of we hier ooit kunnen komen en hoe wenselijk dat is.

Dit begrip van AI is zeker ten dele afgeleid van, en ontstaan uit deze media. Wanneer we daarnaast ook kijken naar wet- en regelgeving, ontstaat er een brede praktische definitie van AI. In plaats daarvan zien we de ontwikkeling van complexe algoritmen waarvan wij niet langer makkelijk de uitkomst kunnen voorstellen. Deze toepassingen zijn inmiddels wel gemeengoed geworden, zowel in de zakelijke markt als in dagelijks privégebruik. Bijvoorbeeld de diverse generatieve AI’s, zoals Gemini en ChatGPT en de afbeelding-, audio- en videogeneratoren (ook wel ‘deep fakes’).

Technische aspecten van AI



Bij AI zijn er dan vaak drie componenten aanwezig: trainingsdata, modellen en algoritmen.

Trainingsdata – Trainingsdata kan van alles zijn, van teksten, geluid tot afbeeldingen. Hoe beter de data beschreven zijn in metadata, hoe waardevoller deze data is om AI mee te trainen. Bij deze trainingsdata kunnen allerlei rechten van toepassing zijn, zoals het auteursrecht en het privacyrecht. Deze trainingsdata is het hoofdonderwerp van dit kennisdocument.

Datasets – Het trainen van AI met een algoritme levert een model op – een groot bestand met getallen. Het algoritme vertaalt patronen uit trainingsdata naar dit model. Dit model kan gezien worden als een zeer vereenvoudigde wiskundige weergave van de werkelijkheid, bruikbaar om voorspellingen te maken hoe (tekstuele informatie) verder loopt. Aan de hand van de trainingsdata kan een model bijvoorbeeld voorspellen wat het volgende deel is van deze zin “het regent vandaag, ik moet niet vergeten om ...”. waarop het kan zeggen “een regenjas aan te trekken.”, of “een paraplu mee te nemen.”

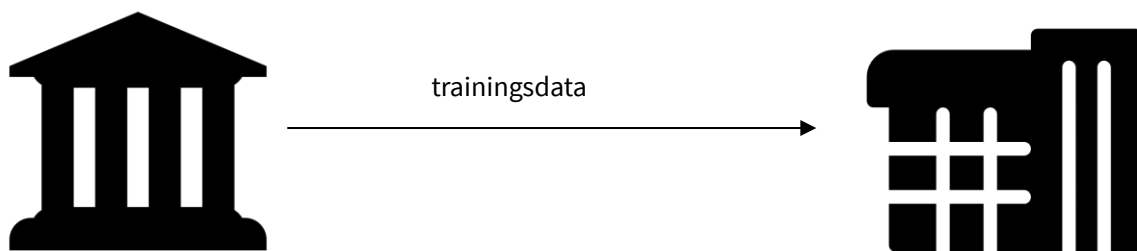
³ Dit hoofdstuk is grotendeels afgeleid van het gelijknamige hoofdstuk in het [Kennisdocument AI en Erfgoed](#).

Algoritme – Een algoritme volgt stapsgewijs een procedure om een specifieke taak uit te voeren of een probleem op te lossen. Het kan zelf een symbolische AI of een statistische analyse zijn die op een dataset van toepassing is. Ook kan het algoritme modellen interpreteren om deze uit te voeren. In een archiefcontext zijn algoritmen onderdeel van informatie-objecten met bewaartermijnen die uit selectielijsten onder de Archiefwet naar voren komen. Ook kunnen implementaties van algoritmen, oftewel computerprogramma's, auteursrechtelijk beschermd zijn.

1.2. Gebruik AI in de erfgoedsector

AI-toepassingen worden in erfgoedinstellingen op verschillende manieren gebruikt. AI wordt als technologie ingezet (denk aan tekstherkenning, het genereren van ondertiteling, herkenning van elementen in beeld en vertalingen)⁴, de uitkomsten van het gebruik van de technologie worden gebruikt en er wordt (onbewust) trainingsmateriaal aangeboden.

Erfgoedinstellingen zijn de hoeders van waardevolle data. Deze data worden bewust – en onbewust – beschikbaar gesteld voor het trainen en vormen van kunstmatige intelligentie. Wanneer materiaal online staat, kunnen derden deze data (zonder expliciete toestemming) kopiëren en gebruiken als input voor tekst- en datamining, voor het trainen van modellen van kunstmatige intelligentie.



Figuur. Derden gebruiken data van erfgoedinstellingen als trainingsdata.

⁴ Zie ook <https://pro.europeana.eu/project/ai4culture-an-ai-platform-for-the-cultural-heritage-data-space> voor meer voorbeelden.

2. Juridisch kader

AI-bedrijven, waaronder bijvoorbeeld Google, Anthropic, Aleph Alpha, Mistral AI en OpenAI maken gebruik van materiaal dat op het open internet beschikbaar is om hun modellen te trainen. Hiervoor maken zij kopieën van onder andere beeldbanken en documenten die zij op internet vinden. Deze worden vervolgens geanalyseerd en in stukjes opgeknipt om gebruikt te kunnen worden voor het trainen van o.a. grote (multimodale) modellen (multimodal large language models of LLMs).

Dit document behandelt het impliciet aanbieden van digitaal en gedigitaliseerd collectiemateriaal voor deze doelstelling door archiefinstellingen. Als materiaal online staat, kunnen derden deze data (zonder expliciete toestemming) kopiëren, opknippen, en gebruiken als input voor tekst- en datamining, voor het trainen van modellen van kunstmatige intelligentie. Expliciet niet behandeld in dit document zijn eigen toepassingen van AI binnen de organisatie en het gebruik van producten van AI. Ook wordt de collectie die (nog) niet gedigitaliseerd is, niet behandeld.

2.1 Relevante juridische termen en wetgeving

De vraag of de instelling wil, kan of mag toestaan dat collectiemateriaal gebruikt wordt voor het trainen van de modellen voor AI-toepassingen, wordt beantwoord door toepasselijke wetgeving. De volgende juridische termen en wetten moeten hierbij in gedachten gehouden worden.

Werk – Het auteursrecht is een recht dat over een intellectuele creatie gaat. Voor het auteursrecht maakt het niet uit of het gaat om het originele werk of een raadpleegkopie, of de tekst machinaal leesbaar is of niet. Een foto van een gedicht of een transcriptie van datzelfde gedicht zijn voor het auteursrecht hetzelfde werk.

Openbaarheid – De Archiefwet, Wet Open Overheid (Woo) en Wet Hergebruik Overheidsinformatie (Who) regelen alle drie de openbaarheid van overheidsinformatie, maar op verschillende momenten. De Woo regelt de openbaarheid van documenten die nog bij een overheidsorganisatie liggen. Zodra archiefbescheiden zijn overgebracht naar een archiefbewaarplaats, is de Woo niet meer van toepassing en geldt het openbaarheidsregime van de Archiefwet. De Woo en de (nieuwe) Archiefwet 20XX sluiten nauw op elkaar aan. De uitzonderingsgronden in de Woo komen vrijwel overeen met de beperkingsgronden in de nieuwe Archiefwet, wat het gemakkelijker maakt om bij de overbrenging te bepalen of er (nog steeds) beperkingen nodig zijn. De Who is een aanvullende wet die zich richt op het **hergebruik** van voornamelijk niet-overgebrachte overheidsinformatie (waaronder uitgeplaatste informatie) en overgebracht openbare overheidsinformatie. De Who stelt dat deze op een technisch herbruikbare manier beschikbaar moet worden gesteld.

Voorwaarden voor hergebruik openbare overheidsinformatie – De voorwaarden die gesteld mogen worden op het hergebruik van openbare overheidsinformatie, zijn beperkt.⁵ Voorwaarden voor het hergebruik van openbare overheidsinformatie moeten objectief, evenredig en niet-discriminerend zijn, en gerechtvaardigd worden door een doel van algemeen belang. Ook mogen deze voorwaarden het hergebruik en mededinging niet nodeloos beperken.

⁵ Volgende uit artikel 17 lid 1 Archiefwet 1995 / Artikel 8.2 lid 5 Archiefwet 20xx is artikel 6 van de Wet Hergebruik Overheidsinformatie van toepassing.

Van belang is ook dat er voorwaarden zijn verbonden aan hergebruik van persoonsgegevens. Onder meer mag hergebruik niet ingezet worden voor het ongedaan maken van pseudonimisering of voor het opnieuw identificeren van betrokkenen. Overheidsinstellingen en andere publieke organisaties zijn verantwoordelijk voor het nemen van voldoende maatregelen om dergelijk verboden hergebruik van persoonsgegevens tegen te gaan.⁶

Auteurswet en Wet op de naburige rechten – De Auteurswet geeft makers het exclusieve recht om hun werk te kopiëren, te verspreiden en te exploiteren. Dit geldt voor originele literaire, wetenschappelijke of artistieke werken. De duur van het auteursrecht is doorgaans 70 jaar na de dood van de maker. Voor anonieme of pseudonieme werken is dit 70 jaar na de eerste rechtmatige openbaarmaking. De **Wet op de naburige rechten** is vergelijkbaar, maar richt zich op de rechten van uitvoerende kunstenaars (zoals acteurs en muzikanten), producenten van fonogrammen, filmproducenten en omroeporganisaties. Dit recht kan naast het auteursrecht bestaan op hetzelfde werk.

Databankenwet – De Databankenwet beschermt de producent van een databank die een substantiële investering heeft gedaan in het verkrijgen, controleren of presenteren van de inhoud ervan. Dit recht geldt 15 jaar en beschermt de systematische of methodische ordening van de databank, niet de individuele elementen erin. De organisaties van metadata en paradata in collectiebeheersystemen kunnen onder het databankenrecht vallen. De hele website van een erfgoedinstelling kan als een databank worden gezien. De Databankenwet kent dezelfde regels voor een TDM-voorbehoud als de Auteurswet.

AI verordening – De Europese AI-verordening is primair bedoeld als productveiligheidswetgeving om de regels vast te stellen waaraan AI-systemen moeten voldoen, vergelijkbaar met de regels voor het op de markt brengen van een auto of een magnetron. Het categoriseert AI-systemen op basis van risico. Het verbiedt systemen met een onaanvaardbaar risico, waaronder bijvoorbeeld het manipuleren van mensen of het willekeurig verzamelen van gezichten voor gezichtsherkenning. Hoog-risico systemen, bijvoorbeeld in biometrie of rechtspraak, moeten aan strikte eisen voldoen. Voor algemene AI-systemen (zoals ChatGPT) en hoog-risico systemen, is er een transparantieverplichting over de herkomst van de trainingsdata, zodat auteursrechthebbenden dit kunnen controleren. Deze laatste transparantieverplichting is om ex post facto (na de feiten) de handhaving van de auteursrechtelijke TDM-uitzondering uit de Auteurswet te vergemakkelijken.

Publiek Domein – Het publiek domein bestaat uit werken waarop het auteursrecht verlopen is, werken die niet creatief genoeg zijn om in aanmerking te komen voor auteursrechtelijke bescherming, waaronder ook ideeën, feiten en informatie, en uitgesloten onderwerpen als wetteksten en administratieve beslissingen⁷. De regels van Tekst- en Datamining zijn alleen van toepassing op materiaal waarop het auteursrecht, naburig recht, of databankenrecht van toepassing is. Bij publiek domein werken betekent dat dan ook dat er geen auteursrechtelijke grond is om het trainen tegen te gaan.⁸ Auteursrechtelijk beschermde informatie die door de openbare macht beschikbaar is gesteld,

⁶ Wet Hergebruik Overheidsinformatie, artikel 6 lid 3

⁷ Artikel 11 Auteurswet

⁸ Er zijn altijd contractuele afspraken mogelijk, de handhaafbaarheid daarvan is problematisch. Daarnaast moet er ook rekening gehouden worden met de AVG, deze wordt niet behandeld van dit document.

mag zonder toestemming verveelvoudigd worden tenzij deze expliciet voorbehouden wordt. Een TDM-voorbehoud is zo'n voorbehoud.

Beperkt openbaar materiaal – Ook de door de archiefwet geregelde openbaarheid heeft indirect impact op deze regels. Materiaal dat beperkt openbaar is mag immers niet (online) publiek toegankelijk zijn en kan daarmee ook niet beschikbaar worden gemaakt voor het trainen van AI (door derde partijen).

2.2 Tekst- en Datamining (TDM)

Tekst- en datamining (TDM) is een geautomatiseerde analysetechniek die digitale teksten en gegevens ontleedt om patronen, trends en verbanden te ontdekken. TDM is de auteursrechtelijke handeling die uitgevoerd wordt bij het trainen van ChatGPT, Gemini, Claude, Perplexity, Mistral, GPT-NL, etc. Om deze analyse uit te kunnen voeren moet een kopie gemaakt worden van trainingsmateriaal. Vaak worden deze kopieën gemaakt zonder dat daarvoor toestemming wordt gevraagd, en in veel gevallen is dat ook niet nodig.

TDM wordt ook gebruikt binnen erfgoedinstellingen. Denk bijvoorbeeld aan Transkribus, ABBYY en Loghi (OCR en HTR), maar ook aan applicaties die Named Entity Recognition (NER) inzetten. In essentie is dit eenzelfde auteursrechtelijke handeling. Er is daar immers ook een kopie nodig. Hieronder wordt uiteengezet dat het wel uitmaakt wie de kopie maakt.

Het uitgangspunt van het intellectueel eigendomsrecht is dat voor een kopie van beschermd materiaal toestemming nodig is. En bij TDM wordt een kopie gemaakt. Om TDM te reguleren zijn er in 2021 twee uitzonderingen in de wetgeving geïntroduceerd, een voor erfgoedinstellingen en onderzoekinstellingen en de andere voor het algemeen publiek.

Culturele erfgoedinstellingen – Het archief, als culturele erfgoedinstelling⁹, mag altijd zonder toestemming tekst- en datamining uitvoeren voor wetenschappelijk onderzoek op werken waartoe het rechtmatige toegang heeft, zelfs als dat recht wordt voorbehouden – dat de rechthebbende geen toestemming geeft om de kopie te maken. Dit is dan een uitzondering van 'dwingend recht', contractuele bepalingen die dit verbieden zijn nietig. Het archief mag de informatie die nodig is voor de analyse bewaren, mits het passende beschermingsmaatregelen treft en de toegang beperkt tot gebruikers met onderzoeksdoeleinden. De brondata zelf mag niet gedeeld worden, maar de resultaten van de analyse wel.

Andere partijen – commerciële partijen en andere gebruikers mogen TDM toepassen, maar alleen als de rechthebbende geen voorbehoud heeft gemaakt – de rechthebbende moet expliciet zeggen dat anderen geen kopie mogen maken om te kunnen TDM'en. Dit voorbehoud moet op een passende, machinaal leesbare manier gecommuniceerd worden, bijvoorbeeld via de algemene voorwaarden van een website of een machinaal leesbaar formaat.

Let wel: de *rechthebbende* heeft het recht om dit voorbehoud te maken. Deze rechten gaan in veel gevallen niet over naar het archief. Het archief is zelf niet rechthebbende van een groot deel van de collectie. Strikt genomen kan de archiefinstelling dan ook geen TDM-voorbehoud maken gebaseerd op intellectueel eigendom zonder verzoek van de rechthebbende. Het voorbehoud op TDM moet van de

⁹ Volgende uit artikel 25a lid 4 Auteurswet

rechthebbende van het collectiemateriaal komen. Het is aan het archief om dit goed bij te houden, te registreren en met de archiefvormer te bespreken.¹⁰

Machinaal leesbare voorbehouden – Een TDM-voorbehoud moet uitdrukkelijk op passende wijze worden voorbehouden, zoals door middel van machinaal leesbare middelen.¹¹ Denk hierbij aan de inzet van technische standaarden en algemene voorwaarden. Ofwel, wanneer je niet wilt dat anderen TDM'en dan moet je communiceren op zo'n manier dat het leesbaar is voor de doelgroep.

Voorbehoud op de kopie (materieel voorbehoud) – Een materieel voorbehoud is gericht op het corpus mechanicum, de daadwerkelijke, fysieke digitale kopie van een werk. Dit voorbehoud is contractueel van aard en heeft betrekking op het specifieke exemplaar van het werk dat op een bepaalde locatie (zoals een website) wordt aangeboden. Dit kan bijvoorbeeld in de algemene voorwaarden opgenomen worden en hoeft niet gebaseerd te zijn op enig intellectueel eigendom. Dit betekent alleen wel dat de beperking alleen geldt voor de specifieke kopie, en niet voor alle mogelijke kopieën van het werk. Zelfs als het werk tot het publieke domein behoort, kan zo'n voorbehoud worden toegepast.

Voorbehoud op het werk (immaterieel voorbehoud) – Een immaterieel voorbehoud is daarentegen gericht op het corpus mysticum, het immateriële, abstracte werk zelf. Dit type voorbehoud is direct gerelateerd aan het auteursrecht, naburig recht of databankenrecht en stelt de rechthebbende in staat om het recht op TDM voor te behouden. Dit voorbehoud geldt dan voor alle mechanische reproducties van het werk, ongeacht waar ze zich bevinden of wie de kopie maakt. Tenzij dit uiteraard anders is overeengekomen.

Kenmerk	Materieel Voorbehoud	Immaterieel Voorbehoud
Grondslag	Contractueel (Algemene voorwaarden)	Wettelijk (Auteurswet, art. 15n)
Waarop van toepassing?	Op de kopie van een werk dat op de site staat.	Op het intellectuele werk zelf.
Wie bepaalt?	De aanbieder van de kopie (het archief).	De rechthebbende (bv. fotograaf, schrijver, Pictoright).
Juridische Status	De juridische houdbaarheid is niet getest en moeilijk handhaafbaar.	Sterker wettelijk verankerd. Dit is de officiële "opt-out" die de wetgever heeft gecreëerd.
Voorbeeld	"Het is verboden om een kopie van dit document te gebruiken voor AI-training." (Een regel van het archief).	"De fotograaf doet een voorbehoud in de zin van artikel 15o, Auteurswet. De fotograaf verbiedt het gebruik van de foto voor AI-training." (Een recht van de maker).

¹⁰ In de casus van de Koninklijke Bibliotheek, de Nationale Bibliotheek, wordt hier nog een suggestie over gedaan.

¹¹ Artikel 15n lid 1, Auteurswet

2.3 Publieke raadpleging beperkt openbare archiefbescheiden

Naast de kaders van het intellectueel eigendomsrecht, legt de aanpassing van de **Archiefwet** voor het beschikbaar stellen van beperkt openbare archiefbescheiden die persoonsgegevens bevatten en door de minister aangewezen zijn voor publieke raadpleging – het CABR archief – een specifieke inspanningsverplichting¹² bij de archiefdienst wanneer collecties online beschikbaar worden gesteld wanneer daar persoonsgegevens in voorkomen.

De beheerder van een archiefbewaarplaats is verplicht¹³ om passende maatregelen te nemen zodat de rechten en vrijheden van betrokkenen niet onevenredig worden geschaad door online beschikbaarstelling. In de praktijk betekent dit dat de beheerder in ieder geval maatregelen moet nemen om het volgende te voorkomen:

- Ongeoorloofde indexerings;
- Ongeoorloofde tekst- en datamining zoals bedoeld in de Auteurswet;
- **Grootschalig kopiëren** van archiefbescheiden of gegevens daaruit.

Interpretatie: 'Zoveel mogelijk' De memorie van toelichting bij de wet verduidelijkt dat de zinsnede "zoveel mogelijk" duidt op een **inspanningsverplichting**. Van de beheerder wordt verwacht dat hij – voor zover dat technisch redelijkerwijs mogelijk is – alle redelijke maatregelen neemt om ongeoorloofde mining of kopiëren tegen te gaan. Hiermee erkent de wetgever dat absolute preventie technisch niet altijd haalbaar is, maar dat een passieve houding niet volstaat.

Samenhang met de zorgplicht Deze verplichting uit de Archiefwet sluit aan bij de algemene zorgplicht van het archief om de belangen van betrokkenen en rechthebbenden te beschermen. Waar de Auteurswet de nadruk legt op het economisch belang van de maker, richt de Archiefwet zich hier op de bescherming van de burger wiens gegevens in het archief zijn opgenomen.

Voor de uitvoering van deze hele specifieke categorie collecties betekent dit dat het implementeren van machinaal leesbare voorbeholden (zoals beschreven in Hoofdstuk 3) niet alleen een auteursrechtelijke keuze is, maar voor bepaalde collecties ook een **wettelijke opdracht** vormt om aan de eisen van de Archiefwet te voldoen.

2.4 Juridische analyse

Het antwoord op de hoofdvraag "Welk beleid moeten archiefinstellingen hanteren om tekst- en datamining (TDM) op hun collecties conform het intellectueel eigendomsrecht en technisch uitvoerbaar te faciliteren?" valt uiteen in meerdere mogelijke antwoorden. Afhankelijk van wie de rechthebbende is.

Overheid als rechthebbende. Een overheid is, met uitzondering van wetten, besluiten en verordeningen, rechterlijke uitspraken en administratieve beslissingen¹⁴, auteursrechthebbende op literaire en artistieke werken die zij zelf gemaakt heeft. Een overheid is ook rechthebbende over alle rapporten en ingekochte auteursrechtelijk beschermde werken die onder de ARVODI vallen.¹⁵ Een

¹² [Wetsvoorstel 36.889](#) – Verruiming mogelijkheden raadplegen en gebruik van archiefbescheiden die persoonsgegevens bevatten – Volgende uit de memorie van toelichting

¹³ [Wetsvoorstel 36.889](#) – Verruiming mogelijkheden raadplegen en gebruik van archiefbescheiden die persoonsgegevens bevatten – Artikel 17, zesde lid, onder a

¹⁴ Artikel 11

¹⁵ Artikel 23 lid 1 ARVODI-2025

overheid is ook rechthebbende over geproduceerde databanken, waaronder de collectiesite van archieven. De overheid is hiermee auteursrechthebbende van vele werken.

Een rechthebbende mag een voorbehoud maken op tekst- en datamining.¹⁶ Op deze werken kan zij een TDM-voorbehoud doen.¹⁷ Doet zij niets, dan doet zij ook geen voorbehoud en mag er getraind worden op het materiaal. Of een dergelijk voorbehoud passend is ten aanzien van Woo/Who is de vraag. De Woo is maar beperkt toepasbaar. Op moment van schrijven¹⁸ vallen grote categorieën werken nog niet onder de Woo. Het lijkt erop dat deze in 2026 toegevoegd gaan worden.¹⁹ Om toekomstbestendig te adviseren wordt van dit scenario uitgegaan. Daarnaast richt dit document zich niet op de vraag of een document openbaar is of niet, maar indien het openbaar is of het dan ook gebruikt mag worden voor tekst- en datamining.

De Wet Hergebruik Overheidsinformatie kan de bewegingsruimte van de overheid ten aanzien van het maken van een voorbehoud verder beperken. Zo mag onder de Who hergebruik niet aan voorwaarden worden onderworpen, tenzij die voorwaarden objectief, evenredig en niet-discriminerend zijn, en gerechtvaardigd worden door een doel van algemeen belang²⁰, of aan voorwaarden die de mogelijkheden tot hergebruik nodeloos beperken of die bedoeld zijn om de mededinging te beperken²¹.

Een TDM-voorbehoud op informatie van de overheid moet gerechtvaardigd worden. Onder een doel van algemeen belang wordt in ieder geval begrepen de bescherming van de goede uitvoering en continuïteit van een publieke taak. Een mogelijk auteursrechtelijk voorbehoud zou bijvoorbeeld gemaakt kunnen worden omdat de systemen waarop de informatie staat overbelast raken.²² Daarmee ontstaat wel de beleidsvraag of overheden een TDM-voorbehoud mogen inzetten op overheidsinformatie. Zij kunnen ook een materiaal voorbehoud maken zonder dat dit rust op het auteursrecht.

Derde als rechthebbende In voornamelijk de erfgoedcollecties van archiefinstellingen is de instelling zelf geen rechthebbende. Hier zijn de Who en de Woo minder van toepassing, het antwoord moet dan voornamelijk gevonden worden in de auteurswet en aanverwante wetten. Wil de rechthebbende een immaterieel voorbehoud instellen dan moet de archiefinstelling deze ook implementeren – of het werk niet online beschikbaar stellen. In deze situatie moet de archiefinstelling wel van deze opt-out op de hoogte gesteld worden.

Het grootste voorbeeld waarbij derden aan erfgoedinstellingen om een dergelijk immaterieel voorbehoud in te zetten is de koepelovereenkomst Periodieken. Alle instellingen die de koepelovereenkomst Periodieken getekend hebben moeten zich houden aan regels rondom het beschikbaar stellen van periodieken die geregistreerd staan in het portaal van de EUIPO.²³ In deze

¹⁶ Artikel 15o Auteurswet

¹⁷ Artikel 15o

¹⁸ September 2025

¹⁹ Zie [Woo-informaticategorieën en definities | Open Overheid](#)

²⁰ Artikel 6 lid 1, WHO & Artikel 5 lid 2 Datagovernanceverordening (2022/868)

²¹ Artikel 6 lid 2, WHO & Artikel 5 lid 2 Datagovernanceverordening (2022/868)

²² Zie [Are AI Bots Knocking Cultural Heritage Offline?](#)

²³ <https://euipo.europa.eu/out-of-commerce>

overeenkomst tussen de KB, nationale bibliotheek, en Pictoright en Lira zijn regels afgesproken om periodieken die niet langer in de handel zijn voor niet commerciële doeleinden te gebruiken.

Zij zijn gebonden aan artikel 8.2 van de overeenkomst:

“De KB en de toegetroeden Erfgoedinstellingen zullen door middel van een passend machinaal leesbaar rechtenvoorbehoud kenbaar maken dat de Periodieken niet mogen worden gebruikt voor tekst- en datamining met een commercieel oogmerk in de zin van artikel 15o Auteurswet en artikel 4 DSM-richtlijn, daaronder mede begrepen gebruik voor AI trainingsdoeleinden.”

Deelnemers aan de koepel zijn verplicht om een materieel voorbehoud in te stellen voor periodieken die onder het convenant beschikbaar worden gesteld.

2.5 Rechtszaken

Er is veel te doen rondom het trainen van AI. Elke week zijn er ergens op de wereld nieuwe uitspraken die verdere kaders geven aan wat AI-trainers mogen doen. Dit document zal daarmee voor het einde van dit jaar weer achterhaald zijn. Drie rechtszaken worden uitgelicht. Deze geven nog wat additionele kaders over hoe we TDM kunnen interpreteren.²⁴ Let op, al deze rechtszaken zijn van lagere gerechtshoven. Ze geven een richting wat je kan verwachten maar ze geven geen breed geaccepteerde consensus van de gerechtshoven.

LAION v Robert Kneschke

De LAION v Robert Kneschke²⁵ was de eerste grote rechtszaak rondom AI-training in Europa. LAION, een onderzoeksbureau, maakt tijdelijke kopieën van afbeeldingen die vindbaar zijn op het internet en analyseert deze (Named Entity Recognition). LAION deelt een dataset met uitkomsten met het publiek. Volgens de rechter is een TDM opt-out niet van toepassing, LAION is een onderzoekinstelling, het hoeft zich daardoor niet aan de opt-out te houden.

DPG Media v. HowardsHome

De Nederlandse rechtszaak tussen HowardsHome en DPG Media. Howards Home maakt, persoverzichten maakt, gebaseerd op materiaal van o.a. DPG Media.²⁶ Een van de aspecten van de rechtszaak was hoe een TDM opt-out gecommuniceerd kan worden. De uitspraak maakt dit duidelijk: dit moet machinaal leesbaar zijn, maar de rechter ziet de algemene voorwaarden van een website ook als machinaal leesbaar. Let wel, deze opt-outs moeten precies zijn. een opt-out for ChatGPT, is geen opt-out voor Anthropic.

Boligportal v. Redata

De Deense rechtszaak lijkt op DPG Media v. HowardsHome. Boligportal is een organisatie voor het vinden van huurhuizen in Denemarken. In dit geval gebruikte Redata de informatie van Boligportal om een eigen service aan te bieden aan banken en investeerders. Boligportal maakte geen gebruik van robots.txt. Volgens de rechter was dit ook niet nodig. De algemene voorwaarden zouden een voldoende opt-out communiceren.

²⁴ Andere rechtszaken rondom AI zoals de Duitse GEMA v. OpenAI over memorisatie en Hongaarse over TDM en zoekmachines zijn niet meegenomen.

²⁵ https://drive.google.com/file/d/1A_vnSJUwlrVovhIqA4rKIFaOktR4TvBt/view

²⁶ zie <https://uitspraken.rechtspraak.nl/details?id=ECLI:NL:RBAMS:2024:6563>

Aansluitovereenkomst GPT-NL

De Europese en nationale wetgeving rondom TDM-uitzonderingen en opt-outs geven een basis om zonder expliciete toestemming materiaal te gebruiken om TDM te doen. Je kan natuurlijk ook toestemming vragen, zoals het Nederlandse initiatief GPT-NL dat doet. Zij hanteren een contractueel model gebaseerd op actieve deelname (opt-in).

Instellingen die data bijdragen aan dit Nederlandse taalmodel, tekenen een Content Contributor Agreement.²⁷ Dit is een overeenkomst die de afspraken maakt tussen de data-leverancier, de erfgoedinstelling, en TNO, als ontwikkelaar van GPT-NL.

De kernpunten van deze overeenkomst voor erfgoedinstellingen zijn:

- **Opt-in** – met de overeenkomst geeft de instelling expliciet toestemming voor gebruik van data.
- **Licentie, geen overdracht** – De instelling blijft eigenaar van de data (auteursrechthebbende). Er wordt een licentie verstrekt aan GPT-NL om de data te gebruiken voor het trainen, evalueren en testen van het model.
- **Vrijwaring en Rechten** – De leverancier moet garanderen dat zij bevoegd is om de data te delen en de licentie te verstrekken.
 - **Let op:** Als een archief materiaal van derden (waarvan zij niet de auteursrechthebbende is) inbrengt, moet het archief zeker weten dat zij hiervoor een licentie mag verstrekken. Sommige instellingen kiezen er dan ook voor om alleen publiek domein materiaal aan te leveren.
- **Tegenprestatie en Governance** – In ruil voor de data kunnen leveranciers toetreden tot de 'Content Board', waarmee zij inspraak krijgen in het beheer en de toekomst van het model. Ook bevat de overeenkomst clausules over een mogelijke verdeling van netto-opbrengsten (revenue sharing).
- **Privacy en AVG** – De overeenkomst stelt strikte eisen aan het opschonen van data. Hoewel GPT-NL processen heeft voor anonimisering, ligt er een verantwoordelijkheid bij de leverancier om persoonsgegevens en schadelijke inhoud zo veel mogelijk vooraf te filteren of te markeren.

Meer informatie hierover is te vinden op gpt-nl.nl

²⁷ https://gpt-nl.nl/publish/pages/7064/template_content_contributor_agreement_oktober_2025.pdf

Hoofdstuk 3. Standaarden machinaal voorbehoud

Een voorbehoud moet ook machinaal leesbaar zijn, bij een immaterieel voorbehoud is dat een wettelijk vereiste, bij een materieel voorbehoud is dat praktisch. De technische standaarden maken geen onderscheid tussen een materieel of immaterieel voorbehoud. Ze geven alleen aan dat op dat specifieke platform niet gemined mag worden.

Hieronder worden twee machinaal leesbare technieken geïntroduceerd. Robots.txt en TDMRep. Robots.txt is een breed ondersteunde standaard (robots.txt), maar ongeschikt voor de nieuwe, genuanceerde juridische realiteit. De nieuwe, specifiek ontworpen standaard (TDMRep) is nog niet breed genoeg ondersteund om effectief te zijn. Geadviseerd wordt om beide te onderzoeken en indien mogelijk beide te implementeren. Andere standaarden zoals Altags.org, en robot specifieke meta-tags lijken de race om de standaard te worden niet te halen.

3.1 Robots.txt

Robots.txt is een tekstbestand²⁸ dat op websites wordt geplaatst om robots te instrueren welke delen ze wel en niet mogen crawlen en indexeren. Het werkt als een verkeersbord voor robots: het geeft aan waar ze wel en niet mogen komen. Deze toestemming of verboden kunnen voor alle robots of voor specifieke robots (user-agents) worden geconfigureerd. In de praktijk volgen niet alle robots deze regels. Robots.txt is dus geen firewall of harde beperking, maar communiceert alleen een verbod op machinaal leesbare wijze.

Het bestand specificiert via eenvoudige regels, zoals User-agent en (Dis)allow, welke delen van de website wel of niet door robots mogen worden bezocht en geïndexeerd. In het volgende voorbeeld mogen alle robots (*) alles doen, behalve de robot met de user-agent 'ChatGPT-User' die mag niet in de map /media/ komen.

```
⌘User-agent: *
```

```
Allow: /
```

```
User-agent: ChatGPT-User
```

```
Disallow: /media/*
```

```
⌘
```

3.2 TDM reservation protocol

Een ander voorbeeld van een standaard voor machinaal voorbehoud is het TDM Reservation Protocol (TDMRep). Deze is nog in ontwikkeling en is voorgesteld als standaard door W3C. TDMRep definieert een webprotocol waarmee het behoud van rechten met betrekking tot tekst- en datamining (TDM) kan worden uitgedrukt. Het protocol biedt een technische oplossing voor de vereisten van TDM-wetgeving. Het voordeel van TDMRep ten opzichte van robots.txt is dat het fijnmaziger kan blokkeren. Robots.txt geeft een blokkade voor een gehele webpagina, TDMRep kan objecten binnen een webpagina

²⁸ Zie <https://en.wikipedia.org/wiki/Robots.txt>

blokkeren. Daarnaast kan ook een pagina met meer informatie opgenomen worden, een tdm-policy. Dat doen zij via de headers van de bestanden.

Het TDM Reservation Protocol (TDMRep)²⁹ is een W3C-specificatie die een machinaal leesbare methode beschrijft waarmee rechthebbenden hun auteursrechtelijke reserveringen voor tekst- en datamining (TDM) kunnen aangeven. Het protocol gebruikt twee belangrijke kenmerken: een binaire vlag (tdm-reservation) die aangeeft of de rechten zijn gereserveerd, en een URL (tdm-policy) die verwijst naar een gedetailleerd machinaal leesbaar beleid over de toegestane licenties en voorwaarden. Het doel is om de juridische onzekerheid voor TDM-gebruikers te verminderen en een efficiënte en legale manier te creëren om op grote schaal data te verwerken, in lijn met auteursrechtelijke wetgeving.

☒ HTTP/1.1 200 OK

Content-type: image/jpeg

tdm-reservation: 1

tdm-policy: <https://jouw-site.nl/licentie-voor-deze-foto.json>

☒

²⁹ Zie <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/>

Hoofdstuk 4. Praktische voorbeelden uit de sector

Voor deze sectie zijn beleidsmedewerkers van het Nationaal Archief, het ministerie van Onderwijs Cultuur en Wetenschap, het ministerie van Justitie en Veiligheid³⁰ en beleidsmedewerkers van de Gemeente Amsterdam benaderd met de vraag of zij bekend zijn met enig beleid rondom tekst- en datamining van overheidsinformatie.

Er is gekeken³¹ of er in de algemene voorwaarden en technieken van de websites van rechthebbenden met veel werken in het archief een teken is van rechtenuitvoering onder de tekst- en datamining-uitzondering. Hiervoor is gekeken naar het Nederlands Fotomuseum, Het Maria Austria Instituut, ANPFoto, en AVIODROME. Geen van deze rechthebbenden heeft het archief benaderd met vragen ten aanzien van TDM. Daarnaast kon bij AVIODROME geen beleid ten aanzien van hun collectie gevonden worden. Tenslotte is er gekeken naar de literatuur rondom TDM en uitingen van collectieve beheersorganisaties. Ook is er gekeken naar de bestaande en ver ontwikkelde praktijk van de KB, de nationale bibliotheek.

De aanpak van andere organisaties toont een divers landschap, variërend van passief (Rijksoverheid) tot actief (KB). De inconsistenties, met name in de technische uitvoering, maken het belang duidelijk voor het archief om een beleid te kiezen dat niet alleen juridisch is onderbouwd, maar ook technisch correct wordt geïmplementeerd.

4.1 Stadsarchief Amsterdam

Het Stadsarchief Amsterdam heeft een algemeen immaterieel voorbehoud geplaatst op een algemene disclaimer op haar website:

“[...] Daarnaast behoudt het Stadsarchief Amsterdam zich uitdrukkelijk het recht voor als bedoeld in artikel 2, eerste lid van de Databankenwet (Stb.1999, 303), voor zover deze wet niet anders bepaalt.”³²

Hierbij moet worden opgemerkt dat deze tekst niet op goede locatie staat. Dit soort teksten horen in de algemene voorwaarden van de website. De vraag is dan ook hoe handhaafbaar dit is. Deze tekst is waarschijnlijk een gevolg van de Databankenverordening Amsterdam uit 2001.³³ Het databankenrecht is alleen van toepassing op de openbare macht wanneer dit uitdrukkelijk is voorbehouden. In 2001 hebben de gemeenten Amsterdam en Rotterdam dat recht expliciet ingeroepen. Na de invoering van de TDM-regelgeving in 2019 is deze regel ook rechtstreeks van toepassing.³⁴

De colofon van het archief geeft additionele beperkingen:

“De inhoud van deze site is bedoeld voor persoonlijk, niet-commercieel gebruik. Niets uit de teksten en het beeldmateriaal mag worden verveelvoudigd, opgeslagen of openbaar gemaakt in enige vorm of op enige

³⁰ Beleidseigenaar van de auteurswet

³¹ Op moment van schrijven, najaar 2025

³² <https://archieff.amsterdam/disclaimer>

³³ <https://lokaleregelgeving.overheid.nl/CVDR73042>

³⁴ De Koninklijke Bibliotheek, nationale bibliotheek, stelt dat hun collectiewebsites bijvoorbeeld zelf databanken zijn en dat zij op grond daarop TDM kunnen voorbehouden.

wijze, hetzij elektronisch, mechanisch, via opname of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de gemeentearchivaris van Amsterdam.”³⁵

Door een beperkte lijst met toegestane handelingen te stellen worden andere handelingen, waaronder tekst- en datamining, niet toegestaan. Door geen verwijzing te maken naar het auteursrecht of aanverwante rechten kan dit gezien worden als een materieel voorbehoud.

Maria Austria Instituut

Het Maria Austria Instituut beheert foto's van ruim 50 fotografen, zij zijn ook de rechthebbende van het gehele oeuvre van een deel van deze fotografen.³⁶ Het Maria Austria Instituut doet een materieel TDM-voorbehoud in hun algemene voorwaarden:³⁷

“harvesting van foto's of metadata is uitdrukkelijk verboden.”

Tegelijkertijd staat de robots.txt³⁸ wel alle robots toe op de website:

⊠User-agent: *

Disallow:

⊠

ANPFoto

Fotopersbureau ANPFoto beheert zo'n 75 miljoen beelden. Er wordt vanuit gegaan dat het ook rechthebbende is van al deze beelden. De algemene voorwaarden van ANP geven voldoende handvatten voor een algemeen voorbehoud op TDM.³⁹

Dit wordt verder bekrachtigd door de robots.txt⁴⁰ van ANPFoto. Deze staat geen enkele robot toe:

⊠User-agent: *

...

Disallow: /

⊠

³⁵ <https://www.amsterdam.nl/stadsarchief/colofon/>

³⁶ Een vergelijkende steekproef tussen de vertegenwoordigde fotografen van het instituut en de ledenlijst van Pictoright laat geen overlap zien.

³⁷ Zie <https://www.maibeeldbank.nl/algemene-voorwaarden> (Versie 17 juni 2024)

³⁸ <https://www.maibeeldbank.nl/robots.txt> (opgehaald 11 september 2025)

³⁹ Artikel 12.2 van de algemene voorwaarden geven de beperkte mogelijkheden van het gebruik aan.

⁴⁰ <https://www.anpfoto.nl/robots.txt> (opgehaald op 11 september 2025)

Nederlands Fotomuseum

Het archief beheert ook werken van het Nederlands Fotomuseum.⁴¹ Ook dat heeft een TDM-voorbehoud in zijn algemene voorwaarden opgenomen.

“9.3 Elk gebruik anders dan voor eigen oefening, studie of gebruik dat niet uitdrukkelijk en voorafgaand met Museum is overeengekomen wordt beschouwd als een inbreuk op het auteursrecht van het Museum, dan wel haar volmachtgever.”

De implementatie van robots.txt van het museum staat daarentegen weer alle gebruik toe.

User-agent: *

Disallow:

4.2 Rijksoverheid

Op de vraag of het Nationaal Archief, Ministerie van Justitie en Veiligheid en het Ministerie van OCW een TDM-voorbehoud heeft gemaakt op overheidsinformatie wordt niet bevestigend gereageerd. Voor zover bekend is overheidsinformatie niet onderhevig aan enig voorbehoud. De algemene voorwaarden van de Rijksoverheid maken dan ook geen voorbehoud op TDM.

Een doorloop van de implementatie van 1720 websites van de Rijksoverheid⁴² laat zien dat slechts een kleine 80 unieke domeinen enig voorbehoud hebben op TDM. In totaal worden de volgende selectie bots door robots.txt tegengehouden:

- Alle robots (*), 66x
- ClaudeBot, 65x
- AmazonBot, 61x
- GPTBot, 21x
- GoogleExtended, 9x
- CCBot, 7x

4.3 Pictoright

Pictoright, de collectieve beheersorganisatie voor beeldmakers, heeft samen met de Federatie Beeldrechten⁴³ een collectief voorbehoud ontwikkeld. Ongeveer 4.000 beeldmakers⁴⁴ hebben zich via een collectieve 'opt-out' ingeschreven⁴⁵, wat betekent dat zij zich verzetten tegen het gebruik van hun werk voor TDM.

⁴¹ Opgemerkt moet worden dat een deel van de collectie publiek domein is, waaronder werken van Richard Tepe:

https://collectie.nederlandsfotomuseum.nl/collectie/?mode=gallery&view=list&rows=45&page=1&fq%5B%5D=search_s_fotograaf:%22Tepe,%20Richard%22&sort=order_i_relevantie%20asc&filterAction&reverse=0

een TDM-voorbehoud is hierop niet van toepassing.

⁴² Zie <https://www.communicatierijk.nl/documenten/2016/05/26/websiteregister>

⁴³ <https://pictoright.nl/thema/auteursrecht-en-ai/>

⁴⁴ Uit een steekproef van deze auteur blijkt dat 9 op de tien van deze beeldmakers geen lid is van Pictoright.

⁴⁵ <https://federatiebeeldrechten.nl/optout>

Een collectief voorbehoud is problematisch zonder actieve communicatie naar partijen die werken beschikbaar stellen van deze makers. Het archief kan moeilijk achterhalen of een veel voorkomende naam als ‘Bert de Vries’ daadwerkelijk dezelfde persoon is als de vervaardiger in de eigen collectie. Het is in de sector niet bekend of beeldmakers van dit collectief contact hebben opgenomen met een archief met betrekking tot hun TDM-voorbehoud.

4.4 Koninklijke Bibliotheek, nationale bibliotheek

De Koninklijke Bibliotheek, nationale bibliotheek (KB), heeft een specifiek beleid met een TDM-voorbehoud. In haar AI-statement kondigde de KB aan de toegang tot auteursrechtelijk beschermd materiaal op haar digitale collecties te beperken voor commerciële AI-bedrijven. Dit besluit is genomen omdat de KB van mening is dat deze bedrijven haar ethische principes niet naleven, met name door zonder expliciete toestemming websites te crawlen. Deze maatregel heeft tot doel de rechten van auteurs en uitgevers te beschermen en misbruik van hun werk te voorkomen. De toegankelijkheid voor wetenschappelijk onderzoek blijft onaangetast.

De KB implementeert de TDM-beperkingen via verschillende technische en juridische middelen:

- **Robots.txt** – Deze standaard wordt ingezet om specifieke webcrawlers, zoals GPTBot en ChatGPT-User, de toegang tot Delpher te ontzeggen. Deze methode wordt ook om andere redenen gebruikt, zoals het blokkeren van Googlebot-image en het tegengaan van overbelasting van de site door crawlers als dataforseo.com.
- **Gebruiksvoorwaarden** – Een specifieke clause is toegevoegd aan de gebruiksvoorwaarden van DBNL en Delpher, die commerciële partijen verbiedt kopieën te maken voor TDM en training van AI-modellen. Dit vormt een formeel voorbehoud conform relevante wetgeving.

“Het downloaden van grote aantallen objecten die (mogelijk) auteursrechtelijk beschermd zijn, is niet toegestaan. Onder voorwaarden is bulkgebruik van deze objecten wel mogelijk voor wetenschappelijke onderzoeksdoeleinden. Lees hierover meer in Data in Delpher.

Voor commerciële partijen is het verboden om kopieën van auteursrechtelijk beschermde werken op onze website te maken ten behoeve van tekst- en datamining (TDM) waaronder training van AI-modellen. Dit verbod vormt een voorbehoud als bedoeld in artikel 15o Auteurswet, artikel 4a sub b Databankenwet en artikel 4(3) CDSM-richtlijn.”⁴⁶

De koepelovereenkomst periodieken tussen de KB, de Stichting Lira en de Stichting Pictoright bevat sectorbrede afspraken voor out-of-commerce periodieken. Artikel 8.2 van de overeenkomst verplicht de KB en de aangesloten erfgoedinstellingen om een passend, machinaal leesbaar TDM-voorbehoud te implementeren. Dit voorbehoud maakt duidelijk dat de periodieken niet mogen worden gebruikt voor tekst- en datamining met een commercieel oogmerk, inclusief het trainen van AI-modellen.

“De KB en de toetredende Erfgoedinstellingen zullen door middel van een passend machinaal leesbaar rechtenvoorbehoud kenbaar maken dat de Periodieken niet mogen worden gebruikt voor tekst- en datamining met een commercieel oogmerk in de zin van artikel 15o Auteurswet en artikel 4 DSM-richtlijn, daaronder mede begrepen gebruik voor AI trainingsdoeleinden.”⁴⁷

⁴⁶ <https://www.delpher.nl/over-delpher/gebruiksvoorwaarden> (11 september 2025)

⁴⁷ <https://www.kvan.nl/wp-content/uploads/2024/07/315.2-Koepelovereenkomst-Out-of-commerce-Periodieken-juli-2024.pdf> (11 september 2025)

4.5 Nationaal Archief

Op navraag bij het Nationaal Archief lijkt er geen eenduidig TDM-beleid aanwezig te zijn. De algemene voorwaarden van de website stimuleren hergebruik van het materiaal:

“Artikel 4 –Het Nationaal Archief moedigt hergebruik van de informatie op zijn website aan. In principe is alle informatie op de Website voor hergebruik beschikbaar, tenzij anders vermeld. Zie hiervoor de pagina Open Data. Op afbeeldingen (foto's) kunnen rechten van derden gelden. Alleen afbeeldingen die (ook) in hoge resolutie gedownload kunnen worden, zijn voor hergebruik beschikbaar.”

Hier staat geen voorbehoud van rechten voor TDM in te lezen. De inrichting van het robots.txt bestand is wat meer onduidelijk. Deze is lang en uitgebreid. Een kleine selectie laat zien dat robots niet mogen zoeken en niet op de Engelstalige variant van de zoekindexen mogen komen. De Nederlandse zoekpagina's zijn niet uitgesloten.

```
⊠Disallow: /book/*
Disallow: /en/book/*
Disallow: /en/research/archive/*
Disallow: /en/research/photo-collection/*
Disallow: /en/research/map-collection/*
Disallow: /en/research/index/*
Disallow: /onderzoeken/transcriptie/*
Disallow: /services/*
Disallow: /zoek/*
Disallow: /index.php/book/*
Disallow: /index.php/en/book/*
Disallow: /index.php/en/research/archive/*
Disallow: /index.php/en/research/photo-collection/*
Disallow: /index.php/en/research/map-collection/*
Disallow: /index.php/en/research/index/*
Disallow: /index.php/onderzoeken/transcriptie/*
Disallow: /index.php/services/*
Disallow: /index.php/zoek/*
Disallow: /archievenoverzicht-suriname/*
```

Hoofdstuk 5. Communicatie

De archiefinstelling juridisch niet bevoegd is om namens derden een immaterieel TDM-voorbehoud te maken zonder dat deze dat verzoekt. Het initiatief voor een 'opt-out' ligt bij de rechthebbende. Maar de archiefinstelling kan een faciliterende rol aannemen, in lijn met hun rol binnen de samenleving. Dit proces kan worden ingericht langs drie lijnen: informatievoorziening, meldingsprocedure en verwerking.

1. Actieve informatievoorziening Rechthebbenden zijn zich er vaak niet van bewust dat hun online materiaal gebruikt wordt voor AI-training, noch dat zij de mogelijkheid hebben dit te verbieden. Het archief kan daarom op de website – bijvoorbeeld op de pagina over auteursrechten of in de footer bij de disclaimer – duidelijk uitleggen wat het beleid is. De kernboodschap hierbij is: "Het archief stelt materiaal beschikbaar voor hergebruik". Wilt u als rechthebbende niet dat uw werk gebruikt wordt voor AI-training? Dan communiceren wij uw wettelijke recht op een opt-out bij de werken die wij beschikbaar stellen." Hiermee toont het archief zich een betrouwbare partner die de belangen van makers serieus neemt, zonder zelf op de stoel van de rechthebbende te gaan zitten.

2. Procedure voor afmelding (Opt-out) Het is verstandig een duidelijk kanaal open te stellen waar rechthebbenden hun voorbehoud kenbaar kunnen maken. Denk hierbij bijvoorbeeld aan een mailadres. Communiceer dat zolang er geen melding wordt gemaakt, de wettelijke standaard van toepassing is. Dit betekent dat zonder bericht commerciële TDM op rechtmatig openbaar gemaakt materiaal is toegestaan.

3. Verwerking Na ontvangst en validatie van de melding, moet het archief het voorbehoud ook daadwerkelijk aan het object koppelen. In de praktijk betekent dit dat het archief voor de betreffende objecten of collecties in het beheersysteem een vinkje 'TDM-voorbehoud' aanzet. Dit zorgt er vervolgens voor dat de technische standaarden (zoals het TDM Reservation Protocol) voor die specifieke objecten op 'restricted' worden gezet. Op deze manier wordt de juridische wil van de rechthebbende direct vertaald naar een machineleesbaar signaal dat AI-bots kunnen begrijpen.

Hoofdstuk 6. Conclusies

De leidende vraag van dit kennisdocument is: "Welk beleid moeten archiefinstellingen hanteren om tekst- en datamining (TDM) op hun collecties conform het intellectueel eigendomsrecht en technisch uitvoerbaar te faciliteren?"

Het voorbehoud dat hierop gemaakt kan worden komt voor uit het auteursrecht en uit het contractenrecht. De auteursrechtelijke status en de rechthebbende zijn hierin bepalend wat voor soort voorbehoud mogelijk is.

De vraag of een immaterieel voorbehoud van toepassing is hangt af van de vragen of

1. Is het werk auteursrechtelijk beschermd? En
2. Wie is de rechthebbende?

Als het werk niet auteursrechtelijk beschermd is – **publieke domein werken** – waaronder wetten, besluiten en verordeningen, rechterlijke uitspraken en administratieve beslissingen of wanneer dat recht is verlopen, is een immaterieel voorbehoud niet mogelijk. Een materieel voorbehoud is wel mogelijk.

Wanneer het werk nog wel auteursrechtelijk beschermd is, maar **een derde is rechthebbende** dan moet het voorbehoud van de rechthebbende gevolgd worden. Is daarentegen het **de archiefinstelling zelf rechthebbende** dan mag het archief daar ook zelf over beslissen. Een bijzondere categorie van werken zijn **werken van openbare overheidspublicaties**. Hierbij is de openbare macht vaak zelf de rechthebbende, maar de toepassing van enig immaterieel voorbehoud is beperkt.

6.1 Principes

Wees daarom spaarzaam met een TDM voorbehoud. Zet deze alleen in wanneer een derde rechthebbende hierom vraagt. Maak gebruik van de volgende principes:

1. Een immaterieel voorbehoud op TDM wordt alleen ingezet bij beperkt openbare overheidspublicaties en de rechthebbenden die TDM voorbehouden.
2. Een immaterieel TDM-voorbehoud wordt in een besluit vastgelegd en beperkt zich tot het in stand kunnen houden van dienstverlening van het archief.
3. Het archief maakt geen voorbehoud op materiaal waarvan zij rechthebbende is.

	Materieel voorbehoud	Immaterieel voorbehoud
Publiek domein	Nee	NVT
Derden rechthebbenden	NVT	Volg beslissing rechthebbende
Zelf rechthebbende	Nee*	Nee*
Overheidspublicaties (openbaar)	Nee	Nee
Zwijgende rechthebbende	Nee	Nee

Een voorbehoud moet uiteraard gecommuniceerd worden voor het effectief kan zijn. De plek om een voorbehoud te communiceren is onder andere de algemene voorwaarden van de website waarop het materiaal wordt gepresenteerd. Daarnaast wordt er aanbevolen om ook beide technisch standaarden (robots.txt en TDMRep) te implementeren en verwijzingen naar deze implementaties op te nemen in de algemene voorwaarden, zodat bezoekers ook weten welke standaarden gehanteerd worden.

Dit is een uitgave van de Koninklijke Vereniging Archiefsector Nederland

