

# Webharvesting

Samenvatting (p. 3) / Summary (p. 11)



*Instituut voor Informatierecht*  
*Institute for Information Law*



UNIVERSITEIT VAN AMSTERDAM

Martin R.F. Senftleben

Stef J. van Gompel

Anne Helmond

Luna D. Schumacher

Jef Ausloos

Joris V.J. van Hoboken

João Pedro Quintais

20 september 2021

© 2021 Martin R.F. Senftleben, Stef J. van Gompel, Anne Helmond, Luna D. Schumacher, Jef Ausloos, Joris V.J. van Hoboken, João Pedro Quintais.

Onderzoek in opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum, Ministerie van Justitie & Veiligheid. Auteursrechten voorbehouden. Niets uit dit rapport mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm, digitale verwerking of anderszins, zonder voorafgaande schriftelijke toestemming van de auteurs.

# Samenvatting

## Beleidsmatige wenselijkheid

Webharvesting kan beleidsmatig wenselijk worden geacht. Dit blijkt uit een analyse van het Europese en Nederlandse grondwettelijke kader. In het licht van de waarborging van de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap, is het legitiem en wenselijk om maatregelen te nemen die webharvesting mogelijk maken. Met de creatie van webarchieven levert webharvesting een belangrijke bijdrage aan de verwezenlijking van deze fundamentele rechten en vrijheden. Dit geldt niet alleen voor specifieke harvestingsactiviteiten, zoals event harvestings, maar ook voor overkoepelende nationale domeincrawls (3.2.6).

Daar staat tegenover dat vanuit grondrechtelijk perspectief tevens rekening moet worden gehouden met fundamentele rechten die in het kader van webharvesting aangetast zouden kunnen worden: auteurs- en naburige rechten, databankrechten, het recht op privacy en het recht op gegevensbescherming. Beleidsmakers dienen dus een oplossing te vinden die ruimte biedt voor webharvesting zonder intellectuele eigendomsrechten, privacyrechten en het recht op gegevensbescherming te zeer in het gedrang te brengen.

Beleidsmatige wenselijkheid betekent dus geen carte blanche. De Nederlandse wetgever is ten aanzien van de keuze van maatregelen ter ondersteuning van webharvesting niet geheel vrij. Veeleer laten ontwikkelingen in de rechtspraak van het HvJ EU zien dat de vereiste balans – tussen grondrechten die door webharvesting ondersteund worden en grondrechten die door webharvesting aangetast zouden kunnen worden – bij voorkeur tot stand moet worden gebracht *binnen* het wettelijke kader dat voor de genoemde rechten bestaat op het niveau van secundaire Europese wetgeving. Dit geldt met name voor het in vergaande mate geharmoniseerde systeem van exclusieve rechten en beperkingen in het Europese auteursrecht (3.2.6).

## Implementatie in de praktijk

Nederlandse erfgoedinstellingen die middels webcrawling en andere harvestings technieken een archief wensen op te bouwen van het Nederlandse digitale erfgoed lopen tegen verscheidene typen praktische problemen aan. Deze zijn bijvoorbeeld organisatorisch (onderlinge afstemming), financieel (capaciteit) en technisch (gebrek aan kennis, snelle ontwikkelingen). Voorts rijzen conceptuele vragen omtrent de afbakening van het nationale domein in het kader van een nationale domeincrawl (2.2.2). In het licht van deze uitdagingen is onder Nederlandse erfgoedinstellingen weinig animo voor een wettelijke verplichting en verantwoordelijkheid om het nationale domein volledig te archiveren. De voorkeur gaat uit naar een systeem dat enige beleidsvrijheid biedt om harvestingsactiviteiten aan te passen aan de ter beschikking staande capaciteit, financiële ruimte en technische expertise. De KB heeft wel belangstelling voor het uitvoeren van een nationale domeincrawl. Zij heeft op dat terrein ook reeds uitgebreide kennis en expertise opgebouwd.

Vanuit juridisch perspectief kan daaraan worden toegevoegd dat het opt-out systeem dat tegenwoordig in het kader van webharvesting gebruikt wordt om toestemming te vragen aan rechthebbenden van online content zeer tijdrovend is (2.2.3.1). Bovendien kan de rechthebbende niet altijd getraceerd worden en kan zich op een webpagina ook content van derden bevinden waarvoor de beheerder geen toestemming kan geven. Tegen deze achtergrond kan geconcludeerd worden dat de bestaande harvestingspraktijk bijzondere problemen oproept. De focus op individuele toestemming belemmert de praktische uitvoering van webharvesting in Nederland.

Dit geldt in het bijzonder voor omvangrijke harvestingsprojecten, zoals een nationale domeincrawl. Een alternatieve benadering – op basis van een wettelijke bevoegdheid of maatregelen die juridische risico's minimaliseren – zou het bestaande webarchiveringsproces in belangrijke mate kunnen versimpelen en versnellen. In een ideaal scenario zou het de huidige opt-out praktijk overbodig maken.

Een alternatief scenario dat uitgaat van de ontwikkeling van specifieke wetgeving roept wel de vervolgvraag op welke instellingen een harvestingsbevoegdheid toebedeeld zouden moeten krijgen. Te denken valt enerzijds aan een *gecentraliseerd systeem* waarin de wetgever de webharvestingtaak opdraagt aan enkele centrale erfgoedinstellingen, zoals de KB, Beeld en Geluid en het Nationaal Archief. Om ook aan de wensen en collectiestrategieën van kleinere instellingen tegemoet te komen zou in dit scenario een constructie kunnen worden opgezet, waarbij de kleinere erfgoedinstellingen bij één van deze centrale instellingen verzoeken tot archivering zouden kunnen insturen. In dat geval wordt het harvestingsproces uitbesteed aan deze instellingen, maar blijft de selectie decentraal en dus bij de specifieke kennisbron gebeuren. De combinatie van een gecentraliseerd systeem met de mogelijkheid voor kleinere instellingen om individuele harvestingsverzoeken in te dienen, vereist wel nader onderzoek naar de capaciteit en financiële ruimte die bij centrale erfgoedinstellingen met een harvestingstaak ter beschikking staan.

Voorts is het oprichten van een *gedecentraliseerd systeem* denkbaar waarbij een brede webharvestingsbevoegdheid globaal wordt toegekend aan een grote(re) groep culturele erfgoedinstellingen, zoals alle archieven en publiek toegankelijke bibliotheken. Om overlapping te voorkomen en kennis uit te wisselen zouden in dit systeem onderlinge afspraken omtrent de inhoud en ontsluiting van webarchieven tot stand moeten komen op basis van reeds bestaande samenwerkingsverbanden, zoals het Netwerk Digitaal Erfgoed, het Nationaal Register Webarchieven en het landelijk overleg webarchieven. In een dergelijk geval zouden dus alle erfgoedinstellingen een eigen webcollectie aanleggen. De individuele collecties zouden mogelijk wel centraal toegankelijk of raadpleegbaar gemaakt kunnen worden.

Een laatste optie is een *universeel systeem* dat de webarchiveringsbevoegdheid toekent aan alle rechtssubjecten. In dat geval zouden ook individuen, NGO's of bedrijven die juist wél een commercieel doel nastreven hierop een beroep kunnen doen en een eigen webarchief kunnen aanleggen.

## **Juridische oplossingsrichtingen**

In feite hangt de ontwikkeling van één van deze beleidsopties af van de mogelijkheden die het wettelijke kader op nationaal en Europees niveau biedt. In dat opzicht zijn in de analyse de volgende oplossingsrichtingen ten aanzien van het auteursrecht, de naburige rechten en het databankenrecht, het privacy- en gegevensbeschermingsrecht, en andere civiel-, straf- en bestuursrechtelijke rechtsposities naar voren gekomen.

### **Auteursrecht, naburige rechten en databankenrecht**

De beleidsopties op het terrein van het auteursrecht, de naburige rechten en het databankenrecht kunnen op basis van een matrix worden weergegeven die de parameters centrale/decentrale bevoegdheid en nationale domeincrawl/beredeneerde selectie onderscheidt. In het volgende overzicht betekent "NEE" dat het ontwikkelen van wetgeving op dit terrein niet raadzaam is. "JA" betekent dat beleidsruimte bestaat die de Nederlandse wetgever als basis voor een regeling zou kunnen gebruiken:

	Nationale domeincrawl	Beredeneerde selectie
Bevoegdheid: decentraal	NEE	JA
Bevoegdheid: centraal	JA	JA

### *Decentrale bevoegdheid – nationale domeincrawl*

De analyse laat zien dat het bestaande wettelijke kader voor beperkingen van het auteursrecht, de naburige rechten en het databankenrecht een te onzekere grondslag biedt om daarbinnen een algemene beperking voor het uitvoeren van nationale domeincrawls vorm te geven waarop alle culturele erfgoedinstellingen een beroep zouden kunnen doen.

De meeste ruimte om reproductiehandelingen als het scrapen, opslaan, indexeren en archiveren van webpagina's in het kader van een nationale domeincrawl toe te staan, biedt de beperking van het auteursrecht en de naburige rechten ten aanzien van door culturele erfgoedinstellingen vervaardigde reproducties "in welbepaalde gevallen" (3.3.1.5). Deze regeling is opgenomen in de Europese lijst van toegestane beperkingen die de nationale wetgever optioneel kan implementeren (art. 5 lid 2 sub c InfoSoc-richtlijn). De Nederlandse wetgever heeft van deze mogelijkheid tot nu toe geen gebruik gemaakt. De neutrale voorwaarde van "welbepaaldheid" zou ruimte kunnen bieden om een nauwkeurig omschreven en duidelijk afgebakende beperking te introduceren die het maken van reproducties van websites in het kader van webharvesting door niet-commerciële archieven en openbaar toegankelijke bibliotheken toelaat.

Op grond van Overweging 40 InfoSoc-richtlijn rijzen ten aanzien van deze oplossing wel verschillende vragen. Overweging 40 sluit de toepassing van de beperking uit wanneer het uiteindelijke doel is de geharveste materialen online beschikbaar te stellen aan het publiek. Deze wens – beschikbaarstelling online, bijvoorbeeld via een bestaande online infrastructuur als CLARIAH – werd door Nederlandse erfgoedinstellingen geuit (2.6). Voorts dient de beperking volgens Overweging 40 beperkt te blijven tot bepaalde bijzondere gevallen welke onder het reproductierecht vallen. Dit doet de vraag rijzen of een beperking in voldoende mate specifiek kan worden geacht wanneer webharvesting diverse reproductiehandeling met zich brengt en globaal wordt vrijgesteld voor alle niet-commerciële archieven en openbaar toegankelijke bibliotheken in Nederland. Ook al lijkt de driestappentoets in dat opzicht geen onoverkomelijke obstakels op te werpen (3.3.4.2), het begrip "welbepaald" zou op een mate van specificering kunnen duiden die het niet toelaat om de harvestingebevoegdheid globaal aan alle erfgoedinstellingen toe te wijzen. De analyse van buitenlandse systemen laat in dit opzicht zien dat in geen van de onderzochte landen art. 5 lid 2 sub c InfoSoc-richtlijn als een basis werd gebruikt om een gedecentraliseerde aanpak ten aanzien van de nationale domeincrawl te rechtvaardigen. Voorts heeft het HvJ EU de beperking in *TU Darmstadt* slechts toegepast op boeken die reeds onderdeel waren van de collectie van een universiteitsbibliotheek (3.3.4.1). Uit het arrest in *TU Darmstadt* kan dus niet worden afgeleid dat een beroep op art. 5 lid 2 sub c InfoSoc-richtlijn mogelijk is om beschermd materiaal – zoals auteursrechtelijk beschermd webpagina's – in de collectie van een erfgoedinstelling te integreren.

Een ruimere interpretatie in Nederland zou dus rechtsonzekerheid met zich brengen. Het valt niet uit te sluiten dat een globale domeincrawlbevoegdheid buitensporig en niet "welbepaald" wordt geacht wanneer alle culturele erfgoedinstellingen in staat worden gesteld om web content onderdeel van hun collecties te maken. Daar komt bij dat het databankenrecht geen corresponderende beperking bevat (3.3.3.2). Een op "welbepaalde gevallen" gebaseerde regeling zou databankrechten dus niet omvatten.

Een gedecentraliseerde bevoegdheid lijkt dus niet raadzaam ten opzichte van nationale domeincrawls. Dit geldt in nog sterkere mate voor de boven geschetste universele aanpak waarbij naast culturele erfgoedinstellingen ook individuen, NGO's en bedrijven de mogelijkheid zouden hebben om een nationale domeincrawl uit te voeren.

#### *Centrale bevoegdheid – nationale domeincrawl*

Daarentegen lijkt een gecentraliseerde nationale domeincrawlbevoegdheid met focus op specifieke, hiervoor geselecteerde culturele erfgoedinstellingen wel goed verdedigbaar wanneer een wettelijke depotverplichting aan het harvesten ten grondslag ligt. Deze route hebben Denemarken, Duitsland, Frankrijk en het VK gekozen (Hoofdstuk 4). Een wettelijke depotverplichting zorgt ervoor dat web content deel gaat uitmaken van de collectie van culturele erfgoedinstellingen. Zodra deze stap is genomen wordt het mogelijk om een beroep te doen op beperkingen die gekoppeld zijn aan de collectiestatus, zoals de regeling ten opzichte van preserveringskopieën voor de archivering van webpagina's (3.3.1.2) en de leesterminalexceptie voor het ontsluiten van het webarchief (3.3.2.1). Voorts is een beroep op beperkingen mogelijk die reproducties en openbaarmakingshandelingen ten behoeve van wetenschappelijk onderzoek toelaten (3.3.1.4). In dit verband dient wel te worden vermeld dat het databankenrecht op grond van ontbrekende beperkingen voor de beschikbaarstelling aan onderzoekers en via leesterminals (3.3.3.2) als een onzekerheidsfactor kan worden beschouwd. In het licht van de bijzondere inbreukcriteria die onder meer in de rechtspraak van het HvJ EU gestalte hebben gekregen, lijkt inbreuk op sui generis databankrechten echter niet aannemelijk. Historische archiefversies van databanken die als gevolg van webharvesting in een webarchief worden opgenomen en volgens specifieke toegangsregels worden ontsloten zullen de rechthebbende meestal niet de mogelijkheid ontnemen om investeringen terug te verdienen (3.3.3.3).

6

Het hanteren van een wettelijke depotplicht heeft in alle onderzochte buitenlandse systemen geleid tot een centralisering van webharvestingactiviteiten bij individuele culturele erfgoedinstellingen, zoals de nationale bibliotheek. Dit zorgt voor een duidelijke institutionele afbakening van de bevoegdheid. Voorts kan gezegd worden dat de bundeling van de harvestingbevoegdheid bijdraagt aan een duidelijke en nauwkeurige omschrijving van het gebruiksprivilege. Wetgevers in landen met een wettelijke depotplicht aarzelen dan ook niet om in verband met webharvestingactiviteiten van reproducties “in welbepaalde gevallen” in de zin van art. 5 lid 2 sub c InfoSoc-richtlijn te spreken (4.2.2.3).

Wetgeving die een dergelijke centrale bevoegdheid toekent aan één of meerdere erfgoedinstellingen hoeft niet noodzakelijkerwijs de vorm van depotwetgeving te krijgen. Hoewel dit in de meeste bestudeerde landen wel het geval is, verschillen de uitgangspunten voor deze oplossing van de situatie in Nederland in die zin dat reeds een depotsysteem voor analogo materiaal bestond. Dit systeem is in de onderzochte landen simpelweg uitgebreid naar elektronische publicaties met alle gevolgen van dien: er ontstaat een plicht voor de websitehouder (uitgever) en/of de bibliotheek om het materiaal bijeen te brengen, op grond van de conventies die zijn verbonden aan traditionele depotwetgeving.<sup>1</sup> Nederland hoeft echter niet deze route te bewandelen. Een depotverplichting die ervoor zorgt dat relevante websites onderdeel worden van de collectie van culturele erfgoedinstellingen kan in specifieke wetgeving worden opgenomen

---

<sup>1</sup> De UNESCO heeft in 2000 een hernieuwing van de in 1981 door J. Lunn ontwikkelde *Guidelines for Legal Deposit Legislation* uitgegeven, waarin dieper wordt ingegaan op het depot van elektronisch materiaal. Hierin wordt ook uiteengezet hoe een wettelijk depotsysteem eruit dient te zien. Kiezen voor depotwetgeving betekent dat de wetgever rekening zal moeten houden met dit kader. Hierin wordt een systeem gebaseerd op vrijwilligheid sterk afgeraden. Zie: J. Larivière, *Guidelines for Legal Deposit Legislation*, UNESCO 2000. Beschikbaar via: <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-legal-deposit-legislation/> (laatst gewijzigd: 03-09-2021).

met een focus op webharvesting. Het is niet nodig om in dit verband algemene depotwetgeving voor alle publicaties – analoog en digitaal – in het leven te roepen. Een plek om een specifieke bevoegdheid en depotverplichting voor online bronnen te introduceren en nauwkeurig te beschrijven zou bijvoorbeeld gevonden kunnen worden in de wetten die de taak omschrijven van de gekozen instelling(en). In Nederland zouden dat bijvoorbeeld de Whw zijn voor de KB en de Mediawet voor Beeld en Geluid (2.2.4.3).

Een alternatieve oplossingsrichting voor webharvesting is in Portugal te vinden. Het Portugese systeem is voornamelijk gebaseerd op een maatschappelijke consensus die ook door rechthebbenden wordt gedeeld (4.2.5). In Portugal is op deze basis – zonder wettelijke depotverplichting – een gecentraliseerd harvestingsysteem tot stand gekomen dat qua volume en frequentie van de nationale domeincrawl en de mate van toegankelijkheid van het webarchief omvangrijker lijkt te zijn dan de harvestingsactiviteiten in vele andere landen. Opvallend is in het bijzonder dat de Portugese instelling Arquivo.pt erin is geslaagd om het webarchief – voorzien van zoekfunctionaliteit – ter beschikking te stellen aan het gehele internetpubliek. Ondanks de aanzienlijke omvang van de harvestingsactiviteiten werd in Portugal nog niet tegen Arquivo.pt geprocedeerd. Het systeem lijkt in belangrijke mate het resultaat te zijn van impliciete toestemming en/of stilzwijgend verleende licenties.

Ook dit alternatieve model biedt aanknopingspunten voor nieuwe impulsen in Nederland. In het licht van de ervaringen in Portugal rijst de vraag of erfgoedinstellingen hun harvestingsactiviteiten – zowel het crawlen en archiveren alsook het ontsluiten – zouden kunnen uitbreiden op basis van een analyse van inbreukrisico's en de ontwikkeling van gepaste maatregelen om deze risico's te beperken. In Portugal vormen een embargoperiode en een efficiënt notice-and-takedown systeem centrale pilaren van het harvestingsysteem. De embargoperiode voorkomt directe concurrentie met de exploitatie van meer recente versies van bronmateriaal dat op internet wordt aangeboden. Het notice-and-takedown systeem biedt rechthebbenden de mogelijkheid om inbreuk te melden en de toegang tot problematische bronnen te laten blokkeren. Vergelijkbare bouwstenen zouden ook in Nederland ertoe kunnen leiden dat een overstap wordt gemaakt van het actuele systeem met toestemming en opt-out naar een systeem met embargoperiode en notice-and-takedown. De Nederlandse wetgever zou deze oplossing kunnen ondersteunen door een vrijwaring van aansprakelijkheid te introduceren die vergelijkbaar is met het safe harbour regime voor hostingdiensten (3.5.3). Voorts is het denkbaar om met een uitdrukkelijke wettelijke taakopdracht de legitimiteit van webharvesting (misschien zelfs met ontsluiting van gearchiveerd materiaal voor het gehele internetpubliek) te benadrukken en potentiële twijfels over de omvang van door erfgoedinstellingen ontwikkelde harvestingsactiviteiten weg te nemen.

Een verdere oplossing zou gezocht kunnen worden in het systeem van ECL. Daarvoor is een grondslag te vinden in art. 12 DSM-richtlijn en art. 45 Aw. Art. 12 staat EU-lidstaten toe een ECL-regeling te introduceren voor duidelijk afgebakende doeleinden voor gebruik op het eigen grondgebied. De rechtsfiguur van ECL wordt gekenmerkt door een combinatie van vrijwillig collectief beheer met een wettelijke uitbreiding van het toepassingsbereik van de collectieve licentie tot rechthebbenden die niet bij de betrokken cbo zijn aangesloten (3.3.4.5). Het ECL-systeem mag alleen worden toegepast binnen welomschreven toepassingsgebieden waar de verkrijging van toestemmingen van rechthebbenden op individuele basis in de regel zo hinderlijk en onpraktisch is dat de vereiste licentietransactie onwaarschijnlijk is. In het geval van webharvesting lijkt aan deze voorwaarde te zijn voldaan. De vraag is echter of ten aanzien van alle relevante deelgebieden waarbinnen de rechten moeten worden geregeld wel cbo's bestaan die voldoende representatief zijn.

In Nederland bestaan diverse cbo's, onder andere voor schrijvers, vertalers en freelance journalisten (Lira) en visuele makers (Pictoright). Deze cbo's vertegenwoordigen echter lang niet

alle rechthebbenden wier werken in de context van webharvesting worden gebruikt. Webdesigners bijvoorbeeld zijn niet als zodanig in een cbo vertegenwoordigd. Het systeem van ECL lijkt daarom niet in staat om een volledige oplossing voor webharvesting te bieden. Daarnaast is de vraag of het vanuit cultuurpolitieke beleidsoverwegingen wenselijk is om de toekomst van webharvesting afhankelijk te stellen van collectieve licenties die vrijwillig moeten worden overeengekomen. Wanneer er geen overeenstemming kan worden bereikt, dan zal ook geen webharvesting kunnen plaatsvinden. Zelfs Scandinavische landen die van oudsher zeer goed ontwikkelde ECL-systemen hebben, geven ten aanzien van webharvesting de voorkeur aan een wettelijke depotverplichting. Dat blijkt uit de analyse van het systeem in Denemarken (4.2.1).

#### *Decentrale bevoegdheid – beredeneerde selectie*

Het feit dat gecentraliseerde systemen meer rechtszekerheid voor een nationale domeincrawl bieden hoeft niet tot het resultaat te leiden dat erfgoedinstellingen buiten het gecentraliseerde systeem geen enkele harvestingactiviteit meer mogen uitvoeren. Ook al is het niet raadzaam om een globale bevoegdheid voor nationale domeincrawls toe te kennen aan alle culturele erfgoedinstellingen, de auteurs-, nabuur- en databankrechtelijke regelingen laten wel ruimte voor meer specifieke harvestingprojecten die op een selectie van websites zijn gebaseerd. Naast de beperking ten aanzien van reproducties in welbepaalde gevallen biedt het bestaande wettelijke kader ruimte voor harvestingactiviteiten die een specifieke focus hebben.

De nieuwe regeling ten opzichte van TDM (art. 3 en 4 DSM-richtlijn en art. 15n en 15o Aw, 3.3.1.3) biedt bijvoorbeeld een grondslag voor gerichte harvestingactiviteiten in het kader van wetenschappelijke projecten, indien web content wordt binnengehaald om deze te analyseren teneinde er informatie uit te genereren met het oog op onderzoek.

8

Een vergelijkbare conclusie kan worden getrokken ten opzichte van de mogelijkheid om reproductie- en openbaarmakingshandelingen “ten behoeve van het wetenschappelijk onderzoek” toe te staan (art. 5 lid 3 sub a InfoSoc-richtlijn, 3.3.1.4). Ook dit vereist een voldoende duidelijke connectie met wetenschappelijk onderzoek. Een “beredeneerde selectie” of “event harvesting”, wellicht op basis van gezamenlijke projecten tussen erfgoedinstellingen en onderzoeksorganisaties, zou onder deze beperking mogelijk wel te rechtvaardigen zijn. Thans is hiervoor evenwel nog geen ruimte, omdat de optionele beperking “ten behoeve van het wetenschappelijk onderzoek” niet in Nederland is geïmplementeerd.

Een wettelijke regeling van selectieve harvestingactiviteiten op basis van de nieuwe wetgeving inzake TDM en de beperking ten behoeve van wetenschappelijk onderzoek heeft wel het voordeel dat het databankenrecht – tenminste ten aanzien van reproductiehandelingen – uitdrukkelijk vergelijkbare ruimte biedt (3.3.3.1 en 3.3.3.2). Deze oplossing maakt het voorts mogelijk om ook individuen, NGO's en bedrijven in staat te stellen om in het kader van TDM-activiteiten websites te harvesten. Art. 4 DSM-richtlijn en art. 15o Aw zijn namelijk niet aan culturele erfgoedinstellingen zijn gekoppeld. Deze bepalingen bieden dus een basis voor een bredere regeling die niet op culturele erfgoedinstellingen is toegespitst.

#### *Centrale bevoegdheid – beredeneerde selectie*

De beschreven mogelijkheden voor het rechtvaardigen van webharvesting op basis van een beredeneerde selectie staan ter beschikking voor alle culturele erfgoedinstellingen. Dit omvat dus ook erfgoedinstellingen die in het kader van een gecentraliseerd systeem voor nationale domeincrawls zijn aangewezen als depotinstellingen met domeincrawlbevoegdheid. Zij kunnen dus wettelijk in staat worden gesteld om zowel nationale domeincrawls alsook specifieke harvestingprojecten met een bepaalde focus uit te voeren.



## Privacy- en gegevensbeschermingsrecht

De analyse heeft ook de mogelijkheden op het terrein van het privacy- en gegevensbeschermingsrecht in kaart gebracht. In dit opzicht lijkt de Nederlandse wetgever een bewuste (cf. MvT) zij het impliciete (cf. eigenlijke tekst UAVG) keuze gemaakt te hebben om verwerkingen van gegevens in het kader van archiveringsactiviteiten onder de uitzondering voor journalistieke, academische, artistieke en literaire doeleinden te laten vallen (3.4.2.4). Een mogelijke reden hiertoe is dat de AVG maar beperkte bewegingsvrijheid bood om een breed uitzonderingskader op te zetten voor “archivering in het algemeen belang”. Al zou een expliciete vermelding van (web)archivering door erfgoedinstellingen onder het genereuze uitzonderingsregime in art. 43 UAVG (“Uitzonderingen inzake journalistieke doeleinden of academische, artistieke of literaire uitdrukkingvormen”) meer duidelijkheid scheppen, kan men zich ook afvragen of de Nederlandse wetgever daarmee de grens opzoekt van wat toegelaten is volgens de AVG.

Gezien de ruimte die art. 85 AVG biedt is dit echter een juridisch verdedigbare keuze, waarbij ook de bijdrage die door erfgoedinstellingen wordt geleverd expliciet wordt gekoppeld aan de vrijheid van meningsuiting en de informatievrijheid. Archiveringsactiviteiten door erfgoedinstellingen zoals de KB zijn door de Nederlandse wetgever dus vrijgesteld van heel wat plichten en verantwoordelijkheden in de (U)AVG. Het uitzonderingsregime voor journalistieke, academische, artistieke en literaire doeleinden stelt vrijwel de hele UAVG buiten werking, alsook verschillende bepalingen in de AVG. Als gevolg van deze beslissing heeft de KB – en andere erfgoedinstellingen met vergelijkbare doelstellingen – significant minder werk met het interpreteren en implementeren van de (U)AVG.

Een alternatieve beleidsoptie komt in de bespreking van buitenlandse systemen naar voren. In tegenstelling tot de Nederlandse oplossing op basis van art. 43 UAVG wordt webharvesting in Duitsland, Frankrijk, Portugal en het VK onder de “archivering in het algemeen belang” uitzondering (art. 89 AVG) geschaard. Al deze landen sommen verder ook nog verschillende rechten van de betrokkenen op die onder bepaalde voorwaarden niet ingeroepen kunnen worden (zie overzicht in 4.4).

## Andere civiel-, straf- en bestuursrechtelijke rechtsposities

Tenslotte biedt de analyse ook aanknopingspunten voor een juridische oplossing op het terrein van onrechtmatige en illegale inhoud die in het kader van een nationale domeincrawl als “bijvangst” in een webarchief terecht zou kunnen komen. Het ligt om meerdere redenen voor de hand om in dat opzicht het aansprakelijkheidsprivilege voor hosting (art. 14 REH en art. 6:196c BW) als maatstaf te gebruiken en in het kader van de aansprakelijkheid van publieke erfgoedinstellingen voor onrechtmatige en illegale inhoud uit te gaan van dezelfde beperking van aansprakelijkheid als welke geldt voor hosting, tenzij er sprake is van bijzondere omstandigheden.

Ten eerste zijn culturele erfgoedinstellingen die zich met webharvesting bezighouden onbedoeld de bron van mogelijk onrechtmatige of illegale inhoud (van derden) op internet. Dit kan tenminste worden aangenomen zolang een culturele erfgoedinstelling niet zelf een actieve rol bij de overname in het webarchief heeft gespeeld en de overname het resultaat is van de toepassing van een automatisch harvestingsprocedé. Ten tweede mag niet over het hoofd worden gezien dat webharvesting een belangrijke bijdrage levert aan de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap (3.2.6). Tegen deze achtergrond is het problematisch om culturele erfgoedinstellingen op te zadelen met een aansprakelijkheidsrisico dat harvestingsactiviteiten zou kunnen frustreren. De analoge toepassing van het aansprakelijkheidsprivilege voor hosting neemt dit risico weg. Ten derde volgt een vergelijkbare

beperking ook uit de algemene beginselen van het aansprakelijkheidsrecht. Ten vierde biedt deze oplossing – vanuit praktisch oogpunt – het meeste houvast voor het opstellen van relevante kaders voor het instellen van een procedure voor meldingen van eventuele problematische inhoud en eventuele verwijdering (of beperking van toegang), namelijk het introduceren van notice-and-takedown procedures die in het licht van toekomstige ontwikkelingen op het terrein van de Digital Services Act verfijnd zouden kunnen worden.

# Summary

## Desirability from a policy perspective

From a policy perspective, web harvesting can be considered desirable. This follows from an analysis of the European and Dutch constitutional framework. In light of the need to protect freedom of expression, freedom of information and the right to research, it is legitimate and advisable to take measures that make web harvesting possible. With the creation of web archives, web harvesting contributes substantially to the realisation of these fundamental rights and freedoms. This applies not only to specific harvesting activities, such as event harvesting, but also to overarching national domain crawls (3.2.6).

From a fundamental rights perspective, however, it is also necessary to take rights into account that could be affected by web harvesting: copyright and related rights, database rights, the right to privacy and the right to data protection. Policy makers must therefore find a solution that allows web harvesting without unduly compromising intellectual property rights, privacy rights and data protection rights.

Hence, desirability from a policy perspective does not mean *carte blanche*. In the choice of measures to support web harvesting, the Dutch legislator is not entirely free. Developments in the case law of the Court of Justice of the European Union show that the required balance – between fundamental rights that are supported by web harvesting and fundamental rights that could be affected by web harvesting – should preferably be created *within* the legal framework that exists for these rights at the level of secondary EU legislation. This applies in particular to the system of exclusive rights and limitations in EU copyright law (3.2.6).

---

11

## Implementation in practice

Dutch cultural heritage institutions that wish to use web crawling and other harvesting techniques to build an archive of Dutch digital heritage face various practical problems. These include problems of an organisational (coordination of activities with other institutions), financial (limited capacity because of limited budget) and technical (lack of knowledge, need to keep pace with rapid technological developments) nature. In addition, conceptual questions arise with regard to the demarcation of the national domain when embarking on a national domain crawl (2.2.2). In the light of these challenges, there is little appetite among Dutch heritage institutions for a legal obligation and responsibility to archive the national domain in its entirety. They prefer a system that offers sufficient room to adjust harvesting activities to the capacity, financial resources and technical expertise available. The Dutch National Library (*Koninklijke Bibliotheek*) is interested in carrying out a national domain crawl. It has already developed extensive knowledge and expertise in this area.

From a legal perspective, it can be added that the opt-out system which is currently used in the Netherlands to request permission from rightholders of online content in the context of web harvesting is very time-consuming (2.2.3.1). Moreover, the rightholder of website content cannot always be traced. A webpage may contain third-party content, in respect of which the website manager is not in a position to give consent for web harvesting purposes. Against this background, it can be concluded that the existing harvesting practice poses particular problems. The focus on individual consent hinders the practical implementation of web harvesting in the Netherlands. This applies in particular to large-scale harvesting projects, such as a national domain crawl. An alternative approach – based on a statutory web harvesting entitlement or measures that

minimise legal risks – could simplify and speed up the existing web harvesting process to a significant degree. Ideally, it would make the current opt-out practice redundant.

An alternative scenario based on the development of specific legislation raises the further question of which institutions should be given the power to harvest the web. One option is a *centralised system* in which the legislator assigns the task of web harvesting to a few central heritage institutions, such as the Dutch National Library, The Netherlands Institute for Sound and Vision and the National Archive. In order to also meet the wishes and support collection strategies of smaller institutions, a system could be developed that allows smaller heritage institutions to submit web archiving requests to one of these central institutions. While the harvesting process would be outsourced to central institutions in this scenario, the selection would remain decentralised and, therefore, in the hands of specialised professionals in different cultural heritage institutions with specific knowledge. The implementation of this combined approach – a centralised system with the possibility for smaller institutions to submit individual harvesting requests – would require a careful further examination of the capacity and financial resources that are necessary to enable the dedicated central institutions to carry out the harvesting task not only with regard to their own collection strategy but also in reaction to requests received from smaller institutions.

It is also possible to envisage the creation of a *decentralised system* whereby a broad web-harvesting competence would be assigned globally to a large(r) group of cultural heritage institutions, such as all archives and publicly accessible libraries. To foster the exchange of knowledge and avoid the duplication of website copying, archiving and making available to the public, this system would require agreement about the scope of harvesting activities of the different institutions and the accessibility of resulting web archives. Existing collaboration initiatives, such as the *Netwerk Digitaal Erfgoed*, the *Nationaal Register Webarchieven* and national web archive consultation meetings, could serve as platforms to coordinate web archiving activities. In this scenario, each individual heritage institution would thus create its own web collection. Mutual agreement on archiving and distribution plans, however, could lead to central access and consultation facilities.

A final option is a *universal system* that offers all legal entities the opportunity to carry out web harvesting activities. In this case, individuals, NGOs and companies that pursue a commercial goal would also be able to invoke the web harvesting rules and create their own web archives.

## Legal solutions

The development of any of these policy options depends on the possibilities offered by the legal framework at the national and European level. In that respect, the analysis has provided the following guidelines for legal solutions with respect to copyright, related rights and database rights, privacy and data protection law, and issues arising in other fields of civil, criminal and administrative law.

### Copyright, related rights and database rights

The policy options in the areas of copyright, related rights and database rights can be presented on the basis of a matrix that distinguishes between the parameters of central/decentralised web harvesting competence and general national domain crawl/specific thematic selection. In the following overview, “NO” means that it is not advisable to develop legislation that permits web harvesting in this area. “YES” means that policy space exists which the Dutch legislator could use as a basis for the regulation of web harvesting:

	National domaincrawl	Specific selection
Competence: decentralised	NO	YES
Competence: centralised	YES	YES

#### *Decentralised competence - national domain crawl*

The analysis shows that the existing legal framework for the protection of copyright, related rights and database rights does not provide sufficient room for the adoption of a limitation of exclusive rights that would allow all cultural heritage institutions to carry out national domain crawls.

The option in the field of copyright and related rights to exempt “specific acts of reproduction” made by publicly accessible cultural heritage institutions offers the broadest basis for allowing acts of reproduction, such as the scraping, storing, indexing and archiving of web pages in the context of a national domain crawl (3.3.1.5). This rule is included in the European list of permitted exceptions and limitations which the national legislator is free to implement (Article 5(2)(c) InfoSoc Directive). The Dutch legislator has not made use of this possibility so far. The neutral condition of “specific” acts could offer room to introduce a clearly defined exception allowing non-commercial archives and publicly accessible libraries to make reproductions of websites in the context of web harvesting.

However, Recital 40 of the InfoSoc Directive raises several questions regarding this solution. Recital 40 excludes the application of the exception where the ultimate purpose is to make the harvested materials available in the context of online delivery. This wish – making archived websites available online, for example via an existing online infrastructure such as CLARIAH – was expressed by Dutch cultural heritage institutions (2.6). Recital 40 also clarifies that the exception should be limited to certain special cases that are covered by the reproduction right. This raises the question whether an exception can be deemed sufficiently specific when web harvesting involves various acts of reproduction and encompasses harvesting activities of all non-commercial archives and publicly accessible libraries in the Netherlands. Even if the three-step test does not seem to present insurmountable obstacles in this respect (3.3.4.2), the notion of “specific” acts could indicate a level of specification that would not allow the creation of a general harvesting competence for all heritage institutions. In this regard, the analysis of foreign legal systems shows that no examined country invoked Article 5(2)(c) InfoSoc Directive as a basis to justify a decentralised approach entitling all cultural heritage institutions to carry out national domain crawls. Furthermore, the CJEU applied Article 5(2)(c) in *TU Darmstadt* only to books that were already part of the collection of a university library (3.3.4.1). Therefore, it cannot readily be inferred from the *TU Darmstadt* decision that Article 5(2)(c) of the Infosoc Directive is applicable in cases where protected material, such as a copyright-protected web page, is not already part of the collection of a cultural heritage institution, but must first be integrated into the collection.

A broader interpretation of Article 5(2)(c) of the InfoSoc Directive in the Netherlands would therefore create legal uncertainty. It cannot be ruled out that a global domain crawl competence would be considered excessive and not “well-defined” when enabling all cultural heritage institutions to make web content part of their collections. Moreover, database law does not contain a corresponding exception (3.3.3.2). Thus, a regime based on “specific acts” of reproduction would not cover database rights.

In sum, it does not seem advisable to establish a decentralised competence entitling all cultural heritage institutions to carry out national domain crawls. This applies even more strongly to the universal approach outlined above, whereby in addition to cultural heritage institutions, individuals, NGOs and companies would have the option of conducting a national domain crawl.

In contrast to the described legal uncertainties arising from a broad decentralised approach, a centralised national domain crawl competence for selected cultural heritage institutions seems defensible when this harvesting possibility is based on a statutory deposit obligation covering online material. Denmark, Germany, France and the UK have chosen this route (Chapter 4). A legal obligation to deposit online material ensures that web content becomes part of the collection of cultural heritage institutions. Once this step has been taken, it becomes possible to rely on limitations of copyright, neighbouring rights and database rights that are linked to the collection status, such as the exemption of preservation copies for the archiving of web pages (3.3.1.2) and the reading terminal exception for providing access to the web archive (3.3.2.1). Furthermore, it is possible to introduce limitations of protection that allow for reproductions and acts of making available for the purpose of scientific research (3.3.1.4). In this respect, it should be mentioned that the database right can be considered as a potential source of legal uncertainty because of the lack of exceptions for acts of making available to researchers and via reading terminals (3.3.3.2). However, in the light of the special infringement criteria that have evolved, inter alia, in the case law of the CJEU, infringement of sui generis database rights does not seem plausible. Historical archival versions of databases that are incorporated into a web archive as a result of web harvesting and accessed according to specific access rules will usually not prevent the rightholder from recovering investments (3.3.3).

In all foreign legal systems under examination, the application of a statutory obligation to deposit online material led to a centralisation of web-harvesting activities in the hands of individual cultural heritage institutions, such as the national library. This ensures a clear institutional demarcation of the web-harvesting competence. Furthermore, it can be said that the bundling of the harvesting authority contributes to a clear and precise definition of the use privilege. Legislators in countries with a legal deposit obligation do not hesitate to refer to web harvesting activities as “specific acts” of reproduction in the sense of Article 5(2)(c) of the InfoSoc Directive (4.2.2.3).

Legislation that centralises the web-harvesting competence in the hands of one or more heritage institutions need not take the form of deposit legislation. Although this is the case in most of the foreign legal systems that have been studied, the starting points for this solution in other countries differ from the situation in the Netherlands. In the other countries under examination, a legal deposit system already existed for analogue material prior to the adoption of web-harvesting legislation. This pre-existing system has simply been extended to online publications, with all associated consequences: a legal obligation arises for the website holder (publisher) and/or the library to deposit the material, based on the rules and guidelines that come along with traditional legal deposit legislation.<sup>2</sup> However, the Netherlands need not follow the same approach. A statutory deposit obligation ensuring that relevant websites become part of the collection of cultural heritage institutions can be included in specific web-harvesting legislation. It is not necessary to create general legal deposit legislation for all publications – analogue and digital – in this context. A place to introduce and accurately describe a specific competence and deposit obligation for online sources could be found, for example, in the laws that define the tasks of selected web-harvesting institution(s). In the Netherlands, for example, this would be the Higher Education and Scientific Research Act (*Wet op het hoger onderwijs en wetenschappelijk*

---

<sup>2</sup> In 2000, UNESCO issued a renewal of the *Guidelines for Legal Deposit Legislation* developed by J. Lunn in 1981, which addresses the deposit of electronic material in more detail. The guidelines also set out what a legal deposit system should look like. Choosing for deposit legislation means that the legislator will have to take this framework into account. It strongly advises against a voluntary system. See: J. Larivière, *Guidelines for Legal Deposit Legislation*, UNESCO 2000, available at: <http://www.unesco.org/new/en/communication-and-information/resources/publications-and-communication-materials/publications/full-list/guidelines-for-legal-deposit-legislation/>.

*onderzoek*) for the Dutch National Library and the Media Act (*Mediawet 2008*) for the Institute for Sound and Vision (2.2.4.3).

An alternative solution for web harvesting can be found in Portugal. The Portuguese system is primarily based on a societal consensus that is also shared by rightholders (4.2.5). In Portugal, a centralised harvesting system has emerged on this basis – without a legal deposit obligation – which, in terms of volume and frequency of the national domain crawl and the degree of accessibility of the web archive, appears to be more extensive than the harvesting activities in many other countries. It is particularly noteworthy that the Portuguese institution Arquivo.pt has managed to make the web archive – with search functionality – available to the entire Internet public. Despite the considerable volume of harvesting activities, Arquivo.pt has not yet been confronted with infringement claims in Portugal. The system appears to be largely the result of implicit consent and/or tacitly granted licences.

This alternative model may offer new impulses for the discussion in the Netherlands as well. In the light of the experiences in Portugal, the question arises whether heritage institutions could expand their harvesting activities – both acts of reproduction (crawling and archiving) as well as acts of making available – based on an analysis and assessment of infringement risks and the development of appropriate measures to reduce those risks. In Portugal, an embargo period and an efficient notice-and-takedown system are central pillars of the harvesting system. The embargo period prevents direct competition with the exploitation of recent versions of source material offered on the Internet. The notice-and-takedown system offers rightholders the possibility to report infringements and ensure the blocking of access to problematic archive material. In the Netherlands, similar building blocks could also allow a switch from the current burdensome system with permissions and opt-out mechanisms, to a system with an embargo period and notice-and-takedown options. The Dutch legislator could support this solution by introducing an indemnity against liability comparable to the safe harbour regime for hosting services (3.5.3). It is also conceivable to give specific cultural heritage institutions an explicit legal mandate (perhaps even reflecting the desirability of giving the entire Internet public access to archived material) to emphasise the legitimacy of web harvesting and dispel potential doubts about the scope and reach of harvesting activities developed by heritage institutions.

---

15

A further solution could follow from a system of extended collective licensing (ECL). A basis for this can be found in Article 12 of the DSM Directive and Article 45 of the Dutch Copyright Act (*Auteurswet*). Article 12 allows EU Member States to introduce an ECL scheme for clearly circumscribed purposes of use on their own territory. The legal form of ECL is characterised by a combination of voluntary collective management with a statutory extension of the scope of the collective license to rightholders who are not affiliated with the collecting society in question (3.3.4.5). The ECL system should only be applied in well-defined fields of application, where obtaining permissions from rightholders on an individual basis is, as a rule, cumbersome and impractical to such an extent that the required licensing transaction is unlikely to take place. In the case of web harvesting, this condition appears to be fulfilled. The question is, however, whether sufficiently representative collecting societies exist in the Netherlands with regard to all protected elements of web sites that would require rights clearance.

Various collecting societies exist in the Netherlands, including those for writers, translators and freelance journalists (*Lira*) and visual artists (*Pictoright*). However, these collecting societies do not represent all rightholders whose works are used in the context of web harvesting. For example, web designers are not represented as such in any Dutch collecting society. The ECL system, therefore, does not seem capable of providing a complete solution for web harvesting. In addition, it is questionable whether it is desirable from a cultural policy perspective to make the future of web harvesting dependent on collective licenses that have to be agreed upon voluntarily. If no agreement with a collecting society can be reached, no web

harvesting will take place. Even Scandinavian countries, which traditionally have very well-developed ECL systems, have given preference to a solution based on a statutory deposit obligation to enable web harvesting. The analysis of the system in Denmark (4.2.1) attests to this policy choice.

#### *Decentralised competence - specific selection*

The fact that centralised systems provide greater legal certainty for a national domain crawl need not lead to the conclusion that heritage institutions outside the centralised system can no longer undertake any harvesting activity. While, as described, it is not advisable to give all cultural heritage institutions general authority to carry out national domain crawls, the legal regimes for the protection of copyright, neighbouring rights and database rights leave room for more specific harvesting projects, covering only pre-selected web sites. The existing legal framework offers room for harvesting activities with this specific focus. The available breathing space goes beyond the option to exempt specific acts of reproduction.

For example, the new rules relating to text and data mining (Articles 3 and 4 of the DSM Directive and Articles 15n and 15o of the Dutch Copyright Act, 3.3.1.3) provide a basis for targeted harvesting activities in the context of scientific research projects, if web content is retrieved for the purpose of analysing it and generating information for research purposes.

A similar conclusion can be drawn with regard to the possibility of permitting acts of reproduction and disclosure “for the sole purpose of [...] scientific research” (Article 5(3)(a) of the InfoSoc Directive, 3.3.1.4). As the text and data mining rules, this copyright limitation requires a sufficiently clear connection with scientific research. Web harvesting based on a “reasoned selection” of web content or following from an “event harvesting” initiative, might be justifiable under this provision, when it is carried out in the framework of joint projects between heritage institutions and research organisations. However, there is currently no room for this solution in Dutch law because the optional limitation of exclusive rights for the “sole purpose” of scientific research has not been implemented in the Netherlands.

A legal regulation of selective harvesting activities on the basis of the new legislation on text and data mining and the exemption of use for the purposes of scientific research has the advantage that sui generis database law – at least with regard to acts of reproduction – explicitly offers comparable breathing space (3.3.3.1 and 3.3.3.2). This solution also makes it possible for individuals, NGOs and companies to harvest websites in the context of text and data mining activities. The scope of Article 4 of the DSM Directive and Article 15o of the Dutch Copyright Act is not limited to cultural heritage institutions. Hence, these provisions offer a basis for a broader regulation including other interested parties.

#### *Central competence - specific selection*

The described possibilities for justifying web harvesting with a specific focus on pre-selected online content are available to all cultural heritage institutions. This includes those heritage institutions that are specifically entitled to carry out general national domain crawls in a centralised system. They can thus conduct both national domain crawls and specific harvesting projects with a particular focus.

### **Privacy and data protection law**

The analysis also explored the possibilities in the field of privacy and data protection law. In this respect, the Dutch legislator seems to have made a conscious (see the Explanatory Memorandum) albeit implicit (see the actual text of the Dutch implementing legislation of the General Data



Protection Regulation (GDPR), the *Uitvoeringswet Algemene verordening gegevensbescherming* (UAVG)) choice to bring the processing of personal data in the context of web archiving under the umbrella of the exception regime for journalistic, academic, artistic and literary purposes (3.4.2.4). A possible reason for this is that the GDPR provides only limited room to set up a broad exemption framework for “archiving purposes in the public interest”. Although an explicit mention of web archiving by heritage institutions under the generous exception regime in Article 43 UAVG (“for journalistic purposes and the purposes of academic, artistic or literary expression”) would create more clarity, one can also wonder whether the Dutch legislator is thereby pushing the limit of what is allowed under the GDPR.

However, given the scope offered by Article 85 GDPR, this is a legally defensible choice that also explicitly links the contribution made by heritage institutions to freedom of expression and freedom of information. Consequently, the Dutch legislator has exempted heritage institutions such as the Dutch National Library from many duties and responsibilities under the GDPR and the UAVG. The exemption regime for journalistic, academic, artistic and literary purposes practically exempts web harvesting project from the entire spectrum of obligations in the UAVG, as well as various provisions in the GDPR. As a result of this decision, the Dutch National Library – and other heritage institutions with similar objectives – have significantly less work in interpreting and implementing the GDPR and the UAVG.

An alternative regulatory option emerged from the examination of foreign legal systems. Unlike the Dutch solution on the basis of Article 43 UAVG, Germany, France, Portugal and the UK include web harvesting in the exemption for “archiving purposes in the public interest” (Article 89 GDPR). All these countries also list various rights of data subjects that cannot be invoked under certain conditions (see overview in 4.4).

## **Other civil, criminal and administrative legal problems and possibilities**

17

---

Finally, the analysis also offers starting points for a legal solution in the area of unlawful content that may end up in a web archive as “by-catch” in the context of a national domain crawl. For several reasons, it makes sense to use the liability privilege for hosting (Article 14 of the E-Commerce Directive and Article 6:196c of the Dutch Civil Code (*Burgerlijk Wetboek*)) as a yardstick in this respect and assume that cultural heritage institutions can benefit from the same liability privilege with regard to unlawful content that is available for hosting services, unless there are special circumstances.

Firstly, cultural heritage institutions that engage in web harvesting are unintentionally collecting unlawful or illegal content which third parties make available on the Internet. This can at least be assumed as long as a cultural heritage institution has not itself played an active role in the selection of content that finally becomes a part of the web archive and the inclusion of problematic content is the result of an automatic harvesting process. Secondly, it should not be overlooked that web harvesting makes an important contribution to freedom of expression, freedom of information and the right to science (3.2.6). Against this background, it would be problematic to burden cultural heritage institutions with a liability risk that could frustrate harvesting activities. The analogous application of the hosting liability privilege removes this risk. Third, a similar limitation of liability can be deduced from general principles of liability law. Fourth, from a practical point of view, a solution based on the liability regime for hosting – known from the E-Commerce Directive – provides maximum guidance for establishing mechanisms for notifying potentially problematic content and ensuring the takedown (or blocking of access), namely the introduction of notice-and-takedown procedures that could be refined in the light of future developments in the field of the proposed Digital Services Act.