



PBLQ

Het faciliteren van webharvesting

rapportage
16 april 2025

Managementsamenvatting

Het Nederlandse erfgoed bestaat niet meer enkel uit fysieke objecten, zoals kranten of tijdschriften. Ook digitale objecten zoals websites en webcontent zijn onderdeel van het Nederlands cultureel erfgoed. Webcontent kan tegenwoordig worden meegenomen bij het vormen van een beeld van de Nederlandse samenleving ten behoeve van bijvoorbeeld onderzoek. Echter, het web is constant in beweging en online content wordt regelmatig aangepast, toegankelijk gemaakt of zelfs verwijderd. Daardoor bestaat de mogelijkheid dat materiaal verloren gaat.

Dit is de reden dat Nederlandse erfgoedstellingen wijzen op het belang van webarchivering: het verzamelen en op systematische wijze ordenen en bewaren van (delen van) het nationale internetdomein voor toekomstig onderzoek en voor het maatschappelijk belang van het doorgeven van erfgoed aan toekomstige generaties alsmede de rechten ten aanzien van de vrijheid van informatie. Om dit te kunnen doen is webharvesting nodig. Webharvesting is een middel om webarchieven te vormen en die voor toekomstige generaties veilig te stellen. In een urgentiebrief aan de commissie van Onderwijs, Cultuur en Wetenschap van de Tweede Kamer hebben de Koninklijke Bibliotheek (KB), Koninklijke Vereniging van Archivarissen in Nederland (KVAN), het Nationaal Archief (NA) en het Nederlands Instituut voor Beeld en Geluid (Beeld en Geluid) aangegeven dat in de huidige situatie er veel webarchieven verloren gaan vanwege het gebrek aan een wettelijke grondslag om websites te harvesten. Daarom pleiten deze partijen voor een wettelijke facilitering van webharvesting door erfgoedinstellingen. Het gaat hierbij om het harvesten/crawlen, acquireren, conserveren en toegankelijk maken (presenteren via een leesterminal) van het Nederlandse webdomein, zowel in zijn geheel (nationale domeincrawl) als meer selectief (beredeneerde selecties). Ook is op 24 oktober 2023 in de Tweede Kamer de motie Dekker-Abdulaziz (Kamerstuk 26 643, nr. 1076) aangenomen, waarin de regering wordt verzocht om in kaart te brengen wat de wettelijke, technische en financiële obstakels zijn die in de weg staan van een mogelijke archivering van het nationale webdomein en hoe deze te overkomen.

Het Instituut voor Informatierecht (IViR) heeft een rechtsvergelijkend onderzoek uitgevoerd naar webharvesting in Europese lidstaten. In het IViR rapport wordt benoemd dat webharvesting bijdraagt aan het waarborgen van de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap¹. Het IViR rapport concludeert dat beleidsmakers een juridische oplossing dienen te vinden die ruimte biedt voor webharvesting, zonder conflicterende rechten en belangen te zeer in het gedrang te brengen. De potentieel conflicterende rechten en belangen vinden hun grond op basis van bestaande fundamentele rechten. Het gaat hierbij om het spanningsveld tussen het auteursrecht, de naburige rechten en het sui generis databankenrecht, de bescherming van persoonsgegevens en fundamentele rechten zoals het informatie- en erfgoedbelang.

Doelstelling en opzet van dit onderzoek

Het onderzoek van IViR geeft inzicht in de juridische aspecten van webharvesting, maar beschrijft nog geen oplossingsrichting om dit in de praktijk te faciliteren. In het (voorliggende) onderzoek wordt gekeken naar een aantal scenario's in een breder perspectief waarbij ook organisatorische, technische en financiële aspecten zijn onderzocht. Het doel van dit onderzoek is om deze scenario's tegen elkaar te kunnen afwegen, zodat het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW)

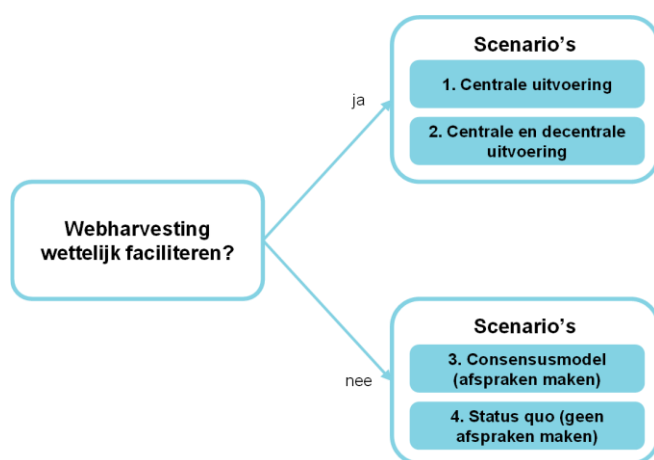
¹ Webharvesting, IViR, (2021), p. 55.

een zo goed mogelijk geïnformeerde en onderbouwde keuze kan maken over het al dan niet faciliteren van webharvesting. Het onderzoek moet daarbij bijdragen aan het maken van een keuze voor een oplossingsrichting op hoofdlijnen, waarna er binnen deze oplossingsrichting nog meer keuzes gemaakt moeten worden over de exacte juridische invulling en de uitvoering van webharvesting die in dit onderzoek niet zijn uitgewerkt.

Ten aanzien van het faciliteren van webharvesting zijn er in ieder geval keuzes te maken op drie onderdelen:

1. **Juridische keuzes** (of routes): de keuzes die te maken zijn vanuit juridisch perspectief, waaronder het al dan niet wettelijk faciliteren van webharvesting en de specifieke route die hiervoor gekozen kan worden;
2. **Scenario's**: een van de vier scenario's om webharvesting uit te voeren, namelijk (1) centrale uitvoering, (2) centrale én decentrale uitvoering, (3) consensusmodel, (4) status quo; en
3. **Afbakening van het nationaal webdomein**: de wijze waarop kan worden bepaald welke webdomeinen tot het nationaal webdomein behoren ten behoeve van harvestingsactiviteiten.

De juridische keuzes bespreken we gekoppeld aan de scenario's, waarbij een onderscheid wordt gemaakt voor *wettelijk faciliteren* en *niet wettelijk faciliteren*. Scenario 1 en 2 vragen om het wettelijk faciliteren van webharvesting. Scenario 3 en 4 zijn gebaseerd op andersoortige juridische afspraken (niet wettelijk). Dit staat weergegeven in het onderstaande figuur:



De te maken keuzes hebben verschillende gevolgen (die we in ons onderzoek *impactcategorieën* noemen), namelijk:

1. **Informatie- en erfgoedbelang**: De mate waarin het belang van informatievrijheid en het belang van Nederlands erfgoed en het doorgeven daarvan aan toekomstige generaties wordt gediend (of niet);
2. **Belang van auteursrechthebbenden**: De mate waarin het belang van (auteurs)rechthebbenden wordt gediend of geschaad;
3. **Belang van betrokkenen**: De mate waarin het belang van betrokkenen en de bescherming van hun persoonsgegevens wordt gediend of juist geschaad;
4. **Juridische risico's**: De mate waarin een juridische route binnen bereik is (haalbaar) en toekomstvast is (houdbaar);
5. **Kosten**: De kosten die bij een keuze horen;

6. **Draagvlak:** Het draagvlak voor een keuze bij de betrokken partijen;
7. **Uitvoerbaarheid:** De mate waarin of wijze waarop een keuze uitvoerbaar is, anders dan in juridische zin.

Per keuze hebben we de effecten in kaart gebracht en beschreven. De juridische route beoordelen we op de juridische haalbaarheid en houdbaarheid (onderdeel 4). De scenario's en de afbakening van het nationaal webdomein beoordelen we op de overgebleven impactcategorieën (1 t/m 3 en 5 t/m 7).

1. Juridische keuzes

De belangrijkste keuze ten aanzien van de juridische route is of webharvesting wel of niet wettelijk wordt gefaciliteerd. Het IViR heeft aanbevolen de webharvestingactiviteiten juridisch te waarborgen. Er zijn ook oplossingsrichtingen zonder wettelijke waarborging. Beide juridische routes worden hieronder toegelicht.

Wettelijk faciliteren

Binnen het wettelijk waarborgen dan zijn er twee hoofdrichtingen:

- het opnemen van een uitzondering in het auteursrecht; of
- het vastleggen van de bevoegdheid tot webharvesten in een nadere wettelijke regeling.

Beide mogelijkheden kennen juridische risico's ten aanzien van de houdbaarheid en haalbaarheid van de te treffen (wettelijke) regeling.

Onder experts bestaan grote twijfels bij de juridische houdbaarheid van de route via het auteursrecht, omdat de voorgestelde uitzondering in het auteursrecht juridisch gezien niet bedoeld is voor de gewenste toepassing ten behoeve van webharvesting en ook niet door andere lidstaten zo wordt toegepast. Met andere woorden, de beoogde route via het auteursrecht past niet goed binnen de bestaande (Europese) regelgeving. Dit maakt het in beperkte mate toekomstvast, omdat het kwetsbaar is voor juridische procedures en negatieve beoordeling van (juridische) adviesorganen. De beperkte houdbaarheid maakt tevens dat er weerstand is onder experts tegen deze juridische route, waardoor er ook twijfels zijn over de haalbaarheid ervan.

De route om webharvesting juridisch te waarborgen via het vastleggen van een expliciete wettelijke taak of bevoegdheid voor een of meer erfgoedinstellingen om deze activiteiten uit te voeren blijkt op basis van dit onderzoek het meest houdbaar en haalbaar. Deze route is het meest toekomstvast en biedt een grote mate van rechtszekerheid voor rechthebbenden, erfgoedinstellingen en betrokkenen. Ook geeft deze route een verduidelijking in de grondslag voor de verwerking van persoonsgegevens en de daarbij komende verantwoordelijkheden. Vanuit juridisch perspectief geeft een combinatie van een wettelijke depotverplichting met een beroep op de beperkingen op de verschillende intellectuele eigendomsrechten een hoge mate van rechtszekerheid. Deze variant is het meest toekomstvast omdat het de meeste duidelijkheid kan verschaffen ten aanzien van webharvesting: auteursrechthebbenden kunnen voorzien welke inbreuken op hun rechten kunnen worden gemaakt (en voor welk doeleinde en onder welke voorwaarden), erfgoedinstellingen weten daarmee welke juridische ruimte ze hebben voor hun (webharvesting)activiteiten en voor betrokkenen geldt dat voorzienbaar is voor welke doeleinden en onder welke voorwaarden hun persoonsgegevens verwerkt kunnen worden.

Erfgoedinstellingen hebben vraagtekens gezet bij de haalbaarheid van het wettelijk vastleggen van een taak of bevoegdheid, met als belangrijkste vraagpunt de mogelijkheid voor regionale instellingen

om ook aan webharvesting te kunnen doen. Een wetgevingstraject moet duidelijkheid verschaffen over de exacte invulling van deze route, waarbij er keuzes te maken zijn welke instellingen de wettelijke bevoegdheid krijgen, met welk bereik en voor welk materiaal.

Niet wettelijk faciliteren

Ten aanzien van het niet wettelijk faciliteren van webharvesting zijn er twee varianten:

- Het maken van afspraken met auteursrechthebbenden over het harvesten van materiaal waar auteursrechten op rusten, en
- Het in stand houden van de huidige praktijk (niets doen).

Het maken van afspraken is gebaseerd op een informeel systeem. Dit biedt een mate van rechtszekerheid ten aanzien van auteursrechthebbenden, maar niet als het gaat om de verwerking en bescherming van persoonsgegevens. Deze route is daarom beperkt juridisch houdbaar omdat het kwetsbaar is voor juridische argumenten en procedures ten aanzien van de bescherming van persoonsgegevens. Ook is er geen zekerheid dat (sommige) rechthebbenden geen bezwaar zullen maken tegen het harvesten van hun materiaal, omdat gemaakte afspraken niet bindend zijn. Gezien het draagvlak bij auteursrechthebbenden lijkt deze route wel binnen bereik en juridisch haalbaar, hoewel dit enige inspanning zal vragen. Een belangrijk risico is dat niet alle auteursrechthebbenden vooraf deel hoeven te nemen in de te maken afspraken, wat de onzekerheid vergroot.

Aangaande de haalbaarheid van scenario 4 geldt dat deze overeenkomt met de huidige praktijk, en dus technisch gezien haalbaar is. Wel zijn er zorgen over de voortzetting van bestaande webharvestingactiviteiten gezien het gebrek aan rechtszekerheid. Daarnaast is de houdbaarheid van dit scenario zeer beperkt, gezien het ontbreken van een wettelijke grondslag voor de verwerking van persoonsgegevens en het ontbreken van een grond voor inbreuk op auteursrechten. Ook is de huidige situatie niet transparant en voorzienbaar voor rechthebbenden en betrokkenen. Deze route is het minst houdbaar van alle routes.

2. Scenario's voor uitvoering van webharvesting

In dit onderzoek zijn vier scenario's uitgewerkt voor de uitvoering van webharvesting, rekening houdend met de hierboven beschreven juridische keuzes. Twee van deze scenario's gaan uit van het wettelijk faciliteren van webharvesting (scenario 1 en 2) en twee scenario's gaan uit van niet (scenario 3 en 4). In alle scenario's wordt er rekening gehouden met de twee verschillende vormen van webharvesting: de *nationale domeincrawl* en *beredeneerde selecties*. Een nationale domeincrawl legt een momentopname van het nationale webdomein vast op relatief oppervlakkig niveau, zodat er een algemeen beeld van het Nederlandse internetdomein ontstaat. Bij een beredeneerde selectie wordt een vooraf vastgestelde lijst van domeinen verdiepend geharvest, zodat alle relevante kenmerken (bijvoorbeeld met betrekking tot een onderwerp of gebeurtenis) van een domein worden gepreserveerd.

De vier scenario's zijn:

1 Centrale uitvoering door één instelling met loketfunctie voor andere instellingen.

In dit scenario is er één centraal gekozen instelling die uitvoering geeft aan zowel de nationale domeincrawl als beredeneerde selecties. Decentrale instellingen kunnen een verzoek doen bij de centrale instelling om specifieke websites te harvesten, waartoe de centrale instelling een loketfunctie heeft om deze verzoeken te behandelen en uit te voeren.

2 Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

Net als in het eerste scenario voert ook in dit scenario één centraal gekozen instelling nationale domeincrawl (en beredeneerde selecties) uit, maar nu kunnen ook decentrale archief- en erfgoedinstellingen beredeneerde selecties (zelfstandig) uitvoeren.

3 Afspraken maken over uitvoering van webharvesting (consensus)

In dit scenario worden er expliciete en duidelijke afspraken gemaakt met recht- en belanghebbenden over webharvesting, aangevuld met een notice-and-takedown-systeem, maar wel zonder wettelijke basis. Het uitvoeren van de beredeneerde selecties blijft zoals nu (zonder wettelijke basis).

4 Status quo (geen afspraken maken)

Dit is de situatie zoals deze momenteel is. Er vindt geen nationale domeincrawl plaats. Wel worden er – de huidige praktijk – door een aantal partijen beredeneerde selecties uitgevoerd op basis van een opt-out voor auteursrechthebbenden.

Uit onze analyse blijkt dat de huidige situatie (scenario 4) de meeste relatief negatieve effecten met zich meebrengt. Alleen de kosten zijn dit scenario lager dan in andere scenario's, wat voornamelijk veroorzaakt wordt door het niet uitvoeren van de nationale domeincrawl en het aantal uitgevoerde beredeneerde selecties. Dit heeft tevens een relatief negatief effect op het informatie- en erfgoedbelang. Het belang van auteursrechthebbenden, de bescherming van persoonsgegevens en het draagvlak vallen in de status quo slechter uit, wat te maken heeft met het gebrek aan juridische waarborgen en rechtszekerheid voor alle betrokken partijen.

In scenario 3, het consensusmodel, is de verwachting dat auteursrechthebbenden er vanwege specifieke afspraken en mogelijke bijbehorende compensatie beter vanaf zullen zijn. Op andere aspecten waaronder het draagvlak onder erfgoedinstellingen en de uitvoerbaarheid van het maken en beheren van afspraken beoordelen wij dit scenario minder goed. Ook zijn er in dit scenario geen juridische waarborgen voor de bescherming van persoonsgegevens aanwezig, zoals een wettelijke grondslag om te harvesten.

Scenario 1 en scenario 2 worden op basis van dit onderzoek relatief gelijk beoordeeld, behalve dat scenario 2 beter uitvoerbaar is vanwege het ontbreken van een centrale loketfunctie die verschillende vraagstukken met zich meebrengt in de uitvoering en vanwege het draagvlak bij erfgoedinstellingen.

Gekeken naar de beoordeling per impactcategorie, dienen scenario 1 en 2 het beste het informatie- en erfgoedbelang vanwege de grootste dekkingsgraad van harvestingactiviteiten. Dit komt door de uitvoering van de nationale domeincrawl en een groter aantal beredeneerde selecties. In andere scenario's is er (zeer waarschijnlijk) geen haalbare mogelijkheid voor de uitvoering van de nationale domeincrawl.

De belangen van auteursrechthebbenden worden voor belangrijke mate bepaald door enerzijds gemaakte afspraken of juridische/organisatorische zekerheid en waarborgen en anderzijds de omvang van harvestingactiviteiten, in het bijzonder de uitvoering van een nationale domeincrawl. Scenario 1 en 2 bieden de meeste waarborgen en zekerheid, maar ook in scenario 3 zijn de belangen van auteursrechthebbenden geborgd. Wel geldt voor alle scenario's dat webharvesting in algemene zin inbreuk maakt op auteursrechten.

Als het gaat om de bescherming van persoonsgegevens dan is het verschil tussen scenario's de juridische basis, grondslag en bijbehorende voorzienbaarheid² die er is ten aanzien van de verwerking van persoonsgegevens. In scenario 1 en 2 zijn er de meeste wettelijke waarborgen, hoewel dit wel afhankelijk is van de te maken juridische keuzes (waarbij het vastleggen van een wettelijke bevoegdheid de meeste waarborgen geeft). In scenario 3 en 4 zijn deze waarborgen er niet.

Er zijn verder beperkte verschillen in kosten tussen de scenario's. Scenario 1 heeft lagere kosten dan scenario 2 wanneer er meer dan elf erfgoedinstellingen aan webharvesting doen (break-even punt), omdat bij dit aantal er sprake is van schaalvoordelen van de loketfunctie. In scenario 3 zijn relatief gezien de meeste kosten te verwachten, wat voornamelijk komt door de benodigde notice-and-takedown voorziening. In scenario 4 zijn er de minste kosten.

Scenario 2 geniet het meeste draagvlak onder belanghebbenden, voornamelijk omdat dit scenario ten opzichte van scenario 1 en 3 het beste aansluit bij de huidige uitvoeringspraktijk, er wettelijke waarborgen mogelijk zijn (afhankelijk van de juridische route) ten aanzien van de belangen van verschillende partijen en er sprake is van een hoge mate van rechtszekerheid. Deze laatste onderdelen zijn afwezig in scenario 4.

Vanwege dezelfde reden is de uitvoerbaarheid van scenario 2 het hoogst beoordeeld: het past het beste bij de huidige uitvoeringspraktijk. Het belangrijkste verschil in uitvoerbaarheid ten opzichte van scenario 1 is de loketfunctie.

In de onderstaande tabel staat onze beoordeling van de scenario's weergegeven:

	Scenario 1 (centrale uitvoering)	Scenario 2 (centrale en decentrale uitvoering)	Scenario 3 (consensus)	Status quo (geen afspraken maken)	Toelichting
Informatie- en erfgoedbelang					Hoe meer webarchieven geharvest (kunnen) worden, hoe beter het informatie- en erfgoedbelang geborgd is (scenario 1 en 2).
Belang van auteursrecht- hebbenden					Er zijn geen grote verschillen tussen scenario 1, 2 en 3: Hoe meer rechtszekerheid, voorzienbaarheid en/of compensatie, hoe beter de belangen van auteursrechtelhebbenden geborgd zijn.
Bescherming persoons- gegevens					Hoe meer waarborgen, zoals een grondslag voor verwerking, hoe beter de bescherming van persoonsgegevens geborgd is (scenario 1 en 2).
Kosten					De kosten verschillen niet in grote mate tussen de scenario's, maar zijn wel afhankelijk van de hoeveelheid opslag en hoeveelheid uitvoerende

² Voorzienbaarheid is gerelateerd aan de indicator met betrekking tot de inbreuk op persoonsgegevens en geeft aan of het voor een betrokkenen voorzienbaar is dat zijn/haar persoonsgegevens worden verwerkt (in het kader van webharvesting) bij het doen van webuitingen.

					partijen. Scenario 3 heeft een p.m. post voor de notice-and-takedown, waardoor deze waarschijnlijk (met een mate van onzekerheid) hoger uitvalt dan de andere scenario's.
Draagvlak	■	■	■	■	Het draagvlak is het grootst bij het wettelijk faciliteren van de huidige praktijk (waarbij erfgoedinstellingen zelf al harvesten) en het waarborgen en voorzienbaar maken van harvestingactiviteiten (scenario 1 en 2).
Uitvoerbaarheid	■	■	■	■	Hoe minder onzekerheden er zijn bij het uitvoeren van webharvesting (vanwege praktische en/of wettelijke aspecten; het kleinst bij scenario 2 en in mindere mate scenario 1), hoe beter uitvoerbaar dit is voor erfgoedinstellingen.

Afbakening van het nationaal webdomein

Een algemene definitie van het nationaal webdomein is *alle webactiviteiten die verband houden met Nederland (als natiestaat)*³. Een afbakening van het nationaal webdomein is relevant bij het juridisch verankeren welke webdomeinen kunnen worden geharvest en welke domeinen er binnen de reikwijdte van de nationale domeincrawl en beredeneerde selecties liggen.

De operationalisatie van deze definitie is gebaseerd op criteria om te bepalen of een webdomein bij het nationale webdomein hoort. Deze criteria kunnen onder andere het top-level domein van een website (zoals .nl, .frl of .amsterdam), de taal, de cultuur of de geografische locatie (van Nederlandse servers) zijn.

In dit onderzoek stellen we twee strategieën voor het definiëren van het nationale webdomein, namelijk:

1. Een strategie op basis van vooraf vastgelegde criteria die zoveel mogelijk automatiseerbaar en objectief zijn. Een voorbeeld hiervan is het gebruik van top-level domeinen.
2. Een strategie waarbij criteria niet vooraf vastgelegd zijn, maar via een vastgesteld proces worden geformuleerd. Naast automatiseerbare en objectieve criteria gaat het hierbij ook over andere criteria die om meer specifieke of individuele beoordeling vragen, zoals cultuur en erfgoed.

Ten aanzien van de afbakening van het nationaal webdomein hebben beide strategieën voor- en nadelen. Geen van de twee strategieën heeft over de impactcategorieën in grote mate meer voor- of nadelen ten opzichte van de andere strategie. Een strategie met criteria die via een vastgesteld proces worden bepaald geeft voordelen op het gebied van het informatie- en erfgoedbelang omdat de dekkingsgraad groter is, en op het gebied van draagvlak bij uitvoerende partij(en) omdat het meer flexibiliteit biedt. Een strategie op basis van vastgestelde criteria heeft relatief gezien meer voordelen op andere impactcategorieën, zoals de kosten en de voorzienbaarheid van webharvestingactiviteiten

³ Dit is de definitie zoals ook door het IViR gehanteerd wordt. Zie ook N. Brügger, D. Laursen & J. Nielsen, Exploring the domain names of the Danish web. In: N. Brügger & R. Schroeder (eds) The Web as History: Using Web Archives to Understand the Past and the Present. London: UCL Press 2017, pp. 64.

voor auteursrechthebbenden en betrokkenen omdat er waarschijnlijk minder webdomeinen worden geharvest en vooraf duidelijker is welke domeinen geharvest gaan worden.

Conclusie

Het doel van dit onderzoek was niet om te adviseren over een te maken keuze, maar om de verschillende opties van argumenten te voorzien. Het is aan de opdrachtgever om de onderzoeksbevindingen te delen met betrokkenen en om de bevindingen ten aanzien van de juridische routes en scenario's voor uitvoering te wegen. In deze conclusie geven we nogmaals kort en bondig weer wat de belangrijkste verschillen zijn in effecten van de te maken keuzes, te beginnen bij de juridische routes.

De huidige situatie (status quo) is juridisch gezien het minst houdbaar en toekomstvast, omdat er geen recht wordt gedaan aan de afwegingen van verschillende belangen en grondrechten. Van de andere opties is het bij wet creëren van een publieke taak of bevoegdheid in combinatie met depotverplichting juridisch gezien de meest haalbare en houdbare optie. Binnen deze hoofdroute zijn er vervolgkeuzes te maken over de wijze waarop deze taak of bevoegdheid wordt toegewezen aan één of meerdere instellingen (namelijk attribueren, delegeren of mandateren). Deze vervolgvarianten zijn in dit onderzoek niet in detail uitgewerkt maar worden wel toegelicht in Hoofdstuk 4.

Naast de keuze met betrekking tot het wettelijk faciliteren van webharvesting, is de belangrijkste vraag op welke manier webharvesting gaat worden uitgevoerd. Hiervoor zijn vier scenario's uitgewerkt.

Het eerste scenario (centrale uitvoering) heeft als voordeel dat de belangen van auteursrechthebbenden beter beschermd kunnen worden via waarborgen en passende technische en organisatorische maatregelen door een centrale instelling met relatief veel middelen. Daarnaast kan dit scenario kostenbesparingen met zich meebrengen als er elf of meer instellingen zijn die op reguliere basis een verzoek voor webharvesting doen. Het scenario wijkt echter wel af van de huidige praktijk waarin ook decentrale instellingen al webharvestingactiviteiten uitvoeren.

Het tweede uitgewerkte scenario (centrale en decentrale uitvoering) sluit het meest aan bij de huidige praktijk waarbij erfgoedinstellingen zelfstandig harvesten. Wel kunnen er meer kosten mee zijn gemoeid als er een groter aantal instellingen dan in de huidige praktijk gaan webharvesten (meer dan elf) vergeleken met het eerste scenario. Ook is er het risico dat kleinere instellingen minder effectief waarborgen en passende maatregelen kunnen nemen om auteursrechtelijk beschermde webarchieven en persoonsgegevens te beschermen, hoewel dat in de huidige praktijk nog geen probleem is.

Wanneer er gekozen wordt voor een route waarbij webharvesting niet wettelijk gefaciliteerd wordt, dan is er de keuze om afspraken te maken met belanghebbenden of de huidige praktijk (status quo) in stand te houden. In het geval van de eerste optie zijn de kosten iets hoger dan de twee scenario's waarin er wel sprake is van wettelijk faciliteren, en zijn er negatieve effecten op de juridische houdbaarheid en het draagvlak.

Na het wegen van de verschillende opties en het maken van de in dit onderzoek beoordeelde keuzes zijn er nog verschillende vervolgstappen te zetten. Deze vervolgstappen betreffen zowel het maken van nadere (detail)keuzes als het verder onderzoeken en uitwerken van juridische en uitvoering-gerelateerde aspecten. Welke vervolgstappen nodig zijn en welke inspanning daarmee gemoeid is, is afhankelijk van de gemaakte keuzes. Het gaat onder andere om een wetgevings- of juridische traject

waarbij een groot aantal belanghebbenden bij betrokken is. In het vervolgtraject zullen nog meer keuzes moeten worden gemaakt ten aanzien van wet- en regelgeving, uitvoering en waarborgen.

Leeswijzer

- In Hoofdstuk 1 geven we een algemene inleiding op het vraagstuk en het onderzoek.
- In Hoofdstuk 2 staat de opzet van onderzoek toegelicht, inclusief toelichting en definities van de te maken keuzes, scenario's en door ons beoordeelde categorieën van impact. Dit hoofdstuk is relevant voor lezers die meer willen weten over onze onderzoeksopzet.
- In Hoofdstuk 3 staan de belangrijkste uitkomsten van ons onderzoek, welke relevant zijn voor alle lezers.
- In Hoofdstuk 4 tot en met 8 staan detailanalyses op juridisch vlak, uitvoerbaarheid, kosten, draagvlak en ten aanzien van de afbakening van het nationaal webdomein. Deze hoofdstukken zijn relevant voor lezer die verdieping willen op de uitkomsten die in Hoofdstuk 3 worden gepresenteerd.

Inhoudsopgave

Managementsamenvatting	2
1. Inleiding	13
1.1 Aanleiding en achtergrond	13
1.2 Doel- en vraagstelling van het onderzoek	16
1.3 Aanpak	17
1.4 Leeswijzer	17
2. Onderzoekskader en indicatoren	19
2.1 Opzet van het onderzoekskader	19
2.2 Omschrijving van webharvesting	21
2.3 Juridische routes	22
2.4 Scenario's	24
2.5 Nationaal webdomein	25
2.6 Impactcategorieën en indicatoren	28
3. Conclusies en scenario's voor uitvoering van webharvesting	33
3.1 Scenario 1 en 2: wettelijk faciliteren	33
3.2 Scenario 3 en 4: niet wettelijk faciliteren	44
3.3 Nationaal webdomein	51
3.4 Conclusie	53
4. Juridische analyse	58
4.1 Intellectuele eigendomsrechten	58
4.2 Privacy, gegevensbescherming en fundamentele waardenspanningen	61
4.3 Mogelijkheden juridische invulling oplossingsrichtingen	66
5. Uitvoerbaarheid	70
5.1 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen	70
5.2 Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties	71
5.3 Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)	72
5.4 Status quo (geen afspraken maken)	73
6. Kosten	74
6.1 Uitgangspunten kostenraming	74
6.2 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen	75

6.3	Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties.....	81
6.4	Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)	84
6.5	Status quo (geen afspraken maken)	87
6.6	Vergelijking kosten verschillende scenario's	88
6.7	Gevoeligheidsanalyses.....	90
6.8	Kostenontwikkeling van opslag van de nationale domeincrawl	95
7.	Draagvlak	97
7.1	Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen	97
7.2	Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties.....	98
7.3	Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)	98
7.4	Status quo (geen afspraken maken)	98
8.	Het nationaal webdomein	99
8.1	Beredeneerde selecties.....	99
8.2	Nationale domeincrawl	100
8.3	Europese voorbeelden	101
8.4	Strategieën voor operationaliseren	103
8.5	Aanvullende opmerkingen over het nationale webdomein.....	106
Bijlage A	Overzicht gesprekken	107
Bijlage B	Deelnemers expertgroepen	108
Bijlage C	Roadmaps voor uitvoering van webharvesting.....	109
	Roadmap scenario 1 en 2.....	110
	Roadmap scenario 3.....	112

1. Inleiding

1.1 Aanleiding en achtergrond

Het Nederlandse erfgoed bestaat niet meer enkel uit kranten, tijdschriften en andere fysieke objecten. Ook webcontent kan tegenwoordig worden meegenomen bij het vormen van een beeld van de Nederlandse samenleving op een bepaald moment en in onderzoek. Het web is echter constant in beweging en online content wordt regelmatig aangepast, ontoegankelijk gemaakt of kan verwijderd worden. Daardoor bestaat de mogelijkheid dat materiaal verloren gaat. Dit is de reden dat Nederlandse erfgoedstellingen wijzen op het belang van webarchivering: het opslaan van (delen van) het nationale internetdomein voor toekomstig onderzoek. Om dit te kunnen doen is webharvesting nodig. Webharvesting is een middel om webarchieven te vormen en digitaal erfgoed voor toekomstige generaties veilig te stellen, door deze met behulp van technologie te verzamelen, te conserveren en daarna toegankelijk te maken (te presenteren).

De Koninklijke Bibliotheek (KB), het Nationaal Archief (NA), de Koninklijke Vereniging van Archivarissen in Nederland (KVAN) en het Nederlands Instituut voor Beeld en Geluid (Beeld en Geluid; BenG) pleiten voor een wettelijke facilitering van webharvesting door erfgoedinstellingen. Het gaat hierbij om het crawlen, bewaren en toegankelijk maken (presenteren) van het Nederlandse webdomein, zowel in zijn geheel als meer selectief en diepgaander. Momenteel worden webarchieven verzameld als een rechthebbende geen gebruik heeft gemaakt van een 'opt-out'. Dit betekent dat aan rechthebbenden en webmasters van te harvesten webdomeinen gevraagd wordt of zij *niet* willen dat hun website wordt geharvest. De KB, KVAN, NA en Beeld en Geluid, partijen die nu al webharvestingactiviteiten uitvoeren (of dergelijke partijen vertegenwoordigen) vinden dit niet praktisch uitvoerbaar. Bovendien stellen ze dat ze met deze werkwijze opereren binnen een juridisch grijs gebied.

Met deze achtergrond heeft de toenmalig Minister van Rechtsbescherming het Instituut voor Informatierecht (IViR) verzocht om een adviesrapport ('IViR rapport Webharvesting'), dat in 2021 is verschenen. In het IViR rapport wordt benoemd dat webharvesting bijdraagt aan het waarborgen van de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap⁴. Het IViR concludeert dat beleidsmakers een oplossing dienen te vinden die ruimte biedt voor webharvesting, zonder de conflicterende rechten en belangen te zeer in het gedrang te brengen. Het gaat daarbij om rechten en belangen zoals het auteursrecht, de naburige rechten en het sui generis databankenrecht. Hier doet het IViR rapport, evenals een artikel dat uit hetzelfde onderzoek is voortgekomen, ook concrete suggesties voor.

In de reactie op het beleidsplan van de KB 2023-2026 heeft het Ministerie van Onderwijs, Cultuur en Wetenschap (OCW) toegezegd om de haalbaarheid en wenselijkheid van het wettelijk mogelijk maken van webharvesting te zullen onderzoeken. In 2023 hebben de KB, KVAN en Beeld en Geluid een brief⁵ verzonden naar de vaste Kamercommissie van OCW, met daarin het verzoek om webharvesting wettelijk te faciliteren. Ook is op 24 oktober 2023 in de Tweede Kamer de motie Dekker-Abdulaziz aangenomen, waarin de regering wordt verzocht om in kaart te brengen wat de wettelijke, technische en financiële obstakels zijn die in de weg staan van een mogelijke archivering van het nationale webdomein en hoe deze te overkomen, en om de Kamer hierover te informeren. De meeste van deze

⁴ Webharvesting, IViR, (2021), p. 55.

⁵ [Dringende noodzaak om behoud van Nederlands online erfgoed wettelijk mogelijk te maken](#)

obstakels zijn onder andere door het IViR al in kaart gebracht. In het huidige onderzoek wordt er gekeken naar de mogelijke vervolgstappen en scenario's voor het (wettelijk) faciliteren van webharvesting.

1.1.1 Tijdlijn

2019 DSM-richtlijn

Bij de implementatie van de Richtlijn auteursrechten en naburige rechten in de digitale eengemaakte markt ('DSM-richtlijn') in 2019 is de mogelijkheid tot webharvesting niet opgenomen, omdat het volgens de wetgever een afzonderlijk onderwerp betreft dat binnen zijn eigen specifieke context dient te worden behandeld.

2019 Gezamenlijke reactie van Nederlandse erfgoedinstellingen op het DSM-implementatiewetsvoorstel

Namens een groot aantal Nederlandse erfgoedinstellingen is er een reactie geschreven op het DSM-implementatiewetsvoorstel. Hierin uiten zij hun wens om online erfgoed te preservareren door middel van webharvesting. De erfgoedinstellingen stellen voor om een exceptie in de DSM-richtlijn toe te voegen waarmee het mogelijk wordt om werken die op het Nederlandse openbaar toegankelijke web staan te kopiëren en als digitaal erfgoed te bewaren. Deze exceptie valt binnen de kaders van het Auteursrecht.⁶

2021 IViR rapport

In 2021 verscheen het eindrapport van het onderzoek 'Webharvesting', uitgevoerd door het Instituut voor Informatierecht (IViR) van de Universiteit van Amsterdam.⁷

24 mei 2022 urgentiebrief aan ministerie van OCW

De KB, de KVAN, het NA en Beeld en Geluid stuurden een brief aan OCW waarin ze pleitten voor een wettelijke kader om webharvesting mogelijk te maken, naar aanleiding van de resultaten van het IViR onderzoek.⁸

Juli 2023 OCW doet toezegging voor onderzoek

Het ministerie van OCW zegde in zijn reactie op het beleidsplan van de KB 2023-2026 toe om de haalbaarheid en wenselijkheid van het wettelijk mogelijk maken van webharvesting te zullen onderzoeken.⁹ Ook informeerde het ministerie de KB, de KVAN, het NA en Beeld en Geluid dat er een projectteam binnen het ministerie is opgezet.

29 augustus 2023 brief aan vaste Kamercommissie van OCW

De KB, de KVAN, het NA en Beeld en Geluid stuurden gezamenlijk een brief aan de vaste Kamercommissie van OCW waarin zij aandacht vragen voor de dringende noodzaak om webharvesting door Nederlandse erfgoedinstellingen wettelijk mogelijk te maken. Ook riepen zij de Kamercommissie op om het ministerie van OCW aan te sporen om met spoed tot een wetsvoorstel hiervoor te komen.¹⁰ Daarbij refereerden de auteurs aan de urgentiebrief die zij op 24 mei 2022 aan het ministerie van OCW hebben verstuurd.

28 oktober 2023 motie Dekker-Abdulaziz

Tweede Kamerlid Dekker-Abdulaziz diende een motie in waarin de regering verzocht wordt om in

⁶ Gezamenlijke reactie van Nederlandse erfgoedinstellingen op het DSM-implementatiewetsvoorstel, Nederlandse erfgoedinstellingen, (30 augustus 2019).

⁷ Webharvesting, IViR, (2021). Geraadpleegd via: https://www.ivir.nl/publicaties/download/Webharvesting_WODC.pdf

⁸ Urgentiebrief Wettelijke facilitering van webharvesting voor erfgoedinstellingen, KB, KVAN, NA en Beeld en Geluid (24 mei 2022). Geraadpleegd via: <https://www.kvan.nl/nieuws/kvan-blijft-druk-uitoefenen-voor-wettelijke-faciliteiten-webharvesting/>

⁹ Kamerstukken II 2022/23, 36200-VIII nr. 244.

¹⁰ Dringende noodzaak om behoud van Nederlands online erfgoed wettelijk mogelijk te maken, KB, KVAN en Beeld en Geluid (2023). Geraadpleegd via: <https://www.kvan.nl/nieuws/kvan-blijft-druk-uitoefenen-voor-wettelijke-faciliteiten-webharvesting/>

kaart te brengen wat de wettelijke, technische en financiële obstakels zijn die in de weg staan van een mogelijke archivering van het .nl-domein en hoe deze te overkomen. Ook verzocht zij om de Kamer hierover te informeren in 2024. De motie werd unaniem aangenomen.¹¹

1.1.2 Het belang en de behoefte van webharvesting

Erfgoedinstellingen hebben de wens om Nederlands online erfgoed, waaronder websites, duurzaam te kunnen behouden. Zij stellen dat steeds meer informatie enkel nog in digitale vorm op internet wordt gepubliceerd en dat de informatie vluchtig is vanwege de frequente actualisatie.

Erfgoedinstellingen willen online bronnen kunnen bewaren. Zij vinden webharvesting van groot belang om online erfgoed te kunnen behouden voor toekomstige generaties en in de huidige situatie worden slechts enkele tienduizenden webdomeinen door erfgoedinstellingen bewaard.¹² De Koninklijke Bibliotheek (KB), Nederlands Instituut voor Beeld & Geluid, Nationaal Archief en KVAN stellen dat webharvesting essentieel is om online erfgoed te kunnen behouden vanwege een drietal elementen¹³:

Waarom webharvesting belangrijk is voor Nederland¹⁴

1. **Snel verdwijnende internetcultuur.**
Webpagina's zijn relatief kort online en worden frequent gewijzigd of verwijderd. Informatie is daardoor kort en/of beperkt raadpleegbaar.
2. **Geen papieren equivalent**
Veel informatie wordt enkel online op het web gepubliceerd, zoals online kranten, opinieblogs en onderzoek platforms. Maatschappelijke tendensen manifesteren zich in sommige gevallen hoofdzakelijk online. Webdomeinen behoren daarom tot het nationaal erfgoed en dienen te worden bewaard in de collecties om een compleet en inclusief beeld te kunnen bewaren.
3. **Nederland is een digitale samenleving**
In Nederland wordt relatief veel gebruik gemaakt van het internet, het kan daardoor gezien worden als een representatie van de samenleving. Het behoud van dit online erfgoed zorgt in de toekomst voor essentiële informatiebronnen.
4. **Webuitingen zijn cultureel erfgoed met kwalitatieve waarde voor de samenleving.**
Naast de informatiewaarde van webarchieven is er ook een culturele waarde in webuitingen die niet door het IViR rapport wordt benoemd, maar in de brede context van webharvesting wel relevant is.

parafrasering uit: Wettelijke facilitering van webharvesting voor erfgoedinstellingen, p 2-3.

In het wetenschappelijke artikel naar aanleiding van het IViR-onderzoek schetsen Van Gompel, Schumacher en Senftleben (2022, p. 4) dat naast een beleidsmatige wenselijkheid van webharvesting om webarchieven tot stand te brengen en digitaal erfgoed te bewaren een belangenafweging moet worden gemaakt die naast de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap ook rekening houdt met de mogelijke aantasting van andere rechten zoals het

¹¹ Motie van het lid Dekker-Abdulaziz over in kaart brengen wat de wettelijke, technische en financiële obstakels zijn die een mogelijke archivering van het .nl-domein in de weg staan, (18 oktober 2023). Geraadpleegd via: [Informatie- en communicatietechnologie \(ICT\) | Tweede Kamer der Staten-Generaal](#)

¹² Wettelijke facilitering van webharvesting voor erfgoedinstellingen, KB in nauw overleg met Beeld & Geluid, KVAN, Nationaal Archief, 24 mei 2022.

¹³ Wettelijke facilitering van webharvesting voor erfgoedinstellingen, KB in nauw overleg met Beeld & Geluid, KVAN, Nationaal Archief, 24 mei 2022.

¹⁴ Wettelijke facilitering van webharvesting voor erfgoedinstellingen, KB in nauw overleg met Beeld & Geluid, KVAN, Nationaal Archief, 24 mei 2022. p.2 -3.

auteursrecht. Deze mogelijk conflicterende belangen worden hier geschetst vanuit het recht en het juridische perspectief. Dat was ook de insteek van het IViR-onderzoek.

In het huidige (voorliggende) onderzoek wordt gekeken naar een aantal scenario's in een breder perspectief, waarin ook organisatorische, technische en financiële aspecten worden onderzocht.

1.1.3 Huidige praktijk

Verschillende erfgoedinstellingen doen reeds aan webharvesten op een relatief kleine schaal in lijn met hun expertise en werkgebied¹⁵. Er is echter geen exceptie voor deze activiteiten binnen het auteursrecht. Binnen het auteursrecht is het daarom zonder voorafgaande instemming van rechthebbenden niet toegestaan om webdomeinen te kopiëren/reproducen ten behoeve van preserving/archivering. Erfgoedinstellingen maken daarom gebruik van opt-out mechanismen om webmasters en/of rechthebbenden in te laten stemmen met het harvesten van hun materiaal. Deze praktijk wijkt af van de niet-digitale situatie, waarin er geen of minder beperkingen zijn op kopiëren (indien er materiaal 'aan de balie' wordt afgeleverd) voor hetzelfde doel van preserving en archivering.

Het Nationaal Register Webarchieven is een belangrijk initiatief in de huidige praktijk. Het Netwerk Digitaal Erfgoed is initiatiefnemer van dit Nationaal Register Webarchieven, waarin wordt bijgehouden welke websites in Nederland zijn gearchiveerd. Het register wordt gebruikt door erfgoedinstellingen om inzicht te krijgen in wat er al gearchiveerd is, zodat dubbel werk wordt voorkomen en samenwerkingsverbanden kunnen worden aangegaan. Er zijn twaalf instellingen¹⁶ die het register gebruiken, waaronder de KB, Beeld en Geluid en een aantal regionale en gemeentelijke archieven. Samen archiveren zij ongeveer 26.000 websites, eenmalig en/of periodiek.

In Nederland vormt onder andere de Auteurswet een belemmering voor het op grote schaal uitvoeren van webharvestingactiviteiten. Kennis en technologie zijn al beschikbaar en in gebruik bij (sommige) erfgoedinstellingen, maar erfgoedinstellingen die websites willen opnemen in hun collecties moeten dus vooraf instemming vragen. In de praktijk is dit zeer tijdrovend en daardoor niet goed uitvoerbaar voor erfgoedinstellingen. Er wordt vanuit het veld al langer gepleit voor het moderniseren van wetgeving om webharvesting mogelijk te maken en rechtszekerheid te bieden. In diverse andere landen gebeurt dit ook al.

Het vertrekpunt van ons onderzoek is dat er een belang is bij webharvesting in brede zin en diepe zin zoals geuit door erfgoedinstellingen in de urgentiebrief en is beschreven in de voorgaande paragraaf. Ook is er in de huidige praktijk al technologie beschikbaar voor webharvesting en wordt deze al gebruikt. Dit vormt ook de basis voor de formulering van de te onderzoeken scenario's.

1.2 Doel- en vraagstelling van het onderzoek

Tegen de hiervoor beschreven achtergrond voeren PBLQ, Ecorys en Confider in opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) een onderzoek uit naar het (wettelijk) faciliteren van webharvesting. Het doel van dit onderzoek is om inzicht te geven in **de mogelijkheden en voorwaarden om te komen tot het (wettelijk) faciliteren van webharvesting in het nationale**

¹⁵ <https://www.registerwebarchieven.nl/>

¹⁶ Ten tijde van dit onderzoek zijn dit de volgende instellingen: KB, Beeld en Geluid, Archiefweb.eu, Archipol (RUG), Groninger Archieven, Haags Gemeentearchief, Provincie Overijssel, Regionaal Archief Dordrecht, Regionaal Archief Nijmegen, Stadsarchief Rotterdam, Streekarchief Midden-Holland en Zeeuws Archief.

domein, door een vergelijking van een aantal scenario's op juridische, organisatorische, technische en financiële aspecten, alsmede het draagvlak van deze scenario's bij belanghebbenden.

Bij deze doelstelling hebben we de volgende onderzoeksvraag geformuleerd:

Welke scenario's zijn er om webharvesting (wettelijk) te faciliteren en wat zijn de verschillen tussen deze scenario's op juridisch, organisatorisch, technisch en financieel vlak?

De te onderzoeken scenario's zijn toegelicht in paragraaf 2.4 en de te onderzoeken effecten van deze scenario's en de te maken keuzes zijn toegelicht in paragraaf 2.6.

1.3 Aanpak

Tijdens dit onderzoek hebben we gesprekken gevoerd met de belangrijkste stakeholders, zoals de KB, Beeld en Geluid, KVAN en het NA. Daarnaast hebben we enkele regionale archieven gesproken over hun ervaringen met webharvesting. Ook hebben we gesproken met vertegenwoordigers van verschillende perspectieven die in dit onderzoek aan bod komen¹⁷:

- De Federatie Auteursrechtbelangen, VOICE en Platform makers over het perspectief van auteursrechtelbenden.
- Privacy First en de Autoriteit Persoonsgegevens over het perspectief van de belangen van betrokkenen en de bescherming van persoonsgegevens.
- Open State Foundation over het perspectief van vrijheid van en toegang tot informatie.

Gedurende het onderzoek hebben er vier expertgroepbijeenkomsten plaatsgevonden¹⁸: één over de technische, organisatorische en financiële aspecten van het vraagstuk en drie over de juridische aspecten, omdat de juridische aspecten de meeste uitwerking behoeften.

Naast de juridische, organisatorische, technische en financiële analyses hebben we ook een gevoeligheidsanalyse uitgevoerd om de effecten van wijzigende aannames op de kosten in kaart te brengen. In Hoofdstuk 2 worden het onderzoekskader en de indicatoren die gebruikt zijn in de analyses nader toegelicht.

1.4 Leeswijzer

De opbouw van dit onderzoeksrapport maakt dat lezers met verschillende behoeften zich kunnen focussen op verschillende hoofdstukken van het rapport. Hieronder staat toegelicht welke onderdelen van het rapport voor welke doelgroepen relevant zijn.

Lezers geïnteresseerd in de uitkomsten van het onderzoek

In Hoofdstuk 3 lichten we onze beoordeling van de effecten van de keuzes voor de verschillende scenario's voor de uitvoering van webharvesting toe, waarbij we per scenario aangeven welke verwachte impact er is op de belangen van verschillende stakeholders, de juridische haalbaarheid en houdbaarheid, de uitvoerbaarheid en de kosten. Ook gaan we in op twee aanpakken om te komen tot een afbakening van het nationaal webdomein.

¹⁷ Bijlage A bevat een overzicht van alle gesprekken die in het kader van dit onderzoek gevoerd zijn.

¹⁸ Bijlage B bevat een overzicht van de deelnemers aan de expertgroepen.

Lezers geïnteresseerd in de opzet van het onderzoek

In Hoofdstuk 2 lichten we ons onderzoekskader toe. We beschrijven en definiëren de verschillende scenario's, de te maken keuzes en hoe we de effecten van deze keuzes geoperationaliseerd hebben.

Lezers geïnteresseerd in de detailanalyses van het onderzoek

In Hoofdstuk 4 tot en met 8 staan de detailanalyses van ons onderzoek en gaan we in op de verschillende categorieën van impact van de scenario's voor de uitvoering van webharvesting.

- Lezers die geïnteresseerd zijn in de juridische analyse kunnen zich richten tot Hoofdstuk 4.
- Lezers die geïnteresseerd zijn in de uitvoerbaarheid van de verschillende scenario's kunnen zich richten tot Hoofdstuk 5.
- Lezers die specifiek geïnteresseerd zijn in de kosten van de verschillende scenario's kunnen zich richten tot Hoofdstuk 6.
- Lezers die meer willen weten over het draagvlak van verschillende partijen ten aanzien van de te maken keuzes en scenario's kunnen zich richten tot Hoofdstuk 7.
- Ten slotte kunnen lezers meer lezen over onze analyse van de afbakening van het nationaal webdomein zich richten tot Hoofdstuk 8.

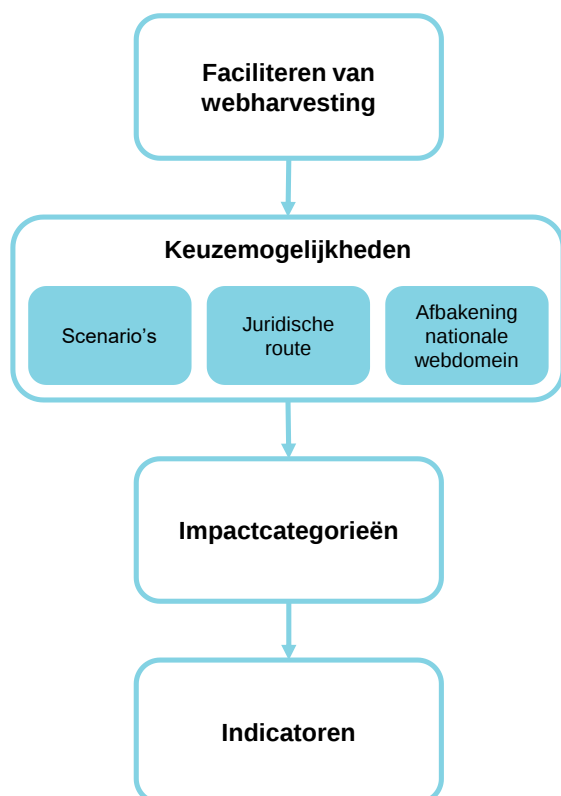
2. Onderzoekskader en indicatoren

Om de onderzoeksvraag te beantwoorden hanteren we een onderzoekskader. Ten aanzien van het faciliteren van webharvesting zijn er verschillende keuzes te maken. Dit zijn de scenario's voor uitvoering, de juridische route en de afbakening van het nationaal webdomein. Deze keuzes kunnen effecten hebben op verschillende belangen. Ook kunnen sommige keuzes meer of minder uitvoerbaar zijn. In dit hoofdstuk lichten wij ons onderzoekskader toe in termen van te maken keuzes en de mogelijke impact van deze keuzes op belangen en uitvoerbaarheid.

2.1 Opzet van het onderzoekskader

Rondom het faciliteren van webharvesting zijn er ten eerste drie onderwerpen waar keuzes over kunnen worden gemaakt (*keuzemogelijkheden bij het vormgeven van oplossingen*): de juridische route, de scenario's (in termen van welke organisaties welke activiteiten uitvoeren) en de afbakening van het nationaal webdomein.

Elk van deze keuzes heeft mogelijke effecten. Deze effecten kunnen op verschillende gebieden spelen, zoals de belangen van rechthebbende, de bescherming van persoonsgegevens en de uitvoerbaarheid. We hebben de effecten daarom ingedeeld op *categorieën van impact*. Deze impactcategorieën worden geoperationaliseerd met één of meerdere *indicatoren* om de impact in kaart te brengen. In figuur 2.1 is dit weergegeven:



Figuur 2.1: Het onderzoekskader

De drie **keuzemogelijkheden** zijn de 'bewegende delen' in ons onderzoekskader: het zijn de elementen gerelateerd aan webharvesting waarover keuzes gemaakt moeten worden. We onderscheiden de volgende keuzemogelijkheden:

- De juridische route om webharvesting (al dan niet wettelijk) te faciliteren;
- De scenario's in termen van *organisatie* (wie harvest) en *activiteiten* (hoe die partijen harvesten);
- De afbakening van het nationaal webdomein (dat wat geharvest wordt).

Deze keuzemogelijkheden hangen sterk met elkaar samen. Zo bepaalt de keuze voor een juridische route onder meer welke scenario's er mogelijk zijn en heeft de definitie van het nationaal webdomein invloed op de juridische route en oplossingsrichting. In paragrafen 2.3, 2.4 en 2.5 gaan we respectievelijk verder in op de verschillende keuzes voor juridische routes, scenario's en het nationaal webdomein.

In het onderzoekskader onderkennen we de volgende **impactcategorieën**, die later in dit hoofdstuk nader worden toegelicht (inclusief de indicatoren die gebruikt worden):

- Het informatie- en erfgoedbelang
- Het belang van (auteurs)rechthebbenden
- Het belang van de bescherming van persoonsgegevens (betrokkenen)
- De kosten
- De juridische risico's (in termen van juridische houdbaarheid en haalbaarheid)
- Het draagvlak bij partijen
- De uitvoerbaarheid (in technische en organisatorische zin)

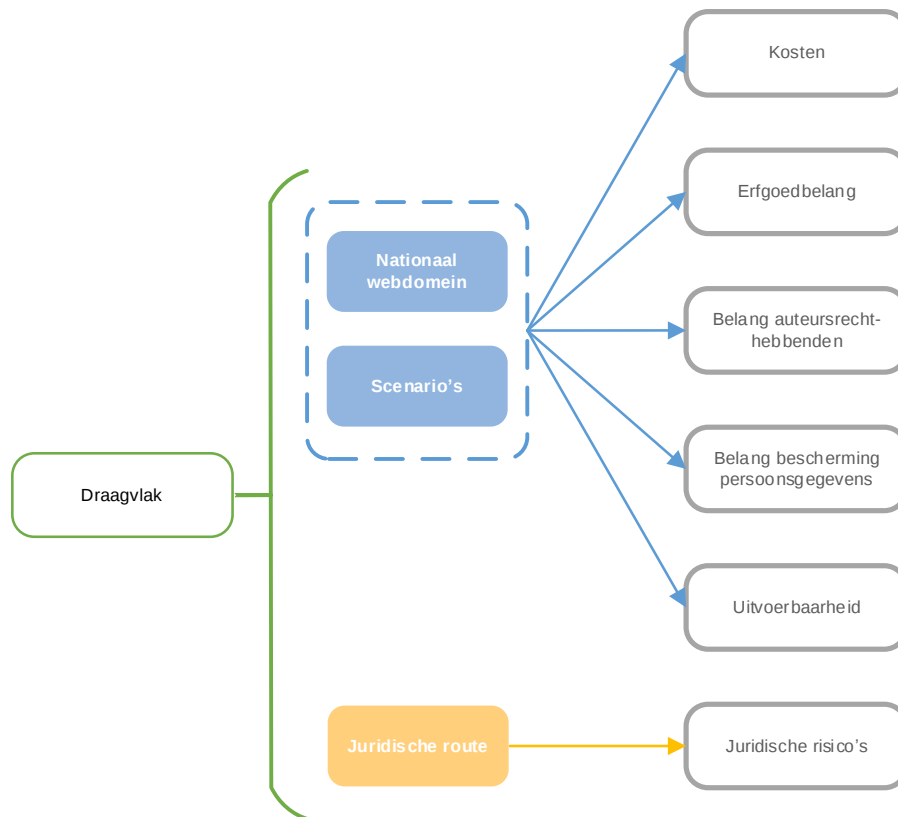
Naast deze impactcategorieën hebben we ook de uitvoering door de Rijksoverheid in kaart gebracht in de vorm van een *roadmap* per scenario. Deze roadmaps geven aan welke stappen er per scenario zouden moeten worden gezet om dit in de praktijk te brengen. De roadmaps zijn opgenomen in Bijlage C.

Niet alle te maken keuzes hebben in ons onderzoekskader effect op alle impactcategorieën. Afhankelijk van de keuzemogelijkheden passen we verschillende impactcategorieën toe. De **juridische route** beoordelen we op basis van de juridische risico's (houdbaarheid en haalbaarheid). In relatie tot de haalbaarheid kijken we of een juridische route binnen bereik ligt, waar draagvlak een belangrijke rol speelt. De **scenario's** beoordelen we op alle impactcategorieën, behalve op de juridische risico's. De afbakening van het nationaal webdomein beoordelen we op alle impactcategorieën behalve de juridische houdbaarheid/haalbaarheid en het draagvlak.

Het draagvlak van keuzes bij partijen heeft een specifieke plaats in ons onderzoekskader die toelichting behoeft. Het draagvlak bij partijen (nader toegelicht in paragraaf 2.6.6) geeft aan welke houding zij hebben ten opzichte van een keuzemogelijkheid. De houding is veelal gebaseerd op een argument die te relateren is aan een andere impactcategorie¹⁹.

¹⁹ Bijvoorbeeld (hypothetisch): Een partij staat negatief tegenover de keuze voor een scenario, omdat het hoge kosten met zich meebrengt en moeilijk uitvoerbaar is.

In figuur 2.2 staat ons onderzoekskader schematisch weergegeven:



Figuur 2.2: Samenhang in het onderzoekskader

In de rest van dit hoofdstuk worden de verschillende onderdelen van het onderzoekskader nader toegelicht, te beginnen bij de keuzemogelijkheden. Daarna wordt ingegaan op de impactcategorieën inclusief de gebruikte indicatoren. We vervolgen dit hoofdstuk eerst met de omschrijving van het begrip webharvesting.

2.2 Omschrijving van webharvesting

In dit onderzoek definiëren wij webharvesting als een proces waarbij verschillende handelingen worden uitgevoerd ten aanzien van websites en webdomeinen. Dit webharvestingproces bestaat uit verschillende fases:

1. In de **harvestingfase** worden websites en webdomeinen gecrawld²⁰ en (deels) gekopieerd.
2. Bij het **acquireren** wordt het materiaal opgenomen in een collectie van een erfgoedinstelling. Onderdeel van deze fase is ook het verkrijgen van toe- of instemming (in de huidige Nederlandse praktijk via een opt-out procedure).
3. In de **preservingfase** wordt materiaal opgeslagen en beheerd. Het beheer van het materiaal heeft als doel het duurzaam toegankelijk te houden.
4. De **presentatiefase** betreft het toegankelijk maken van het materiaal, bijvoorbeeld in de studiezaal van een instelling. Wanneer het in dit onderzoek gaat over presenteren, bedoelen

²⁰ Webcrawlen (kort: crawlen) is het systematisch bezoeken en kopiëren van websites en webpagina's op gespecificeerde domeinen. Dit wordt gedaan door specifieke software: een webcrawler.

we enkel deze vorm van presenteren via een leesterminal (tenzij anders aangegeven). Het op een andere wijze digitaal ontsluiten van materiaal voor een bepaalde doelgroep (bijvoorbeeld wetenschappers) of zelfs de hele samenleving wordt in dit onderzoek niet meegenomen.

We hanteren hiermee een brede definitie van webharvesting die niet alleen de specifieke actie 'harvesten' inhoudt, maar een proces beslaat dat bestaat uit meerdere fasen.

2.2.1 Twee vormen van webharvesting

Wanneer het in dit onderzoek gaat over webharvesting onderscheiden we twee soorten webharvesting: breed harvesten door middel van een nationale domeincrawl en selectief en diepgaand harvesten door middel van een beredeneerde selectie.

Bij de **nationale domeincrawl** wordt er een momentopname van het Nederlandse webdomein vastgelegd. Dit is een oppervlakkige crawl, waarbij slechts één of enkele pagina's van een website worden geharvest.

Bij een **beredeneerde selectie** wordt een vooraf vastgestelde lijst van domeinen geharvest. Dit zijn vaak domeinen die gerelateerd zijn aan een bepaald onderwerp of evenement (zogenaamde 'hotspots'). Bij beredeneerde selectie is het mogelijk om een domein in de diepte te harvesten waarbij alle relevante content wordt verzameld (bijvoorbeeld alle webpagina's op een website). Er kan ook sprake zijn van het harvesten van een specifiek type content, zoals video's en geluidsfragmenten.

De varianten kunnen complementair zijn aan elkaar en bij gecombineerd gebruik het belang van het behoud van Nederlands erfgoed beter dienen dan bij de toepassing van enkel één van de twee varianten²¹. Een nationale domeincrawl kan inzicht geven in de belangrijkste onderwerpen en webactiviteiten in het Nederlandse webdomein en de verschuivingen daarin. Aan de hand van een nationale domeincrawl kunnen blinde vlekken worden ondervangen die niet met selectieve harvesting door middel van beredeneerde selectie worden behouden. Omgekeerd kunnen de beredeneerde selecties een aanvulling zijn op het oppervlakkigere materiaal zoals bij de nationale domeincrawl wordt opgehaald.

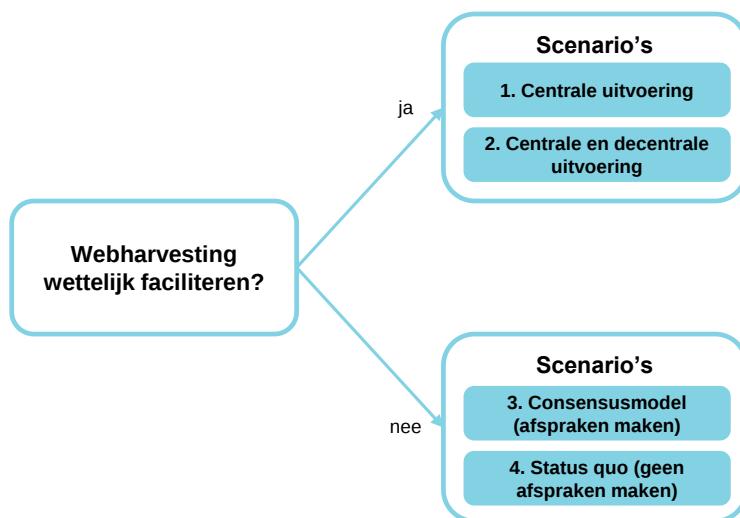
2.3 Juridische routes

Er zijn verschillende juridische routes waarop webharvesting (wettelijk) gefaciliteerd kan worden. Deze routes kunnen zelf weer verschillende (detail)keuzes en uitwerking bevatten, evenals juridische randvoorwaarden.

De eerste keuze die voorligt is of de voorkeur ligt bij het wettelijk faciliteren óf dat afspraken daarover aan de betrokken webharvestingorganisaties en vertegenwoordigers van rechthebbenden moet worden overgelaten. Deze keuze bepaalt mede welke scenario's (zie paragraaf 2.4) er daarna gekozen kunnen worden. Indien wordt gekozen voor het wettelijk faciliteren van webharvesting vallen scenario 3 (consensusmodel) en het nulscenario (scenario 4) af. Indien wordt gekozen voor het niet wettelijk inregelen van webharvesting, vallen scenario 1 en 2 af.²² Dit staat in het onderstaande figuur weergegeven:

²¹ IViR, [Web harvesting: auteursrechtelijke implicaties en oplossingen](#) (2022), p. 3.

²² In alle scenario's moeten er technische en organisatorische maatregelen worden getroffen op om ervoor zorg te dragen dat webharvesting plaatsvindt met in achtname van de geldende privacyregelgeving. Dit wordt in 2.4 verder besproken.



Figuur 3: Juridische routes en scenario's

Wettelijk faciliteren

In scenario 1 moet wettelijk worden vastgelegd dat een centrale instelling een wettelijke basis heeft om activiteiten voor webharvesting uit te voeren. In scenario 2 moet er voor meerdere partijen een wettelijke basis zijn om webharvesting toe te staan.

Het wettelijke faciliteren van webharvesting kan op twee manieren worden ingericht:

1. Door de Auteurswet aan te passen en een beperking op te leggen voor webharvesting (waarmee een uitzondering hiervoor wordt gecreëerd in het Auteursrecht)²³, of
2. Door een publieke taak of bevoegdheid (of taken/bevoegdheden) toe te kennen aan een of meerdere organisaties binnen bestaande wetgeving om geharvest materiaal op te nemen in een collectie en/of een depotverplichting voor makers.

Bij het aanpassen van de Auteurswet gaat het om het opnemen van uitzonderingen voor het verzamelen en bewaren van webdomeinen. Voor het presenteren kan dan op dit moment de al bestaande leesterminaluitzondering worden toegepast.

Het toekennen (of attribueren) van een publieke taak of bevoegdheid kan plaatsvinden via bestaande wettelijke regelingen, zoals de Wet op het hoger onderwijs en wetenschappelijk onderzoek of de Erfgoedwet. Er zijn aanvullend mogelijkheden voor een instelling met een publieke taak of bevoegdheid om deze via lagere regelgeving te delegeren aan andere instellingen, bijvoorbeeld door middel van een ministeriële regeling. De attributie van een publieke taak of bevoegdheid kan hand in hand gaan met een depotverplichting voor rechthebbenden, om hen te verplichten hun werk in een depot op te laten nemen (deponeren op verzoek).

Een depotverplichting houdt in dat makers verplicht zijn hun werk in een depot op te laten nemen, actief (op eigen beweging in een depot plaatsen/overhandigen) danwel passief (op verzoek overhandigen). Een partij die webdomeinen wil archiveren op basis van haar publieke taak of bevoegdheid kan zich beroepen op deze depotverplichting om de activiteiten voor webharvesting uit te voeren. Het betreft geen verplichting voor de archiverende partij. De webinhoud wordt dan onderdeel

²³ Deze route is in de urgentiebrief genoemd als een mogelijke oplossingsrichting. Wij hebben deze route op zichzelf beoordeeld naar de houdbaarheid en haalbaarheid.

van de collectie van de instelling, waardoor de bewaar- en terminaluitzondering uit de Auteurswet van toepassing zijn.

Niet wettelijk faciliteren

Binnen de optie om webharvesting niet wettelijk te faciliteren is er een vervolgkeuze te maken:

- Het maken van afspraken met verschillende belang- en rechthebbenden (consensusmodel).
- Status quo in stand houden.

In het consensusmodel worden afspraken gemaakt tussen marktpartijen over de voorwaarden en verantwoordelijkheden rondom webharvesting. In deze optie komen partijen onderling overeen (via afspraken) hoe materiaal wordt verzameld, bewaard en gepresenteerd, met specifieke aandacht voor toegang, bewaartermijnen en gebruiksrechten. Het uitgangspunt van deze optie is verder dat afspraken worden gemaakt met (vertegenwoordigers van) auteursrechthebbenden, uitvoerders van webharvesting en maatschappelijk betrokken organisaties, om een breed draagvlak te creëren en de belangen van alle betrokkenen te waarborgen.²⁴

2.4 Scenario's

Naast de juridische oplossingsrichting zijn er verschillende manieren waarop webharvestingactiviteiten kunnen worden belegd bij erfgoedinstellingen in organisatorische termen. Dit heeft zoals in de vorige paragraaf beschreven een relatie met de juridische oplossingsrichtingen omdat deze oplossingsrichtingen de mogelijke (organisatorische) scenario's mede bepalen, maar ze beschrijven op zichzelf andere aspecten van het faciliteren van webharvesting. Het gaat hierbij namelijk om *welke* instellingen dit kunnen uitvoeren en *hoe* ze dit uitvoeren. De keuzemogelijkheid 'scenario's' beschrijft de varianten hierin ten aanzien van de uitvoering van webharvesting. De scenario's zijn op te splitsen in twee scenario's waarbij er sprake is van het wettelijk faciliteren van webharvesting en twee scenario's waarbij webharvesting niet *wettelijk* wordt gefaciliteerd, zoals in de vorige paragraaf toegelicht. De scenario's met betrekking tot webharvesting zijn als volgt:

Scenario's

1 Centrale uitvoering door één instelling met loketfunctie voor andere instellingen

In dit scenario is er één centraal gekozen instelling die de nationale domeincrawl en beredeneerde selecties uitvoert. Decentrale instellingen kunnen een verzoek doen bij de centrale instelling om specifieke websites (op basis van beredeneerde selecties) te harvesten ten behoeve van behoud van content voor volgende generaties. Dat betekent in de praktijk dat de centrale instelling een loketfunctie krijgt.

2 Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

Net als in het eerste scenario voert ook in dit scenario één centraal gekozen instelling de nationale domeincrawl uit. Het verschil met het eerste scenario is dat decentrale archief- en erfgoedinstellingen ook beredeneerde selecties (kunnen) uitvoeren. Zij verzamelen daarmee webarchieven en behouden deze, zonder dat zij daarvoor afhankelijk zijn van een centrale instelling zoals in scenario 1.

²⁴ In algemene zin kan er een vergelijking worden gemaakt met de verschillende regelingen voor Out of Commerce werken, waarbij er, weliswaar in dit specifieke geval op basis van een wetswijziging, regelingen met rechthebbenden worden getroffen. Een algemene toelichting inclusief verwijzing naar specifieke regelingen is te vinden op de [website van KVAN](#).

3 Afspraken maken over uitvoering van webharvesting (consensus)

Het consensusmodel wordt door het IViR geschetst als model naar Portugees voorbeeld, waarbij er een inbreukanalyse moet worden gedaan door erfgoedinstellingen en er gepaste maatregelen ontwikkeld moeten worden om risico's te beperken. Wij gaan er – anders dan in het Portugese model – vanuit dat er in dit scenario expliciete en duidelijke afspraken worden gemaakt met recht- en belanghebbenden, aangevuld met een notice-and-takedown-systeem. De reden hiervoor is dat zowel het draagvlak als de haalbaarheid van een systeem zónder afspraken vooraf als zeer laag wordt ingeschat. Dit is bevestigd in de gesprekken die in dit onderzoek zijn gevoerd.

Daarnaast moeten er waarborgen worden ingebouwd om te zorgen dat deze afspraken voldoen aan de geldende privacywetgeving en auteursrechtelijke kaders. Door gezamenlijke standaarden en protocollen vast te stellen, kunnen marktpartijen een uniforme aanpak hanteren die zowel de belangen van rechthebbenden als de publieke toegankelijkheid van informatie respecteert. Ook moet toezicht worden georganiseerd om naleving van deze afspraken te waarborgen, bijvoorbeeld door een onafhankelijke toezichthouder aan te stellen.

4 Status quo (geen afspraken maken)

Dit is de situatie zoals deze momenteel is. Er vindt geen nationale domeincrawl plaats. Wel worden er door een aantal partijen beredeneerde selecties uitgevoerd. Voordat de content geharvest wordt, geven de partijen de website-eigenaren de mogelijkheid tot een opt out. Wanneer die eigenaren kiezen voor een opt out, zal het webdomein niet geharvest worden.

2.5 Nationaal webdomein

Een afbakening of definitie van het nationaal webdomein heeft als doel om aan te geven welke webdomeinen binnen het bereik van webharvestingactiviteiten door Nederlandse erfgoedinstellingen vallen. Hoe breder de afbakening, hoe meer webdomeinen geharvest (zouden kunnen) worden. Het vastleggen van deze definitie (in wet- en regelgeving of op een andere wijze) heeft als doel om aan te geven welke webharvestingactiviteiten (ten opzichte van webdomeinen) er beoogd zijn en mogelijk zijn. Dit heeft een beperkende werking, omdat de afbakening aan erfgoedinstellingen voorschrijft op welke domeinen zij zich kunnen richten. Maar daarnaast heeft de afbakening ook een werking ten aanzien van de rechtszekerheid van harvestende partijen: de afbakening geeft aan waar zij zich op kunnen richten. In deze zin biedt de afbakening ook rechtszekerheid en waarborgen bij het uitvoeren van (publieke) taken.

In dit onderzoek hanteren we een benadering van de afbakening of definitie van het nationaal webdomein gericht op de uitvoering. Daarmee bedoelen we dat we niet kijken naar theoretische of fundamentele aspecten voor een definitie, maar vooral op zoek zijn gegaan naar een perspectief dat aansluit op de uitvoeringspraktijk en bestaande processen. Ons doel is dat deze benadering praktische handvatten geeft voor het vervolg. We maken daarbij ten eerste onderscheid tussen de nationale domeincrawl en beredeneerde selecties als het gaat om de afbakening van het nationaal webdomein.

We beginnen met beredeneerde selecties. In de praktijk is een beredeneerde selectie van een webdomein onderdeel van een bestaand proces van collectievorming (op basis van een verzamelbeleid en collectieplan waarin onderbouwde keuzes worden gemaakt over het opnemen van

een webdomein in een collectie. Er is al een bestaand proces om invulling te geven aan welke webdomeinen relevant zijn voor en welke tot de reikwijdte van de collectie behoren. Met andere woorden, erfgoedinstellingen hanteren in de uitvoeringspraktijk al definities voor hun collectievorming. Een afbakening van het nationaal webdomein zou in onze ogen moeten aansluiten op deze bestaande processen, om zo rechtszekerheid te bieden aan de uitvoerders van webharvesting. We gaan uit van de betrouwbaarheid van deze processen en stellen vast dat dit vraagt om een afbakening die aansluit bij de oplossingen die er in de praktijk al zijn.

Voor de nationale domeincrawl is de situatie anders. Ten eerste gaat het hier om een nog niet bestaand proces: er wordt nog geen nationale domeincrawl uitgevoerd. De nationale domeincrawl is een nieuwe activiteit met een (zeer) brede potentiële reikwijdte. Er is nog geen bestaand (verzamel)beleid voor deze activiteit. Een definitie van het nationaal webdomein heeft daarmee de functie om de (potentiële) reikwijdte van de nationale domeincrawl te bepalen. Tegelijkertijd vraagt de nationale domeincrawl om een precieze afbakening zodat deze zoveel mogelijk geautomatiseerd kan worden uitgevoerd.

Gegeven het bovenstaande richten wij ons ten aanzien van de afbakening van het nationaal webdomein voor beredeneerde selecties op het bieden van rechtszekerheid binnen de bestaande uitvoeringspraktijk en voor de nationale domeincrawl op handvatten voor het bepalen van de reikwijdte van deze nieuwe taak.

In het IViR-onderzoek wordt een brede definitie genoemd voor het nationaal webdomein (elke webactiviteit die verband houdt met de betreffende natiestaat), maar er wordt ook geschetst dat deze verdere afbakening vereist om daadwerkelijk hanteerbaar te zijn voor het archiveren van het nationale webdomein. Daarvoor zijn verschillende selectiecriteria opgesteld door de KB²⁵:

- 1) het top-level domein²⁶: het top-level domein van Nederland is .nl, maar er zijn ook andere relevante top-level domeinen zoals .frl en .amsterdam);
- 2) de geografie van een bepaald domein: is de publicatie (het webdomein) in Nederland gepubliceerd? Deze vraag kan bijvoorbeeld worden beantwoord op basis van het adres van de geregistreerde website, of het IP-adres;
- 3) de taal: Is de publicatie in het Nederlands geschreven?;
- 4) de cultuur: wat vindt de gemeenschap zelf dat tot hun digitale cultuur behoort? Hierbij kan de gemeenschap betrokken worden; een mogelijk voorbeeld is het in kaart brengen van de linkstructuur van de Nederlandse Wikipedia om belangrijke Nederlandse websites in kaart te brengen;
- 5) de netwerkstructuur: Naar welke andere pagina's verwijzen pagina's die, op basis van andere criteria dan deze, tot het nationale domein horen? De gedachte hierachter is dat pagina's waar vaak naar wordt verwezen vanuit het nationale domein, mogelijk zelf ook tot dit domein kunnen horen;
- 6) de geschiedenis: Hoe hebben pagina's die, op basis van andere criteria dan deze tot het nationaal domein behoren, zich ontwikkelt? Heeft er bijvoorbeeld een migratie plaatsgevonden van .com domeinen naar .nl domeinen toen deze beschikbaar werden gesteld aan individuen?

Gekeken naar toekomstige ontwikkelingen is het ook van belang om vast te stellen dat het begrip 'Nederlands webdomein' niet vaststaat en aan verandering onderhevig is. Dit kan op verschillende

²⁵ De criteria zijn overgenomen uit *Webharvesting*, IViR (2021), p.20 en voorzien van verdere toelichting.

²⁶ De Engelse term voor top-level domein is *country code top-level domain*, afgekort als ccTLD.

manieren gebeuren. Zo kunnen er nieuwe domeinen bijkomen (.denhaag, .middelburg of .limburg). Maar ook ontwikkelingen in de wereld kunnen relevant zijn bij het bepalen wat het Nederlandse internetdomein is, waarbij gedacht kan worden aan de gemeenschap van Oekraïense vluchtelingen die in Nederland verblijft. Verder worden er in de huidige praktijk al webdomeinen geharvest buiten nationale top-level domeinen (zoals .com), omdat deze via collectievormende processen een (vastgestelde) relatie hebben met Nederland.

Er kleven ook risico's aan het toepassen van bovenstaande criteria, voornamelijk in de vorm van het harvesten van domeinen die niet noodzakelijkerwijs relevant zijn voor het nationale webdomein²⁷. Voorbeelden hiervan zijn het taal-criterium (Nederlandstalige websites van buiten de landsgrenzen, bijvoorbeeld uit Vlaanderen) of het netwerk-criterium (hyperlinks naar (internationaal) veel gebruikte pagina's die geen relatie met Nederland hebben). Vanwege de schaal van een nationale domeincrawl is een (volledige) kwaliteitscontrole niet haalbaar; bij beredeneerde selecties zijn hier wel mogelijkheden voor omdat het gaat om minder webdomeinen tegelijkertijd én omdat deze preciezer geselecteerd worden.

Het veelvoud aan criteria en de veranderlijkheid van het begrip maken het lastig om een exacte definitie te geven die de tand des tijds doorstaat. Ten behoeve van het faciliteren van webharvesting is het wel van belang om de definitie te hebben die gehanteerd kan worden bij het uitvoeren van webharvesting taken. Dit is belangrijk vanwege een aantal redenen:

- De mate waarin de definitie cultureel relevante webuitingen omvat bepaalt ook de mate waarin met webharvesting activiteiten aan de achterliggende doelen kan worden voldaan ten aanzien van het behoud van (cultureel) erfgoed, en de informatievrijheid en vrijheid van wetenschap.
- De bescherming van of inbreuk op belangen van betrokkenen en auteursrechthebbenden op het vlak van:
 - De uit te voeren taken moeten (juridisch) voldoende *specifiek en concreet* zijn.
 - De *reikwijdte* van het nationaal webdomein bepaalt de (potentiële) inbreuk op het recht van auteursrechthebbenden en de bescherming van persoonsgegevens.
 - De *voorzienbaarheid en transparantie* welke webdomeinen worden geharvest voor betrokkenen (persoonsgegevens) en auteursrechthebbenden (werken waar auteursrecht op rust).
- De kosten voor de nationale domeincrawl, in bijzonder gerelateerd aan de benodigde opslag. Dit punt is tevens gerelateerd aan de hierboven genoemde reikwijdte.

Gegeven de bovenstaande toelichting op de veranderlijkheid enerzijds en het belang van een definitie anderzijds zien we twee mogelijke strategieën die gekozen kunnen worden om te komen tot een in de praktijk te hanteren definitie van het nationaal webdomein. Dit zijn:

- Een strategie op basis van vooraf vastgelegde criteria die zoveel mogelijk automatiseerbaar en objectief zijn. Een voorbeeld hiervan zijn het gebruik van top-level domeinen.
- Een strategie waarbij criteria niet vooraf vastgelegd zijn, maar via een vastgesteld proces worden vastgelegd. Naast automatiseerbare en objectieve criteria gaat het hierbij ook over andere criteria die om meer specifieke of individuele beoordeling vragen, zoals cultuur.

Als laatste zien we een relatie tussen de afbakening van het nationaal webdomein en de gekozen juridische route (paragraaf 2.3) ten aanzien van de wijze van vastlegging van de definitie. De gekozen

²⁷ IViR (*Webharvesting*, 2021, p.34) spreekt over 'bijvangst', maar doet dit in het bijzonder in relatie tot strafbare of laakbare content.

juridische route bepaalt mede de wijze waarop de definitie vastgelegd *kan* worden. Zo sluit het *niet* wettelijk faciliteren van webharvesting uit dat definitie in wet- en regelgeving zal worden vastgelegd.

In paragraaf 3.3 worden deze keuzes en hun effecten toegelicht. In Hoofdstuk 8 staat de gedetailleerde analyse, worden deze strategieën nader toegelicht en wordt er stilgestaan bij voorbeelden uit andere Europese landen met betrekking tot het toepassen van selectiecriteria.

2.6 Impactcategorieën en indicatoren

De keuzes uit de vorige paragrafen hebben effecten op verschillende impactcategorieën. Deze impactcategorieën hebben we geoperationaliseerd met behulp van indicatoren. In de onderstaande paragrafen beschrijven we elke impactcategorie met een korte definitie van de impactcategorie zelf en een definitie van de relevante indicatoren.

Niet alle impactcategorieën worden afgezet tegen alle keuzemogelijkheden, omdat niet alle keuzes effect hebben op alle categorieën.

2.6.1 Informatie- en erfgoedbelang

De impactcategorie informatie- en erfgoedbelang gaat over de mate waarin een keuze bijdraagt aan rechten ten aanzien van de vrijheid van meningsuiting, de informatievrijheid en het recht op het bedrijven van wetenschap (informatiebelang) alsmede het behoud van Nederlands erfgoed en het belang van het doorgeven daarvan aan toekomstige generaties waar dat van waarde is (erfgoedbelang). Het doel van deze categorie is om de effecten van keuzes op het gebied van de juridische route, de scenario's en de afbakening van het nationaal webdomein uit te drukken in termen van informatievrijheid en erfgoed. Een belangrijk uitgangspunt daarbij is dat als materiaal (webarchieven) verloren gaan, er een risico is dat materiaal dat op een gegeven moment in de toekomst belangwekkende informatie kan bevatten of dat potentieel als erfgoed wordt gezien, ook verloren gaat.

Deze impactcategorie beslaat verschillende invalshoeken die zeker niet gelijk zijn aan elkaar, maar wel te maken hebben met elkaar als het gaat om de (achterliggende) doelstellingen van webharvesting.

We onderkennen twee perspectieven die relevant zijn ten aanzien van het informatie- en erfgoedbelang:

- Ten eerste is er het perspectief gericht op (cultuur)historische betekenis van webdomeinen, in termen van waarden, kennis, overtuigingen en tradities van de samenleving, die voor huidige en toekomstige generaties van waarde kunnen zijn omdat ze een rol kunnen spelen in bijvoorbeeld processen van (individuele of collectieve) identiteitsvorming of het vormgeven de toekomst in relatie tot een gedeeld verleden. Dit 'brede' **erfgoedbelang** is gerelateerd aan het verdrag van Faro²⁸.
- In het IViR rapport wordt verder benoemd dat webharvesting wenselijk kan worden geacht gezien vanuit de vrijheid van meningsuiting, de informatievrijheid en het recht op het bedrijven van wetenschap. Dit perspectief richt zich op de rechten die er zijn ten aanzien van **informatie**.

²⁸ [Verdrag van Faro in het Nederlands](#)

De verschillende perspectieven van cultureel erfgoed, het preserveren van webarchieven die in de urgentiebrief, het IViR rapport en de door ons gevoerde gesprekken worden geuit vatten we onder de noemer informatie- en erfgoedbelang.

Ten aanzien van ons onderzoek hebben we een specifieke operationalisatie gemaakt van dit informatie- en erfgoedbelang, gericht op algemene indicatoren die zo objectief mogelijk (kwalitatief) in kaart te brengen zijn. De belangrijkste indicatoren voor de impactcategorie informatie- en erfgoedbelang staan in de onderstaande tabel:

Indicatoren	Toelichting
Dekkingsgraad webharvestingactiviteiten	De mate waarin webharvestingactiviteiten het Nederlandse webdomein betreffen: worden de vanuit een erfgoedperspectief relevante uitingen bewaard voor toekomstig gebruik.
Beschikbaarheid van webarchieven	De mate waarin geharveste webarchieven beschikbaar zijn voor verschillende vormen van gebruik.

2.6.2 Belang van (auteurs)rechthebbenden

De impactcategorie belang van (auteurs)rechthebbenden heeft betrekking op diverse intellectuele eigendomsrechten (o.a. auteursrechten, naburige rechten, databankenrecht). Deze rechten staan mogelijk op gespannen voet met webharvestingactiviteiten, waarbij de mate van inbreuk verschilt per type activiteit. Auteursrechten bieden rechthebbenden het exclusieve recht om hun werken te verveelvoudigen en openbaar te maken, wat impliceert dat webharvesting – een proces dat inherent gebaseerd is op het kopiëren en archiveren van online content – vaak inbreuk maakt op deze rechten zonder voorafgaande toestemming.

Naburige rechten beschermen de rechten van uitvoerende kunstenaars, producenten van fonogrammen en omroeporganisaties. Dit impliceert dat bij webharvesting waarbij audiovisueel materiaal, opnames of uitzendingen worden gekopieerd en gearcheerd, erfgoedinstellingen mogelijk inbreuk maken op deze rechten zonder de vereiste toestemming van de rechthebbenden.

Het databankrecht beschermt de producent van een databank tegen het opvragen en hergebruiken van substantiële delen van de databank zonder toestemming. Daar waar met webharvestingactiviteiten een databank wordt gecrawld, is er een mogelijke inbreuk op het databankrecht.

De belangrijkste indicatoren voor de impactcategorie belang van (auteurs)rechthebbenden staan in de onderstaande tabel:

Indicatoren	Toelichting
Inbreuk op auteursrechten	De mate waarin de webharvestingactiviteiten inbreuk (kunnen) maken op auteursrechten.
Inbreuk op naburige rechten	De mate waarin de webharvestingactiviteiten inbreuk (kunnen) maken op naburige rechten.

Inbreuk op databankenrecht

De mate waarin de webharvestingactiviteiten inbreuk (kunnen) maken op het databankenrecht.

2.6.3 Belang met betrekking tot bescherming van persoonsgegevens (betrokkenen) en andere grondrechten

De impactcategorie belang van betrokkenen heeft betrekking op diverse fundamentele rechten (privacy, gegevensbescherming en grondrechtelijke waardenspanningen). De analyse van privacy en gegevensbescherming in relatie tot webharvesting toont aan dat er goederinstellingen te maken hebben met juridische beperkingen, voornamelijk door de regels van de Algemene Verordening Gegevensbescherming (AVG) en de bredere bescherming van het recht op privacy zoals vastgelegd in artikel 8 van het Europees Verdrag voor de Rechten van de Mens (EVRM).

Webharvesting roept naast het privacyvraagstuk ook verschillende grondrechtelijke vragen op, vooral in de context van de Europese bescherming van fundamentele rechten zoals vastgelegd in het Handvest van de Grondrechten van de Europese Unie (Handvest) en het Europees Verdrag voor de Rechten van de Mens (EVRM). Grondrechten zoals de vrijheid van meningsuiting (artikel 10 EVRM), de informatievrijheid (artikel 11 Handvest) en het recht op eigendom (artikel 1 Protocol nr. 1 EVRM) moeten zorgvuldig worden afgewogen tegen elkaar in het kader van het belang van webharvesting.

De gebruikte indicatoren voor de impactcategorie belang van betrokkenen staan in de onderstaande tabel:

Indicatoren	Toelichting
Inbreuk op privacy en gegevensbescherming	De mate waarin er inbreuk kan worden gemaakt op de regels van de AVG.
Inbreuk op grondrechtelijke waarden	De mate waarin privacy als grondrechtelijke waarde onder druk kan komen te staan.

2.6.4 Juridische risico's

De impactcategorie juridische risico's heeft betrekking op juridische risico's die komen kijken bij de uitvoering van webharvestingactiviteiten, in termen van haalbaarheid en houdbaarheid van de gekozen oplossing voor het faciliteren van webharvesting. De gekozen route moet juridisch haalbaar zijn om uit te voeren en voldoen aan wettelijke eisen en regelgeving. Ook moet de oplossing juridisch houdbaar zijn en een toekomstbestendige, rechtszekere regeling bieden.

De belangrijkste indicatoren voor de impactcategorie juridische risico's staan in de onderstaande tabel:

Indicatoren	Toelichting
Haalbaarheid	De mate waarin de gekozen manier om webharvestingactiviteiten uit te voeren juridisch binnen bereik ligt: kan een wetgevings- of juridisch traject succesvol worden afgerond?
Houdbaarheid	De mate waarin de gekozen manier om webharvestingactiviteiten uit te kunnen voeren juridisch houdbaar is en toekomstbestendig is tegen rechtszaken van bijvoorbeeld betrokkenen of rechthebbenden.

2.6.5 Kosten

De impactcategorie kosten heeft betrekking op alle kosten die komen kijken bij de uitvoering van webharvestingactiviteiten.

De belangrijkste indicatoren voor de impactcategorie kosten staan in de onderstaande tabel:

Indicatoren	Toelichting
Opslagkosten	De hoogte van de kosten voor het opslaan van geharvest materiaal.
Uitvoeringskosten (personeel)	De hoogte van de personeelskosten voor het uitvoeren van het webharvesting(sproces).
Uitvoeringskosten (techniek)	De hoogte van de technische kosten (software en hardware) voor het webharvesten en webharvestingproces.

Er worden, in lijn met het uitgangspunt dat is toegelicht in paragraaf 2.2, geen specifieke kosten voor de presentatiefase meegenomen.

2.6.6 Draagvlak

De impactcategorie draagvlak heeft betrekking op de houding van alle betrokken partijen (archief- en erfgoedinstellingen, rechthebbenden en andere betrokkenen) tegenover het uitvoeren van webharvestingactiviteiten. Daarbij zijn de argumenten en redenen waarom een partij een bepaalde houding heeft ten aanzien van een specifieke uitvoering van webharvesting veelal gebaseerd op andere impactcategorieën.

De indicatoren voor de impactcategorie draagvlak staan in de onderstaande tabel:

Indicatoren	Toelichting
Juridische basis	De mate waarin men het eens is dat er een gedegen juridische basis is voor de uitvoering van webharvestingactiviteiten, die voldoende rechtszekerheid biedt.

Praktische uitvoerbaarheid	De mate waarin men het eens is dat de webharvestingactiviteiten praktisch uitvoerbaar zijn door de instelling(en) die ze uitvoeren.
Organisatie	De mate waarin partijen tevreden zijn over wie er webharvestingactiviteiten mag/kan uitvoeren.

Wij verwachten dat de effecten van keuzes op andere onderwerpen ook van invloed zijn op het draagvlak. Zou zouden de effecten van keuzes op het belang van auteursrechthebbenden iets kunnen betekenen voor het draagvlak dat er voor deze keuze is bij de belanghebbenden.

2.6.7 Uitvoerbaarheid

Met de impactcategorie uitvoerbaarheid worden de effecten van de scenario's en (de afbakening van) het nationaal webdomein uitgedrukt. Het gaat daarbij om de organisatorische en technische effecten die een keuze heeft op de erfgoedinstellingen die webharvesting uitvoeren. Dit kan bijvoorbeeld betrekking hebben op de benodigde (technische) middelen, personeel, administratieve lasten of benodigde processen en procedures voor de uitvoering.

De indicatoren waarmee de uitvoerbaarheid wordt beoordeeld staan in de onderstaande tabel:

Indicatoren	Toelichting
Middelen (technologie)	De mate waarin de benodigde technische middelen voor uitvoerende instellingen beschikbaar zijn.
Middelen (personeel en expertise)	De mate waarin de benodigde (personele) expertise voor uitvoerende instellingen beschikbaar zijn.
Administratieve lasten	De administratieve lasten voor uitvoerende instellingen.

2.6.8 Uitvoering door de (Rijks)overheid

De impactcategorie uitvoering door de (rijks)overheid heeft betrekking op de rol die de (rijks)overheid heeft in het faciliteren van webharvesting. Deze categorie wordt niet per keuzemogelijkheid beoordeeld, maar is als roadmap opgenomen in Bijlage C. Bij het opstellen van de roadmap kijken we naar een aantal relevante acties:

1. Acties met betrekking tot wetgeving
2. Acties met betrekking tot draagvlak en stakeholdermanagement
3. Acties met betrekking tot inrichten van de uitvoering
4. Acties met betrekking tot risico's en onzekerheden

3. Conclusies en scenario's voor uitvoering van webharvesting

In dit hoofdstuk beschrijven we de beoordeling van de te maken keuzes (keuzemogelijkheden) op basis van de effecten die deze keuzes hebben (impactcategorieën en bijbehorende indicatoren). Dat doen we voor alle drie de keuzemogelijkheden: de juridische route, de scenario's en de afbakening van het nationaal webdomein. De juridische route bespreken we gekoppeld aan de scenario's, waarbij een onderscheid wordt gemaakt tussen *wettelijk faciliteren* (scenario 1 en 2) en *niet wettelijk faciliteren* (scenario 3 en 4). Ieder scenario beoordelen we per impactcategorie. Dit doen we ook voor de afbakening van het nationaal webdomein.

Voor scenario 1 en 2 bespreken we allereerst de juridische routes en geven we een toelichting op de juridische houdbaarheid en haalbaarheid per route, omdat de juridische route voor deze twee scenario's in grote lijnen gelijk is. Voor scenario 3 en 4 bespreken we de juridische route apart per scenario. In paragraaf 3.1 wordt het wettelijk faciliteren van webharvesting besproken, waarna in paragraaf 3.1.1 en paragraaf 3.1.2 respectievelijk scenario 1 (centrale uitvoering) en scenario 2 (centrale én decentrale uitvoering) worden toegelicht. In paragraaf 3.2.1 en paragraaf 3.2.2 worden respectievelijk scenario 3 (afspraken maken in een consensusmodel) en scenario 4 (status quo, nulscenario) toegelicht.

Naast de scenario's gaan we in paragraaf 3.3 in op de afbakening van het nationaal webdomein. We sluiten dit hoofdstuk af met een conclusie (paragraaf 3.4).

3.1 Scenario 1 en 2: wettelijk faciliteren

De keuze voor het wel of niet wettelijk faciliteren van webharvesting heeft effect op verschillende impactcategorieën, waaronder het informatie- en erfgoedbelang. Op basis van de gesprekken met de potentiële uitvoerders van webharvesting wordt verwacht dat wanneer er geen wettelijke verduidelijking komt van de rechtmatigheid van webharvesting dit een negatief effect heeft op het informatie- en erfgoedbelang: er wordt minder geharvest omdat er meer risico's zijn. Ook zal er geen nationale domeincrawl worden uitgevoerd, waardoor een groot deel van het Nederlandse web niet gepreserveerd wordt.²⁹

De onzekerheid over de juridische status van webharvesting en de potentiële aansprakelijkheid bij eventuele onrechtmatige handelingen zorgen ervoor dat erfgoedinstellingen aangeven nu zeer terughoudend te zijn om dergelijke activiteiten uit te voeren. Zonder een vorm van wettelijke regeling vormt iedere handeling in het webharvestingproces, wanneer voorafgaande toestemming van alle rechthebbenden ontbreekt, een inbreuk op de intellectuele eigendomsrechten en is er een risico op het handelen in strijd met de goede bescherming van persoonsgegevens.

Voor iedere verveelvoudiging en openbaarmaking van een werk geldt dat voorafgaande toestemming van de maker of zijn rechtverkrijgenden is vereist. Voor het uitvoeren van de handelingen die nodig zijn bij webharvesting is in beginsel steeds toestemming van de rechthebbenden nodig. Op deze hoofdregel zijn beperkingen mogelijk, maar die moeten dan wel bij wet zijn vastgelegd.

²⁹ Ten aanzien van het informatie- en erfgoedbelang geven we een nadere beoordeling per scenario in paragraaf 3.1.1 en 3.1.2.

Bij de keuze voor het juridisch waarborgen van webharvesting zijn er twee routes met verschillende deeltwerkingen te onderscheiden: (1) het vastleggen van de mogelijkheid in een nadere wettelijke regeling of (2) het opnemen van een uitzondering in het auteursrecht, beiden worden in het hiernavolgende – zij het in omgekeerde volgorde – uitgewerkt. In hoofdstuk 4 staat een verdieping op de juridische onderwerpen van dit onderzoek.

Implementatie van artikel 5, tweede lid InfoSoc-Richtlijn³⁰ in de Auteurswet

Deze bepaling introduceert een uitzondering voor erfgoedinstellingen, die hen toestaat reproducties van auteursrechtelijk beschermd materiaal te maken onder strikte voorwaarden. Deze uitzondering maakt het mogelijk om een kopie te maken van een werk. De uitzondering biedt (volgens het IViR rapport) geen grondslag voor verdere openbaarmaking (anders dan de bestaande leesterminal exceptie) van het geharveste materiaal aan het publiek. De hiervoor in de juridische expertsessie aangedragen argumenten zijn, kort samengevat:

- De beperking geldt niet voor alle auteurs- en nabuurrechtelijk beschermde prestaties. Software en databanken zijn uitgesloten.
- De beperking van rechten van rechthebbenden maakt het alleen mogelijk om een kopie te maken, er is geen grondslag om het geharveste materiaal openbaar te maken.
- De beperking is geclausuleerd tot bepaalde bijzondere gevallen. Een algemene vrijstelling van webharvesting door erfgoedinstellingen voldoet daar niet aan. Er bestaat onder experts gerede twijfel of de toepassing van deze beperking de drie-stappen-toets (artikel 5, vijfde lid InfoSoc-richtlijn) wel zal doorstaan: er mag geen afbreuk worden gedaan aan de normale exploitatie van het materiaal, de belangen van de rechthebbende mogen niet onredelijk worden geschaad, behoudens specifieke gevallen zoals in de InfoSoc-richtlijn benoemd. Beperkingen en uitzonderingen op auteursrechten mogen niet te breed worden geformuleerd, conform de drie-stappen-toets.
- In andere lidstaten van de EU wordt dit artikel niet gebruikt als grondslag voor webharvesting: er is geen precedent bekend onder de experts.
- De beoogde exceptie is bij het ontwerpen van de Europese regelgeving volgens bij die ontwikkeling betrokken experts niet bedoeld voor het doeleinde van webharvesting.

Houdbaarheid

Samengevat: Onder experts bestaan grote twijfels bij de juridische houdbaarheid van de route via het auteursrecht, omdat het volgens hen niet de gewenste rechtszekerheid zal bieden die wordt gezocht. Daarnaast is de voorgestelde uitzondering in het auteursrecht niet bedoeld voor de beoogde toepassing ten behoeve van webharvesting en wordt het zo ook niet door andere lidstaten toegepast. Met andere woorden, de beoogde route via het auteursrecht past niet goed binnen de bestaande (Europese) regelgeving. Dit maakt het in beperkte mate toekomstvast, omdat het kwetsbaar is voor juridische procedures en negatieve beoordeling van (juridische) adviesorganen. Wanneer deze route wordt gekozen is er onder experts gerede twijfel dat deze juridisch stand zal houden bij een rechter.

Haalbaarheid

Op ambtelijk niveau bij het eerstverantwoordelijke ministerie voor auteursrecht (Justitie en Veiligheid; JenV) is vanwege de hiervoor genoemde argumenten met betrekking tot rechtszekerheid en houdbaarheid een inschatting gemaakt dat er beperkt draagvlak is voor het uitvoeren van deze route. In de eerder genoemde urgentiebrief is deze route ook naar voren gebracht. Op basis van ons

³⁰ Richtlijn 2001/29/EG van het Europees Parlement en de Raad van 22 mei 2001 betreffende de harmonisatie van bepaalde aspecten van het auteursrecht en de naburige rechten in de informatiemaatschappij.

onderzoek stellen wij vast dat zonder juridische houdbaarheid het draagvlak onder alle betrokken partijen beperkt is, wat een negatief effect heeft op de haalbaarheid: zonder draagvlak ligt deze route niet of beperkt binnen bereik.

Wettelijke taak of bevoegdheid

De andere route om webharvesting juridisch te waarborgen is het vastleggen van een expliciete wettelijke taak of bevoegdheid voor erfgoedinstellingen om deze activiteiten uit te voeren. Door een wettelijke basis te creëren die specifiek webharvesting benoemt als een taak voor erfgoedinstellingen, wordt de rechtmatigheid van het verzamelen en archiveren van webinhoud formeel vastgelegd. Dit zou erfgoedinstellingen juridische zekerheid geven en hen beschermen tegen mogelijke aansprakelijkheid bij het kopiëren van auteursrechtelijk beschermd materiaal zonder individuele toestemming van de rechthebbende. Via een digitale depotplicht kan de juridische zekerheid verder worden vergroot, omdat op deze manier webinhoud onderdeel wordt van de collectie van de instelling. Ingevolge de DSM-richtlijn is het noodzakelijk dat een lidstaat voorziet in een uitzondering zodat instellingen in hun collecties content kunnen reproduceren en permanent onderdeel kunnen maken van hun collecties. Dit onderdeel maken van de permanente collectie kan door overdracht van eigendom of licentieovereenkomst, regelingen inzake permanente uitleen of een wettelijke depotplicht.³¹

Houdbaarheid

De meest houdbare variant is van een wettelijke taak of bevoegdheid waarin er een 'digitale depotverplichting' in een regeling wordt meegenomen. Door webcontent wettelijk te classificeren als onderdeel van de collectie, krijgen erfgoedinstellingen toegang tot specifieke uitzonderingen in de Auteurswet, zoals de preservingsexceptie en de leesterminalexceptie. Daardoor kan het materiaal toegankelijk worden gemaakt zonder dat daarvoor individuele toestemming van rechthebbenden nodig is. In de juridische expertsessie is naar voren gekomen dat deze invulling van de juridische route de meest sterke variant is in de zin van houdbaarheid. Ook het overdragen van eigendom of het sluiten van een licentieovereenkomst is een juridische houdbare variant binnen de genoemde regeling, maar deze zijn minder haalbaar zoals hieronder toegelicht.

Haalbaarheid

De varianten waarbij er sprake is van het overdragen van eigendom of het sluiten van licentieovereenkomst zijn in de praktijk minder goed haalbaar, omdat er sprake is van een zeer groot aantal auteursrechthebbenden. Niet alleen beslaat het Nederlandse internetdomein waarschijnlijk miljoenen webpagina's, er zijn ook meerdere rechthebbenden per webpagina. Dit maakt het overdragen van eigendom of het sluiten van licentieovereenkomsten met deze personen (of met hun vertegenwoordigers) niet of zeer beperkt haalbaar.

Ten aanzien van de haalbaarheid van een digitale depotverplichting vraagt dit nog om nader onderzoek en uitwerking van. De keuze voor welke instellingen een wettelijke bevoegdheid krijgen, met welk bereik voor webharvesting en de nationale domeincrawl, en voor welk type content zijn daarbij van belang. Met betrekking tot de keuze voor het beleggen van een bevoegdheid aan een instelling zijn ook varianten te bedenken van delegatie of mandatering tussen de organisaties onderling. Over de juridische houd- en haalbaarheid van deze varianten moeten tussen de betrokken

³¹ Zie ook overweging 27 en 29 DSM-richtlijn (Richtlijn (EU) 2019/790 van het Europees Parlement en de Raad van 17 april 2019 inzake auteursrechten en naburige rechten in de digitale eengemaakte markt en tot wijziging van Richtlijnen 96/9/EG en 2001/29/EG)

organisaties nadere gesprekken worden gevoerd. Voorop staat: hoe vager de afbakening van de taak en inhoud van de depotplicht, hoe minder sterk de houdbaarheid van deze oplossing.

Vanuit de erfgoedorganisaties zijn er vraagtekens gezet bij de haalbaarheid van het wettelijk vastleggen van de bevoegdheid. De vraag is opgeworpen tot op welk detailniveau deze verplichting kan worden uitgewerkt en of dit nog mogelijkheid biedt voor regionale erfgoedinstellingen om aan webharvesting te doen.

Neveneffect wettelijke taak/depotplicht op gegevensbescherming

De keuze voor het inregelen van een wettelijke bevoegdheid heeft een positief effect op de houdbaarheid in de zin van het goed beschermen van persoonsgegevens. Het scherp afbakenen en het vastleggen van een wettelijke verplichting opent de deur voor het beroep kunnen doen op de grondslag 'wettelijke verplichting' vanuit de AVG. Hoe scherper de taak is vastgelegd bij wet, hoe aannemelijker het is dat een beroep kan worden gedaan wettelijke verplichting als grondslag voor de rechtmatige verwerking vanuit de AVG.³²

3.1.1 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen

In de onderstaande tabel staat per impactcategorie een toelichting van de beoordeling van het scenario op de bijbehorende indicatoren:

Impactcategorie	Indicatoren
Informatie- en erfgoedbelang	- Vanwege het uitvoeren van een nationale domeincrawl is er een aanzienlijk grotere dekkinggraad dan in scenario 4. - Er zullen meer beredeneerde selecties worden uitgevoerd vanwege het (wettelijk) faciliteren hiervan vergeleken met scenario 3 en 4, waardoor de dekkinggraad hoger is dan in deze scenario's. - Vergeleken met scenario 2 zijn er geen grote verschillen in de dekkinggraad, omdat verwacht wordt dat nagenoeg een gelijk aantal erfgoedinstellingen (zelfstandig of via de loketfunctie) aan webharvesting zullen doen.
Belang van auteursrechthebbenden	- Er is een (grotere) inbreuk op de belangen van auteursrechthebbenden ten opzichte van de status quo (scenario 4) omdat er meer auteursrechtelijk beschermd materiaal wordt geharvest. Er is geen verschil met de andere scenario's. - Door het wettelijk faciliteren is er wel een duidelijkere onderbouwing van de inbreuk (grotere mate van rechtszekerheid) vergeleken met scenario 3 en 4.
Belang van betrokkenen (bescherming persoonsgegevens)	Een grotere set aan geharveste webdomeinen betekent een grotere (mogelijke) inbreuk op de bescherming van persoonsgegevens, waarbij dit scenario en scenario 2 een grotere inbreuk betekent vergeleken met scenario 3 en 4.

³² Een soortgelijke redenering over grondslagen lezen we ook terug in de waarschuwing van de AP over het CBAR. Deze redenering is door ons geverifieerd bij de AP.

	➤ Een wettelijke grondslag draagt bij aan de bescherming van persoonsgegevens, omdat een wettelijke taak ten behoeve van archiverings- en webharvestingdoeleinden een belangrijke waarborg is ten aanzien van de bescherming van persoonsgegevens. Bij een sterkere wettelijke taak of verplichting is het aannemelijker dat er een beroep kan worden gedaan op de uitzonderingsgronden voor de verplichtingen en verantwoordelijkheden die uit de AVG volgen. Dit geeft betere waarborgen dan scenario 3 of 4. ➤ De uitvoering door één centrale partij vergroot de mogelijkheden voor transparantie en de voorzienbaarheid (vergeleken met alle andere scenario's).
Kosten	➤ De kosten zijn afhankelijk van de hoeveelheid partijen die webharvestingactiviteiten uitvoeren en de hoeveelheid materiaal dat wordt geharvest, waarbij we aannemen dat er in dit scenario relatief veel instellingen zullen harvesten. ➤ Naast de kosten voor webharvestingactiviteiten worden er ook kosten voor de loketfunctie gemaakt. ➤ Ten opzichte van scenario 2 zijn de kosten in dit scenario hoger dan wanneer een beperkt aantal partijen harvesten (gebruik maken van het loket). Het break-even punt voor de loketfunctie betreft elf of meer partijen³³ (vanaf dat aantal zijn de kosten van dit scenario <i>lager</i>).
Draagvlak	➤ Het draagvlak voor het wettelijk faciliteren van zowel een nationale domeincrawl als beredeneerde selecties is in het algemeen groot, en in het bijzonder bij erfgoedinstellingen (gelijk in scenario 1 en 2). ➤ Het wettelijk faciliteren van webharvesting heeft bij alle belanghebbenden de voorkeur (ten opzichte van scenario 3 en 4), omdat dit zorgt voor grotere voorzienbaarheid en (rechts)zekerheid over webharvestingactiviteiten. ➤ Ten aanzien van de bescherming van persoonsgegevens bij de nationale domeincrawl zijn er zorgen bij sommige maatschappelijke partijen, maar die zijn niet onoverkomelijk met voldoende waarborgen. Andere belanghebbenden zien juist de voordelen van een nationale domeincrawl, bijvoorbeeld zodat werken voor langere tijd gepreserveerd worden. ➤ Het uitvoeren van een nationale domeincrawl door één partij is goed uitvoerbaar en hier is ook draagvlak voor bij de erfgoedinstellingen en andere belanghebbenden. ➤ Het (enkel) centraal uitvoeren van beredeneerde selecties wijkt af van de huidige praktijk en voorziet niet in de behoefte van erfgoedinstellingen en de huidige uitvoerende instellingen zijn hier geen voorstander van.

³³ Wanneer 11 of meer organisaties webharvestingactiviteiten uitvoeren geldt deze stelling niet meer en zijn de kosten juist lager (het break-even punt). Dit wordt in paragraaf 6.6 verder toegelicht.

Uitvoerbaarheid

- Het aanbieden van de loketfunctie brengt een grote uitvoeringslast met zich mee voor de centrale partij in het ontvangen, beoordelen en verwerken van verzoeken voor selecties vergeleken met andere scenario's (en in het bijzonder scenario 2).
- Schaarse technische expertise en middelen kunnen (deels) worden gebundeld bij het loket, wat de uitvoerbaarheid vergroot vergeleken met andere scenario's.
- Decentrale partijen zijn voor het kunnen uitvoeren van beredeneerde selecties afhankelijk van een centrale partij, wat mogelijk voor vertraging door aanvullende procedures zorgt (vergeleken met scenario 2).

De kosten zijn opgedeeld in twee hoofdactiviteiten: *de nationale domeincrawl* en *de beredeneerde selecties*.

Nationale domeincrawl

De technische kosten voor de nationale domeincrawl zijn in alle scenario's gelijk, omdat dezelfde kostenstructuur en dezelfde uitgangspunten worden gehanteerd. Wel zijn in scenario 3 de uitvoeringskosten voor een notice-and-takedown procedure meegenomen als een p.m. post. Er is één partij die de domeincrawl uitvoert en de domeincrawl heeft in alle scenario's dezelfde afbakening. De kosten zijn onderverdeeld in technische personeelskosten en de kosten voor de benodigde opslagcapaciteit.

- Technische personeelskosten: Deze blijven jaarlijks constant na een opstartperiode;
- Opslagkosten: Deze nemen jaarlijks lineair toe. Hierin doen wij in eerste instantie de aanname dat de nationale domeincrawl 150 TB aan opslagruimte nodig heeft per jaar over een periode van 10 jaar.³⁴ In de hoofdanalyse blijven de opslagkosten per TB constant, maar deze aanname wordt getest met een gevoeligheidsanalyse in paragraaf 6.7.2.

In het onderstaande figuur staat de kostenraming over 10 jaar voor de nationale domeincrawl, op basis van bovenstaande aannames. De volledige raming is verder uitgewerkt in Hoofdstuk 6.

³⁴ Onze kosteninschatting beperkt zich tot een periode van 10 jaar, maar het is aannemelijk dat ook daarna er sprake is van een vergelijkbare stijging in opslag.

Figuur 3.1: Raming kosten nationale domeincrawl 150 TB per jaar, centraal scenario

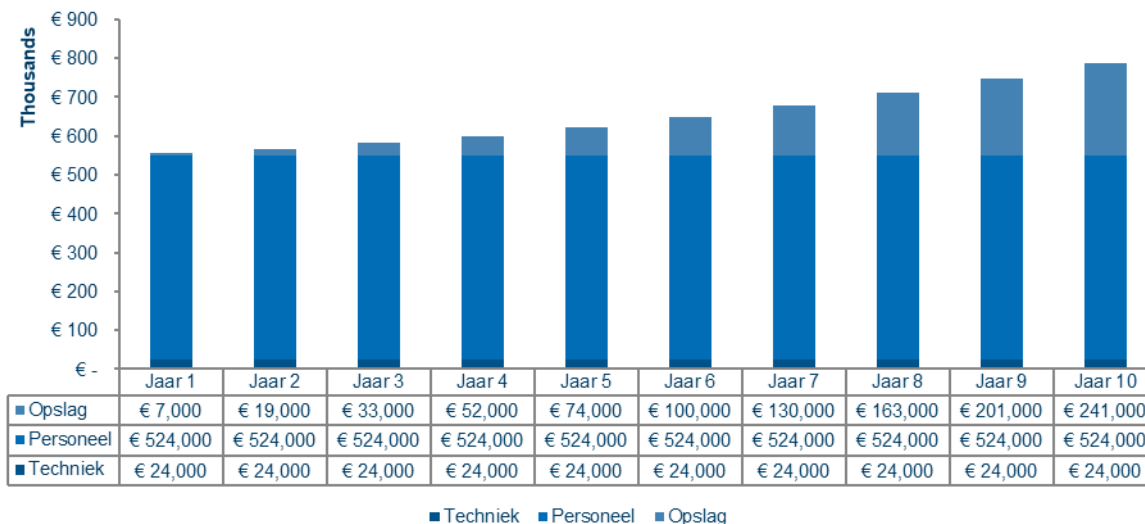


Beredeneerde selecties

In het centrale scenario heeft slechts één instelling (de centrale instelling) personeel en techniek nodig specifiek voor het uitvoeren van de beredeneerde selecties. De andere erfgoedinstellingen zullen nog wel tijd besteden aan het selecteren van websites (collectievorming, hun reguliere werkzaamheden) maar niet aan het technische aspect van het harvesten en opslaan.

Voor de beredeneerde selecties worden jaarlijks websites geharvest door de centrale instelling, waarbij sommige websites eenmalig worden opgeslagen en andere jaarlijks opnieuw worden geharvest. Dit leidt ertoe dat het aantal opgeslagen versies van websites over de jaren *exponentieel toeneemt*. Als gevolg hiervan nemen ook de opslagkosten exponentieel toe. In dit scenario gaan we uit van een jaarlijks aantal van 3.300 nieuw geharveste websites. Deze aannames worden verder toegelicht in hoofdstuk 6. De jaarlijkse kosten van de beredeneerde selecties wordt weergegeven in de onderstaande tabel en figuur.

Figuur 3.2: Jaarlijkse kosten beredeneerde selectie centrale uitvoering



3.1.2 Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

In de onderstaande tabel staat per impactcategorie een toelichting van de beoordeling van het scenario op de bijbehorende indicatoren:

Impactcategorie	Indicatoren
Informatie- en erfgoedbelang	- Vanwege het uitvoeren van een nationale domeincrawl is er een aanzienlijk grotere dekkinggraad dan bij scenario 4. - Er zullen meer beredeneerde selecties worden uitgevoerd vanwege het (wettelijk) faciliteren hiervan vergeleken met scenario 3 en 4, waardoor de dekkinggraad hoger is dan in deze scenario's. - Vergeleken met scenario 1 zijn er geen grote verschillen in de dekkinggraad, omdat verwacht wordt dat een gelijk aantal erfgoedinstellingen (zelfstandig of via de loketfunctie) aan webharvesting zullen doen.
Belang van auteursrechthebbenden	- Er is een (grotere) inbreuk op de belangen van auteursrechthebbenden ten opzichte van de status quo (scenario 4) omdat er meer auteursrechtelijk beschermd materiaal wordt geharvest. Er is geen verschil met de andere scenario's. - Door het wettelijk faciliteren is er wel een duidelijkere onderbouwing van de inbreuk (grotere mate van rechtszekerheid) vergeleken met scenario 3 en 4.

	➤ Vergeleken met scenario 1 is er minder zekerheid voor auteursrechthebbenden dat alle harvestende instellingen adequate waarborgen en beveiligingsmaatregelen nemen ter bescherming van hun werken. Hiervoor zijn wel mitigerende maatregelen (zoals het stellen van eisen over beveiliging en andere waarborgen door de wetgever) mogelijk.
Belang van betrokkenen (bescherming persoonsgegevens)	➤ Een grotere set aan geharveste webdomeinen betekent een grotere inbreuk op de bescherming van persoonsgegevens, waarbij dit scenario en scenario 2 een grotere inbreuk betekent vergeleken met scenario 3 en 4. ➤ Een wettelijke grondslag draagt bij aan de bescherming van persoonsgegevens, omdat een wettelijke taak voor archiverings- en webharvestingdoeleinden een belangrijke waarborg is ten aanzien van de bescherming van persoonsgegevens. Bij een sterkere wettelijke taak of verplichting is het aannemelijker dat er een beroep kan worden gedaan op de uitzonderingsgronden voor de verplichtingen en verantwoordelijkheden die uit de AVG volgen. Dit geeft betere waarborgen dan scenario 3 of 4.
Kosten	➤ De kosten zijn afhankelijk van de hoeveelheid partijen die webharvestingactiviteiten uitvoeren en de hoeveelheid materiaal die geharvest wordt. ➤ Ten opzichte van scenario 1 zijn de kosten in dit scenario lager als een beperkt aantal partijen harvesten. Het break-even punt betreft meer dan 11 partijen³⁵ (vanaf dat aantal zijn de kosten van dit scenario <i>hoger</i>, de kosten voor de loketfunctie in acht genomen).
Draagvlak	➤ Er is wat betreft draagvlak een sterke basis voor dit scenario omdat vanwege de introductie van een wettelijke facilitering van webharvesting voor erfgoedinstellingen. Het draagvlak voor het wettelijk faciliteren van zowel een nationale domeincrawl als beredeneerde selecties is in het algemeen groot, en in het bijzonder bij erfgoedinstellingen (gelijk in scenario 1 en 2). ➤ Het wettelijk faciliteren van webharvesting heeft bij alle belanghebbenden de voorkeur, omdat dit zorgt voor grotere voorzienbaarheid en (rechts)zekerheid over webharvestingactiviteiten. ➤ Ten aanzien van de bescherming van persoonsgegevens bij de nationale domeincrawl zijn er zorgen bij sommige maatschappelijke partijen, maar die zijn niet

³⁵ Wanneer 11 of meer organisaties webharvestingactiviteiten uitvoeren geldt deze stelling niet meer en zijn de kosten juist lager (het break-even punt). Dit wordt in paragraaf 6.6 verder toegelicht.

	onoverkomelijk met voldoende waarborgen. Andere belanghebbenden zien juist de voordelen van een nationale domeincrawl. ➤ Het uitvoeren van een nationale domeincrawl door één partij is goed uitvoerbaar en hier is ook draagvlak voor bij de erfgoedinstellingen en andere belanghebbenden. ➤ Door zowel een nationale domeincrawl als beredeneerde selecties uit te voeren is de dekkinggraad hoog (zie informatie- en erfgoedbelang), wat het draagvlak van dit scenario bij erfgoedinstellingen groter maakt dan in alle andere scenario's. ➤ Auteursrechthebbenden hebben minder voorkeur voor dit scenario ten opzichte van scenario 1 omdat er zorg is dat kleinere erfgoedinstellingen minder goede waarborgen ten aanzien van de bescherming van hun rechten verzorgen.
Uitvoerbaarheid	➤ Erfgoedinstellingen kunnen zelf en met bestaande processen en informatiesystemen webharvesting uitvoeren, wat de uitvoerbaarheid vergroot ten opzichte van scenario 1. ➤ Ook is de uitvoerbaarheid groter ten opzichte van scenario 1 vanwege het ontbreken van een loketfunctie met bijbehorende uitvoeringslast. ➤ Elke erfgoedinstelling moet zelf investeren in expertise en technologische middelen, wat voor kleinere instellingen een drempel kan zijn, waardoor de uitvoerbaarheid lager is vergeleken met scenario 1 (maar gelijk met andere scenario's). Dit geldt in het bijzonder voor instellingen die nu nog geen webharvestingactiviteiten uitvoeren. ➤ Het uitvoeren van een nationale domeincrawl door één partij is goed uitvoerbaar vanwege de bundeling van benodigde kennis en expertise.

De kosten zijn opgedeeld in twee hoofdactiviteiten: de *nationale domeincrawl* en de *beredeneerde selecties*.

Nationale domeincrawl

De raming van de kosten voor de nationale domeincrawl blijft in dit scenario hetzelfde als in het centrale scenario. De opbouw van deze kosten is al eerder besproken en verandert niet afhankelijk van het gekozen uitvoeringsmodel, behalve ten aanzien van de notice-and-takedown procedure in scenario 3.

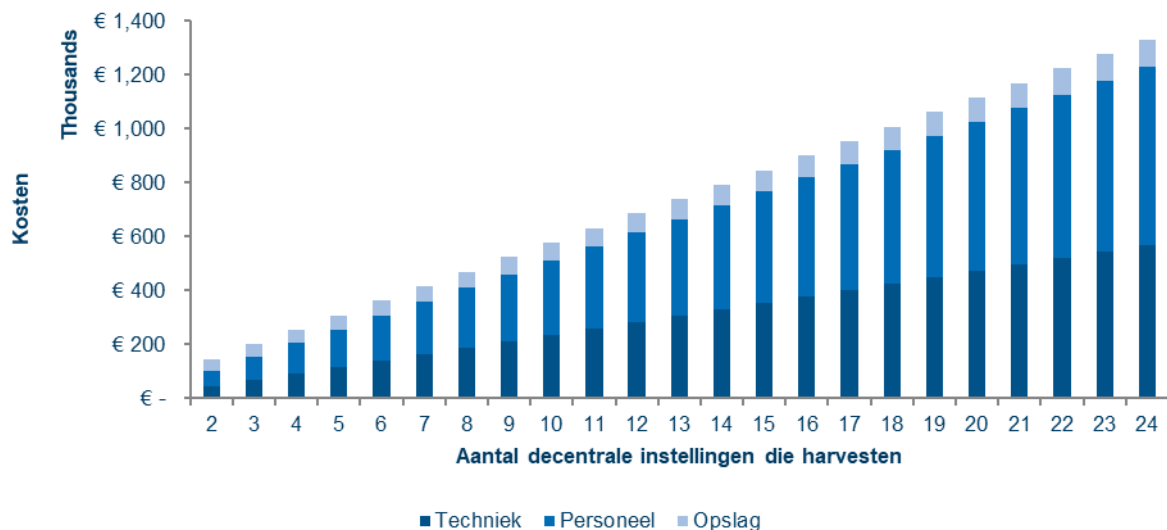
Beredeneerde selecties

In het scenario waarin er geen centrale bevoegdheid is voor het uitvoeren van de beredeneerde selecties, is het minder duidelijk hoeveel decentrale instellingen daadwerkelijk beredeneerde selecties gaan harvesten. Dit komt door de significante financiële drempel die genomen moet worden door de desbetreffende instelling. Een aanname in dit scenario is dat er mogelijk minder erfgoedinstellingen

gaan harvesten, vanwege de financiële en technische drempels. Op dit moment (en ook in het status-quo scenario) zorgt de bestaande financiële drempel er immers ook voor dat niet alle erfgoedinstellingen aan webharvesting doen. In dit scenario *moeten instellingen zelf techniek aanschaffen en expertise opbouwen*. Het gaat hierbij onder andere om het verkrijgen van licenties voor software, de benodigde hardware en het aantrekken van specialistisch personeel.

De behoefte om te harvesten is er naar waarschijnlijkheid wel, maar de kosten zijn dusdanig hoog voor één instelling dat het onzeker is hoeveel instellingen deze keuze zullen maken. Dit is een belangrijke bepalende factor in de kostenberekening van dit scenario. Hierom brengen wij alle scenario's in kaart betreft de hoeveelheid instellingen die besluit beredeneerde selecties uit te gaan voeren. Deze aannames hebben een significante invloed op de uiteindelijke kosten van dit scenario, maar dus ook op de uiteindelijke hoeveelheid materiaal dat geharvest wordt.

Figuur 3.3: Gemiddelde jaarlijkse kosten decentrale instellingen (10 jaar)



In de onderstaande figuur is een situatie uitgewerkt waarin wordt uitgegaan van een situatie waarin 13 erfgoedinstellingen beredeneerde selecties zullen uitvoeren. Net als in het centrale scenario nemen de opslagkosten exponentieel toe over de tijd, aangezien het aantal versies van websites toeneemt. De details hiervan worden verder uitgelegd in Hoofdstuk 6.

Figuur 3.4: Jaarlijkse kosten bereedeneerde selectie decentrale uitvoering, voor 13 instellingen



3.2 Scenario 3 en 4: niet wettelijk faciliteren

De keuze voor het niet wettelijk faciliteren valt samen met een keuze voor scenario 3 of 4. Bij scenario 3 (afspraken maken) en scenario 4 (geen wettelijk regeling) horen verschillende juridische routes, die per scenario zullen bespreken ingaande op haalbaarheid en houdbaarheid.

3.2.1 Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)

In het derde scenario is er geen sprake van een wettelijke juridische borging van de mogelijkheid tot webharvesting, maar worden om ieders belangen te borgen onderlinge (expliciete) afspraken gemaakt tussen erfgoedinstellingen en belanghebbenden, aangevuld met een embargoperiode en notice-and-takedown procedure. Het bereiken van deze consensus kan door middel van het vastleggen van onderlinge afspraken in een convenant. In Portugal, waar het consensusmodel wordt uitgevoerd, is er sprake van een maatschappelijke consensus zonder een wettelijke depotverplichting, gebaseerd op impliciete toestemming en/of stilzwijgend verleende licenties.³⁶ In Portugal wordt geen gebruik gemaakt van een convenant of van expliciet vastgelegde afspraken. Wij achten deze situatie in Nederland niet haalbaar, voornamelijk omdat erfgoedinstellingen aangeven géén nationale domeincrawl uit te willen voeren zonder voldoende rechtszekerheid. Daarom gaan we in dit scenario wel uit van vast te leggen afspraken.

Houdbaarheid

Het scenario van consensus is gebaseerd op een informeel systeem. Het biedt een geringe mate van rechtszekerheid, omdat de grondslag op basis waarvan webharvestingactiviteiten plaatsvinden onduidelijk is en rechthebbenden hun rechten moeten uitoefenen via een notice-and-takedown procedure (achteraf). Een embargoperiode zou directe concurrentie met de exploitatie van meer recente versies van materiaal voorkomen. In het kader van de goede bescherming van persoonsgegevens biedt dit scenario geen versterking van de grondslag waaronder de verwerking

³⁶ Zie voor nadere toelichting ook het IViR-rapport, pagina 131 en 138.

plaatsvindt. Volgens de onderzoekers van het IViR-rapport zou de Nederlandse wetgever de oplossing in dit scenario kunnen ondersteunen door een vrijwaring van aansprakelijkheid te introduceren die vergelijkbaar is met het aansprakelijkheidsprivilege voor hostingdiensten dat voortvloeit uit de regels in de Richtlijn elektronische handel (art. 14 REH en art. 6:196c BW).

Haalbaarheid

Dit scenario is gebaseerd op het maken van onderlinge afspraken tussen rechthebbenden en erfgoedinstellingen, eventueel aangevuld met een wettelijke vrijwaring van aansprakelijkheid. Over de juridische haalbaarheid van deze variant moeten tussen de betrokken organisaties nadere gesprekken worden gevoerd, waarbij de vertegenwoordiging van rechthebbenden verspreid is over verschillende organisaties. Dit kan een complicerend element zijn bij het komen tot gezamenlijke afspraken. Een variant van dit scenario gebaseerd op een maatschappelijke consensus zonder vastlegging van afspraken lijkt op basis van de gesprekken met vertegenwoordigers van auteursrechthebbenden niet haalbaar en binnen bereik. Hierbij speelt mee dat zonder vastlegging van afspraken er onduidelijk is ten aanzien van de scope en het domein van webharvestingactiviteiten (waarbij dit in andere scenario's wél wordt vastgelegd).

In de onderstaande tabel staat per impactcategorie een toelichting van de beoordeling van het scenario op de bijbehorende indicatoren:

Impactcategorie	Indicatoren
Informatie- en erfgoedbelang	➤ Ook op basis van afspraken wordt er zeer waarschijnlijk geen (of wellicht een beperkte) nationale domeincrawl uitgevoerd vanwege het beperkte draagvlak bij uitvoerende partijen, waardoor de dekkinggraad van webharvestingactiviteiten beperkt blijft ten opzichte van scenario's 1 en 2. ➤ Op basis van afspraken worden er waarschijnlijk meer beredeneerde selecties uitgevoerd vergeleken met het nulscenario (4), vanwege de grotere mate van rechtszekerheid, wat het informatie- en erfgoedbelang ten goede komt.
Belang van auteursrechthebbenden	➤ Belangen van auteursrechthebbenden kunnen via afspraken worden gecompenseerd. Dit geeft auteursrechthebbende meer rechtszekerheid vergeleken met scenario 4. ➤ De mate van rechtszekerheid is beperkter ten opzichte van scenario 4 vanwege het invoeren van bescherming via een notice-and-takedown procedure. Een opt-out geeft een mogelijkheid vooraf om rechten uit te oefenen. ➤ Hoewel er sprake kan zijn van compensatie, is de rechtszekerheid lager dan in scenario 1 en 2 waar er sprake is van wettelijk faciliteren en van brede en expliciete vastlegging van publieke taak of bevoegdheid en bijbehorende voorzienbaarheid.

Belang van betrokkenen (bescherming persoonsgegevens)	➤ Het ontbreken van een wettelijke grondslag draagt <u>niet</u> bij aan de bescherming van persoonsgegevens, omdat bij een gebrek aan wettelijke taak of verplichting ten behoeve van archiveringsdoeleinden het minder aannemelijk is dat er een beroep kan worden gedaan op de uitzonderingsgronden voor de verplichtingen en verantwoordelijkheden die uit de AVG volgen. ➤ Vergeleken met scenario 1 en 2 is de verwachting dat er minder webdomeinen worden geharvest in dit scenario, wat een kleinere inbreuk op persoonsgegevens betekend in schaal.
Kosten	➤ Naast de kosten voor de webharvestingactiviteiten worden er ook kosten gemaakt voor het instellen van een notice-and-takedown voorziening. ➤ De kosten zijn afhankelijk van de hoeveelheid partijen die webharvestingactiviteiten uit zullen gaan voeren.
Draagvlak	➤ Er is wat betreft draagvlak geen sterke basis voor dit scenario, met als belangrijkste argument dat er geen sprake is van het <i>wettelijk</i> faciliteren voor erfgoedinstellingen. ➤ Er wordt waarschijnlijk geen nationale domeincrawl uitgevoerd en er vindt slechts in beperkte mate beredeneerde selectie plaats, waardoor de mate van volledigheid zeer beperkt is. Dit zorgt voor weinig draagvlak bij erfgoedinstellingen, zoals onder andere geuit in de urgentiebrief en in dit onderzoek is vastgesteld op basis van gesprekken. ➤ Auteursrechthebbenden hebben meer draagvlak voor dit scenario dan voor de huidige situatie (status quo), maar ze hebben dezelfde houding voor dit scenario als voor scenario 1 en 2. ➤ Maatschappelijke organisaties hebben minder draagvlak voor scenario 3 ten opzichte van scenario 1 of 2 (maar meer dan voor scenario 4), vanwege de voorzienbaarheid en rechtszekerheid.
Uitvoerbaarheid	➤ Het maken en beheren van afspraken vraagt om significante inspanning van de beherende partij, onder andere in het onderhouden van contact met verschillende groepen rechthebbenden. Dit maakt de uitvoerbaarheid ten opzichte van scenario 2 en 4 lager. De uitvoerbaarheid is ook enigszins lager dan in scenario 1 (loketfunctie) vanwege de hoeveelheid afspraken en partijen die beheerd moeten worden. ➤ Wanneer afspraken over materiaal wijzigen kan dit ook invloed hebben op bestaande collecties. Dit vraagt daarom

om administratie over de afspraken in de praktijk, wat de uitvoerbaarheid ten opzichte van andere scenario's relatief laag maakt.

- In de praktijk zal er minder zekerheid zijn bij erfgoedinstellingen bij het uitvoeren van webharvesting en zijn er meer risico's voor de uitvoering vanwege deze onzekerheid. Het webharvesten op basis van beredeneerde selecties is voor kleinere partijen (die het nu nog niet uitvoeren) minder goed uitvoerbaar (de exacte behoefte hiervoor is onbekend).

De kosten zijn opgedeeld in twee hoofdactiviteiten: *de nationale domeincrawl* en *de beredeneerde selecties*.

Nationale domeincrawl

De raming van de kosten voor de nationale domeincrawl blijft in dit scenario hetzelfde als in het centrale scenario. De opbouw van deze kosten is al eerder besproken en verandert niet afhankelijk van het gekozen uitvoeringsmodel. Echter, is de notice-and-takedown voorziening ook voor de nationale domeincrawl relevant. Hiervoor is er dus een aanvullende kostenpost.

Er is nog geen inzicht in hoe een notice-and-takedown voorziening vormgegeven wordt. Het dichtstbijzijnde kengetal dat beschikbaar is komt van de KB, die aangeeft dat op dit moment ongeveer 0,2 fte nodig is voor opt-out aan de voorkant in de huidige situatie. Het verschil tussen een opt-out en een notice-and-takedown voorziening is dat in de eerste situatie er sprake is van een benadering vooraf (mogelijkheid om niet in te stemmen, met inspanning om rechthebbenden te vinden en benaderen), en in de tweede situatie er sprake is van een mogelijkheid om achteraf bezwaar te maken als een partij geen webdomeinen wil laten harvesten. In de praktijk zal een opt-out voorziening dus veel meer tijd en moeite met zich meebrengen omdat er correspondentie met alle relevante partijen nodig is.

De Portugese uitvoerder van de nationale domeincrawl (Arquivo.pt) geeft aan dat zij beperkte notice-and-takedown verzoeken ontvangen. In Portugal wordt het consensusmodel gehanteerd en vindt ook een nationale domeincrawl plaats. Er wordt aangegeven dat het notice-and-takedown voorziening een beperkte belasting vormt, met afgelopen jaar slechts twaalf verzoeken die weinig tijd in beslag hebben genomen.

Desalniettemin nemen wij aan dat de centrale instelling iets moet regelen om deze notice-and-takedown mogelijk te maken. Het is echter niet mogelijk om een betrouwbare raming te geven, omdat het aantal verzoeken op jaarbasis moeilijk te schatten is: de inspanning zal veel minder zijn dan de opt-out procedure, maar onze inschatting is dat het wel degelijk capaciteit zal vragen om verzoeken af te handelen. Omdat het niet mogelijk is een exacte raming te maken, waarderen wij deze post met P.M. +. De post kan met concrete plannen of invulling van de vormgeving van een dergelijk voorziening concreter gemaakt worden.

Raming gemiddelde jaarlijkse kosten nationale domeincrawl (10 jaar), consensusmodel

Domeincrawl	150 TB (raming KB)
Personeel & techniek	€ 117 duizend
Opslag	€ 413 duizend
Notice-and-takedown voorziening	P.M. +
Totaal	€ 530 duizend +

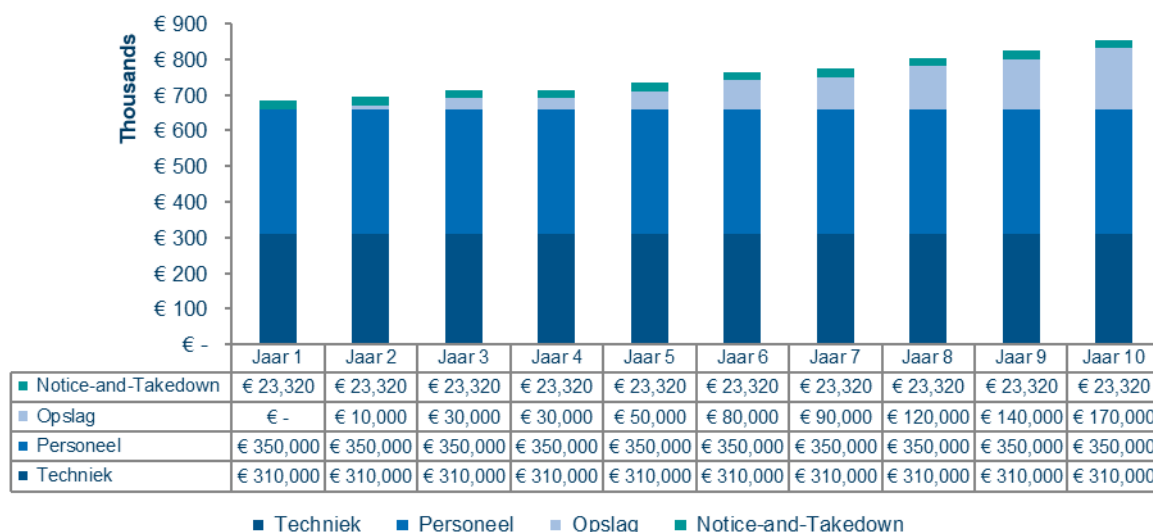
Beredeneerde selecties

In het scenario van het consensusmodel is de uitvoering van de beredeneerde selecties grotendeels hetzelfde als in het scenario met decentrale bevoegdheid. Dat wil zeggen:

- Niet alle instellingen hoeven te gaan harvesten, wat de schaal van de activiteiten beïnvloedt.
- Instellingen moeten zelf technologie aanschaffen door middel van licenties voor webharvestingstechnologie.
- Opslagkosten nemen exponentieel toe, net als in de andere scenario's, vanwege het toenemende aantal opgeslagen versies van websites.

Wat dit scenario onderscheidt van het decentrale model, is de toevoeging van een extra kostenpost: *het inrichten van een notice-and-takedown systeem*. Dit systeem is bedoeld om op een gestructureerde manier om te gaan met verzoeken om inhoud te verwijderen. De verdere uitwerking van deze kostenpost en de impact ervan op het totale kostenplaatje is te vinden in paragraaf 6.3.

Figuur 3.5: Jaarlijkse kosten beredeneerde selectie consensusmodel, voor 13 instellingen



3.2.2 Scenario 4: Status quo (geen afspraken maken)

In dit vierde scenario is er geen sprake van een wettelijke juridische borging van de mogelijkheid tot webharvesting en worden er ook geen onderlinge afspraken gemaakt tussen de betrokkenen. Het bouwt voort op de bestaande opt-out regeling.

Houdbaarheid

Gelijk aan de overwegingen bij het derde scenario biedt deze variant een geringe mate van

rechtszekerheid, nu de grondslag op basis waarvan webharvestingactiviteiten plaatsvinden onduidelijk is, er ook geen onderlinge aanvullende afspraken zijn gemaakt en rechthebbenden hun rechten niet goed kunnen uitoefenen behoudens een opt-out procedure, vanwege het uitblijven van wettelijke waarborgen of andere juridische afspraken. In het kader van de goede bescherming van persoonsgegevens biedt dit scenario eveneens geen versterking van de grondslag waaronder de verwerking plaatsvindt.

Haalbaarheid

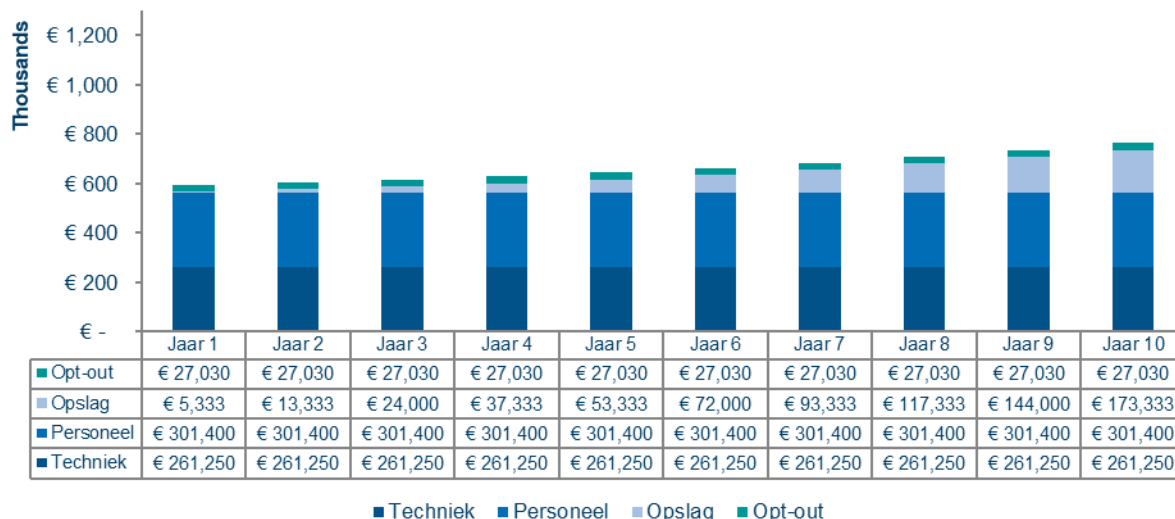
Dit scenario is een voortzetting van de situatie zoals deze nu is. Gelet op de gevraagde urgentie voor het beschermen van het informatie- en erfgoedbelang zoals deze is aangegeven door de erfgoedinstellingen is de haalbaarheid van dit scenario zeer gering. Bij de keuze voor dit scenario blijft rechtszekerheid rondom het uitvoeren van webharvestingactiviteiten voor alle betrokken partijen beperkt. Daarbij hebben erfgoedinstellingen aangegeven dat wanneer de rechtszekerheid beperkt blijft, dit ook effect kan hebben op de schaal en omvang van harvestingactiviteiten die nu worden uitgevoerd.

Impactcategorie	Indicatoren
Informatie- en erfgoedbelang	Er is een beperkte dekkingsgraad met de webharvestingactiviteiten. Er wordt geen nationale domeincrawl uitgevoerd. Het informatie- en erfgoedbelang wordt in dit scenario het laagst beoordeeld.
Belang van auteursrechthebbenden	- Auteursrechthebbenden hebben een opt-out en erfgoedinstellingen spannen zich in om hen hiervan op de hoogte te stellen, maar niet alle rechthebbenden worden bereikt. De voorzienbaarheid van inbreuken vanwege webharvestingactiviteiten is beperkt. - Vergeleken met andere scenario's worden de belangen van auteursrechthebbenden in dit scenario het minst gewaarborgd. - In dit scenario wordt naar verwachting het minste aantal webdomeinen geharvest (minder inbreuk op auteursrechten).
Belang van betrokkenen (bescherming persoonsgegevens)	- Het ontbreken van een wettelijke grondslag draagt <u>niet</u> bij aan de bescherming van persoonsgegevens, omdat bij een gebrek aan wettelijke taak of verplichting ten behoeve van archiveringsdoeleinden het minder aannemelijk is dat er een beroep kan worden gedaan op de uitzonderingsgronden voor de verplichtingen en verantwoordelijkheden die uit de AVG volgen. - Vergeleken met scenario 1 en 2 is de verwachting dat er minder webdomeinen worden geharvest in dit scenario, wat een kleinere inbreuk op persoonsgegevens betekent in schaal.

Kosten	- Naast de kosten voor de webharvestingactiviteiten worden er ook enige kosten gemaakt voor het uitvoeren van de opt-out procedure. - In dit scenario zijn de kosten naar verwachting het laagst, in acht genomen dat er geen nationale domeincrawl wordt uitgevoerd.
Draagvlak	- Er is wat betreft draagvlak geen sterke basis voor dit scenario omdat er geen sprake is van het <i>wettelijk</i> faciliteren voor erfgoedinstellingen. - Er wordt geen nationale domeincrawl uitgevoerd en er vindt slechts in beperkte mate beredeneerde selectie plaats, waardoor de mate van volledigheid zeer beperkt is. Dit is in strijd met het draagvlak bij erfgoedinstellingen, zoals onder andere geuit in de urgentiebrief. - Het draagvlak bij auteursrechthebbenden en maatschappelijke organisaties is lager dan andere scenario's, omdat de rechtszekerheid en voorzienbaarheid van dit scenario lager is dan andere scenario's.
Uitvoerbaarheid	De uitvoerbaarheid blijkt uit de huidige praktijk. Webharvesting volgens de huidige werkwijze is voor veel (kleine) instellingen niet praktisch uitvoerbaar.

In het nulscenario wordt geen nationale domeincrawl uitgevoerd, maar ligt de nadruk op een decentrale aanpak met beredeneerde selecties. Hierbij wordt uitgegaan van een beperkte inzet van elf instellingen (minder dan in andere scenario's) en met lagere volumes dan in de andere scenario's. De personeels- en technische kosten worden berekend op basis van aannames uit de bestaande praktijk. Dit is een modelmatige weergave en dus geen één-op-één weergave van de kosten die nu gemaakt worden door de desbetreffende instellingen. Net als in de andere scenario's nemen de opslagkosten exponentieel toe door de groei van gearriveerd materiaal. Daarnaast ontstaat een aanvullende kostenpost voor de uitvoering van de 'opt-out'-procedure. Deze procedure is arbeidsintensief en vormt een relatief grote kostenpost ten opzichte van andere scenario's.

Figuur 3.6: Kosten beredeneerde selectie nulscenario, over 10 jaar



3.3 Nationaal webdomein

De effecten van een keuze op het gebied van de afbakening het nationaal webdomein worden bepaald door de strategie die gekozen wordt voor het vaststellen van die afbakening. Het kan hierbij gaan om een strategie die gebaseerd is op vooraf vastgestelde en enkel objectiveerbare en automatiseerbare criteria (bijvoorbeeld het top-level domein .nl) of een strategie die op basis van een proces of procedure de te harvesten domeinen vaststelt, waarbij ook objectiveerbare en automatiseerbare criteria gebruikt kunnen worden.

In dit onderzoek maken we de aanname dat bij het vooraf vastleggen van criteria er minder webdomeinen worden vastgelegd dan als er gekozen wordt voor een strategie die uitgaat van een proces om te selecteren webdomeinen vast te leggen. De reden hiervoor is omdat bij het vooraf vastleggen er *enkel* gekeken wordt naar objectief vast te leggen en automatiseerbare criteria en bij een procesgerichte strategie naar *meer* criteria gekeken wordt, die verder gaan dan alleen automatiseerbare criteria. We gaan er daarbij vanuit dat er in beide strategieën in ieder geval gekeken wordt naar het harvesten van het .nl-domein, wat naar verwachting de grootste groep van webdomeinen is. De aanname dat een vooraf vastgelegde, objectiveerbare en automatiseerbare strategie in het algemeen minder webdomeinen omvat is in grote mate van invloed in de beoordeling van de impactcategorieën en indicatoren.

De effecten van de twee opties staan in de onderstaande tabel:

Impactcategorie	Vooraf vastleggen	Proces om vast te leggen
Informatie- en erfgoedbelang	- Enkel webdomeinen die binnen objectiveerbare en automatiseerbare criteria vallen, dus een beperkt aantal criteria is toepasbaar. - De huidige praktijk van beredeneerde selecties bevat	- Een groter deel van nationale webdomein wordt geharvest omdat er meer criteria gebruikt kunnen worden. - Aansluiting op de huidige praktijk: In de huidige praktijk

	webdomeinen die buiten vooraf vast te leggen en objectiveerbare criteria vallen, wat betekent dat deze domeinen buiten de afbakening vallen.	worden geharveste webdomeinen via een (collectievormend) proces bepaald, waarbinnen informatie- en erfgoedbelang al worden afgewogen.
Bescherming persoonsgegevens	✔ ■ Voorzienbaar welke domeinen worden geharvest omdat de criteria om dit te bepalen zijn vooraf zijn vastgelegd.	✔ ■ Grotere inbreuk op de bescherming van persoonsgegevens omdat er naar verwachting meer webdomeinen worden geharvest. Minder voorzienbaar en transparant welke domeinen worden geharvest, hoewel hier wel (mitigerende) maatregelen voor genomen kunnen worden ten aanzien van transparantie.
Belang van auteursrechthebbenden	✔ ■ Beperkte en voorzienbaardere inbreuk op auteursrechtbelangen omdat de criteria voor harvesten vooraf zijn vastgelegd en zijn in te zien.	✔ ■ Groter inbreuk op belangen van auteursrechthebbenden, omdat er zeer waarschijnlijk meer domeinen worden geharvest. ✔ ■ Grotere kans op bijvangst en materiaal dat beschermd is onder andere jurisdicties bij de nationale domeincrawl.
Kosten ³⁷	✔ ■ Lagere opslagkosten vanwege een lager aantal domeinen om te harvesten.	✔ ■ Relatief grotere kosten voor opslag omdat er meer domeinen worden geharvest.
Draagvlak	✔ ■ Er is minder flexibiliteit in het uitvoeren van de nationale domeincrawl.	✔ ■ Het uitvoeren van de nationale domeincrawl heeft meer flexibiliteit, omdat er ruimte is om keuzes te maken over relevante te harvesten webdomeinen.
Uitvoerbaarheid	✔ ■ De uitvoerbaarheid is hoger dan de andere strategie, omdat criteria vooraf worden vastgelegd en	✔ ■ De flexibiliteit zorgt voor meer wijzigingen in criteria. Ook de toepassing van de criteria zorgt voor meer

³⁷ Het verschil in kosten is indicatief en uitgewerkt als een gevoeligheidsanalyse op de kosten voor de nationale domeincrawl.

enkel periodiek zouden moeten worden herijkt.

uitvoeringslast, hoewel dit ten goede kan komen aan andere impactcategorieën.

- De flexibiliteit kan daarentegen ook zorgen voor een betere aansluiting op de huidige uitvoeringspraktijk en bestaande collectievormende processen, wat zorgt voor een betere uitvoerbaarheid voor huidige harvestende instellingen.

Naast een beoordeling van de strategieën is het ook van belang om te kijken naar de mogelijke juridische oplossingsrichtingen, omdat de definitie en afbakening van het nationaal webdomein ook een relatie heeft met de uiteindelijke juridische vastlegging. De belangrijkste keuze daarbij is of er al dan niet een wettelijke facilitering van webharvesting komt. In het geval van een wettelijke facilitering van webharvesting (scenario 1 of 2) is het mogelijk om via juridische bepalingen de reikwijdte van het nationale webdomein vast te leggen, zodat er grotere zekerheid is dat het nationale webdomein wordt gepreserveerd. In onze ogen kunnen beide strategieën hier worden toegepast: naast objectiveerbare criteria kan een proces om te komen tot vaststelling ook worden vastgelegd.

In het algemeen zijn er nog een aantal opmerkingen te maken over het (vastleggen van) het nationale webdomein:

- Het vastleggen van het nationale webdomein in technologie-specifieke termen brengt het risico met zich mee dat wetgeving snel verouderd en gedateerd raakt. Een voorbeeld hiervan zijn het vastleggen van specifieke criteria of top-level domeinen: dit is aan relatief snelle verandering onderhevig is en niet robuust als criterium.
- Door middel van een technologie onafhankelijke vastlegging in wet- en regelgeving kunnen instellingen inspelen op nieuwe ontwikkelingen die op dit moment nog niet voorzien zijn. Er is echter wel een spanningsveld als het gaat om de specificiteit van wet- en regelgeving ten opzichte van andere belangen, zoals de bescherming van persoonsgegevens: hoe specifieker wetgeving is, hoe beter het de belangen van betrokkenen in lijn met de AVG borgt. In een wetgevingstraject kunnen deze belangen tegenover elkaar worden afgewogen.

3.4 Conclusie

Op basis van het bovenstaande bespreken we in deze paragraaf onze conclusies in twee stappen. Ten eerste laten we samenvattend zien welke beoordeling we per scenario geven op de verschillende impactcategorieën. Daarna volgt een conclusie over de te maken keuzes, waarbij we de analyse van de juridische routes in samenhang met de scenario's bespreken.

3.4.1 Uitkomsten van de beoordeling van de scenario's

In deze paragraaf geven we een toelichting op de beoordeling van de keuzes. De tabel verderop in deze paragraaf bevat een samenvatting van de effecten van de impactcategorieën per scenario, zoals

in de paragrafen hierboven beschreven. Hierbij zijn de kwalitatieve beoordelingen per categorie vertaald naar een samenvattende beoordeling van het effect. Per effect geven we aan hoe deze (relatief gezien) scoort ten opzichte van andere scenario's op basis van de volgende coderingen:

- Relatief positief effect
- Effect niet positief/niet negatief
- Relatief negatief effect

Uit de tabel blijkt dat de huidige situatie de meeste negatieve effecten met zich meebrengt. Alleen de kosten zijn dit scenario relatief gezien beter dan in andere scenario's. Het informatie- en erfgoedbelang, het belang van auteursrechthebbenden, de bescherming van persoonsgegevens en het draagvlak vallen in de status quo slechter uit. In scenario 3, het consensusmodel, is de verwachting dat auteursrechthebbenden er, vooral vanwege specifieke afspraken en mogelijke bijbehorende compensatie, er beter vanaf zullen zijn, maar op andere aspecten waaronder draagvlak en uitvoerbaarheid scoort dit scenario minder goed. Scenario 1 en scenario 2 worden op basis van dit onderzoek relatief gelijk beoordeeld, behalve dat scenario 2 beter uitvoerbaar is vanwege het ontbreken van een centrale loketfunctie die verschillende vraagstukken met zich meebrengt in de uitvoering en vanwege het draagvlak bij erfgoedinstellingen.

De belangen van auteursrechthebbenden worden in belangrijke mate bepaald door enerzijds gemaakte afspraken of juridische/organisatorische zekerheid en anderzijds de omvang van harvestingactiviteiten, in het bijzonder de uitvoering van een nationale domeincrawl. Als het gaat om de bescherming van persoonsgegevens dan is het verschil tussen de scenario's de juridische basis, grondslag en bijbehorende voorzienbaarheid die er is ten aanzien van de bescherming van persoonsgegevens: in de scenario's waarbij er sprake is van wettelijk faciliteren, is de bescherming van persoonsgegevens beter.

De kosten laten beperkte verschillen tussen scenario's zien, behalve in de status quo waar er relatief lagere kosten zijn. Zoals in paragraaf 3.1.2 benoemd is, is de hoogte van de kosten in scenario 2 afhankelijk van het aantal partijen dat webharvestingactiviteiten uitvoert. De kosten zijn lager in scenario 2, mits er elf of meer instellingen webharvestingactiviteiten uitvoeren.

	Scenario 1 (centrale uitvoering)	Scenario 2 (centrale en decentrale uitvoering)	Scenario 3 (consensus)	Status quo (geen afspraken maken)	Toelichting
Informatie- en erfgoedbelang	■	■	■	■	Hoe meer webarchieven geharvest (kunnen) worden, hoe beter het informatie- en erfgoedbelang geborgd is (scenario 1 en 2).
Belang van auteursrecht- hebbenden	■	■	■	■	Er zijn geen grote verschillen tussen scenario 1, 2 en 3: Hoe meer rechtszekerheid, voorzienbaarheid en/of compensatie, hoe beter de belangen van auteursrechthebbenden geborgd zijn.
Bescherming persoons- gegevens	■	■	■	■	Hoe meer waarborgen, zoals een grondslag voor verwerking, hoe beter de bescherming van persoonsgegevens geborgd is (scenario 1 en 2).

Kosten					De kosten verschillen niet in grote mate tussen de scenario's, maar zijn wel afhankelijk van de hoeveelheid opslag en hoeveelheid uitvoerende partijen. Scenario 3 heeft een p.m. post voor de notice-and-takedown, waardoor deze waarschijnlijk (met een mate van onzekerheid) hoger uitvalt dan de andere scenario's.
Draagvlak					Het draagvlak is het grootst bij het wettelijk faciliteren van de huidige praktijk (waarbij erfgoedinstellingen zelf al harvesten) en het waarborgen en voorzienbaar maken van harvestingactiviteiten (scenario 1 en 2).
Uitvoerbaarheid					Hoe minder onzekerheden er zijn bij het uitvoeren van webharvesting (vanwege praktische en/of wettelijke aspecten; het kleinst bij scenario 2 en in mindere mate scenario 1), hoe beter uitvoerbaar dit is voor erfgoedinstellingen.

3.4.2 Conclusie over de te maken keuzes

In deze conclusie geven we nogmaals kort en bondig weer wat de belangrijkste verschillen zijn in effecten van de te maken keuzes, te beginnen bij de juridische routes.

De huidige situatie (status quo) is juridisch gezien het minst houdbaar en toekomstvast, omdat er geen recht wordt gedaan aan de afwegingen van verschillende belangen en grondrechten. Van de andere opties is het bij wet creëren van een publieke taak of bevoegdheid in combinatie met depotverplichting juridisch gezien de meest haalbare en houdbare optie. Binnen deze hoofdroute zijn er vervolgkeuzes te maken over de attributie van een taak of bevoegdheid aan één of meerdere instellingen of het delegeren/mandateren van deze taak door een centrale instelling met de taak/bevoegdheid. Deze vervolgvarianten zijn in dit onderzoek niet in detail uitgewerkt om te komen tot een advies over een keuze.

Naast de keuze met betrekking tot het wettelijk faciliteren van webharvesting, is de belangrijkste vraag op welke manier webharvesting gaat worden uitgevoerd. Het eerste scenario (centrale uitvoering) heeft als voordeel dat de belangen van auteursrechthebbenden beter beschermd kunnen worden via waarborgen en passende technische en organisatorische maatregelen door een centrale instelling met relatief veel middelen. Daarnaast kan dit scenario kostenbesparingen met zich meebrengen als er elf of meer instellingen zijn die op reguliere basis een verzoek voor webharvesting doen. Het scenario wijkt echter wel af van de huidige praktijk waarin ook decentrale instellingen al webharvestingactiviteiten uitvoeren.

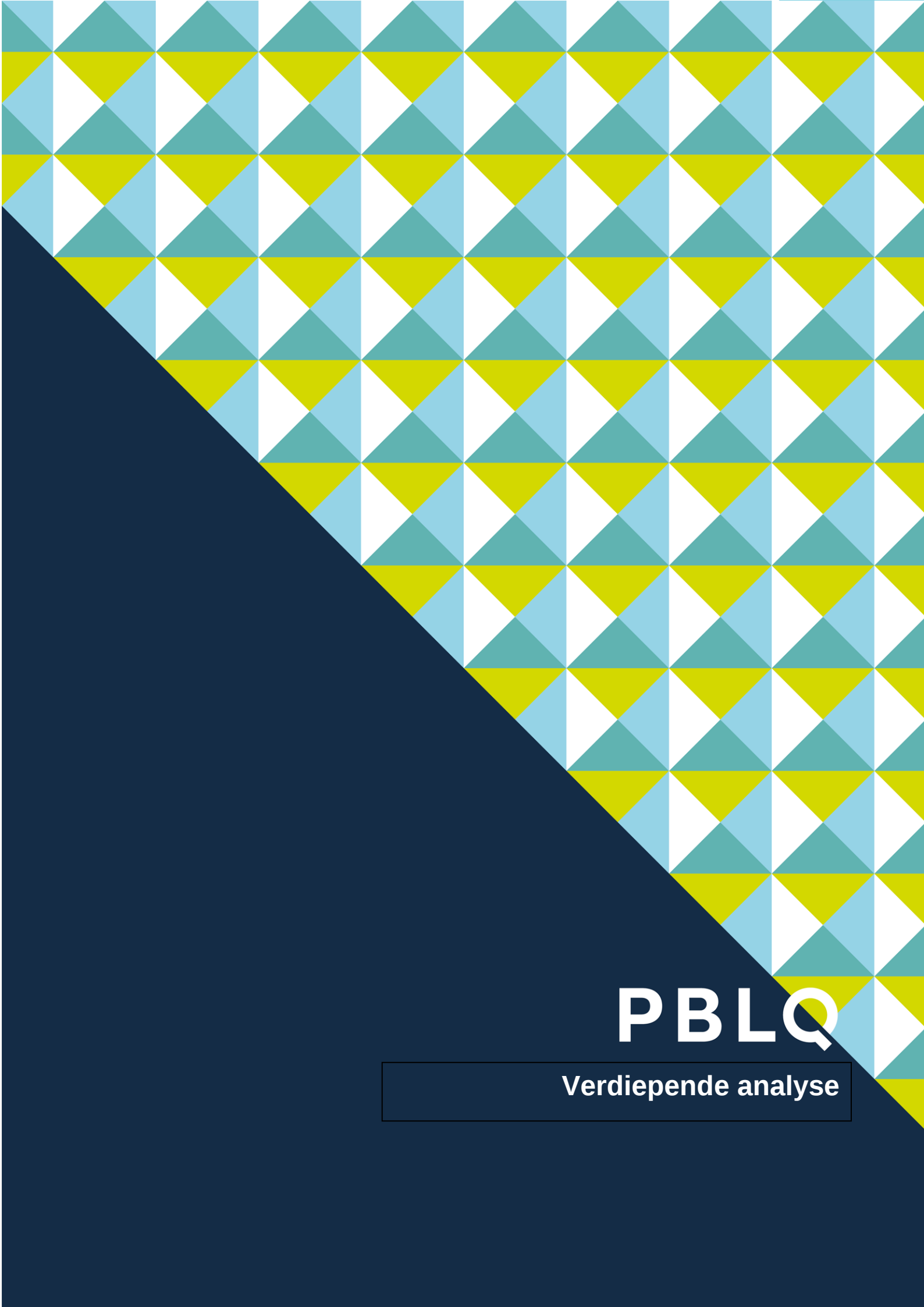
Het tweede uitgewerkte scenario (centrale en decentrale uitvoering) sluit het meest aan bij de huidige praktijk waarbij erfgoedinstellingen zelfstandig harvesten. Wel kunnen er meer kosten mee zijn gemoeid als er een groter aantal instellingen dan in de huidige praktijk gaan webharvesten (elf of meer) vergeleken met het eerste scenario. Ook is er het risico dat kleinere instellingen minder effectief

waarborgen en passende maatregelen kunnen nemen om auteursrechtelijk beschermde webarchieven en persoonsgegevens te beschermen, hoewel dat in de huidige praktijk nog geen probleem is.

Wanneer er gekozen wordt voor een route waarbij webharvesting niet wettelijk gefaciliteerd wordt, dan is er de keuze om afspraken te maken met belanghebbenden of de huidige praktijk (status quo) in stand te houden. In het geval van de eerste optie zijn de kosten iets hoger dan de twee scenario's waarin er wel sprake is van wettelijk faciliteren, en zijn er negatieve effecten op de juridische houdbaarheid en het draagvlak.

Ten aanzien van de afbakening van het nationaal webdomein hebben de twee strategieën voor- en nadelen. Geen van de twee strategieën heeft over de impactcategorieën meer voor- of nadelen ten opzichte van de andere strategie. Een strategie met criteria die via een vastgesteld proces worden bepaald geeft voordelen op het gebied van het informatie- en erfgoedbelang en het draagvlak bij uitvoerende partij(en). Een strategie op basis van vastgestelde criteria heeft relatief gezien meer voordelen op andere impactcategorieën, zoals de kosten en de voorzienbaarheid van webharvestingactiviteiten voor auteursrechthebbenden en betrokkenen.

Na het wegen van de verschillende opties en het maken van de in dit onderzoek beoordeelde keuzes zijn er nog verschillende vervolgstappen te zetten. Deze vervolgstappen zijn zowel het maken van aanvullende keuzes als het verder onderzoeken en uitwerken van juridische en uitvoering-gerelateerde aspecten. Welke vervolgstappen nodig zijn en welke inspanning daarmee gemoeid is, is afhankelijk van de gemaakte keuzes. Het gaat onder andere om een wetgevings- of juridische traject waarbij een groot aantal belanghebbenden bij betrokken is. In het vervolgtraject zullen nog meer keuzes moeten worden gemaakt ten aanzien van wet- en regelgeving, uitvoering en waarborgen.



PBLQ

Verdiepende analyse

4. Juridische analyse

In dit hoofdstuk staat de juridische analyse van de verschillende scenario's. We hebben daarbij gekozen voor een indeling op basis van juridische (deel)onderwerpen, en niet op basis van scenario's. Bij de analyse van de verschillende scenario's spelen de juridische hoofdonderwerpen rondom het auteursrecht, privacy- en het informatierecht in brede zin. Voor de effecten van de te kiezen scenario's zijn de volgende juridische kaders van invloed op de uitvoering en haalbaarheid:

- Intellectuele eigendomsrechten (auteursrechten, naburige rechten, databankenrecht)
- Fundamentele rechten (privacy, gegevensbescherming en grondrechtelijke waardenspanningen)

Beide kaders worden in dit onderzoek uitgewerkt langs de fases van verkrijgen (harvesten en acquireren), preservareren en presenteren van materiaal. Hieronder staan de belangrijkste conclusies op het gebied van deze juridische kaders.

De juridische kaders zijn gebaseerd op het IViR rapport, bureaustudie door de onderzoekers en input vanuit gesprekken met experts van stakeholders. Het hiernavolgende overzicht is bedoeld ter introductie op de verschillende juridische aspecten die op de uitvoering van webharvesting een rol spelen. Het zijn complexe juridische vraagstukken die veel nuances kennen. Die nuances zijn door de onderzoekers van het IViR gedetailleerd beschreven in hun onderzoek. Onderstaande weergave van de conclusies en de uitgebreide juridische analyse zelf, is niet bedoeld volledig te zijn. In de juridische analyse ten behoeve van dit onderzoek is de focus gelegd op de kern van het probleem, de beperking, respectievelijk uitzonderingsmogelijkheden en die elementen die van invloed zijn op de mogelijke oplossingsrichtingen.

Opgemerkt moet worden dat de in dit hoofdstuk benoemde vraagstukken niet uniek zijn voor het digitale erfgoedmateriaal, ze spelen ook bij het vergaren van de collecties en het archiveren daarvan in fysieke vorm.

4.1 Intellectuele eigendomsrechten

Bij het uitvoeren van webharvestingactiviteiten komen erfgoedinstellingen in aanraking met diverse intellectuele eigendomsrechten: het auteursrecht, naburige rechten en het databankenrecht. Auteursrechten bieden de makers van creatieve werken exclusieve rechten, waaronder het recht om hun werk te kopiëren, verspreiden en openbaar te maken. Webharvesting, dat vaak inhoudt dat grote hoeveelheden online content systematisch worden gekopieerd en gearcheveerd, kan deze exclusieve rechten van de makers schenden. Het juridische spanningsveld ontstaat doordat webharvesting enerzijds noodzakelijk is voor het behoud van cultureel en historisch erfgoed, maar anderzijds mogelijk inbreuk maakt op de rechten van rechthebbenden. De keuze die uit de auteursrechtelijke oplossingsrichtingen wordt gemaakt is essentieel om juridisch houdbare oplossingen te ontwikkelen die zowel de rechten van auteursrechthebbenden respecteren als het cultureel belang van het behoud van digitaal erfgoed waarborgen.

Auteursrechten

Voor de context van dit rapport is het nuttig om de verschillende handelingen die te maken hebben met webharvesting apart van elkaar te analyseren, omdat iedere handeling een andere juridische impact heeft. Vanuit het uitvoeringsperspectief van de erfgoedinstellingen en de taak die zij hebben

zou je ook een aantal handelingen kunnen samenvoegen, omdat voor hen het wegnemen van de juridische belemmeringen in de fasen van het verkrijgen en conserveren prioriteit heeft. Vanuit het perspectief van de makers en rechthebbenden zijn de verschillende handelingen in iedere fase relevant. Vanwege de organisatorische en juridische effecten worden ze voor de auteursrechten hier apart benoemd. Hiervoor volgen we de analyses uit het IViR rapport en worden geen zelfstandige juridische conclusies getrokken.

De analyse van auteursrechten in relatie tot webharvesting laat zien dat de huidige wetgeving aanzienlijke beperkingen oplegt aan erfgoedinstellingen, in alle fasen van het webharvestingproces. Binnen het webharvestingproces vinden verschillende soorten reproductiehandelingen plaats. Auteursrechten bieden rechthebbenden het exclusieve recht om hun werken te verveelvoudigen en openbaar te maken, wat impliceert dat webharvesting – een proces dat inherent gebaseerd is op het uitvoeren van reproductiehandelingen zoals het kopiëren en archiveren van online content – vaak inbreuk maakt op deze rechten zonder voorafgaande toestemming.

- In de fase van **harvesting**, waarin websites worden gecrawld en content wordt gekopieerd, kan dit worden gezien als een verveelvoudiging van auteursrechtelijk beschermd materiaal. Dit betekent dat erfgoedinstellingen in deze fase het risico lopen inbreuk te maken op het exclusieve recht van de maker. De huidige beperkingen, zoals die in artikel 16n van de Auteurswet, bieden erfgoedinstellingen beperkt ruimte om handelingen uit te voeren zonder voorafgaande toestemming van de rechthebbende. Tijdens het harvesten worden systematische handelingen uitgevoerd waarbij (tijdelijke) kopieën worden gemaakt in de tussenstappen zoals het crawlen. Deze handelingen zijn in meer en mindere mate van tijdelijke aard, mede gelet op het oogmerk deze langdurig te behouden. Niet voor alle handelingen die nodig zijn voor webharvesting wordt in art 5 lid 1 InfoSoc-richtlijn ruimte gevonden.³⁸
- Bij het **acquireren** van materiaal, oftewel het verkrijgen van toestemming of rechten, stuiten erfgoedinstellingen vaak op praktische problemen. Het verkrijgen van expliciete toestemming van alle rechthebbenden, gezien de schaal van webharvesting, is praktisch onuitvoerbaar. Zonder ruimere beperkingen blijven erfgoedinstellingen gebonden aan het complexe proces van toestemming vragen, wat hun mogelijkheden om content te archiveren ernstig beperkt.
- In de **preserveringsfase**, waarin materiaal wordt opgeslagen en beheerd, blijft het risico bestaan dat erfgoedinstellingen auteursrechtelijk beschermd werk bewaren zonder de juiste toestemming. Er is nog weinig duidelijkheid over hoe deze regels moeten worden toegepast op de lange termijn van opslag van digitaal erfgoed binnen de excepties op het auteursrecht vallen.
- De **presentatiefase**, waarin de gearchiveerde content openbaar toegankelijk wordt gemaakt, roept de meeste juridische risico's op. Het openbaar maken van auteursrechtelijk beschermd materiaal zonder toestemming kan een directe schending van het exclusieve recht van de maker zijn, wat ernstige juridische consequenties kan hebben. Er bestaan beperkingen voor bijvoorbeeld de toegang via fysieke terminals in bibliotheek of archief die daartoe een expliciete grondslag hebben wanneer het deel is gaan uitmaken van de eigen collectie, maar deze is beperkt tot de eigen leesterminal. Dit is nu het geval voor de KB.

De conclusie is dat het huidige auteursrechtelijke kader onvoldoende ruimte biedt voor de specifieke uitdagingen van webharvesting in elke fase van het proces. Zonder verduidelijking of uitbreiding van de bestaande beperkingen, blijven erfgoedinstellingen juridische risico's lopen en wordt hun vermogen om digitaal erfgoed effectief te archiveren aanzienlijk beperkt. Dit onderstreept de noodzaak van een

³⁸ IViR rapport, pagina 51.

beter afgestemd juridisch kader dat rekening houdt met de unieke dynamiek van digitale content en de maatschappelijke noodzaak om deze voor de toekomst te bewaren.

Naburige rechten

De analyse van naburige rechten in relatie tot webharvesting toont aan dat erfgoedinstellingen ook hier tegen aanzienlijke juridische beperkingen aanlopen, ook in iedere fase van het webharvestingproces. Naburige rechten beschermen de rechten van uitvoerende kunstenaars, producenten van fonogrammen en omroeporganisaties. Dit impliceert dat bij webharvesting waarbij audiovisueel materiaal, opnames of uitzendingen worden gekopieerd en gearchiveerd, erfgoedinstellingen mogelijk inbreuk maken op deze rechten zonder de vereiste toestemming van de rechthebbenden.

- In de fase van **harvesting**, waarin websites worden gecrawld en grote hoeveelheden data automatisch worden gekopieerd, kunnen werken die onder naburige rechten vallen, zoals muziek, video's of radio-uitzendingen, (bedoeld of onbedoeld) worden verveelvoudigd. Dit verveelvoudigen valt onder het exclusieve recht van de producent of uitvoerende kunstenaar volgens art. 7 van de Wet op de naburige rechten (Wnr). Net als bij auteursrechten biedt het huidige juridische kader weinig tot geen uitzonderingen voor deze grootschalige activiteiten door erfgoedinstellingen.
- Tijdens de **acquisitiefase**, waarin toestemming van rechthebbenden wordt verkregen, stuiten erfgoedinstellingen vaak op soortgelijke problemen als bij auteursrechten. Het verkrijgen van toestemming van de rechthebbenden van naburige rechten – vaak meerdere rechthebbenden per werk – is een complex en tijdrovend proces, vooral op grote schaal. Zonder wettelijke aanpassingen blijven erfgoedinstellingen gedwongen toestemming te vragen voor elk individueel werk, wat het archiveren van audiovisuele werken ernstig belemmert.
- In de **preserveringsfase**, waarin materiaal wordt opgeslagen en beheerd, geldt hetzelfde risico voor inbreuk op naburige rechten als bij auteursrechten. De naburige rechten blijven van kracht, zelfs wanneer het materiaal alleen voor archiveringsdoeleinden wordt bewaard. Hoewel er enige ruimte is voor uitzonderingen, biedt het huidige juridische kader onvoldoende bescherming voor instellingen die systematisch materiaal willen archiveren zonder toestemming.
- De **presentatiefase**, waarin het gearchiveerde materiaal beschikbaar wordt gesteld aan het publiek, is bijzonder problematisch wanneer het gaat om werken beschermd door naburige rechten. Het exclusieve recht van de rechthebbenden om hun werken openbaar te maken, zoals vastgelegd in art. 12 Wnr, maakt het voor erfgoedinstellingen moeilijk om dit materiaal zonder toestemming beschikbaar te stellen. Uitzondering hierop is dat de KB heeft aangegeven dat voor de KB vanwege het deel uit gaan maken van de collectie er wel een exceptie is waardoor op de eigen locaties rechtmatig gepresenteerd kan worden. De beperkingen, zoals de leesterminalexceptie, zijn onvoldoende om aan de behoeften van moderne, digitale archivering te voldoen.

De conclusie is dat ook het huidige juridische kader rond naburige rechten grote beperkingen oplegt aan erfgoedinstellingen in iedere fase van webharvesting. Ook hier betekent dat door een gebrek aan ruime beperkingen dat instellingen aanzienlijke juridische risico's lopen wanneer ze audiovisueel materiaal willen archiveren en presenteren. Zonder aanpassingen aan de wetgeving blijft het praktisch onhaalbaar voor erfgoedinstellingen om hun maatschappelijke taak met betrekking tot het behoud van digitaal erfgoed goed uit te voeren op een juridisch rechtmatige manier. Dit benadrukt de noodzaak om aanpassingen te doen die effect hebben op de naburige rechten en te zorgen voor een flexibelere

toepassing in het kader van webharvesting, zodat erfgoedinstellingen hun rol effectief kunnen vervullen zonder inbreuk te maken op de rechten van uitvoerende kunstenaars en producenten.

Databankenrecht

Anders dan bij het auteursrecht en de naburige rechten is de uitsplitsing naar de verschillende fasen van webharvesting minder relevant voor het databankenrecht. De analyse van databankenrecht in relatie tot webharvesting laat zien dat ook hier aanzienlijke beperkingen gelden voor erfgoedinstellingen. Het databankrecht beschermt de producent van een databank tegen het opvragen en hergebruiken van substantiële delen van de databank zonder toestemming. Voor erfgoedinstellingen die online databanken willen archiveren, zorgt dit voor juridische obstakels in iedere fase van het webharvestingproces. In het kader van webharvesting kan een databank worden gezien als een verzameling van gestructureerde gegevens, bijvoorbeeld de database van een e-commerceplatform, een sociale mediasite of een wetenschappelijke databank, waarin aanzienlijke investeringen zijn gedaan bij het verzamelen, ordenen en presenteren van de gegevens. Ook een website kan worden gekwalificeerd als een databank. Dit betekent dat bij het systematisch crawlen en kopiëren van grote databanken mogelijk inbreuk wordt gemaakt op de rechten van de databankproducent, tenzij hiervoor expliciete toestemming is verkregen.

Voor iedere fase waarin online databanken worden gekopieerd geldt dat het systematisch opvragen van gegevens uit een databank als een inbreuk worden beschouwd, als substantiële delen van de databank worden verwerkt. Artikel 2 van de Databankenwet (Dw) beschermt deze handelingen, waardoor erfgoedinstellingen beperkt worden in hun mogelijkheden om databanken te archiveren zonder toestemming van de producent. De vraag is wel in welke mate de webharvesting ook daadwerkelijk substantiële delen van databanken betreft, wat de impact van dit recht aanzienlijk beperkt.

Volgens Van Gompel is op basis van het *CV-Online Latvia*-arrest van het HvJ EU echter wel het argument te maken dat databankrechtelijk gebruik in het kader van webharvesting de markt voor de originele databankversie niet raakt en de investeringen van de samensteller van de databank dus ook niet frustreert, waardoor inbreuk op het databankenrecht mogelijk kan worden afgewend omdat het gebruik geen schade met zich meebrengt voor databankproducenten. Van Gompel acht het niet aannemelijk dat aan de voorwaarde is voldaan dat hergebruik van substantiële delen bij webharvesting het risico met zich meebrengt dat de investering in de databank niet kan worden terugverdiend. De verwachting is dat historisch archiefversies die opgenomen zijn in een webarchief de makers niet zullen hinderen.

4.2 Privacy, gegevensbescherming en fundamentele waarden spanningen

Privacy en gegevensbescherming

De analyse van privacy en gegevensbescherming in relatie tot webharvesting toont aan dat erfgoedinstellingen te maken hebben met juridische beperkingen, voornamelijk door de regels van de Algemene Verordening Gegevensbescherming (AVG) en de bredere bescherming van het recht op privacy zoals vastgelegd in artikel 8 van het Europees Verdrag voor de Rechten van de Mens (EVRM). Artikel 8 EVRM beschermt het recht op privéleven, wat ook de bescherming van persoonsgegevens omvat.

Webharvesting, waarbij vaak persoonsgegevens worden verwerkt, raakt direct aan dit recht, wat specifieke verplichtingen met zich meebrengt voor erfgoedinstellingen in iedere fase van het proces. Er is door de Nederlandse wetgever gekozen voor het invullen van een uitzonderingsgrond als benoemd in artikel 85 AVG en niet voor de uitzondering die in de AVG opgenomen voor erfgoedinstellingen (artikel 89 AVG), die kent ook zijn begrenzing. Artikel 85 AVG stelt de verwerkingsverantwoordelijke niet volledig vrij en moeten de AVG-basisbeginselen (art. 5 AVG) in acht worden genomen bij de webharvestingactiviteiten.

Artikel 43 Uitvoeringswet Algemene Verordening Gegevensbescherming (uAVG) bevat een lijst van bepalingen waar verwerkingsverantwoordelijken die zich op een uitzondering beroepen, geen rekening mee moeten houden. De vraag is wel wat de reikwijdte is van deze uitzondering. De positie van de KB wordt expliciet benoemd vanwege haar werkzaamheid en mandaten daartoe in sectorwetgeving op het gebied van het bibliotheekwezen en informatievoorziening in de Memorie van Toelichting bij artikel 43 uAVG.³⁹ Het wettelijke mandaat dat de KB heeft, is belangrijk bij de inkleuring van deze uitzondering.

De IViR-onderzoekers halen in hun rapport aan dat men *mutatis mutandis* hetzelfde zou kunnen verwachten met betrekking tot andere erfgoedinstellingen. De uitzondering zoals deze is neergelegd is in artikel 45 uAVG geldt alleen voor persoonsgegevens die deel uitmaken van archief bescheiden als bedoeld in art. 1. onderdeel c. Archiefwet. Dat is een andere scope dan de webharvestingactiviteiten als bedoeld in dit rapport en heeft in dit kader geen toepassing.

- In de fase van **harvesting en acquireren**, waarbij websites worden gecrawld en persoonsgegevens zoals IP-adressen en namen worden verzameld, kan dit een inbreuk vormen op het recht op privacy uit artikel 8 EVRM. Erfgoedinstellingen moeten ervoor zorgen dat de verwerking van deze gegevens in overeenstemming is met de AVG. Zonder geldige rechtsgrond, kan de verwerking van persoonsgegevens als een inbreuk op dit fundamentele recht worden gezien.
 - Expliciete toestemming lijkt, zoals het IViR-onderzoek ook benadrukt, een onrealistische grondslag. Het is nagenoeg onmogelijk om van iedere betrokkene vooraf toestemming te verkrijgen.
 - De grondslag ‘gerechtvaardigde belangen’ kan soelaas bieden voor de organisaties die niet belast zijn met een wettelijke taak. De uitwerking voor deze grondslag is een balans tussen de betrokkene enerzijds en de belangen van de verantwoordelijke en eventuele derden anderzijds. De sectoraal of eventueel wettelijk vastgestelde doeleinden kunnen behulpzaam zijn bij het invullen van die gerechtvaardigde belangen.
 - Indien wordt gekozen voor het wettelijk vastleggen van taken en bevoegdheden zet dit de deur open voor een beroep op de grondslag ‘wettelijke verplichting’. Zodra webharvesting binnen het wettelijk opgelegde takenpakket ligt, kan op deze rechtsgrond worden teruggevallen. Of een taak of (enkel) een bevoegdheid voldoende is voor deze grondslag, hangt af van de exacte juridische invulling. Het gaat dan bijvoorbeeld of de bevoegdheid is opgenomen in een wet of in lagere regelgeving.
 - Alternatief, bij het niet opnemen van een depotplicht kan de grondslag ‘noodzakelijk voor de vervulling van een taak van algemeen belang uitkomst bieden, door invulling te geven aan artikel 89 AVG. Anders dan bij de vorige optie is er geen vereiste van expliciete wettelijke

³⁹ Kamerstukken II 2017/18, 34 851, nr. 3, p. 123.

plicht die verwerking van persoonsgegevens impliceert. De Werkgroep Artikel 29 (WP29), inmiddels opgevolgd door de European Data Protection Board (EDPB) heeft in haar richtsnoeren benadrukt dat de grondslag ‘noodzakelijk voor de vervulling van een taak van algemeen belang’ strikt moet worden geïnterpreteerd.⁴⁰ Dit betekent dat een gegevensverwerking alleen gerechtvaardigd is wanneer deze expliciet noodzakelijk is voor de uitvoering en een specifieke taak van algemeen belang en wanneer aan de vereisten van proportionaliteit en subsidiariteit wordt voldaan. Om op deze grondslag een succesvol beroep te doen is een zeer strikte interpretatie van deze richtsnoeren nodig. Hoe de keuze voor een taak of bevoegdheid tot een depotplicht hangt af van de concrete invulling, wat onderwerp moet zijn van een wetgevingstraject.

- In alle gevallen dient voor gestart kan worden met de verwerking van persoonsgegevens aanvullende maatregelen te worden genomen (verderop in dit hoofdstuk benoemd).
- In de **preserveringsfase** blijft het recht op privacy centraal staan. Dit betekent dat instellingen verplicht zijn om persoonsgegevens te beveiligen en de toegang tot deze gegevens te beperken, vooral als de informatie gevoelig is of betrekking heeft op identificeerbare personen. Ook hier wordt de balans gezocht tussen het belang van archivering en het waarborgen van het recht op privacy zoals vastgelegd in artikel 8 EVRM.
- In de **presentatiefase**, waarin erfgoedinstellingen gearhiveerde content openbaar maken, kan het recht op privacy van betrokkenen opnieuw in het geding komen. Ondanks de uitzonderingsgrond van artikel 85 AVG, blijven instellingen verplicht om te zorgen dat de openbaarmaking van persoonsgegevens voldoet aan de beginselen van noodzakelijkheid en proportionaliteit. In lijn met artikel 8 EVRM moeten instellingen ervoor zorgen dat betrokkenen niet onevenredig worden benadeeld door de openbaarmaking van hun gegevens, wat bijvoorbeeld kan betekenen dat bepaalde gegevens worden verwijderd of geanonimiseerd voordat ze toegankelijk worden gemaakt, maar gelet op de journalistieke uitzondering van artikel 43 uAVG en de belangen die hiermee gediend zijn, moet hier zeer terughoudend mee om worden gegaan.
- Over het geheel geldt dat een geslaagd beroep op de uitzondering van artikel 85 AVG (art. 43 uAVG) van bijzonder belang is, omdat die een uitzondering oplevert voor de in artikel 9 bedoelde bijzondere gegevens, voor zover dit noodzakelijk is voor de beschermde doeleinden van (tot nu toe) de KB.

De conclusie is dat de huidige juridische kaders, inclusief de bescherming van privacy onder artikel 8 EVRM en de uitzonderingen in artikel 85 AVG, erfgoedinstellingen aanzienlijke verplichtingen opleggen bij webharvesting. Hoewel artikel 85 AVG enige ruimte biedt voor de verwerking van persoonsgegevens voor archiveringsdoeleinden, blijven er strikte eisen bestaan om de privacy van betrokkenen te beschermen. Erfgoedinstellingen moeten een zorgvuldige afweging maken tussen hun rol bij het behoud van digitaal erfgoed en de bescherming van het fundamentele recht op privacy, zoals gewaarborgd in artikel 8 EVRM. Verdere verduidelijking van de uitzonderingsgrond in artikel 85 AVG, nadere explicitering in de uitvoeringswet Algemene Verordening Gegevensbescherming (uAVG) of een advies van de Autoriteit Persoonsgegevens (AP) zou kunnen helpen om een beter evenwicht te vinden tussen deze rechten en belangen. Het moet daarvoor meer expliciet worden gemaakt wanneer en in welke mate webharvesting als noodzakelijk kan worden aangemerkt. Dat vereist een meer gedetailleerde beschrijving van de doeleinden en de mate waarin een eventuele verwerking van

⁴⁰ Werkgroep Artikel 29, *Advies 06/2014 over het begrip ‘gerechtvaardigd belang van de voor de verwerking verantwoordelijke’ in de zin van artikel 7 van Richtlijn 95/46/EG*, WP 217, 9 april 2014, beschikbaar via: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp217_nl.pdf.

persoonsgegevens noodzakelijk zijn voor deze doeleinden en welke waarborgen bij de erfgoedinstellingen rusten. Denk hierbij aan anonimiseringsprocedures en toegangsbescherming.

Door het creëren van een expliciete wettelijke grondslag kunnen democratisch gelegitimeerde keuzes worden gemaakt over hoe moet worden omgegaan met webharvesting.

Advies Autoriteit Persoonsgegevens

De Autoriteit Persoonsgegevens (AP) heeft het Nationaal Archief en het Ministerie van OCW in 2024 gewaarschuwd voor de voorgenomen online openbaarmaking van het Centraal Archief Bijzondere Rechtspleging (CABR) wegens mogelijke schendingen van de AVG. De AP stelde dat het openbaar maken van gevoelige persoonsgegevens, waaronder gegevens van verdachten, slachtoffers en getuigen, niet is toegestaan zonder een specifieke wettelijke grondslag. Tevens werd gewezen op de risico's van onbeperkte online toegankelijkheid, waarmee de privacy van betrokkenen in het geding kan komen. Dit advies is van belang voor het onderzoek naar webharvesting, aangezien het onderstreept dat erfgoedinstellingen bij het verzamelen en beschikbaar stellen van webcontent de verwerking van persoonsgegevens moeten baseren op een duidelijke wettelijke grondslag die voorziet in waarborgen ter bescherming van de rechten van betrokkenen. Het ontbreken van dergelijke waarborgen kan leiden tot inbreuken op de AVG.⁴¹

Chilling effect en erosie van het recht op privacy

Het wettelijk mogelijk maken van webharvesting door erfgoedinstellingen kan leiden tot een *chilling effect* op de privacy van individuen. Dit houdt in dat personen zich geremd kunnen voelen in hun onlinecommunicatie en gedrag, uit een angst dat hun gegevens worden verzameld en gearhiveerd. Het is daarom van belang dat bij het ontwikkelen van wetgeving voor webharvesting adequate waarborgen worden ingebouwd waarbij het risico op het *chilling effect* geminimaliseerd wordt. Denk hierbij aan duidelijke kaders voor het toegankelijkheid maken van gegevens.

Daarnaast kan webharvesting bijdragen aan de erosie van het recht op privacy. Door het permanent vastleggen van ontsluiten van persoonlijke en semi-publieke gegevens, die oorspronkelijk mogelijk alleen voor tijdelijke of beperkte verspreiding waren bedoeld, vervagen de grenzen van publiek en privé. Het structureel en grootschalig verzamelen van data kan leiden tot een normalisering van constante gegevenswaring, waardoor het fundamentele recht op en het begrip privacy zelf geleidelijk verwaterd. Het gevaar kan ontstaan dat op termijn de maatschappelijke perceptie van wat als privé wordt beschouwd, kan veranderen.

Dit versterkt de noodzaak voor duidelijke wettelijke kaders en waarborgen die niet alleen duidelijke grenzen stellen, maar ook actief het recht op privacy beschermen.

Grondrechtelijke waardenspanningen

Webharvesting roept verschillende grondrechtelijke vragen op, vooral in de context van de Europese bescherming van fundamentele rechten zoals vastgelegd in het Handvest van de Grondrechten van de Europese Unie (Handvest) en het Europees Verdrag voor de Rechten van de Mens (EVRM). Grondrechten zoals de vrijheid van meningsuiting (artikel 10 EVRM), de informatievrijheid (artikel 11

⁴¹ Autoriteit Persoonsgegevens, *Nationaal Archief gewaarschuwd om online openbaarmaking archief bijzondere rechtspleging*, 29 november 2024, beschikbaar via: <https://www.autoriteitpersoonsgegevens.nl/actueel/nationaal-archief-gewaarschuwd-om-online-openbaarmaking>

Handvest), het recht op privacy (artikel 8 EVRM) en het recht op eigendom (artikel 1 Protocol nr. 1 EVRM) moeten zorgvuldig worden afgewogen tegen elkaar in het kader van het belang van webharvesting. Een keuze maken tussen de scenario's vraagt om het maken van een goede afweging van de verschillende belangen. De hiervoor benoemde rechten staan soms op gespannen voet met elkaar wanneer erfgoedinstellingen online content systematisch verzamelen en presenteren. Tegelijkertijd zijn er ook grote (maatschappelijke) belangen bij het mogelijk maken van webharvesting. Deze rechten en belangen moeten zorgvuldig worden afgewogen in het kader van webharvesting, waarbij het doel is om digitaal erfgoed te bewaren en zo beperkt mogelijk inbreuk te maken op de rechten van individuen of rechthebbenden.

De afwegingen hiertussen in de praktijk hebben effect op de wijze van uitvoering en maatregelen die mogelijk getroffen moeten worden om het voorkomen van niet-proportionele inbreuk op grondrechten te borgen. Primair ligt deze bij de beleidsmakers, maar ook ten aanzien van de wijze van uitvoering zijn erfgoedinstellingen aan zet om goede invulling te geven in het proces om een goede bescherming van grondrechten blijvend te waarborgen. Ook Van Gompel, Schumacher en Senftleben (2022), geven aan dat in het licht van de vrijheid van meningsuiting, de informatievrijheid en het recht op wetenschap er geen twijfel bestaat dat webharvesting wenselijk is.⁴² Bij het mogelijk maken van webharvesting kunnen fundamentele rechten worden aangetast, daarom is het relevant om bij het maken van een keuze binnen de verschillende oplossingsrichtingen, kennis te nemen van die waardenspanningen tussen de verschillende fundamentele rechten.

Waardenspanning vrijheid van meningsuiting en privacy

De vrijheid van meningsuiting, zoals beschermd in art. 10 EVRM en art. 11 van het Handvest, biedt burgers het recht om informatie te delen en te ontvangen. Webharvesting kan als middel dienen om deze vrijheid te waarborgen door online-informatie voor toekomstige generaties beschikbaar te houden. Tegelijkertijd kan het verzamelen van persoonsgegevens, of het bewaren van uitingen van individuen, botsen met het recht op privacy zoals gewaarborgd in artikel 8 EVRM en artikel 7 en 8 van het Handvest. Bij het verwerken van persoonsgegevens tijdens harvesting moeten erfgoedinstellingen rekening houden met de belangen van betrokkenen. Dit leidt tot de vraag of harvesting inbreuk maakt op privacy rechten en of deze inbreuk kan worden gerechtvaardigd in het licht van de publieke taak van erfgoedinstellingen. In het EHRM-arrest *Times Newspapers Ltd. vs. Verenigd Koninkrijk* is het belang van het bewaren van uitingen in het publieke belang bevestigd, maar is ook het belang van privacybescherming bij publicaties benadrukt.

Waardenspanning privacy en informatie- en archiveringsvrijheid

Het acquireren van onlinemateriaal roept vragen op over de proportionaliteit van de inbreuk op privacy in verhouding tot het publieke belang van archivering. Artikel 11 van het Handvest waarborgt de vrijheid van informatie en kan een rechtvaardiging vormen voor erfgoedinstellingen om online content te verzamelen. Van belang is dat dit recht moet worden afgewogen tegen de rechten van individuen onder art. 8 EVRM, die privacybescherming omvat. Voor erfgoedinstellingen is in andere landen bepaald dat zij zich kunnen beroepen op uitzonderingen voor archivering in het algemeen belang, zoals vastgelegd in art. 89 AVG en art. 10 EVRM, om het verzamelen van informatie te rechtvaardigen. Dit is een andere situatie dan de uitzonderingen voor journalistiek doeleinden zoals het in Nederland is geregeld (art 85 AVG). Desondanks moeten zij transparantie betrachten en verantwoording afleggen over hun verzamelpraktijken om te waarborgen dat privacy niet onnodig wordt geschonden.

⁴² S. van Gompel, L. Schumacher en M. Senftleben, 'Webharvesting: auteursrechtelijke implicaties en oplossingsrichtingen', *Auteursrecht* 2022-1, p 4. Zie ook de in dit artikel aangehaalde bronnen.

Waardenspanning opslaan van gegevens, recht op toegang tot informatie versus privacy

De opslag van gearchiveerde informatie roept vragen op over de waarborgen die erfgoedinstellingen moeten bieden om de privacy van individuen te beschermen. Art. 85 resp. 89 AVG biedt enige flexibiliteit voor het bewaren van persoonsgegevens voor archiveringsdoeleinden, maar vereist dat erfgoedinstellingen passende waarborgen implementeren, zoals anonimisering of pseudonimisering. Tegelijkertijd moet het recht op toegang tot informatie worden beschermd, vooral wanneer deze informatie van historisch of wetenschappelijk belang is. Het grondrechtelijk evenwicht ligt hier tussen het recht van het publiek om toegang te krijgen tot informatie (art.10 EVRM) en de bescherming van persoonsgegevens (art. 8 EVRM). De opslag van gegevens die gevoelige informatie bevatten, zoals politieke opvattingen of religieuze overtuigingen, vereist extra zorgvuldigheid.

Waardenspanning grondrechtelijke balans en openbaarmaking van digitaal erfgoedmateriaal

Een relevant grondrechtelijk vraagstuk is de vraag of het recht op toegang tot informatie altijd prevaleert boven de bescherming van privacy. Artikel 8 EVRM stelt duidelijke grenzen aan de openbaarmaking van informatie, vooral wanneer deze informatie schadelijk kan zijn voor de betrokkenen. Archieven die informatie presenteren moeten daarom overwegen of een evenwichtige afweging van belangen heeft plaatsgevonden, vooral in het licht van mogelijke negatieve effecten voor betrokkenen. Het EHRM benadrukt in *Axel Springer AG vs. Duitsland* dat persvrijheid en informatievrijheid belangrijke rechten zijn, maar dat zij in balans moeten worden gebracht met het recht op bescherming van reputatie en het recht op privacy.⁴³

Waardenspanning eigendomsrecht en toegang tot digitale informatie

Naast privacy en informatievrijheid speelt ook het recht op eigendom een belangrijke rol bij webharvesting. Artikel 1 van Protocol nr. 1 bij het EVRM beschermt het eigendomsrecht van natuurlijke en rechtspersonen, waaronder intellectuele eigendomsrechten zoals het auteursrecht. Dit recht kan onder druk komen te staan wanneer erfgoedinstellingen online materiaal archiveren dat auteursrechtelijk beschermd is, zonder voorafgaande toestemming van de rechthebbende. Hergebruik en digitaal toegankelijk maken van dergelijk materiaal, met name wanneer dit publiek toegankelijk wordt gemaakt, kunnen worden gezien als een inmenging in het eigendomsrecht. De waardenspanning ontstaat tussen het publieke belang van digitale toegang tot cultureel erfgoed en het individuele belang van de maker om de exclusieve controle te houden over het werk. Bij de keuze voor een scenario, maar ook bij de keuze van de exacte inrichting van processen dient daarom een zorgvuldige balans gezocht te worden tussen het belang van duurzame toegang en het respecteren van eigendomsrechten, waarbij transparantie en een verantwoorde omgang met rechtenbeheer essentieel zijn.⁴⁴

4.3 Mogelijkheden juridische invulling oplossingsrichtingen

Er zijn op dit moment geen auteursrechtelijke beperkingen of andere juridische instrumenten die de reproductiehandelingen toestaan die nodig zijn voor het uitvoeren van webharvestingactiviteiten.

⁴³ EHRM 7 februari 2012, (*Axel Springer vs Duitsland*), nr. 39954/08

⁴⁴ Zie voor een uitgebreide uiteenzetting van deze waardenspanning ook P.B. Hugenholtz & L. Guibault, 'The Future of the Public Domain: An Introduction', in: L. Guibault & P.B. Hugenholtz (red.), *The Future of the Public Domain: Identifying the Commons in Information Law*, Alphen aan den Rijn: Kluwer Law International 2006, p. 1–6.

De methode zoals deze nu wordt gehanteerd, het informeren de websitebeheerder en na uitblijven van het gebruik maken van de opt-out-regeling overgaan tot webharvesting, biedt weinig juridische zekerheid. Naast dat de websitebeheerder vaak niet de maker en rechthebbende is en daarmee niet de juiste persoon is om het verzoek te doen, bevat de gekozen vorm een van impliciete toestemming, die geen rechtszekerheid biedt. Het voert voor dit onderzoek te ver om alle juridische details ten aanzien van de invulling van de oplossingsrichtingen hier nader uit te werken. Voor de detailuitwerking wordt verwezen naar het IViR-rapport en Van Gompel Schumacher en Senftleben (2022).⁴⁵ De meest belangrijke elementen die van invloed zijn op de oplossingsrichtingen worden hierna besproken.

Balans aanbrengen beperkingen en waarborgen fundamentele rechten

De ruimte in het kiezen tussen de verschillende oplossingsrichtingen is volgens van Van Gompel Schumacher en Senftleben (2022) niet geheel vrij. In de keuzes moet een balans worden gebracht in het systeem van exclusieve rechten en beperkingen zoals die er zijn in het Europese auteursrecht, zo blijkt ook uit het *Pelham*-uitspraak van de HvJ EU.⁴⁶ De bestaande beperkingen in het auteursrecht bieden geen ruimte voor webarchivering op grote schaal. Ook het presenteren aan het publiek is op basis hiervan niet mogelijk.

Artikel 5 lid 2 sub C Info-Soc-richtlijn biedt ruimte om webharvestingactiviteiten met een specifieke focus op te nemen als beperking. Daarvoor is wel vereist dat de te harvesten content 'welbepaald' moet zijn en het is volgens Van Gompel Schumacher en Senftleben (2022) te betwijfelen of een beperking in voldoende mate specifiek kan worden geacht als webharvesting globaal wordt vrijgesteld voor alle niet-commerciële erfgoedinstellingen. Volgens deze auteurs bieden zonder een juridische tussenstap, zoals een wettelijke depotplicht, art. 6 DSM-richtlijn en art. 5 lid 2 sub c InfoSoc-richtlijn (en de nationale tegenhangers in art. 16n Aw, art. 10 sub f Wnr en art. 4a sub d Dw) geen grondslag voor het faciliteren van webharvesting door erfgoedinstellingen.

Depotplicht

De bestaande beperkingen t.a.v. de intellectuele eigendomsrechten van de rechthebbenden vereisen vaak aanvullende flankerende maatregelen, om een goede juridische basis te bieden. Zoals in Hoofdstuk 3 ook benoemd is het ingevolge de DSM-richtlijn noodzakelijk dat een lidstaat voorziet in een uitzondering zodat instellingen in hun collecties content kunnen reproduceren en permanent onderdeel worden van hun collecties. Dit onderdeel maken van de permanente collectie kan door overdracht van eigendom of licentieovereenkomst, regelingen inzake permanente uitleen of een wettelijke depotplicht. Een wettelijke tussenstap, zoals een depotplicht, is nodig om online materiaal onderdeel te maken van de collectie van culturele erfgoedinstellingen, zonder dat een voorgaande instemming is verkregen van de auteursrechthebbende. Pas nadat deze stap is genomen, wordt het mogelijk om beperkingen in te roepen die specifieke reproductie- of openbaarmakingshandelingen toestaan. Deze wettelijke depotplicht omvat een verplichting voor uitgevers en producenten om hun werk beschikbaar te stellen aan een aangewezen erfgoedinstelling, die tot doel heeft deze werken te bewaren als onderdeel van de collectie. Volgens Van Gompel, Schumacher en Senftleben (2022) is het goed verdedigbaar dat ook Nederland deze route kan bewandelen om alsnog een regeling ten aanzien de toevoeging van webcontent aan de collecties van erfgoedinstellingen te introduceren. In andere EU-landen, zoals Denemarken, Duitsland en Frankrijk wordt webharvesting gefaciliteerd door een specifieke regeling voor het wettelijk depot op het internet.

⁴⁵ S. van Gompel, L. Schumacher en M. Senftleben, 'Webharvesting: auteursrechtelijke implicaties en oplossingsrichtingen', *Auteursrecht* 2022-1, p 4.

⁴⁶ HvJ EU 29 juli 2019, C-476/17 (*Pelham*)

Vanuit juridisch perspectief geeft combinatie van een wettelijke depotverplichting met een beroep op de beperkingen op de verschillende intellectuele eigendomsrechten een hoge mate van rechtszekerheid. Op dit moment kent de Nederlandse wetgeving geen bestaande depotwetgeving die uitgebreid kan worden naar webcontent.

Versterking grondslag

Ook vanuit het privacy-rechtelijk perspectief verdient de voorkeur om de taak of bevoegdheid die erfgoedinstellingen hebben te expliciteren in een juridische grondslag. Door expliciet wettelijk vast te leggen welke erfgoedinstellingen onder welke omstandigheden van een uitzonderingsregime kunnen genieten wordt ook ten aanzien van het privacy-thema meer rechtszekerheid geboden.

Een plek om die specifieke taak of bevoegdheid en de depotverplichting voor webcontent vast te leggen en expliciet te maken zouden de wetten zijn waar de taak van de (gekozen) erfgoedinstellingen bij wet is vastgelegd. Voor de KB is dit de Wet op het hoger onderwijs en wetenschappelijk onderzoek en de Mediawet voor Beeld en Geluid. Daarvoor is het noodzakelijk dat de gekozen erfgoedinstellingen een duidelijk wettelijk mandaat krijgen om hun webharvestingactiviteiten uit te voeren en het materiaal onderdeel kan worden van hun collectie.

Mogelijkheden tot uitwerking van taak en bevoegdheden

Bevoegdheden kunnen op drie verschillende manieren worden toebedeeld aan de erfgoedinstellingen: bij attributie, mandaat en delegatie.

Attributie

Attributie is een vorm van bevoegdheidstoedeling waarbij een bevoegdheid die nog niet bestond, door de wetgever in het leven wordt geroepen en toegekend aan een of meerdere instellingen.

Dit betreft het creëren van een nieuwe bevoegdheid door de wetgever en het toekennen daarvan aan een erfgoedinstelling. In het geval van webharvesting zou dit betekenen dat de wetgever een expliciete wettelijke basis creëert die erfgoedinstellingen de bevoegdheid geeft om webcontent te verzamelen en te archiveren.

Mandaat

Mandaat is een vorm van bevoegdheidstoedeling waarbij een reeds bestaande bevoegdheid door de ene organisatie (mandaatgever) een ander (de gemandateerde) machtigt om een bevoegdheid uit te oefenen in naam van (en onder verantwoordelijkheid van) de mandaatgever. Handelingen van de gemandateerde gelden als een handeling van de mandaatgever. De mandaatgever kan de gemandateerde bevoegdheid zelf blijven uitoefenen. Het verlenen van mandaat is in beginsel toegestaan, tenzij bij wettelijk voorschrift anders is bepaald.

Bij mandatering draagt een organisatie een bestaande bevoegdheid over aan een andere partij, die deze bevoegdheid uitoefent in naam van en onder verantwoordelijkheid van de mandaatgever. Omdat er nog geen wettelijke bevoegdheid is voor erfgoedinstellingen, kan mandatering nog niet plaatsvinden.

Delegatie

Delegatie is een vorm van bevoegdheidstoedeling waarbij een reeds bestaande bevoegdheid door de ene organisatie (delegans) wordt overgedragen aan een andere (delegataris). De delegataris

oefent de bevoegdheid vervolgens onder eigen verantwoordelijkheid uit. De delegans kan de gedelegeerde bevoegdheid *niet* meer zelf uitoefenen. Delegatie is alleen mogelijk als de wet dat bepaalt. De delegans:

- kan de delegataris regels opleggen en inlichtingen vorderen over de uitoefening van de gedelegeerde bevoegdheid,
- kan de delegatie inperken en voorbehouden maken,
- kan de delegatie te allen tijde intrekken.

Indien de wetgever specifieke wettelijke bepalingen zou opnemen die een bestaande bevoegdheid om webcontent te archiveren overdraagt aan erfgoedinstellingen, zou dit een vorm van delegatie zijn. Erfgoedinstellingen zouden dan zelfstandig, op basis van hun eigen verantwoordelijkheid, webharvestingactiviteiten uitvoeren zonder directe inmenging van de delegans. Dit vereist echter expliciete wettelijke bepalingen, aangezien delegatie alleen mogelijk is wanneer de wet dit toestaat. Aangezien er in Nederland geen bestaande bevoegdheid is die hiervoor in aanmerking komt, is delegatie momenteel niet toepasbaar.

Taak of bevoegdheid

Vanuit het juridische perspectief van rechtszekerheid is het minder van belang wáár de taak of bevoegdheid precies wordt vastgelegd, maar vooral hóe deze wordt ingevuld. Of iets wordt aangemerkt als een taak of een bevoegdheid, en of deze gepaard gaat met een verplichting of juist een juridische mogelijkheid, hangt af van de bestuursrechtelijke context. Een taak legt doorgaans een verplichting op, terwijl een bevoegdheid ruimte biedt voor discretionaire afweging.⁴⁷ In het kader van webharvesting is het van belang dat erfgoedinstellingen zelf kunnen blijven bepalen welke content zij aan hun collectie willen toevoegen, wat een vorm van discretionaire afweging is. Erfgoedinstellingen geven ook expliciet aan waarde te hechten aan deze keuzevrijheid. Een te strak geformuleerde wettelijke verplichting kan haaks staan op de wens om zelf afwegingen te maken binnen de collectieopbouw. Tegelijkertijd moet de grondslag voor de verwerking van persoonsgegevens ook voldoende concreet zijn om de juiste mate van bescherming te kunnen bieden zoals dat binnen het gegevensbeschermingsrecht bedoeld is. Het is daarom essentieel dat de wetgeving voldoende ruimte biedt voor eigen professionele en culturele afwegingen, terwijl het tegelijk een duidelijke juridische basis verschaft waarbinnen deze activiteiten kunnen plaatsvinden, met inachtneming van de geldende auteursrechtelijke kaders en fundamentele rechten.

⁴⁷ H.D. van Wijk, W. Konijnenbelt & R. van Male, *Hoofdstukken van bestuursrecht*, 16e druk, Deventer: Wolters Kluwer 2014, p. 67

5. Uitvoerbaarheid

Dit hoofdstuk bespreekt de uitvoerbaarheid van de verschillende scenario's in termen van technische middelen, personele middelen en administratieve lasten voor erfgoedinstellingen. Per scenario wordt een toelichting gegeven, waarbij gestart wordt met een opsomming van de benodigde activiteiten per scenario. Deze activiteiten vormen de basis om de uitvoerbaarheid te toetsen, en worden later gebruikt om de kosten te ramen (Hoofdstuk 6) en het draagvlak toe te lichten (Hoofdstuk 7).

5.1 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen

In dit scenario is er één centraal gekozen instelling die zowel de nationale domeincrawl als beredeneerde selecties uitvoert. De nationale domeincrawl vindt op regelmatige basis (jaarlijks) plaats. De beredeneerde selecties kunnen in de praktijk zo vaak als nodig geacht plaatsvinden op verzoek van andere erfgoedinstellingen, in deze kostenanalyse worden aannames gehanteerd over hoe vaak dit zal gebeuren.

Wanneer decentrale instellingen een beredeneerde selectie van webdomeinen wil harvesten, doen zij hiervoor een verzoek bij de centrale instelling (die daarmee een 'loketfunctie' krijgt). De beredeneerde selectie kan bijvoorbeeld bestaan uit webdomeinen die gerelateerd zijn aan een bepaald onderwerp of evenement (zogenaamde hotspots). De centrale instelling beoordeelt het verzoek van de decentrale instelling en voert, na goedkeuring, de harvest uit.

5.1.1 Activiteiten

De volgende activiteiten worden uitgevoerd in dit scenario:

- De centrale instelling gaat in dit scenario de nationale domeincrawl uitvoeren en krijgt een loketfunctie voor het harvesten van materiaal naar aanleiding van beredeneerde selecties van decentrale instellingen.
- Decentrale instellingen voeren het proces van selectie en collectievorming uit voor webarchieven die binnen hun doel- en taakstelling passen.
- Het loket ontvangt verzoeken van decentrale instellingen, controleert deze (bijvoorbeeld op uitvoerbaarheid) en voert de harvest uit. Ook wordt er gecontroleerd of het materiaal kwalitatief in orde is en compleet is. Gedurende het hele proces vindt er afstemming plaats tussen de vragende partij en de harvestende partij.
- Het geharveste materiaal wordt bij de centrale instelling gepreserveerd en op verzoek overgebracht bij de verzoekende partij, waarmee bedoeld wordt dat het in de collectie van de verzoekende partij wordt opgenomen.

5.1.2 Uitvoerbaarheid

Al deze activiteiten kosten tijd en personele inzet, in het bijzonder bij de centrale harvestende instelling, waarmee de uitvoeringslast vooral hoog is voor deze organisatie. De loketfunctie brengt administratieve lasten met zich mee en het afstemmen tussen vragende en harvestende partij kost tijd: verzoeken moeten (tijdig) worden verwerkt en binnen vooraf afgesproken kaders. Ook zal de uitvoerende partij naar verwachting een toets of beoordeling moeten doen op de haalbaarheid en

uitvoerbaarheid van het verzoek, bijvoorbeeld om vast te stellen of en hoe het technisch mogelijk is om de gevraagde domeinen te harvesten (en binnen welke termijn).

Om dit scenario uit te voeren zullen er ook nieuwe processen en technische voorzieningen moeten worden ingericht. De technische consequenties zijn dat de centrale instelling ook meer opslagruimte moet organiseren. Ook moet het opgeslagen materiaal beheerd worden. De opgehaalde data zullen uiteindelijk ook naar de verzoekende partij overgebracht worden, maar dit is volgens de gesproken partijen geen noemenswaardige inspanning en wordt daarom buiten beschouwing gelaten. Het loket zorgt er echter ook voor dat de drempel voor kleine partijen die wel willen harvesten maar daar niet de mensen en middelen voor hebben kleiner wordt. Zij kunnen gebruik maken van de techniek en kennis van de grote centrale partij.

5.2 Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

Net als in het eerste scenario is er in dit scenario één centraal gekozen instelling die op regelmatige basis de nationale domeincrawl uitvoert. Het verschil met het eerste scenario is dat decentrale archief- en erfgoedinstellingen ook beredeneerde selecties (kunnen) uitvoeren. Zij krijgen daarmee de mogelijkheid om collecties aan te leggen om online erfgoed te behouden op basis van beredeneerde selecties, zonder dat zij daarvoor afhankelijk zijn van een centrale instelling zoals in scenario 1. De beredeneerde selecties kunnen zo vaak als nodig geacht worden uitgevoerd op basis van hun eigen collectievormende processen.

5.2.1 Activiteiten

In dit scenario worden de volgende activiteiten uitgevoerd:

- De centrale instelling krijgt de taak van het uitvoeren van de nationale domeincrawl. Daar is dezelfde techniek, opslagruimte en personeel voor nodig als in het voorgaande scenario.
- Alle aangewezen of bevoegde instellingen voeren het proces van selectie en collectievorming uit voor webarchieven die binnen hun doel- en taakstelling passen.
- De instellingen die webdomeinen willen harvesten, hebben daarvoor de juiste techniek en expertise voor nodig. De instellingen die nu al aan webharvesting doen beschikken reeds over deze techniek en expertise.
 - Een aandachtspunt voor instellingen die (nog) niet aan webharvesting doen, is daarbij hoe schaalbaar de systemen en opslagruimte op langere termijn zijn. Met name de benodigde opslagruimte zal gestaag toenemen wanneer er meer geharvest wordt. Ook moet het opgeslagen materiaal beheerd worden. Hier wordt verder op ingegaan in paragraaf 6.3.
- Wanneer meerdere instellingen harvesten is het belangrijk om op een centrale plek bij te houden wie wat doet. Dit gebeurt nu al in het Nationaal Register Webarchieven⁴⁸.

⁴⁸ Het Nationaal Register Webarchieven is een initiatief van het Netwerk Digitaal Erfgoed. In het register wordt bijgehouden welke websites zijn gearchiveerd, door wie en wanneer dat gedaan is en welke software is gebruikt. Ook de reden voor het archiveren wordt vermeld. Er zijn ongeveer elf instellingen die het register gebruiken, waaronder de KB, Beeld en Geluid en een aantal regio- en stadsarchieven.

Bron: <https://www.registerwebarchieven.nl/>

5.2.2 Uitvoerbaarheid

In dit scenario heeft de centrale instelling de techniek en kennis nodig voor het uitvoeren van de nationale domeincrawl, maar hoeft er geen loketfunctie ingericht te worden. Voor de beredeneerde selecties zijn de uitvoeringslasten echter verdeeld over de partijen die ook daadwerkelijk harvesten. Daarmee hebben deze partijen meer vrijheid om zelf te harvesten en zijn ze daarbij niet afhankelijk van een centrale instelling (zoals in het eerste scenario). De keerzijde is dat de decentrale partijen zelf moeten beschikken over de kennis en techniek om de activiteiten uit te voeren. Er zullen partijen zijn voor wie dit niet mogelijk is, waardoor de drempel om aan webharvesting te doen hoog ligt.

5.3 Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)

Binnen het consensusmodel kan iedereen die dat wil aan webharvesting doen. Het scenario is gebaseerd op een maatschappelijke consensus die wordt gedeeld door archiefinstellingen en rechthebbenden. Hierbij is er sprake van impliciete toestemming en/of stilzwijgende verleende licenties, uitgaand van een analyse van inbreukrisico's en gepaste maatregelen om risico's te beperken door de instellingen, met als centrale elementen een embargoperiode en een efficiënt notice-and-takedown-systeem.

5.3.1 Activiteiten

Als onderdeel van het consensusmodel zien we de volgende activiteiten:

- Het maken van duidelijke afspraken tussen betrokken partijen, waaronder ook het vastleggen en beheren van deze afspraken. Dit vraagt een tijdsinvestering, zowel in ureninzet als doorlooptijd.
- Het inrichten van een notice-and-takedown-systeem voor webarchieven, inclusief embargoperiode.
- Het vaststellen van gezamenlijke protocollen en standaarden voor webharvesting, zodat partijen een uniforme aanpak kunnen hanteren die zowel de belangen van rechthebbenden als de publieke toegankelijkheid van informatie respecteren.
- Het organiseren van toezicht om de naleving van deze afspraken te waarborgen, bijvoorbeeld door een onafhankelijke toezichthouder aan te stellen.

5.3.2 Uitvoerbaarheid

De uitvoeringslast voor dit scenario zit vooral in het maken van afspraken. Dit kost inspanning en tijd van alle partijen, zowel de erfgoedinstellingen als de auteursrechthebbenden. Er zijn goede voorbeelden van het maken van afspraken tussen erfgoedinstellingen en auteursrechthebbenden, bijvoorbeeld op het gebied van out-of-commerce werken. Dergelijke afspraken zouden ook voor webdomeinen gemaakt kunnen worden, maar het is de vraag op welke schaal en met welke insteek dit kan worden gedaan. Met specifieke groepen van rechthebbenden zullen waarschijnlijk aparte afspraken gemaakt moeten worden. Bovendien moeten de gemaakte afspraken actief worden beheerd en doorontwikkeld. Hier moet iemand zich verantwoordelijk voor voelen en de coördinatie op zich nemen.

Daarnaast moet elke harvestende partij, net als bij het tweede scenario, beschikken over de benodigde kennis en techniek om te kunnen harvesten. Dat zal niet voor elke partij haalbaar zijn.

5.4 Status quo (geen afspraken maken)

Dit scenario betreft een benadering van autonome ontwikkeling richting de toekomst als er geen afspraken of wetgeving wordt gemaakt. Er vindt geen nationale domeincrawl plaats. Wel worden er door een aantal instellingen beredeneerde selecties uitgevoerd, zoals dat nu al gebeurt. Voordat de content geharvest wordt, geven de partijen de website-eigenaren de mogelijkheid tot een opt out. Wanneer een rechthebbende kiest voor een opt-out, zal het webdomein niet geharvest worden.

5.4.1 Activiteiten

Ten aanzien van de uit te voeren activiteiten gaan we in dit scenario uit van:

- Beredeneerde selectie en collectievorming door ongeveer tien tot twaalf erfgoedinstellingen die samen 26.000 domeinen archiveren, met een groei van ongeveer 2.500 domeinen per jaar.

5.4.2 Uitvoerbaarheid

Instellingen die momenteel al aan webharvesting doen, beschikken ook zelf over de techniek daarvoor.⁴⁹ Een aandachtspunt daarbij is hoe schaalbaar de huidige systemen en opslagruimte op langere termijn zijn. Met name de benodigde opslagruimte zal gestaag toenemen wanneer er meer geharvest wordt. Ook moet het opgeslagen materiaal beheerd worden.

Het proces rondom de opt out is tijdrovend en arbeidsintensief. Ook lukt het niet altijd om de rechthebbende(n) te traceren en een webpagina kan ook content van andere rechthebbenden bevatten waarvoor de website-eigenaar geen toestemming kan geven. De uitvoeringslast voor de decentrale instellingen is hoog, want er is personele inzet nodig om alle website-eigenaren aan te schrijven. Soms is er ook een gesprek nodig om extra toelichting te geven. Om deze redenen wordt er momenteel slechts een zeer beperkt aantal webdomeinen geharvest.

⁴⁹ Een uitzondering hierop is Tresoar. Zij nemen deel aan een pilot met de KB, waarbij Tresoar domeinnamen aanlevert aan de KB die deze vervolgens harvest. Tresoar heeft dus zelf geen technologie, maar maakt gebruik van de technische middelen en expertise van de KB.

6. Kosten

In dit hoofdstuk lichten we de kostenraming toe. De kostenraming is gebaseerd op de activiteiten die in Hoofdstuk 5 uiteen zijn gezet. Per scenario worden de kosten onderbouwd en toegelicht. Dit hoofdstuk beginnen we met een uiteenzetting van de uitgangspunten van de kostenraming, waarna we per scenario de kostenraming bespreken. We eindigen dit hoofdstuk met een vergelijking van de kosten tussen verschillende scenario's en twee gevoeligheidsanalyses die de invloed de afbakening van het nationale webdomein en de opslagkosten laten zien.

6.1 Uitgangspunten kostenraming

Het is belangrijk op te merken dat de kostenanalyse in dit hoofdstuk een raming betreft en geen exacte kostenberekening pretendeert te zijn. De analyse is primair bedoeld om inzicht te geven in de potentiële financiële impact van verschillende scenario's en biedt een hulpmiddel voor het afwegen van keuzes met betrekking tot ordegrrootte van kosten en bij welke partijen de kosten zullen landen. De daadwerkelijke kosten kunnen in de praktijk anders uitvallen, onder andere vanwege de aannames die gedaan zijn en vanwege de variaties in kosten, licenties en te maken keuzes die er in de uiteindelijke uitvoeringspraktijk zullen zijn.

Daarnaast zijn de volgende uitgangspunten van toepassing:

- In alle scenario's gaan we ervan uit dat slechts één partij de nationale domeincrawl uitvoert;
- De kosten van opslag blijven constant over de komende 10 jaar. Wij behandelen eerst de verschillende scenario's in aparte hoofdstukken, met de kosten per terabyte (TB) zoals wij die hebben opgehaald tijdens het onderzoek. In een afsluitende paragraaf doen wij een gevoeligheidsanalyse op de kostenontwikkeling van opslag op de totalen per scenario om de mogelijke ontwikkeling van kosten inzichtelijk te maken;
- De loonkosten worden geacht vast te blijven gedurende de looptijd van de analyse, net als de andere kostenposten, met uitzondering van de opslagkosten waarvoor een gevoeligheidsanalyse is opgenomen;
- Elke (versie van een) website heeft een gemiddelde benodigde opslag. Hier wijken wij alleen van af voor websites die ten behoeve van B&G worden geharvest (deze websites hebben zwaardere opslag);
- Voor de kostenraming nemen we aan dat de erfgoedinstellingen zoals opgenomen in Tabel 6.1 de enige instellingen zijn die beredeneerde selecties uitvoeren (in de praktijk kan dit afwijken);
- De personele inzet om de webarchieven voor beredeneerde selecties te selecteren (collectievorming) wordt buiten de analyse gehouden. Wij kwantificeren alleen de technische personeelskosten en personeel benodigd voor de loketfunctie;
- Geharvest materiaal is alleen in te zien op locatie van de instelling die harvest of de selectie heeft aangevraagd. Er zijn geen kosten geïnventariseerd voor andere vormen van het openbaar maken of presenteren van webarchieven dan nu al gebruikelijk is (de studiezaal/terminalexceptie).

6.2 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen

De centrale instelling heeft in dit scenario de technische functionaliteiten om de nationale domeincrawl en beredeneerde selecties uit te voeren.

Nationale domeincrawl

Voor de nationale domeincrawl hanteren we de afbakening van het nationaal webdomein zoals eerder gespecificeerd. De geschatte kosten voor deze crawl zijn geïnventariseerd op basis van gegevens van de KB. Uit deze gegevens blijkt dat de jaarlijkse opslagbehoefte voor de nationale domeincrawl ongeveer 150 TB bedraagt.

Het feit dat deze kostenberekening gebaseerd is op een inschatting van de KB, betekent dat deze niet één-op-één representatief is voor een andere organisatie. Stel dat er een nieuwe instelling wordt opgericht met als doel het uitvoeren van beredeneerde selecties met aanvraag via een loketfunctie, dan zouden de kosten hoger uitvallen dan getoond in dit hoofdstuk. De KB bouwt namelijk voort op jarenlange investeringen en expertise.

De nationale domeincrawl wordt uitgevoerd als een oppervlakkige crawl, waarbij slechts een beperkte hoeveelheid content per website wordt vastgelegd. Dit betekent dat doorgaans alleen de hoofdpagina en enkele gerelateerde pagina's worden opgeslagen, wat zorgt voor beheersbare opslagvereisten.

Aangezien het begrip 'Nederlands webdomein' dynamisch is en aan verandering onderhevig, kunnen er nog keuzes gemaakt worden met betrekking tot de exacte reikwijdte van de nationale domeincrawl (zie ook paragraaf 3.3). Om rekening te houden met deze onzekerheid, voeren we een gevoeligheidsanalyse uit om de effecten van variaties in het aantal te crawlen domeinen op de kosten en opslagbehoefte inzichtelijk te maken (zie paragraaf 6.7.1).

Activiteit	Kostenpost
Nationale domeincrawl	➤ Personeel: de personele inzet in FTE's die nodig is voor het voorbereiden, uitvoeren, controleren (inclusief kwaliteitscontrole) en beheren van de nationale domeincrawl en het geharveste materiaal. ➤ Techniek: de technische voorziening inclusief licenties die nodig is om de nationale domeincrawl uit te voeren. ➤ Opslag: de opslag van het geharveste materiaal conform de kwaliteitseisen die hieraan gesteld worden door de uitvoerende organisatie.

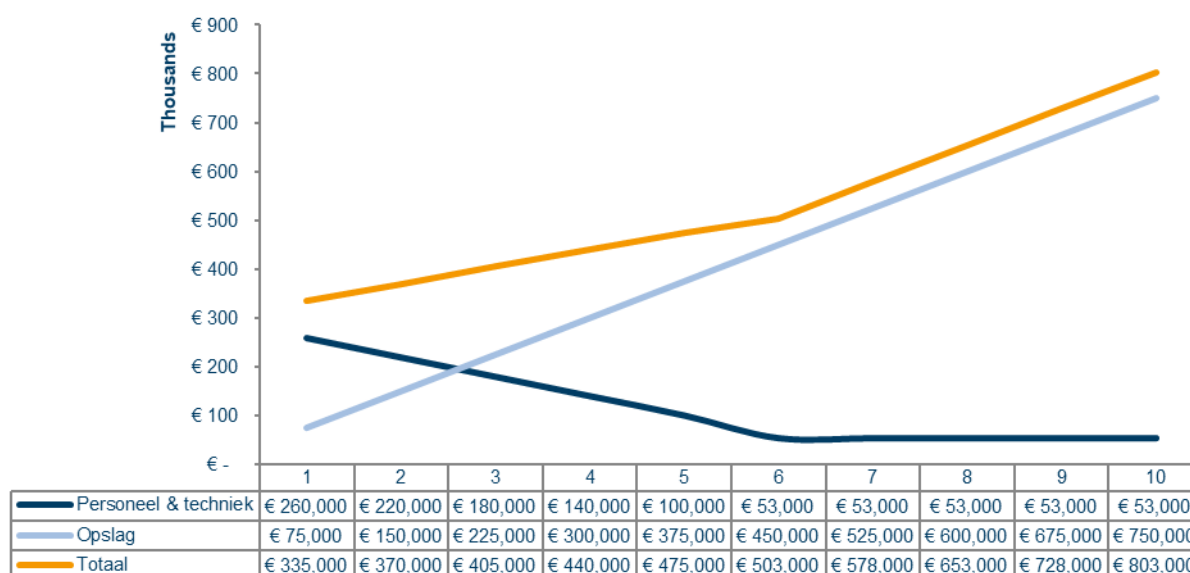
In alle uitgewerkte scenario's gaan we ervan uit dat één partij de nationale domeincrawl zal uitvoeren. De kosten van de nationale domeincrawl zijn geïnventariseerd op basis van de raming van de KB. De KB schat de technische kosten het eerste jaar, exclusief opslag, in op ongeveer € 260.000. Door efficiëntere integratie van werkzaamheden en toenemende productiviteit door opgedane ervaring van personeel wordt verwacht dat de technische kosten jaarlijks met € 40.000 zullen afnemen. Dit afnamepatroon zal doorgaan tot een jaarlijks onderhoudsniveau van 0,5 FTE, wat neerkomt op ongeveer € 50.000. De nationale domeincrawl wordt na een aantal jaren een bekend en repeterend proces, wat relatief weinig personele inzet meer behoeft volgens de raming van de KB.

Belangrijk om op te merken is dat de KB de investeringen in de benodigde technologie voor de domeincrawl al heeft gedaan, waardoor er geen extra opstartkosten voor de domeincrawl-infrastructuur nodig zijn. De geschatte opslagkosten van de nationale domeincrawl bedragen € 500 per TB volgens de KB. Dit is een relatief laag bedrag voor opslag bij webharvesting. In een financieel onderzoek van Improven wordt bijvoorbeeld gerekend met gemiddeld €1.128 per TB⁵⁰ en voor de berekende selecties hanteert de KB-opslagkosten van €3.800 per TB. De kosten per opslag hangen namelijk nauw samen met de kwaliteitseisen van opslag en hoe snel en makkelijk informatie weer opgevraagd kan worden. Met die context is de kosteninschatting van €500 per TB dus relatief aan de lage kant. Dit is te verklaren doordat er sprake is van relatief lage kwaliteitseisen omdat de nationale domeincrawl naar verwachting geen snelle of real-time beschikbaarheid vraagt. Dit maakt dat de kosten voor opslag lager zijn dan de kengetallen voor andere typen opslag (waaronder de gemiddelden die in andere onderzoeken gebruikt worden). Het kengetal dat wij gebruiken is gebaseerd op ons eigen onderzoek, en wijkt met de bovenstaande redenen af van het kengetal dat in het onderzoek van Improven is gebruikt.

Met een jaarlijkse data-opslagbehoefte van 150 TB, betekent dit een jaarlijkse kostenstijging van ongeveer € 75.000, aangezien alle data permanent bewaard blijft. Hierbij nemen wij in de hoofdanalyse nog aan dat de kosten van opslag per TB **niet** afnemen over tijd. In een gevoeligheidsanalyse gaan we in op de ontwikkeling van kosten van opslag (zie paragraaf 6.7.2).

Onderstaande figuur toont de ontwikkeling van de kostenposten over een periode van tien jaar.

Figuur 6.1: Kosten nationale domeincrawl alle scenario's



De eerste twee jaren zullen de personeelskosten de belangrijkste kostenpost vormen, daarna zal de opslag van de jaarlijkse domeincrawl de grootste kostenpost zijn. Vanaf jaar 6 nemen de totale jaarlijkse kosten toe met € 75.000 aan extra opslagkosten.

⁵⁰ Improven (2024). *Financieel onderzoek naar de digitale taken van de Koninklijke Bibliotheek*.

De belangrijkste factor die invloed heeft op de kosten is de omvang van de nationale domeincrawl. De reikwijdte van de domeincrawl, bepaald door de definitie van het nationaal webdomein, is de enige variabele die een substantiële impact heeft, aangezien de personeels- en technische kosten constant blijven. Dit verband tussen de reikwijdte en de kosten wordt verder toegelicht in de gevoeligheidsanalyse aan het einde van het hoofdstuk.

Concluderend worden totale kosten voor de nationale domeincrawl grotendeels bepaald door de permanente opslagbehoefte. In het laatste jaar van onze analyse (jaar 10) zijn de jaarlijkse kosten voor personeel en techniek dus €53.000 en de kosten voor opslag omstreeks €750.000.

Beredeneerde selectie

De centrale instelling zal naast de domeincrawl ook beredeneerde selecties uitvoeren. Voor deze beredeneerde selecties is aanvullende opslag, technische functionaliteiten en personele inzet nodig. Dit is enerzijds om het harvesten daadwerkelijk uit te voeren en de gegevens op te slaan, maar daarnaast ook om selecties te maken of aanvragen te behandelen die binnenkomen via de loketfunctie.

In het centrale scenario is er een centrale uitvoering van zowel de nationale domeincrawl als de beredeneerde selecties. In de berekening van het kostenplaatje gaan we er van uit dat er één organisatie als centrale instelling fungeert voor het harvesten van Nederlandse websites, met daarbij ook de activiteit om specifieke webcollecties samen te stellen op basis van beredeneerde selecties.

Dit betekent dat erfgoedinstellingen, die momenteel zelf webharvesting uitvoeren, hun taken op dit gebied via bij de centrale instelling onderbrengen. Eventuele frictiekosten (het overdragen van bestaand materiaal naar de centrale instelling) houden wij buiten beschouwing.

Wij rekenen wel aanvullende technische kosten voor het uitvoeren van de beredeneerde selecties, namelijk de gemiddelde technische kosten zoals deze ook in de andere scenario's worden gebruikt. Om dit hoofdstuk leesbaar te houden verwijzen wij naar de onderbouwing van deze aannames in paragraaf 6.3 en tabel 6.6.

De nationale domeincrawl zal jaarlijks worden uitgevoerd, terwijl beredeneerde selecties flexibel en op frequente basis kunnen plaatsvinden, afhankelijk van de specifieke behoeften van de erfgoedinstellingen.

Tabel 6.1: Overzicht relevante instellingen beredeneerde selecties

Soort instelling	Aantal
Koninklijke Bibliotheek	1
Beeld en Geluid	1
Regionaal Historische Centra	12
Stedelijke archieven	10
Totaal	24

Binnen dit centrale model zou de centrale instelling een speciaal team oprichten dat als nationaal loket fungeert voor alle deelnemende erfgoedinstellingen in Nederland. Dit team zou verantwoordelijk zijn voor het coördineren van aanvragen voor beredeneerde selecties en zou in samenwerking met de

betrokken instellingen de harvest van webpagina's op structurele wijze uitvoeren (tabel 6.2). Dit centrale loket waarborgt hoogwaardige opslag van geharvest materiaal. Om de werkdruk beheersbaar te houden en te zorgen voor een verantwoorde spreiding van de collecties, zal de centrale instelling kaders stellen voor het aantal en de soorten websites die jaarlijks kunnen worden geharvest. Deze kaders kunnen ook afspraken over financiën en aanleveringstermijnen bevatten.

Uit de gevoerde gesprekken is naar voren gekomen dat een mogelijke indeling van het team er uit kan zien als in tabel 6.2. Dit zal bestaan uit 0,5 fte voor een managementfunctie, 1 fte aan IT-specialisten en 3 fte voor het operationele loketpersoneel. Deze bezetting leidt tot een geschatte jaarlijkse kostenpost van € 524 duizend. Deze aanname is getoetst bij een aantal erfgoedinstellingen en als plausibel beoordeeld in verhouding tot de hoeveelheid beredeneerde selecties (zie Tabel 6.3: Aannames aantal toegestane websites per instelling per jaar centraal scenario) die aangenomen worden in dit scenario.

Tabel 6.2: Aannames team KB beredeneerde selecties – centraal scenario

Aannames	#	€ Jaarlijks gemiddelde
Manager ⁵¹	0,5 fte	€ 69 duizend
IT-personeel ⁵²	1 fte	€ 137 duizend
Personeel loketfunctie ⁵³	3 fte	€ 318 duizend
<i>Totaal</i>	<i>4,5 fte</i>	<i>€ 524 duizend</i>

De opslagbehoefte binnen dit model groeit in vergelijking met de nationale domeincrawl niet lineair, maar (semi-)exponentieel. Dit komt doordat sommige websites slechts eenmaal worden geharvest en opgeslagen, terwijl andere websites elk jaar opnieuw worden geharvest, afhankelijk van de waarde die instellingen aan specifieke collecties toekennen. Naar aanleiding van de gesprekken met de verschillende betrokken instellingen is een inschatting gemaakt van de jaarlijkse behoefte tot beredeneerde webcrawls voor de verschillende instellingen.

Voor de KB is uitgegaan van een blijvende behoefte van ongeveer 1.000 nieuw geharveste websites per jaar, met een gemiddelde opslag van 0,5 GB⁵⁴ per website. Op dit moment harvest de KB ongeveer 2.000 websites per jaar, maar wanneer er ook een nationale domeincrawl zal plaatsvinden zal deze behoefte afnemen naar 1.000 websites per jaar.

Voor Beeld en Geluid ligt deze schatting in hoeveelheid websites lager, namelijk op 100 websites per jaar. De gemiddelde opslag per website zal door de aard van het geharveste materiaal (o.a. audiovisueel) echter veel hoger liggen en is aangenomen op 3,5 GB per versie van een website.⁵⁵

Voor regionale historische centra is aangenomen dat er per centrum jaarlijks gemiddeld 100 websites geharvest zal worden, met dezelfde gemiddelde opslag per website als voor de KB.

⁵¹ Handleiding overheidsstarieven schaal 12, Gemiddelde totale kosten

⁵² Idem manager.

⁵³ Handleiding overheidsstarieven schaal 10, Gemiddelde totale kosten

⁵⁴ Gemiddelde afkomstig van KB

⁵⁵ Gemiddelde afkomstig van B&G

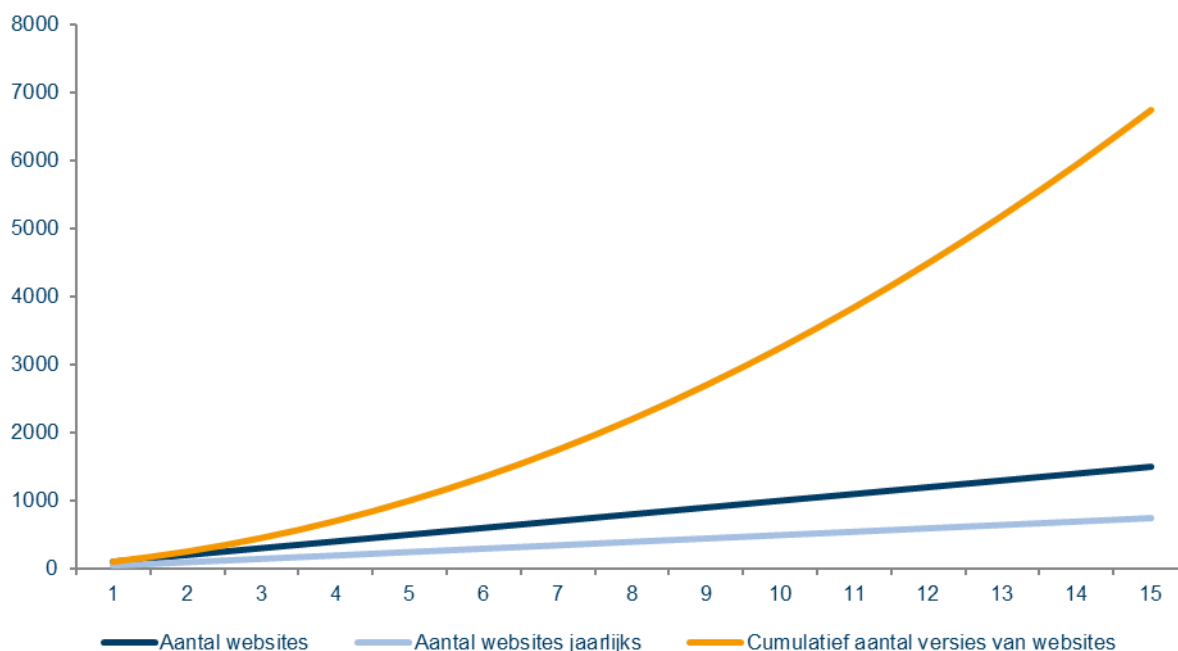
Tabel 6.3: Aannames aantal toegestane websites per instelling per jaar centraal scenario

Instelling	# jaarlijks nieuw opgeslagen websites	Gemiddelde opslag
KB	1000	0,5 GB per versie
B&G	100	3,5 GB per versie
Overige instellingen (22)	100	0,5 GB per versie
Totaal per jaar	3300	-

In dit kostenmodel wordt aangenomen dat de helft van de websites jaarlijks opnieuw wordt geharvest en de andere helft slechts eenmalig wordt opgeslagen. Hoewel dit model een vereenvoudigde weergave biedt, geeft het een inschatting van de exponentiële opslaggroei op lange termijn die specifiek is dan een lineair model.

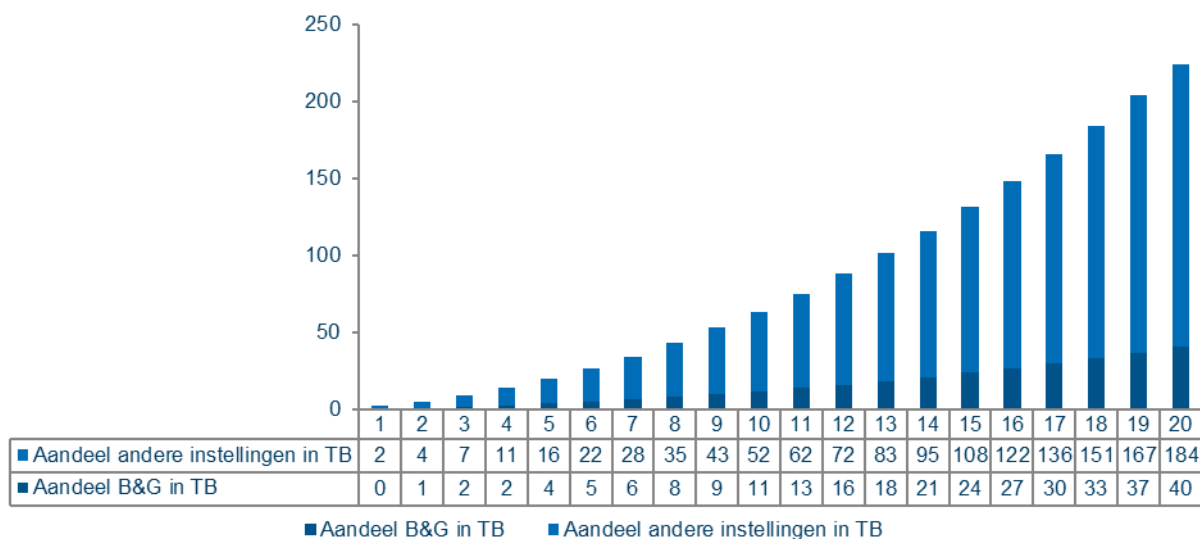
Onderstaande figuur laat het verloop zien voor een theoretische casus waarbij een regionaal archief vanaf de initiële start 100 websites per jaar geharvest worden door middel van een beredeneerde selectie. Hierbij gaan we dus uit van de bovenstaande aanname dat de helft van de jaarlijks nieuw geharveste websites vanaf het moment van de eerste webcrawl op de lange termijn jaarlijks geharvest wordt. Door de exponentiële groei van het aantal opgeslagen versies van deze websites, zal na 10 jaar het aantal verschillende versies zijn uitgebreid tot ongeveer 3.300.

Figuur 6.2: Schematische modelering verloop geharveste websites en versies



Wanneer deze aantallen gecombineerd worden met de aanname dat opslag bij de KB voor beredeneerde selecties ongeveer €3,80 per GB kost (afgeleid uit de stukken van de KB), komen wij uit op een verloop in opslag als zichtbaar in figuur 6.3. De totale kostenpost voor opslag over 10 jaar zou dan €1,0 miljoen zijn, waarvan €0,2 miljoen in jaar 10. Over 20 jaar zijn de totale kosten €6,5 miljoen, met €0,9 miljoen in jaar 20. De kosten nemen dus exponentieel toe over tijd.

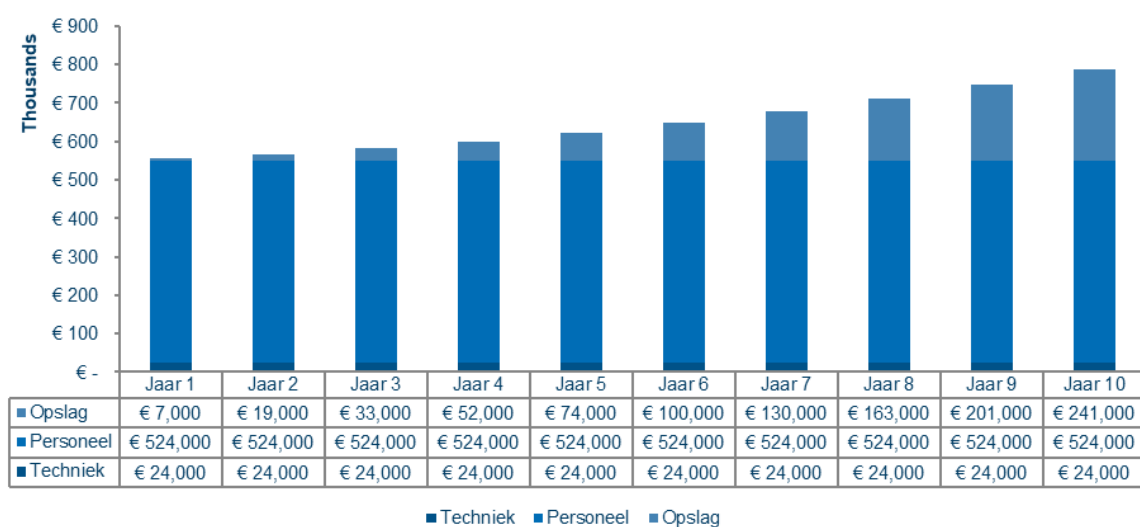
Figuur 6.3: Verloop opslag beredeneerde selecties centraal scenario



In het centrale scenario zal er dus een gespecialiseerd team werkzaam zijn bij de centrale instelling. Erfgoedinstellingen maken in dit scenario dan ook geen eigen technische kosten en opslagkosten, en kunnen (een aantal) crawls per jaar laten uitvoeren door deze aan te vragen bij de centrale instelling. In deze beschrijving van het scenario kijken wij alleen naar te maken technische kosten en gaan nemen wij aan dat het deel selectie van websites onderdeel is van de reguliere collectievorming.

Door het maximum aan FTE dat beschikbaar wordt gesteld bij de centrale instelling, blijven de personeelskosten over de jaren constant. Alleen de opslagkosten stijgen exponentieel, zoals eerder toegelicht. Dit resulteert in een jaarlijkse kostenraming dat er als volgt uitziet:

Figuur 6.4: Kosten beredeneerde selectie centrale uitvoering



On-site beschikbaar maken van materiaal

Uit de gesprekken blijkt dat de kosten voor het presenteren van het geharveste materiaal bij zowel de centrale als decentrale bevoegdheden naar verwachting beperkt zullen zijn. Hoewel dit betekent dat het geharveste materiaal op meerdere locaties on-site beschikbaar wordt gesteld en de kosten voor presentatie daardoor enigszins kunnen oplopen, is gebleken dat deze kosten geen substantiële impact hebben op de totale kostenraming. Daarom laten we de kosten voor beschikbaarstelling van het materiaal in deze analyse buiten beschouwing.

6.3 Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

In dit scenario zijn er zowel voor de centrale instelling als voor de decentrale instellingen kosten voor harvestingsoftware en voor opslagruimte. Daarnaast is ook elke bevoegde instelling zelf bezig met het selecteren van relevante online content ten behoeve van harvesting. Wij kijken in deze raming alleen naar de kosten van techniek, opslag en technisch personeel.

Nationale domeincrawl

Deze kosten verschillen niet met het voorgaand beschreven scenario.

Beredeneerde selecties

In het decentrale scenario voeren erfgoedinstellingen, zoals aangegeven in tabel 6.2, zelfstandig de beredeneerde selecties uit. Dit betekent echter ook dat elke instelling zelf kosten moet maken om de technische faciliteiten en het technisch personeel in te huren om deze beredeneerde selecties uit te voeren. Uit de gesprekken blijkt dat het lastig is in te schatten hoeveel instellingen bereid en in staat zullen zijn om deze drempel te nemen, als zij zelf financieel verantwoordelijk zijn.

Tijdens de gevoerde gesprekken werd duidelijk dat er diverse abonnementsvormen voor webharvesting beschikbaar zijn, elk met een eigen prijsstructuur voor licenties en opslag. Sommige abonnementen kennen een hoger licentiebedrag met lagere opslagkosten per geharveste website, terwijl andere juist lagere licentiekosten hebben en hogere opslagkosten berekenen. De keuze van het type abonnement en de frequentie van webcrawls hebben daarom een directe invloed op de jaarlijkse kosten voor een erfgoedinstelling. Instellingen die op grotere schaal willen harvesten, zouden wellicht baat hebben bij abonnementen met lagere opslagkosten, terwijl instellingen die slechts een beperkt aantal crawls uitvoeren waarschijnlijk beter af zijn met een abonnement met lagere licentiekosten.

Omdat alle instellingen autonoom zijn in hun beslissing om te webharvesten, bepaalt elke instelling zelf hoe vaak en hoeveel er wordt geharvest, afhankelijk van de eigen collectiebehoeften en financiële mogelijkheden. In de praktijk is de belangrijkste beperkende factor de financiële drempel per instelling, in combinatie met het collectiebeleid.

Op basis van de gesprekken onderscheiden we twee modellen die momenteel worden gebruikt om webharvesting uit te voeren. We gaan er vanuit dat beide modellen dezelfde capaciteiten hebben als het aankomt op kwaliteit van opgeslagen websites. Voor een consistente vergelijking met het centrale scenario hanteren we het gemiddelde van deze twee modellen. In model 1 bedragen de jaarlijkse licentiekosten circa €35.000, waarbij opslagkosten tot een bepaald punt gratis zijn en daarna geleidelijk toenemen. Voor de eenvoud gaan we ervan uit dat opslagkosten in dit model verwaarloosbaar zijn, omdat de gesproken partij aangaf dat zij op dit moment geen meerkosten

betalen voor een toename in benodigde opslag. Dit zit nu nog inbegrepen in de jaarlijkse licentiekosten en de verwachting was dat dit niet in de komende jaren zal veranderen.

In model 2 zijn de jaarlijkse licentiekosten lager, ongeveer €12.500, maar bedragen de opslagkosten €25,00 per GB.⁵⁶ Beide modellen rekenen met een gemiddelde personeelsinzet van 0,2 FTE, zoals uit de gesprekken naar voren kwam.

Tabel 6.4: Kostenaannames twee varianten webharvesting software

	Model 1	Rekengemiddelde	Model 2
Jaarlijkse licentiekosten	€35.000 per jaar	€23.750 per jaar	€12.500 per jaar
Opslagkosten	Geen ⁵⁷	€3,80 per GB ⁵⁸	Hoog: €25,00 per GB
Technisch personeel⁵⁹	0.2 fte	0.2 fte	0.2 fte

Voor het rekengemiddelde nemen we het gemiddelde van beide licentiekosten en hanteren we een opslagprijs van €3,80 per GB, gebaseerd op de gemiddelde opslagkosten voor een selectie bij de KB. De aannames over het aantal websites dat jaarlijks per instelling wordt gecrawld zijn gelijk aan die in het centrale scenario, aangezien we veronderstellen dat de behoefte aan het aantal geharveste websites niet scenario-afhankelijk is.

Tabel 6.5: Aantal geharveste websites per jaar – decentraal scenario

Instelling	# jaarlijks nieuw opgeslagen websites	Gemiddelde opslag
KB	1000	0,5 GB per versie
B&G	100	3,5 GB per versie
Overige instellingen (22)	100	0,5 GB per versie

Figuur 6.5 toont de mogelijke nationale kosten voor webharvesting bij decentrale instellingen, gebaseerd op het aantal geharveste websites per jaar (zie ook tabel 6.3). De x-as geeft aan hoeveel erfgoedinstellingen voor het gebruik van beredeneerde webcrawls zouden kiezen. De jaarlijkse kosten worden aan de rechterzijde van de figuur weergegeven voor alle deelnemende partijen, terwijl de situatie waarin er slechts twee instellingen webharvesting uitvoeren aan de linkerzijde zichtbaar is. De berekeningen voor deze figuur zijn gebaseerd op de gegevens uit tabel 6.4 en tabel 6.5. Hierbij hanteren regionale erfgoedinstellingen de kostenstructuur zoals beschreven in tabel 6.4, terwijl de twee initiële instellingen eenzelfde opslagkosten per GB aanhouden als de KB en B&G en geen extra licentiekosten hebben, aangezien zij al actief zijn in webharvesting.

⁵⁶ Dit bedrag is erg hoog in vergelijking met de €0,50 per GB waarmee bij de nationale domeincrawl gerekend wordt. Deze verschillen in prijzen per opslag suggereren enerzijds marktwerking, maar de wijze waarop gegevens opgeslagen worden, bijvoorbeeld hoe makkelijk het is om informatie op te halen, zijn bepalend voor de kosten.

⁵⁷ In dit abonnement waren de abonnementskosten inclusief de opslag, tot een bepaalde (onbekende) hoogte.

⁵⁸ Zelfde bedrag als de KB hanteert bij beredeneerde selecties.

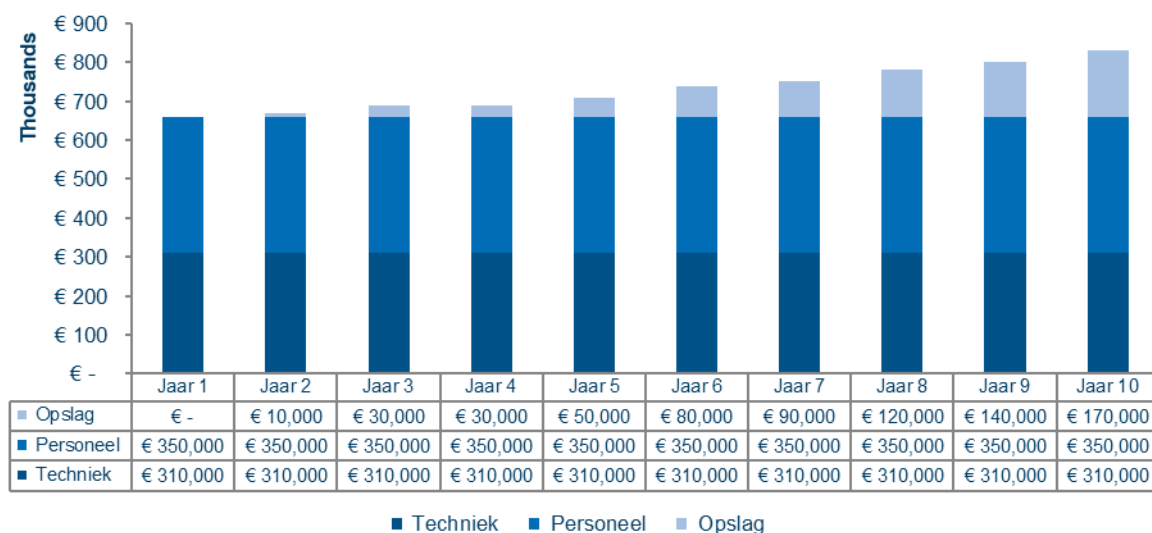
⁵⁹ Handleiding overheidsstarieven schaal 12, Gemiddelde totale kosten

Figuur 6.5: Gemiddelde jaarlijkse kosten decentrale instellingen (10 jaar)



Indien we een scenario uitwerken dat tussen de twee uiterste scenario's ligt, bijvoorbeeld met 13 instellingen die beredeneerde selecties zullen uitvoeren, ziet de kostenontwikkeling over een periode van 10 jaar er als volgt uit:

Figuur 6.6: Kosten beredeneerde selectie decentrale uitvoering, Midden-scenario: 13 instellingen



Ook in dit scenario is te zien dat de opslagkosten (in lichtblauw) exponentieel toenemen.

Om meer inzicht te bieden in de mogelijke variaties binnen het decentrale scenario, laten we drie situaties in tabelvorm zien: het maximumscenario (alle instellingen voeren webharvesting uit), het minimumscenario (twee instellingen met eenzelfde kostenstructuur als Beeld en Geluid en de KB voeren webharvesting uit) en een gemiddeld scenario (13 instellingen voeren webharvesting uit). De

hoeveelheid websites die wordt geharvest volgt uit de hoeveelheid instellingen die harvesten en de aannames van hoeveelheid websites die per instelling per jaar geharvest worden.

Tabel 6.6: Raming gemiddelde jaarlijkse kosten bereedeneerde selecties (10 jaar), decentraal scenario

	Min	Midden	Max
	1.100 websites per jaar	2.200 websites per jaar	3.300 websites per jaar
	2 instellingen	13 instellingen	24 instellingen
Techniek	€48 duizend	€309 duizend	€570 duizend
Personeel	€55 duizend	€356 duizend	€658 duizend
Opslag	€45 duizend	€73 duizend	€102 duizend
Totaal	€147 duizend	€738 duizend	€1.330 duizend

On-site beschikbaar maken van materiaal

Hierbij geldt hetzelfde als in het voorgaande scenario: We laten de kosten voor beschikbaarstelling van het materiaal in deze analyse buiten beschouwing.

6.4 Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)

Nationale domeincrawl

Deze uitvoeringskosten voor de nationale domeincrawl verschillen niet met de voorgaand beschreven scenario's. Echter, is de notice-and-takedown voorziening ook voor de nationale domeincrawl relevant. Hiervoor is dus aanvullende personele inzet nodig.

Er is nog geen inzicht in hoe een notice-and-takedown voorziening vormgegeven wordt. Het dichtstbijzijnde kengetal dat beschikbaar is komt van de KB, wie aangeeft dat op dit moment ongeveer 0,2 fte nodig is voor opt-out aan de voorkant in de huidige situatie. Ter verduidelijking, opt-out betekent dat alle partijen vooraf benaderd worden om toestemming te vragen, terwijl een notice-and-takedown voorziening de mogelijkheid biedt voor die partijen om achteraf bezwaar te maken als een partij haar gegevens niet wil laten crawlen. In de praktijk zal een opt-out voorziening dus veel meer tijd en moeite met zich meebrengen; er is immers correspondentie vooraf en met alle relevante partijen.

Het .nl-domein bevat meer dan 6 miljoen verschillende websites. Wanneer wij het kengetal van de KB van 0,2 fte per 1.000 geharveste websites hanteren voor een nationale domeincrawl, zou dit een jaarlijkse kostenpost betekenen van 1.200 fte, oftewel tientallen miljoenen. Dit bedrag is niet realistisch en dit is niet uitvoerbaar in de praktijk. Bovendien wordt de nationale domeincrawl jaarlijks uitgevoerd, wat betekent dat bedrijven, personen en instellingen die hier bezwaar tegen hebben dat niet elk jaar achtereenvolgens zullen (hoeven) doen. Daarnaast zal er niet voor elke website in het .nl domein een aanvraag bij de notice-and-takedown voorziening gemaakt worden.

Contactpersonen van het Portugese Arquivo.pt geven aan dat zij nauwelijks problemen ondervinden met de domeincrawl. In Portugal wordt een consensusmodel gehanteerd met 'notice-and-takedown'.

procedure en vindt ook een nationale domeincrawl plaats. Er wordt aangegeven dat het 'notice-and-takedown'-proces slechts een minimale belasting vormt, met afgelopen jaar slechts twaalf verzoeken die weinig tijd in beslag hebben genomen.

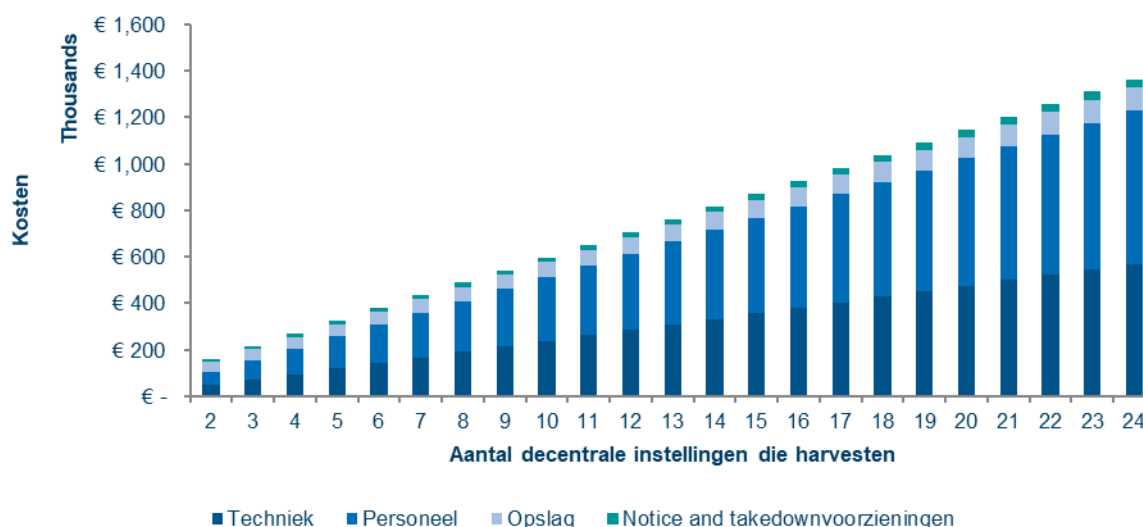
De kostenpost kan naar verwachting echter nog steeds aanzienlijk zijn, omdat de Portugese situatie niet direct te vertalen is naar de Nederlandse situatie. Omdat het niet mogelijk is een exacte raming te maken, waarderen wij deze post met P.M. +. De post kan met concrete plannen of invulling van de vormgeving van een dergelijk voorziening concreter gemaakt worden. Het is onduidelijk hoe een notice-and-takedown voorziening er op deze schaal uit zal zien, hoeveel aanvragen er zullen zijn en hoe deze kosten zich over de tijd zullen ontwikkelen. Echter laten de indicatieve berekening van 0,2 fte per 1.000 websites (1.200 fte tientallen miljoenen op jaarbasis) zien dat deze kosten hoog uit *kunnen* vallen.

Beredeneerde selecties

Binnen het consensusmodel heeft iedere organisatie de mogelijkheid om aan webharvesting te doen. Onze analyse benadert webharvesting vanuit de doelstelling om digitaal erfgoed duurzaam te behouden voor toekomstige generaties (zie ook paragraaf 1.1). Daarom veronderstellen wij dat erfgoedinstellingen, als belangrijkste belanghebbenden bij webharvesting, de primaire uitvoerders van webharvesting zullen zijn. Deze benadering resulteert in een situatie vergelijkbaar met het decentrale scenario.

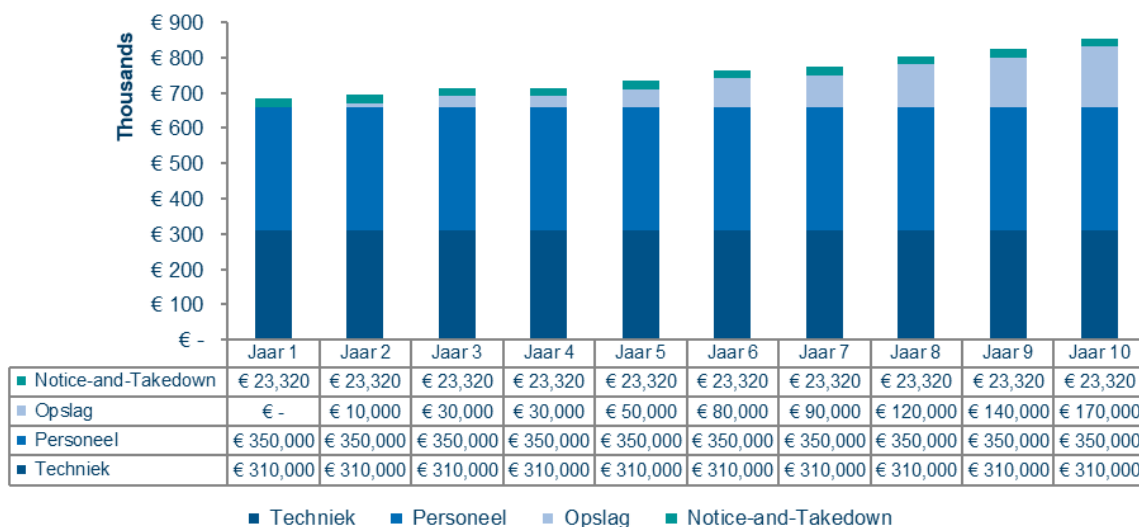
In dit decentrale scenario hanteren we dezelfde uitgangspunten, maar we maken wel gebruik van het opt-out kengetal van de KB omdat dit het enige realistische is kengetal dat beschikbaar is. Als wij deze kostenpost toepassen, leidt dit tot aanvullende kosten in vergelijking met figuur 6.5, zoals gepresenteerd in figuur 6.7. Voor de berekening van de cijfers in figuur 6.7 verwijzen wij naar de methodiek die is toegepast in scenario 2.

Figuur 6.7: Gemiddelde jaarlijkse kosten per aantal harvestende decentrale instellingen (10 jaar), scenario consensusmodel



De genoemde kosten vormen een gemiddelde van de jaarlijkse uitgaven over een periode van 10 jaar. Als we een scenario beschouwen dat zich tussen de twee uiterste scenario's bevindt, bijvoorbeeld met 13 instellingen die gebruikmaken van webharvesting, ziet de kostenontwikkeling over deze 10-jarige periode er als volgt uit:

Figuur 6.8: Kosten berekende selectie consensusmodel, over 10 jaar



Om de mogelijke variaties in het consensusmodel verder te verduidelijken, lichten we drie situaties wederom in tabelvorm toe: het maximumscenario (alle instellingen voeren webharvesting uit), het minimumscenario (twee instellingen met eenzelfde kostenstructuur als Beeld en Geluid en de KB voeren webharvesting uit) en het gemiddelde scenario (in totaal 13 instellingen voeren webharvesting uit).

Tabel 6.7: Raming jaarlijkse kosten berekende selecties, scenario consensusmodel

	Min	Midden	Max
	1.100 websites per jaar	2.200 websites per jaar	3.300 websites per jaar
	2 instellingen	13 instellingen	24 instellingen
Techniek	€48 duizend	€309 duizend	€570 duizend
Personeel	€55 duizend	€356 duizend	€658 duizend
Opslag	€45 duizend	€73 duizend	€102 duizend
Notice and takedown	€12 duizend	€23 duizend	€35 duizend
Totaal	€158 duizend	€762 duizend	€1.365 duizend

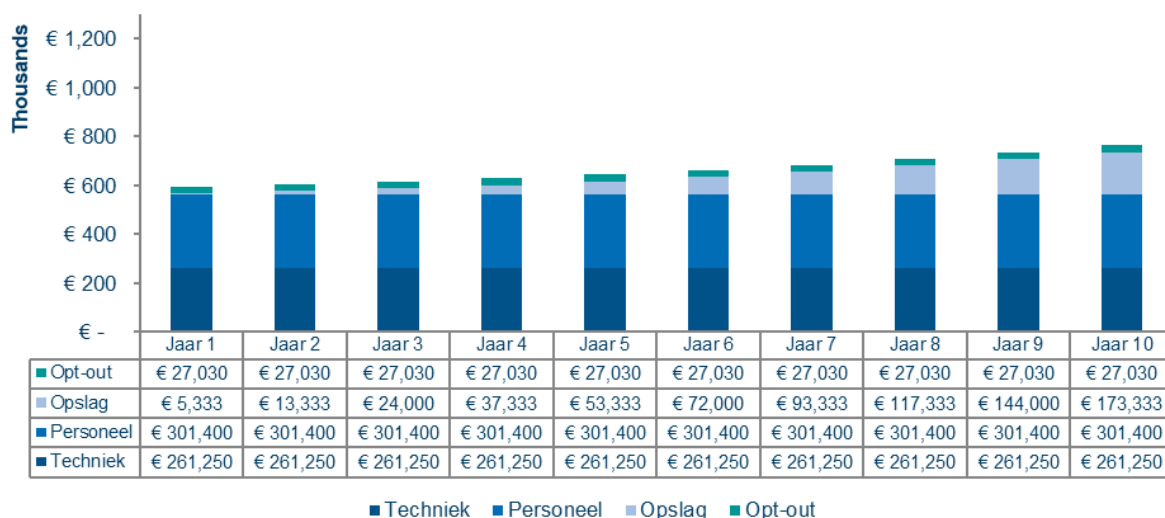
6.5 Status quo (geen afspraken maken)

In het nulscenario wordt uitgegaan van een situatie waarin er geen nationale domeincrawl plaatsvindt. Dit betekent dat de focus op de beredeneerde selecties ligt, waarin jaarlijks circa 2.500 nieuwe websites worden geharvest.

Wij modelleren dat dit wordt gedaan door 11 erfgoedinstellingen, waarbij de KB verantwoordelijk is voor 2.000 websites per jaar, B&G 100 websites per jaar en de overige regionale instellingen ieder 50 websites per jaar harvesten. Voor de uitvoering van de 'opt-out'-regeling wordt uitgegaan van dezelfde benodigde inzet van 0,2 FTE per 1.000 websites per jaar als dat de KB hanteert. Voor de personeels- en technische kosten worden dezelfde aannames gehanteerd als in het consensusmodel en het decentrale scenario.

Dit is een modelmatige weergave en dus geen één-op-één weergave van de kosten die nu gemaakt worden door de desbetreffende instellingen. In de praktijk zijn er ook partijen die nu "kosteloos" harvesten omdat zij dit samen met de KB organiseren. Ook kunnen abonnement structuren afwijken van de aannames die wij hanteren.

Figuur 6.9: Kosten beredeneerde selectie nulscenario



In lijn met de andere scenario's zullen de opslagkosten in het nulscenario exponentieel toenemen. Dit is het gevolg van de groeiende hoeveelheid gearchiveerd materiaal die blijvend moet worden opgeslagen.

Om deze bedragen in context te plaatsen, delen wij een aantal cijfers en inzichten die wij tijdens de interviews hebben opgehaald. De KB geeft aan op dit moment ongeveer 6 fte in dienst te hebben die zich bezighouden met het opslaan van alle digitaal materiaal, wat grofweg een kostenpost van €550.000 per jaar betekent. Dit gaat om de ontwikkeling van software (niet meegenomen in ons kostenmodel), maar ook investeert de KB in kwaliteitscontrole, kennisdeling en -ontwikkeling en het ondersteunen van regionale archieven door middel van diverse onderzoeksprojecten. Projecten als De Digitale Stad, de Web Curator Tool (ontwikkeld in samenwerking met de Nationale Bibliotheek van Nieuw-Zeeland) en het beheer van het Nationaal Register Webarchieven valt hieronder. Naar schatting heeft de KB nu zo'n 23.000 websites geharvest.

Recent onderzoek van Improven onderstreept dat de kosten van de webcollectie en webarchivering slechts een deel vormen van de totale kosten van de KB en voor de kosten voor digitale collectie.⁶⁰ Zelfs wanneer een mogelijke nationale domeincrawl wordt meegenomen, komt uit het onderzoek naar voren dat de kosten voor webarchivering een beperkt onderdeel zijn van de totale kostenbalans. De totale meerkosten van digitale taken richting 2030 (aanvullend op de huidige kosten voor de digitale collectie) worden geraamd op €16,9 miljoen, waarbij het gaat om meer kosten dan de kosten voor webharvesting.

De overige partijen die harvesten zijn diverse erfgoed- en archiefinstellingen. Zij hebben naar schatting in totaal zo'n 2.000 websites geharvest. De kosten van het webcrawlen zijn op dit moment hoog, omdat de licentie voor de techniek en opslag vaak via een commerciële partij loopt en de instellingen veel tijd kwijt zijn met opt-out procedures, in verhouding tot de hoeveelheid websites die wordt opgeslagen.

6.6 Vergelijking kosten verschillende scenario's

In dit hoofdstuk vergelijken we de kosten van de vier verschillende scenario's. Wij doen dit door middel van een overzichtstabel.

Tabel 6.8: Overzichtstabel gemiddelde jaarlijkse kosten alle scenario's (voor 10 jaar, in duizenden euro's)
– voor het decentrale scenario en consensus scenario laten wij de midden-variant zien omdat het niet duidelijk is hoeveel instellingen daadwerkelijk gaan harvesten

Scenario:	1. Centraal	2. Decentraal	3. Consensus	4. Status-quo
Nationale domeincrawl				
Techniek en personeel	€ 117	€ 117	€ 117	€ -
Opslag	€ 413	€ 413	€ 413	€ -
Notice and takedown	€ -	€ -	P.M. +	€ -
<i>Totaal nationale domeincrawl</i>	€ 529	€ 529	€ 529	€ -
<i>Hoeveelheid geharvest materiaal p.j.</i>	<i>150 TB per jaar</i>	<i>150 TB per jaar</i>	<i>150 TB per jaar</i>	<i>0 TB per jaar</i>
Beredeneerde selecties				
Techniek	€ 24	€ 309	€ 309	€ 261
Personeel	€ 524	€ 356	€ 356	€ 301
Opslag	€ 102	€ 73	€ 73	€ 73
Notice and takedown/opt-out	€ -	€ -	€ 23 ⁶¹	€ 29
<i>Totaal beredeneerde selectie</i>	€ 649	€ 738	€ 761	€ 665
<i>Hoeveelheid instellingen die harvesten</i>	<i>24</i>	<i>13</i>	<i>13</i>	<i>11</i>

⁶⁰ Improven (2024). *Financieel onderzoek naar de digitale taken van de Koninklijke Bibliotheek*.

⁶¹ Dit betreft een kostenraming voor notice-and-takedown voor beredeneerde selecties. De notice-and-takedown kosten voor nationale domeincrawl betreft een p.m. post.

Hoeveelheid geharvest materiaal (beredeneerde selecties) ⁶²	3.300 websites per jaar (1.000 KB)	2.200 websites per jaar (1.000 KB)	2.200 websites per jaar (1.000 KB)	2.550 websites per jaar (2.000 KB)
Totaal				
<i>Totale kosten</i>	€ 1178	€ 1267	€ 1290 +	€ 665

Nationale domeincrawl

De nationale domeincrawl zal naar verwachting de hoogste kosten hebben in het consensusmodel, omdat er dan een notice-and-takedown voorziening gerealiseerd moet worden. Gezien de grote schaal van de nationale domeincrawl, kunnen deze kosten erg hoog oplopen, waardoor de uitvoerbaarheid onder druk staat. Wij waarderen deze kostenpost nu met een P.M. +, omdat de precieze vorm van een dergelijke voorziening nog niet is uitgekristalliseerd.

De kosten voor de nationale domeincrawl zijn gelijk in scenario 1 en 2. In het nulscenario zijn er geen kosten omdat er voor dit scenario wordt aangenomen dat er geen nationale domeincrawl plaatsvindt (zoals dat nu ook in de praktijk niet wordt gedaan).

Beredeneerde selecties

Als alle decentrale instellingen beredeneerde selecties gaan uitvoeren, is het centrale scenario economisch het meest voordelig (het nulscenario daargelaten). Dit komt door de economische schaalvoordelen van het harvesten met een centrale voorziening, waardoor in plaats van 24, slechts één voorziening met personeel wordt opgetuigd.

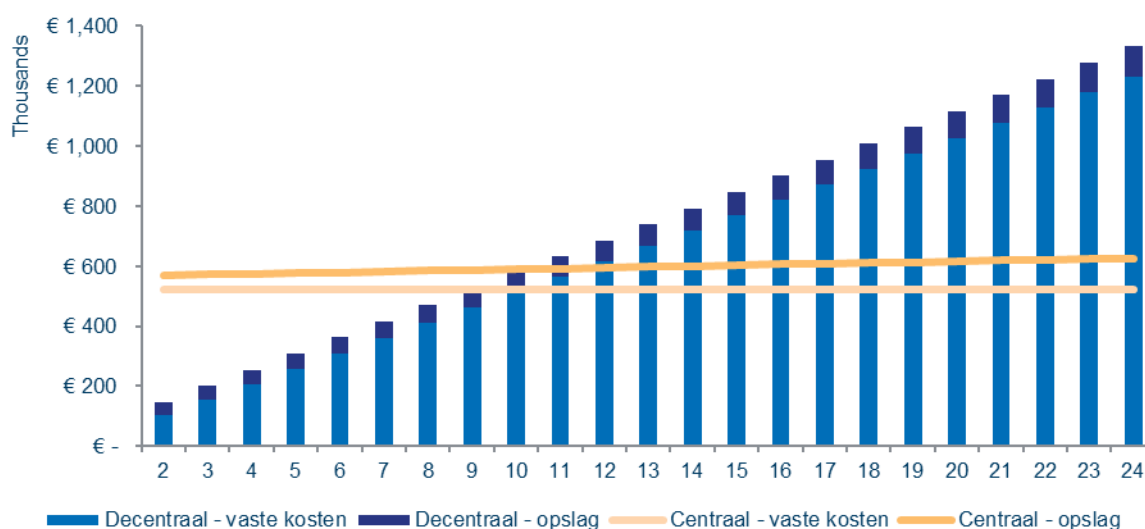
In onze analyse hebben we de decentrale uitvoering en het consensusmodel grotendeels als gelijkwaardig beschouwd, met als enige verschil dat het consensusmodel een hogere kostenpost met zich meebrengt vanwege de additionele 'notice-and-takedown'-voorziening. Hierdoor is het consensusmodel in onze benadering financieel gezien in alle verschillende situaties het minst aantrekkelijk in vergelijking tot de hoeveelheid geharvest materiaal.

De eerste overzichtstabel 6.8 laat de kosten voor alle scenario's zien, waarbij wij aannemen dat in het consensus en decentraal scenario elf van de regionale instellingen zelfstandig gaat harvesten. Zoals in dit hoofdstuk beschreven kan de financiële drempel ervoor zorgen dat niet alle decentrale instellingen ervoor kiezen om beredeneerde selecties uit te gaan voeren. Dit betekent dus dat er mogelijk meer of minder kosten gemaakt worden, maar ook dat er ook meer of minder materiaal geharvest wordt.

In de onderstaande vergelijking analyseren we de totale kosten, bestaande uit vaste kosten (personeels- en technische kosten) en opslagkosten, voor het centrale en decentrale scenario betreffende de beredeneerde selecties. De grafiek toont aan de meest rechterzijde de situatie waarin alle erfgoedinstellingen deelnemen aan de beredeneerde webcrawls, en aan de meest linkerzijde de situatie waarin uitsluitend twee instellingen deelnemen.

⁶² In de huidige situatie nemen we aan dat er *meer* beredeneerde selecties uitgevoerd zullen worden, omdat de KB in scenario 1 en 2 minder beredeneerde selecties uitvoert vanwege het uitvoeren van de nationale domeincrawl.

Figuur 6.10: Confrontatie tussen gemiddelde jaarlijkse kosten incl. opslag decentraal en centraal scenario bereedeneerde selecties (10 jaar)



Figuur 6.10 laat de kostenvergelijking zien tussen het centrale scenario en het decentrale scenario voor de gemiddelde jaarlijkse kosten over een periode van 10 jaar. Bij het centrale scenario blijven de vaste kosten voor de centrale instelling stabiel, aangezien de kosten voor de beschikbare fte-capaciteit constant blijven. De kosten stijgen echter licht bij elke nieuwe erfgoedinstelling die zich aansluit bij de centrale loketfunctie. Dit betekent dat de kostenstijging in het centrale scenario relatief beheerst blijft.

In het decentrale scenario zien we een andere kostenontwikkeling. Hier vormen de vaste kosten, die in grote mate worden beïnvloed door de licentiekosten, de grootste component van de totale kosten. Aangezien elke erfgoedinstelling in het decentrale scenario een eigen licentie moet afsluiten, nemen de totale kosten snel toe wanneer meer instellingen deelnemen. Dit leidt ertoe dat vanaf het moment dat elf instellingen betrokken zijn, de totale kosten van het decentrale scenario de kosten van het centrale scenario overstijgen.

Samenvattend biedt het centrale scenario een kostenefficiëntere oplossing in de situatie waar meer dan elf erfgoedinstellingen bereid zijn om over te gaan op (beredeneerde) webharvesting. De stijging in licentiekosten in het decentrale model leidt tot een aanzienlijke kostenlast bij uitbreiding, waardoor het centrale scenario voordeliger is vanaf een bepaald schaalniveau.

6.7 Gevoeligheidsanalyses

De kosten van opslag zijn een zeer bepalende factor in deze analyse. Daarom heeft het onderzoeksteam ervoor gekozen om doormiddel van een separate gevoeligheidsanalyse de afhankelijkheid van de opslagkosten, maar ook de opslagomvang voor de nationale domeincrawl, te bepalen. We laten eerst de impact van de grootte van de opslagbehoefte voor de nationale domeincrawl zien, en vervolgens de impact van de kostprijs van opslag.

6.7.1 Gevoeligheidsanalyse: Omvang nationale domeincrawl

In deze gevoeligheidsanalyse zijn twee alternatieve scenario's uitgewerkt waarin de benodigde opslag voor de nationale domeincrawl 50% hoger (225 TB) of 50% lager (75 TB) ligt dan de huidige schatting. Dit kan het gevolg zijn van het kiezen van een nauwere omvang van de domeincrawl, of een strenger/minder streng opslagregime wat een lagere benodigde opslag voor de gehele domeincrawl betekent. In beide gevallen nemen we aan dat er geen extra personeelsinzet nodig is voor de nationale domeincrawl, omdat deze scenario's geen inhoudelijke wijziging in de uitvoering vereisen.

Tabel 6.9: Gemiddelde jaarlijkse kosten domeincrawl alle scenario's

Domeincrawl	75 TB (-50%)	150 TB (raming KB)	225TB (+50%)
Personeel & techniek	€ 117 duizend	€ 117 duizend	€ 117 duizend
Opslag	€ 206 duizend	€ 413 duizend	€ 619 duizend
Totaal	€ 320 duizend	€ 530 duizend	€ 740 duizend

Kijkende naar de twee uitersten, blijft de opslag de belangrijkste kostenfactor in de totale kosten van de nationale domeincrawl. De gevoeligheidsanalyse laat zien dat bij een 50% grotere of kleinere opslagbehoefte, de totale kosten aanzienlijk kunnen veranderen, terwijl de personeels- en technische kosten gelijk blijven vanwege de stabiele uitvoeringseisen. Desalniettemin blijven de opslagkosten de grootste kostenpost.

6.7.2 Gevoeligheidsanalyse: Kosten van opslag

In de kostenanalyse van dit hoofdstuk zijn we uitgegaan van een opslagkostprijs van €3,80 per GB voor berekende selecties, en €500 per TB voor de nationale domeincrawl. Deze kosten zijn gebaseerd op gesprekken met erfgoedinstellingen en vervolgens ook gevalideerd⁶³.

De vraag is of deze inschatting richting de toekomst realistisch is. Ten eerste omdat alsmaar stijgende kosten begrotingstechnisch gezien niet praktisch of haalbaar zijn. Daarnaast is het van belang om te erkennen dat de kosten voor opslag historisch gezien een sterke daling hebben laten zien. Volgens 'Our World in Data' is de prijs per opslagunit in de afgelopen 70 jaar met bijna tien orders van grootte afgenomen. De grafiek laat een bijna rechte lijn zien op een logaritmische schaal, wat wijst op een exponentiële kostenreductie.

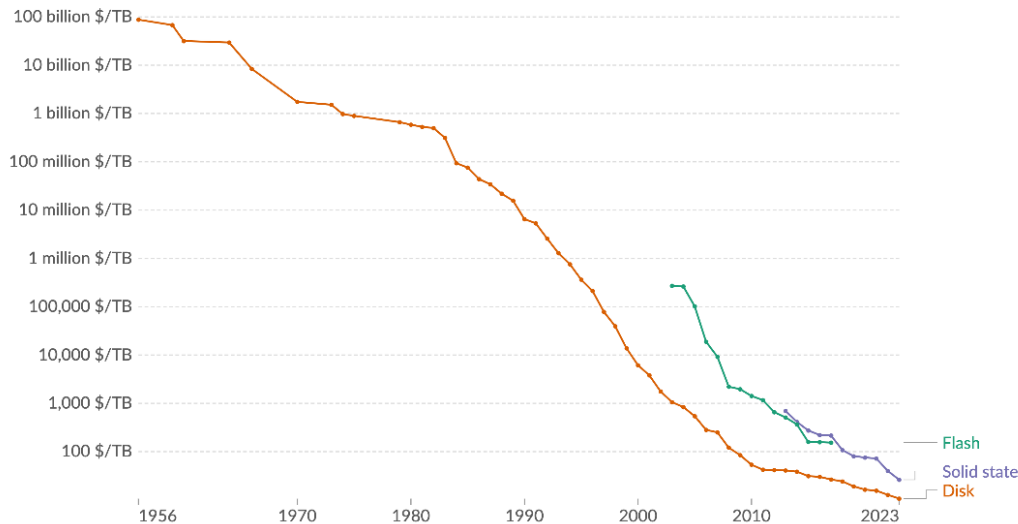
Blijvend stijgende kosten voor webharvesting over de jaren is begrotingstechnisch niet praktisch of haalbaar. De combinatie van een verwachting van dalende kosten, een steady state die wordt bereikt door het volwassen worden van de collectie (waardoor de hoeveelheid extra te harvesten websites mogelijk kan stabiliseren), en maatregelen om de stijging van opslagkosten af te remmen om budgettaire redenen, kan ervoor zorgen dat de kosten in de toekomst stabiliseren. Ons onderzoek is echter bedoeld als een schematische weergave van de ontwikkeling van de kosten gegeven de huidige situatie.

⁶³ Bij de inschatting van dit kengetal is het van belang om de kwaliteitseisen die aan de opslag worden gesteld in acht te nemen. Voor de nationale domeincrawl zijn – op basis van de inschatting van de KB – relatief lage eisen betreffende beschikbaarheid nodig, wat de kosten voor opslag drukt ten opzichte van andere toepassingen.

Figuur 6.11: Historische kostenontwikkeling opslag, Our World in Data

Historical price of computer storage

Expressed in US dollars per terabyte (TB), adjusted for inflation. "Disk" refers to magnetic storage, "flash" to memory used for rapid data access and rewriting, and "solid state" to solid-state drives (SSDs).



Data source: John C. McCallum (2023); U.S. Bureau of Labor Statistics (2024)

OurWorldInData.org/technological-change | CC BY

Note: For each year, the time series shows the cheapest historical price recorded until that year. This data is expressed in constant 2020 US\$.

De prijsontwikkeling voor opslag is niet enkel afhankelijk van technologische veranderingen. Voor afnemers, zoals in ons geval erfgoedinstellingen, kunnen diverse factoren de mate en snelheid van een prijsdaling beïnvloeden:

1. Koppelverkoop: Instellingen die webharvesting uitbesteden aan derden zijn vaak genoodzaakt om ook de opslag van de geharveste content via dezelfde aanbieder af te nemen. Dit kan leiden tot een minder scherpe prijstransitie, omdat opslag hier onderdeel is van een totaalpakket.⁶⁴
2. Prijsrigiditeit door langlopende contracten: Instellingen met lange termijnverplichtingen aan leveranciers van opslagdiensten historische tarieven, maar ervaren daardoor mogelijk vertraging bij kostenbesparingen.

Hiernaast is het van belang welk type opslag wordt gebruikt. Verschillende opslagmethoden, zoals traditionele harde schijven, SSD's, tapes of cloudoplossingen hebben elk hun eigen prijsstructuur, duurzaamheid en geschiktheid voor langdurige opslag. De prijs van deze methoden wordt beïnvloed door technologische verandering, veranderingen in vraag en aanbod, en specifieke marktontwikkelingen. Omdat deze variabelen complex en sterk afhankelijk zijn van externe factoren, valt een gedetailleerde differentiatie buiten de scope van dit onderzoek. In plaats daarvan gebruiken wij een algemene kostentrend gebaseerd op particuliere opslag om inzicht te bieden in mogelijke ontwikkelingen van de opslagkosten op lange termijn.

⁶⁴ [Misbruik van economische machtspositie | ACM.nl](#)

Uit gesprekken blijkt dat veel instellingen voor de komende vijf jaar geen significante prijsdalingen verwachten. Dit is deels te verklaren door de beperkte marktwerking in de sector en de hoge eisen aan de kwaliteit van opslag die door de erfgoedinstellingen gesteld wordt.

Ondanks deze beperkingen is het voor de lange termijn belangrijk om inzicht te bieden in scenario's waarin de opslagkosten wel dalen. Hiervoor maken we gebruik van de data over de particuliere prijs voor opslag die beschikbaar is gesteld door John C. McCallum.⁶⁵ Wanneer we de situatie van medio 2024 vergelijken met de situatie medio 2014, zien we dat de particuliere prijs per GB is gedaald van 8,75 dollar per GB in 2014, naar 1,56 dollar per GB in 2024, een daling van zo'n 82%, of een jaarlijkse daling van zo'n 15,8%.

We hebben daarom vijf groeipaden uitgewerkt die inzicht geven in de impact van prijsveranderingen:

➤ **Groeipad 1: Gelijkblijvende kosten (€3,80/GB)**

Dit groeipad sluit aan bij de hoofdtekst van de analyse, waarbij aangenomen wordt dat de kosten voor dataopslag constant blijven gedurende de analyseperiode van 10 jaar. De argumentatie hiervoor is dat erfgoedinstellingen geen significante prijsdaling verwachten, mede door bovenstaande factoren.

➤ **Groeipad 2: Optimistisch scenario (-15,8%/jaar)**

In dit groeipad projecteren we de jaarlijkse daling van 15,8% in opslagkosten, zoals afgeleid uit historische trends. De prijsdaling is gebaseerd op een vergelijking tussen de particuliere opslagkosten van 2014 (€8,75/GB) en 2024 (€1,56/GB). Dit scenario veronderstelt dat technologische innovaties zich in hetzelfde tempo blijven voortzetten en dat instellingen volledig profiteren van deze trend.

➤ **Groeipad 3: Conservatieve daling (-5% per jaar)**

In dit scenario passen we een meer conservatieve daling toe dan in het optimistische scenario.

➤ **Groeipad 4: Optimistische vertraagde daling (vanaf jaar 5)**

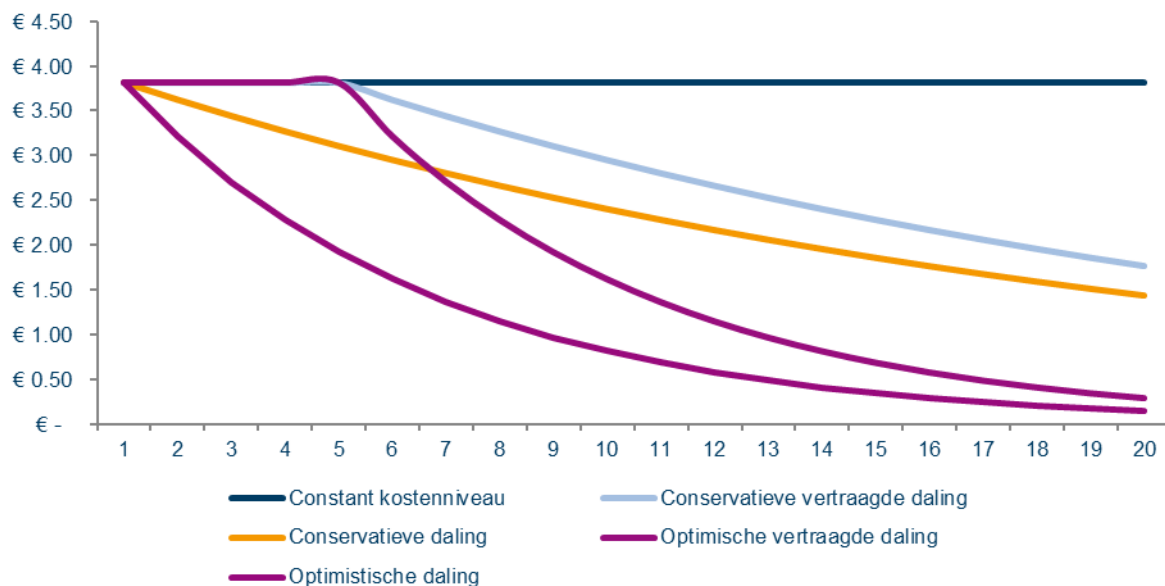
Hier veronderstellen we een vertraagde kostendaling, waarbij de prijs pas na 5 jaar begint te dalen met 15,76% per jaar. Deze vertraging houdt rekening met prijsrigiditeit, bijvoorbeeld door langlopende contracten of trage marktwerking. Vanaf jaar 5 neemt de kostprijs in dit groeipad dezelfde afnamecurve aan als in het optimistische scenario.

➤ **Groeipad 5: Conservatieve vertraagde daling (vanaf jaar 5)**

In dit groeipad passen wij de vertraagde daling toe, maar dan met de conservatieve kostendaling van -5% per jaar na jaar 5.

⁶⁵ Mitchell, J. C. (z.d.). *Memory prices 1957+*. JC Mitchell. Geraadpleegd op 7 maart 2025, van <https://jcmitch.net/memoryprice.htm>

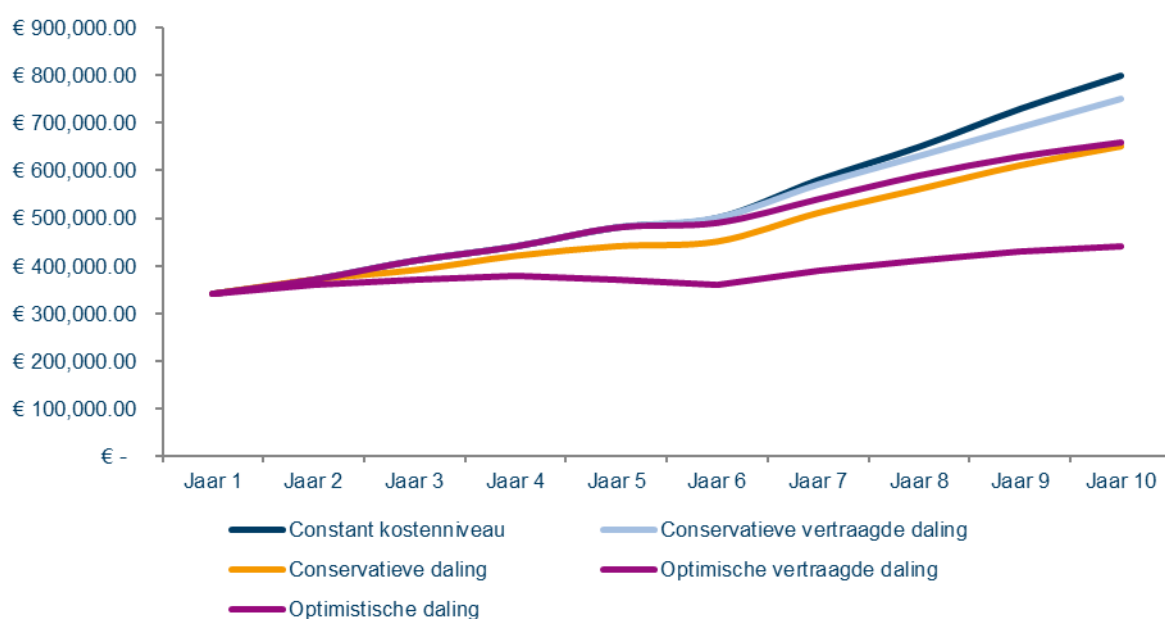
Figuur 6.12: Ontwikkeling opslagkosten per GB – 5 groeipaden



Nationale domeincrawl

De opslagkosten zijn sterk bepalend in de kosten van de nationale domeincrawl, zeker richting de toekomst. In onderstaand figuur zijn de vijf mogelijke kostenontwikkelingen van opslag weergegeven voor de nationale domeincrawl in het centrale en decentrale scenario.

Figuur 6.13: Kostenontwikkeling nationale domeincrawl over 10 jaar, 5 groeipaden



Op de korte termijn hebben de verschillende kostendalingen beperkt effect op de jaarlijkse kosten van de nationale domeincrawl. Echter, over 10 jaar zien we dat de jaarlijkse kosten sterk uiteenlopen. In het geval van de optimistische kostendaling lijken de jaarlijkse kosten zelfs een constant niveau te

bereiken. Dat komt doordat de aanvullende hoeveelheid informatie die jaarlijks wordt opgeslagen (150 TB) op een gegeven moment een kleiner aandeel extra opslag betekent dan de jaarlijkse kostenreductie van 15,8%.

Beredeneerde selecties

De kosten van opslag spelen een minder grote rol bij de beredeneerde selecties, waardoor ook de ontwikkeling van de opslagkosten een minder grote impact heeft op de totale kosten. Onderstaande tabel geeft de kosten voor beredeneerde selecties in alle scenario's weer, met daarin een bandbreedte van opslagkosten. We zien dat het consensus scenario nog steeds het duurste is, maar ook dat een consensus scenario met optimistische kostendaling goedkoper is dan het decentrale scenario met constante opslagkosten. Desalniettemin blijft het centrale scenario het meest goedkoop en kostenefficiënt in relatie tot de hoeveelheid verzameld materiaal.

Tabel 6.10: Gemiddelde jaarlijkse kosten beredeneerde selecties over 10 jaar voor alle scenario's, met een bandbreedte voor de verschillende kostengroeipaden, in duizenden

Scenario	Centraal	Decentraal	Consensus	Nul
Techniek	€ 24	€ 309	€ 309	€ 261
Personeel	€ 524	€ 356	€ 356	€ 301
Opslag	€40 - 102	€ 30 - 73	€ 30 - 73	€ 30 - 73
Notice and takedown/opt-out	€0	€0	€ 23	€ 29
Totale kosten	€ 588 - €650	€ 695 - 738	€718 - 761	€ 621 - 664
Hoeveelheid instellingen die harvesten	24	13	13	11
Hoeveelheid geharvest materiaal	3.300 websites per jaar (1.000 KB)	2.200 websites per jaar (1.000 KB)	2.200 websites per jaar (1.000 KB)	2.550 websites per jaar (2.000 KB)

Ondanks dat de opslagkosten sterk kunnen dalen bij een optimistische daling, verandert het totale kostenplaatje dus relatief beperkt. Hieruit concluderen wij dat de ontwikkeling van opslagkosten een beperkte impact heeft op de totale kosten van de beredeneerde selecties.

6.8 Kostenontwikkeling van opslag van de nationale domeincrawl

In het voorgaande hebben we de kosten van de nationale domeincrawl inclusief enkele gevoeligheidsanalyses uitgewerkt. Uit deze analyses blijkt dat de kosten voor voornamelijk opslag jaarlijks zal groeien: ieder jaar is er een grotere behoefte aan opslag en deze behoefte blijft groeien. Een dergelijke onbeperkte groei van opslagkosten is vanuit een financieringsperspectief een vraagstuk: zijn er jaarlijks extra middelen nodig voor de nationale domeincrawl om dit te kunnen bekostigen? Op basis van ons onderzoek hebben we een aantal nuanceringen op deze kostenontwikkeling opgehaald, die we in deze paragraaf toelichten:

- Ten eerste is er voor de nationale domeincrawl geen noodzaak voor kwalitatief hoge opslag, omdat er in ieder geval op korte termijn geen vragen verwacht worden van gebruikers in de toegang tot geharvest materiaal, zo is de inschatting van de KB.

- Daarnaast zal de uitvoerder van de nationale domeincrawl afwegingen maken binnen haar collectiebeleid. De uitvoering van een nationale domeincrawl kan afgewogen worden tegen ander digitaliseringsprioriteiten.
- De opslagkosten voor de nationale domeincrawl zijn relatief beperkt vergeleken met andere (digitale) collecties.
- Op termijn kan er gekozen worden om delen van de collectie te laten vervallen, indien dit de opslagkosten kan bijdragen.

7. Draagvlak

In dit hoofdstuk lichten we het draagvlak per scenario toe. Daarbij gaan we niet alleen in op de houding van partijen ten opzichte van een scenario, maar lichten we ook toe waarom een partij deze houding heeft.

Er zijn verschillende groepen stakeholders die van invloed zijn op het draagvlak, namelijk de erfgoedinstellingen (centraal en decentraal), de auteursrechthebbenden en maatschappelijke organisaties (zoals de Open State Foundation en Privacy First). De selectie van de stakeholders is gedaan op basis van hun relevantie voor de technische en organisatorische aspecten van de scenario's. Voor de juridische aspecten zijn andere stakeholders, zoals de AP en JenV ook relevant. Deze stakeholders en hun standpunten zijn in de juridische analyse in Hoofdstuk 4 al toegelicht.

7.1 Scenario 1: Centrale uitvoering door één of enkele instelling(en) met loketfunctie voor andere instellingen

Vanuit erfgoedinstellingen is er draagvlak voor het uitvoeren van de nationale domeincrawl door één centrale partij. Wat betreft de beredeneerde selecties en de beoogde loketfunctie is er minder enthousiasme bij de centrale partij(en) die deze functie mogelijk zouden kunnen krijgen. Het vraagt immers meer inspanning en er moeten nieuwe processen ingericht worden. Daarnaast is er sprake van een afhankelijkheid van een centrale organisatie in plaats van een eigenstandige taak/bevoegdheid van erfgoedinstellingen om webharvesting uit te voeren.

Toch geven de potentiële uitvoerders aan dat dit niet onoverkomelijk zou zijn voor hen en staan zij relatief positief tegenover dit voorstel ten opzichte van het nulscenario en het consensusmodel. De decentrale partijen die nu al aan webharvesting doen zijn over het algemeen niet positief over de loketfunctie, die hen afhankelijk zou maken van een andere partij voor het uitvoeren van webharvesting en activiteiten die zij nu al eigenstandig uitvoeren. Die werkwijze wijkt immers af van de huidige werkwijze waarbij een aantal decentrale partijen al zelf aan webharvesting doet. Met name die partijen zien minder in een loketfunctie bij een centrale partij, aangezien ze zelf al de kennis en techniek in huis hebben, en het hen dus extra administratieve lasten zou opleveren om te harvesten via een centrale partij. Wel zijn alle partijen het erover eens dat een loketfunctie handig kan zijn en ontzorgend zou werken voor partijen die anders niet aan webharvesting zouden kunnen doen.

Auteursrechthebbenden zijn blij met een grotere rol voor een centrale organisatie in dit scenario. Het is overzichtelijker als alle activiteiten op één plek plaatsvinden en op één plek alle benodigde waarborgen, zowel procedureel als technisch, zijn ingesteld.

Maatschappelijke organisaties staan over het algemeen positief tegenover dit scenario. Met name de uitvoering van de nationale domeincrawl door één centrale partij wordt positief ontvangen, omdat daarmee op de meest efficiënte en effectieve wijze het nationaal webdomein geharvest wordt. Over de beredeneerde selecties zijn maatschappelijke organisaties minder uitgesproken qua draagvlak, maar er zijn wel wat zorgen en vragen aan met name de privacy kant over de precieze invulling van de loketfunctie. Voor hen is het belangrijk dat er duidelijkheid is over welke instellingen/partijen precies van de loketfunctie gebruik mogen maken en welke voorwaarden daarbij gelden.

7.2 Scenario 2: Centrale uitvoering van de nationale domeincrawl en decentrale uitvoering van beredeneerde selecties

Erfgoedinstellingen staan, nog meer als in het eerste scenario, positief tegenover de uitvoering van de nationale domeincrawl door één centrale partij. Ook zijn erfgoedinstellingen positief over de eigenstandige bevoegdheid voor decentrale partijen om zelf beredeneerde selecties uit te voeren. Dit sluit het meest aan op de bestaande uitvoeringspraktijk en vergt de minste aanpassingen.

Auteursrechthebbenden zijn overwegend positief over dit scenario, mits de juiste waarborgen en eisen aan de decentrale instellingen worden gesteld, onder andere op het gebied van beveiliging. Er moet transparantie zijn over wie wat uitvoert.

Voor maatschappelijke organisaties geldt dat zij, net als bij het eerste scenario, positief staan tegenover de uitvoering van de nationale domeincrawl door één centrale partij. Er is ook draagvlak voor de uitvoering van beredeneerde selecties door decentrale partijen, maar net als de auteursrechthebbenden hechten ook de maatschappelijke organisatie daarbij veel waarde aan de juiste waarborgen en transparantie.

7.3 Scenario 3: Afspraken maken over uitvoering van webharvesting (consensus)

Vanuit de meeste groepen is er weinig draagvlak voor dit scenario. Erfgoedinstellingen staan negatief tegenover het ontbreken van draagvlak voor het uitvoeren van de nationale domeincrawl in dit scenario. In algemeen zin staan auteursrechthebbenden niet negatief tegenover het maken van afspraken, echter zijn ze sceptisch over de reikwijdte en de waarborgen (hoewel beiden onderdeel kunnen zijn van het maken van afspraken). Maatschappelijke organisaties zijn sceptisch over het ontbreken van waarborgen en duidelijkheid over wie wat uitvoert.

7.4 Status quo (geen afspraken maken)

Ook voor dit scenario is er weinig draagvlak bij de betrokken partijen. Er wordt geen nationale domeincrawl uitgevoerd en de partijen die op basis van beredeneerde selecties harvesten vinden zichzelf in een juridisch grijs gebied opereren. Er is onduidelijkheid over wat er mag en wie wat doet, Door het ontbreken van een wettelijke grondslag is ook niet duidelijk of de juiste waarborgen ingesteld zijn. Auteursrechthebbenden en maatschappelijke organisaties hebben behoefte aan meer inzicht en transparantie.

8. Het nationaal webdomein

In dit hoofdstuk bespreken we de afbakening van het nationaal webdomein. Na een algemene inleiding lichten we de functie en het belang van een afbakening en definitie van het nationaal webdomein toe voor beredeneerde selecties en voor de nationale domeincrawl. Daarna bespreken we enkele voorbeelden uit andere Europese lidstaten, gebaseerd op het IViR rapport. We besluiten dit hoofdstuk met de beschrijving van twee strategieën om te komen tot een (geoperationaliseerde) afbakening.

Het doel van het geven van een definitie van het nationaal webdomein is om een afbakening te maken in welke webdomeinen onderwerp kunnen zijn van webharvesting door Nederlandse erfgoedinstellingen. Dit moet worden gezien in de context van een internationaal internet waarbij er vooraf weinig tot geen duidelijke (nationale) grenzen zijn, en er daarom behoefte is aan een afbakening van het nationaal webdomein: niet al het materiaal op het internet is relevant en/of rechtmatig om op te nemen in de collectie van een *Nederlandse* erfgoedinstelling. Een definitie draagt bij aan het vaststellen van relevante grenzen. Met het begrip ‘definitie van het nationaal webdomein’ bedoelen we deze afbakening. In dit onderzoek maken we een onderscheid tussen de relevantie van een afbakening van het nationaal webdomein voor beredeneerde selecties en de nationale domeincrawl, waar we in paragraaf 8.1 en 8.2 verder op ingaan.

Het IViR rapport besteedt al aandacht aan een definitie van het nationaal webdomein in het kader van het selecteren van materiaal om te harvesten. Het IViR beschrijft dat het definiëren een kwestie is die onderzoekers en webarchivarissen al langere tijd bezighoudt.⁶⁶ Er wordt een brede definitie genoemd voor nationaal web (“elke webactiviteit die verband houdt met de betreffende natiestaat”), maar ook geschetst dat deze verdere afbakening vereist om daadwerkelijk op rechtmatige wijze een nationaal webdomein te kunnen archiveren. Er zijn verschillende impactcategorieën relevant voor de definitie van het nationaal webdomein, waaronder het informatie- en erfgoedbelang, het belang van betrokkenen (bescherming van persoonsgegevens) en het belang van (auteurs)rechthebbenden. De beoordeling van deze belangen ten aanzien van verschillende strategieën om te komen tot een geoperationaliseerde definitie van het nationaal webdomein (paragraaf 8.4) is toegelicht in paragraaf 3.3.

8.1 Beredeneerde selecties

In de huidige situatie, zowel ten aanzien van het archiveren van webdomeinen als ten aanzien van het archiveren van ander materiaal, hebben erfgoedinstellingen een verzamelbeleid, collectieplan en/of een acquisitieprofiel om te beoordelen of ze werken al dan niet opnemen in hun collectie. Dit proces van collectievorming geldt zowel voor digitale als papieren werken en is het primaire proces van erfgoedinstellingen. Erfgoedinstellingen die al aan webharvesting doen passen ditzelfde proces toe op webarchieven. Daarbij maken zij de (eigen) afweging dat het verzamelen van webdomeinen binnen hun eigen verzamelbeleid, collectieplan en/of acquisitieprofiel past. Met andere woorden, het collectievormende proces is leidend in het vaststellen of materiaal relevant is om op te nemen in een collectie, ongeacht of het gaat om een webdomein, digitale content of fysiek materiaal.

⁶⁶ Zie voor een overzicht van de discussie: N. Brügger & D. Laursen (eds) *The Historical Web and Digital Humanities: The Case of National Web Domains*. 1st ed. New York, NY: Routledge, 2019. DOI: 10.4324/9781315231662.

In dit onderzoek hebben we vastgesteld dat, daar waar het gaat om beredeneerde selecties, de noodzaak voor een afbakening van het nationaal webdomein voornamelijk gericht is op rechtszekerheid omdat er in de praktijk al oplossingen zijn voor selectie van webdomeinen. Erfgoedinstellingen hebben al beleid (verzamelbeleid, collectieplannen en/of acquisitieprofielen) en uitvoerende processen (collectievorming) om vast te stellen of een werk relevant is om op te nemen in een collectie of niet.

Een afbakening van het nationaal webdomein zou in onze ogen moeten aansluiten op deze bestaande processen, om zo rechtszekerheid te bieden aan de uitvoerders van webharvesting. We gaan uit van de betrouwbaarheid van deze processen en stellen vast dat dit vraagt om een afbakening die aansluit bij de oplossingen die er in de praktijk al zijn.

8.2 Nationale domeincrawl

De nationale domeincrawl is een nieuwe activiteit die (nog) niet past bij een bestaand collectievormend proces. Er is nog geen bestaand kader met betrekking tot collectievorming. Met deze achtergrond is het van belang om vast te stellen op welke webdomeinen de nationale domeincrawl van toepassing is en welke domeinen horen bij het nationale webdomein. Een afbakening en definitie voor het nationaal webdomein is daarom specifiek van belang om in vulling te kunnen geven aan de nationale domeincrawl.

Het Nederlandse webdomein is zo groot (alleen al het .nl-domein omvat meer dan 6,2 miljoen domeinnamen⁶⁷, en er zijn vele Nederlandse webdomeinen buiten het .nl-domein) dat het onmogelijk is om voor allemaal handmatig te beoordelen of ze relevant zijn en op basis van welke criteria. Het is daarom nodig om een selectie te maken op basis van een of meerdere criteria.

Bij deze afbakening gaat specifiek om selectiecriteria voor webdomeinen die op enige wijze gerelateerd zijn aan Nederland. Het IViR rapport noemt dat de *'KB bezig is met het definiëren van de grenzen van een nationaal web voor een eventueel toekomstige domeincrawl. De KB neemt hierbij in acht dat haar webarchief volgens haar content strategie onderdeel is van de Nederlandcollectie. Deze collectie is "een zo compleet mogelijke verzameling van (1) in Nederland of het Nederlands uitgegeven publicaties én (2) publicaties over de Nederlandse taal, cultuur en geschiedenis", inclusief internetpublicaties.*'⁶⁸

Hierin worden verschillende selectiecriteria zoals geografie, taal en content genoemd. Volgens het IViR rapport onderscheidt de KB zes criteria voor het definiëren en afbakenen van het nationale domein.⁶⁹ Deze criteria vormen in de onderstaande volgorde de prioriteit van de KB ten aanzien van het crawlen van het Nederlandse webdomein⁷⁰:

- 1) het top-level domein⁷⁰: het top-level domein van Nederland is .nl, maar er zijn ook andere relevante top-level domeinen zoals .frl en .amsterdam);
- 2) de geografie van een bepaald domein: is de publicatie (het webdomein) in Nederland gepubliceerd? Deze vraag kan bijvoorbeeld worden beantwoord op basis van het adres van de geregistreerde website, of het IP-adres;
- 3) de taal: Is de publicatie in het Nederlands geschreven?;

⁶⁷ <https://www.sidn.nl/>

⁶⁸ Webharvesting, IViR (2021), p.19.

⁶⁹ We baseren ons hierbij op het IViR rapport; er zijn geen nieuwe of gewijzigde criteria gebleken in dit onderzoek.

⁷⁰ De Engelse term voor top-level domein is *country code top-level domain*, afgekort als ccTLD.

- 4) de cultuur: wat vindt de gemeenschap zelf dat tot hun digitale cultuur behoort? Hierbij kan de gemeenschap betrokken worden; een mogelijk voorbeeld is het in kaart brengen van de linkstructuur van de Nederlandse Wikipedia om belangrijke Nederlandse websites in kaart te brengen;
- 5) de netwerkstructuur: Naar welke andere pagina's verwijzen pagina's die, op basis van andere criteria dan deze, tot het nationale domein horen? De gedachte hierachter is dat pagina's waar vaak naar wordt verwezen vanuit het nationale domein, mogelijk zelf ook tot dit domein kunnen horen;
- 6) de geschiedenis: Hoe hebben pagina's die, op basis van andere criteria dan deze tot het nationaal domein behoren, zich ontwikkelt? Heeft er bijvoorbeeld een migratie plaatsgevonden van .com domeinen naar .nl domeinen toen deze beschikbaar werden gesteld aan individuen?⁷¹

Deze criteria geven in het algemeen weer welke webdomeinen er tot het Nederlandse webdomein kunnen behoren, maar ze geven nog niet aan hoe deze in de praktijk gebracht kunnen worden (geoperationaliseerd). Hoe kan er in de praktijk een set aan webdomeinen worden vastgesteld voor de nationale domeincrawl? Deze operationalisatie is van belang om de reikwijdte van de nationale domeincrawl vast te stellen en daarmee te bepalen welke webarchieven tot het Nederlandse webdomein behoren.

Gekeken naar toekomstige ontwikkelingen is het ook van belang om vast te stellen dat het begrip 'Nederlands webdomein' niet vaststaat en aan verandering onderhevig is. Dit kan op verschillende manieren gebeuren. Zo kunnen er nieuwe domeinen bijkomen (.denhaag, .middelburg of .limburg). Maar ook ontwikkelingen in de wereld kunnen relevant zijn bij het bepalen wat het Nederlandse internetdomein is, waarbij gedacht kan worden aan de gemeenschap van Oekraïense vluchtelingen die in Nederland verblijft. Verder worden er in de huidige praktijk al webdomeinen geharvest buiten nationale top-level domeinen (zoals .com), omdat deze via collectievormende processen een (vastgestelde) relatie hebben met Nederland.

In de aard van het internet zit besloten dat het niet gebonden is aan nationale grenzen. Ook is het aan verandering onderhevig (mede door gebrek aan deze nationale grenzen). Dit maakt dat het noodzakelijk is om een specifieke aanpak of strategie te formuleren die recht doet aan de diversiteit van criteria die het Nederlandse webdomein kunnen definiëren, maar tevens een hanteerbare lijst voor de uiteindelijke uitvoering oplevert.

In de praktijk kunnen bovengenoemde criteria ook worden gecombineerd en (semi-)geautomatiseerd worden toegepast om relevante webdomeinen te selecteren (beslisregels).

8.3 Europese voorbeelden

Het IViR rapport geeft inzicht in hoe andere Europese landen kijken naar hun 'nationaal webdomein' en welke definitie of aanpak ze gebruiken. In deze paragraaf geven we een beknopt overzicht van de varianten die er gehanteerd worden. Deze beschrijving gebruiken we als achtergrond om te komen tot twee mogelijke strategieën en moet niet worden gelezen als een eigen (rechts)vergelijkende analyse.

⁷¹ In het genoemde voorbeeld wordt gerefereerd aan het presenteren van .nl (of andere TLD's) aan webmaster en individuen als *nieuwe* TLD's (naast .com). Een ander voorbeeld hiervan is .amsterdam als relatief nieuw TLD.

Voor deze analyse is gekeken naar de informatie in het IViR rapport over Denemarken, Duitsland, Verenigd Koninkrijk (VK), Frankrijk en Portugal.

De voorbeelden en analyses uit het IViR rapport laten zien dat alle onderzochte landen ten minste een top-level domein (zoals .dk of .pt) dat gerelateerd is aan het land binnen de reikwijdte van de nationale domeincrawl plaatsen. In het VK gaat het ook om domeinnamen die verwijzen naar het VK, zoals .london.

8.3.1 Criteria in andere lidstaten

Naast het top-level domein komen de volgende varianten en criteria met betrekking tot de nationale domeincrawl verder ook voor:

Algemeen: Gericht op nationaal publiek

Een kenmerk dat onderdeel kan zijn van de nationale domeincrawl is of het gericht is op nationaal (Deens, Duits, etc.) publiek. In de praktijk moet dit verder worden gespecificeerd. In Denemarken wordt hiervoor bijvoorbeeld gekeken naar taal, eigenaar van het domein, of gerelateerd tot een nationale aangelegenheid of content waarvan de auteur Deens is⁷². De term 'aangelegenheden' wordt niet nader gespecificeerd in het IViR rapport, maar de implicatie is dat het een Deense auteur betreft of op andere wijze betrekking heeft op Denemarken. Hoewel het niet zo wordt genoemd, lijkt het hierbij te gaan om webactiviteiten die verband houden met een natiestaat, in lijn met de algemene definitie die het IViR geeft voor het nationale webdomein⁷³.

Taal-criterium

In Denemarken kan de verantwoordelijk Minister nadere specificaties geven over de nationale domeincrawl. Eén van de invullingen hiervan is een website die de Deense taal gebruikt. In Denemarken krijgt dit praktische invulling door (semi-)handmatige selectie door een redactieraad⁷⁴, aangesteld door de Minister en bestaande uit betrokken instellingen. Andere landen specificeren geen taalcriterium.

Geografie

Een ander criterium is die van geografie: webactiviteiten die plaatsvinden op grondgebied van een land. In Frankrijk wordt dit ingevuld door domeinen in de nationale domeincrawl op te nemen die geregistreerd zijn door een persoon die in Frankrijk woonachtig is of die op Frans grondgebied geproduceerd is. In het VK wordt ook gekeken naar webactiviteiten die plaatsvinden in het VK. Dit krijgt praktische invulling door de contactinformatie die op de website te vinden is, of andere relevante informatie te beoordelen op relatie met het VK⁷⁵. Zo kan worden bepaald of een domein binnen het nationale webdomein hoort.

Nationaliteit en eigenaarschap

In Denemarken wordt er niet gekeken naar geografie maar naar nationaliteit: webactiviteiten van Deense ingezetenen komen in aanmerking voor opname in de nationale domeincrawl. Het gaat dan in

⁷² Webharvesting, IViR (2021), p.95

⁷³ Webharvesting, IViR (2021), p.18.

⁷⁴ Deze redactieraad beoordeelt niet alleen op een taal-criterium, maar kijkt ook naar andere criteria die minder goed te automatiseren zijn.

⁷⁵ Webharvesting, IViR (2021), p.110.

het bijzonder om webdomeinen die eigendom zijn van Deense ingezetenen. Hier wordt ook praktische invulling aan gegeven met behulp van een redactieraad.

8.3.2 Aanvullende bevindingen uit Europese voorbeelden

Proces met betrekking tot het nationale webdomein

Een aanvullend punt om te benoemen heeft niet te maken met de gehanteerde criteria maar met de wijze waarop er tot een invulling van het nationale webdomein voor een domeincrawl wordt gekomen. Het IViR rapport beschrijft specifiek hoe dit in Denemarken wordt gedaan, zoals hierboven al kort is benoemd. In Denemarken kan de Minister nadere specificaties (regelgeving) vaststellen ten behoeve van de nationale domeincrawl. Voor een deel van deze specificaties wordt gebruik gemaakt van een redactieraad om te beoordelen of een domein moet worden opgenomen in de nationale domeincrawl.

Nationale bibliotheken combineren een domeincrawl en beredeneerde selecties

Uit het IViR rapport blijkt verder dat nationale bibliotheken in onderzochte lidstaten een nationale domeincrawl combineren en met beredeneerde selecties. Dat betekent dat er een oppervlakkige harvestingactiviteit plaatsvindt op bijvoorbeeld een aantal top-level domeinen maar er ook ruimte en flexibiliteit is om op specifieke onderwerpen (nationaal) relevante webdomeinen op te nemen in collecties. Op deze manier kunnen bibliotheken recente ontwikkelingen volgen maar ook eventuele gaten in subjectievere selecties dichten met een algemene nationale domeincrawl.

8.4 Strategieën voor operationaliseren

Op basis van onze analyse zien we twee strategieën die kunnen worden gehanteerd om te komen tot een geoperationaliseerde definitie van het nationaal webdomein:

- Een strategie op basis van vooraf vastgelegde criteria die automatiseerbaar en objectief zijn. Een voorbeeld hiervan is het gebruik van top-level domeinen.
- Een strategie waarbij criteria niet vooraf vastgelegd zijn, maar via een vastgesteld proces worden bepaald. Naast automatiseerbare en objectieve criteria kan het hierbij ook gaan over andere criteria die om meer specifieke of individuele beoordeling vragen, zoals cultuur.

We zullen deze twee strategieën hieronder kort bespreken waarbij we een algemene beschrijving geven en een toelichting geven op de activiteiten en uitvoeringslast die ermee gemoeid is. Onze toelichting heeft niet tot doel om exact en uitputtend weer te geven welke (soort) domeinen er binnen de geoperationaliseerde definitie vallen, maar vooral om een weergave te geven van een aanpak om tot een selectie van domeinen te komen.

8.4.1 Vooraf vastgestelde criteria

In deze strategie worden er vooraf een aantal criteria vastgesteld die bepalen of een domein tot het Nederlandse webdomein behoort, waarbij deze criteria ook via automatiseerbare wijze vast te stellen zijn. Een voorbeeld van het automatiseerbaar vaststellen is het gebruiken van een top-level domein als criterium.

Een criterium dat in ieder geval gebruikt kan worden is of een domein tot het .nl-domein (of een ander aan Nederland gerelateerd domein, zoals .amsterdam of .frl) behoort. Dit kan geautomatiseerd worden vastgesteld met hulp van SIDN of andere uitgevers van top-level domeinen. Aan dit criterium kan nog worden toegevoegd dat op basis van IP-adres (geografie) in combinatie met een taalcriterium

aanvullende domeinen (semi-)geautomatiseerd kunnen worden toegevoegd, zo is gebleken uit een gesprek met experts van o.a. de KB.

Voor deze aanpak zijn er in principe geen redactionele handelingen nodig, bijvoorbeeld bij het beoordelen of een webdomein tot de definitie behoort, omdat objectief en geautomatiseerd vast te stellen is of een domein binnen de definitie valt. Dit is een voordeel in termen van uitvoeringslast ten opzichte de procesgerichte strategie. De aanpak is dus relatief eenvoudig uit te voeren (geautomatiseerd).

Een nadeel van deze strategie is dat er ook veel webactiviteiten zijn die buiten de definitie vallen maar wél verbonden zijn met Nederland, bijvoorbeeld op basis van het 'cultuur' criterium. Met andere woorden, door deze strategie toe te passen zal een deel van de webdomeinen die vanuit informatie- en erfgoedbelang idealiter zouden worden geharvest buiten de afbakening voor webharvestingactiviteiten vallen. Dit is in het bijzonder voor beredeneerde selecties een probleem, omdat er in de bestaande collectievormende processen afwegingen kunnen worden gemaakt die buiten de objectieve criteria vallen over welke webdomeinen relevant zijn voor de collectie van een instelling. Dit beperkt de handelingsruimte van instellingen die aan beredeneerde selecties doen, omdat er maar een beperkt aantal criteria is die objectief en automatiseerbaar toe te passen zijn.

Daartegenover staat dat er een beperkt risico is op 'bijvangst' van webdomeinen die niet gerelateerd zijn aan Nederland, zoals bijvoorbeeld wel het geval kan zijn bij het gebruiken van een taal-criterium (waarbij er veel auteurs zijn die in het Nederlands publiceren maar niet tot het Nederlandse webdomein behoren, zoals Vlaamse auteurs). In de huidige praktijk is voor beredeneerde selecties 'bijvangst' geen probleem, aangezien er sprake is van een specifieke selectie-actie van webdomeinen.

De vooraf vastgestelde criteria kunnen ten behoeve van de uitvoering van de nationale domeincrawl ook worden gecombineerd om te komen tot beslisregels (voor het beoordelen van webdomeinen).

8.4.2 Vastgesteld proces voor criteria

In een strategie waarbij er een vastgesteld proces wordt gehanteerd, wordt er niet vooraf een set aan vaste criteria voor de nationale domeincrawl benoemd, maar worden deze criteria in een proces vastgesteld, bijvoorbeeld door een redactieraad of een adviesraad. Het idee hierachter is dat het moeilijk zo niet onmogelijk is om te komen tot een sluitende en precieze definitie van het Nederlandse webdomein zónder enige menselijke beoordeling (ten aanzien van het vaststellen van een definitie). In een procesgerichte aanpak worden er criteria en/of kaders vastgesteld die kunnen worden toegepast door de uitvoerder van de nationale domeincrawl en beredeneerde selecties. Deze criteria en/of kaders kunnen periodiek worden geëvalueerd en bijgesteld om recht te doen aan de veranderlijkheid van het web.

In deze aanpak zijn er twee relevante momenten van vaststellen en gebruiken van de criteria. Ten eerste stelt de redactie- of adviesraad de criteria vast. Deze criteria worden daarna ten tweede toegepast door erfgoedinstellingen om te harvesten webdomeinen te selecteren.

Voorbeelden: Gezamenlijk collectieplan Bibliotheken en Adviesraad van Archieven en Collectie Metamorfoze

Een voorbeeld van een vastgesteld proces is het **gezamenlijk collectieplan**⁷⁶, opgenomen in de Wet stelsel openbare bibliotheekvoorzieningen (Wsob). Het doel van het gezamenlijk collectieplan is om te zorgen voor samenhang tussen de fysieke en digitale collecties van bibliotheken. Bibliotheken zijn zelf verantwoordelijk voor hun collectie, maar het collectieplan geeft een kader voor het samenstellen en beheren van deze collectie.

Het collectieplan wordt voor een periode van vier jaar vastgesteld (Wsob, artikel 10, lid 1). Na deze periode wordt het collectieplan geëvalueerd. Deze evaluatie is inbreng voor het volgende collectieplan.

De Wsob (artikel 10, lid 3) bepaalt dat: “De Koninklijke Bibliotheek stelt het gezamenlijk collectieplan vast in overeenstemming met vertegenwoordigers van de lokale bibliotheken en provinciale ondersteuningsinstellingen.” Voor het huidige collectieplan (2020 – 2024)⁷⁷ is dit in de vorm van een klankbordgroep met vertegenwoordigers van verschillende openbare bibliotheken.

Een tweede voorbeeld van een procesgerichte strategie is de **Adviesraad van Archieven en Collectie Metamorfoze**, die op inhoudelijke gronden en op basis van vooraf vastgestelde voorwaarden projectvoorstellen voor conservering van archieven en collecties (papieren erfgoed) beoordeelt. De adviescommissie bestaat uit experts uit verschillende wetenschapsdisciplines. In dit voorbeeld gaat het niet om het vaststellen van criteria, maar om het beoordelen van voorstellen (anders dan de voorgestelde redactieraad).

Een ander relevant punt van deze aanpak is dat vanwege de breedte van het nationale webdomein een procesgerichte aanpak hoe dan ook zal moeten terugvallen op (semi-)geautomatiseerde selectie van webdomeinen. Daardoor is het niet mogelijk om de precieze effecten van alle criteria (welk domein precies wordt geharvest) vooraf via een proces te beoordelen.

Aanvullend hierop is de rechtszekerheid die deze strategie biedt aan erfgoedinstellingen en andere belanghebbenden mede afhankelijk van de publicatie (openbaarheid) van de selectiecriteria en de mate waarin deze expliciet en eenduidig te interpreteren zijn. Hoe minder eenduidig deze zijn, hoe minder voorzienbaarheid er is over welke webdomeinen geharvest (kunnen) worden.

Een aantal onderwerpen vragen om nadere invulling en precisering in de uitvoering:

- De exacte invulling van de redactie- of adviesraad. Hieronder vallen onder andere de ophanging, de bemensing en de precieze taken. Zo kan worden gedacht aan een redactieraad op uitvoerend niveau met daarin vertegenwoordigers van de rechthebbenden en gebruikers. Ook is het een optie om een raad op beleidsniveau in te richten met wetenschappers en/of experts. Een te verwachten voordeel van een procesgerichte strategie is dat het betrekken van stakeholders ook het draagvlak vergroot.

⁷⁶ Dit voorbeeld dient om te illustreren hoe er in gezamenlijkheid kan worden gekomen tot een (beleids)kader dat via een proces wordt vastgesteld. Het is niet bedoeld als een blauwdruk voor de nationale domeincrawl, omdat het voorbeeld van het collectieplan afwijkt van de nationale domeincrawl, specifiek als het gaat om de benodigde mate van samenwerking en het aantal uitvoerende partijen.

⁷⁷ [Bibliotheeknetwerk.nl](https://bibliotheeknetwerk.nl)

- De mate van praktische hanteerbaarheid van de criteria en/of kaders die aan de uitvoering worden opgelegd. Het doel is om te komen tot een operationalisatie van het Nederlandse webdomein. Daarom zouden de criteria en/of kaders voldoende handvatten moeten geven om beslissingen te nemen over (groepen van) domeinen om op te nemen in de nationale domeincrawl en beredeneerde selecties.
- De exacte invulling van het proces en de procedure om te komen tot een invulling van de criteria.

8.5 Aanvullende opmerkingen over het nationale webdomein

We eindigen dit hoofdstuk met een aantal overige opmerkingen ten aanzien van het nationale webdomein. Ten eerste lijkt het verstandig om vanwege de veranderlijkheid en vluchtigheid van het internet (en bijbehorend nationaal webdomein) de definitie van het nationaal webdomein niet op te nemen in wet- en regelgeving die niet eenvoudig te wijzigen is, maar wel beperkend kan werken op wat er middels een nationale domeincrawl kan worden opgehaald. Zo moet worden voorkomen dat het crawlen van een nieuw top-level domein dat relevant is voor Nederland niet mogelijk is vanwege wet- en regelgeving die te specifiek is of geen ruimte overlaat voor nieuwe ontwikkelingen. In de uitvoering van de nationale domeincrawl lijkt het verstandig om zonder wetswijzigingen uitbreidingen te kunnen doen op de reikwijdte van de nationale domeincrawl. We stellen op basis van de gevoerde gesprekken en expertsessies vast dat de betrokken partijen deze mening ook delen. Een definitie die uitgaat van het beschrijven van regels en een proces om te komen tot een nationaal webdomein kan hier een oplossing voor bieden.

Ten tweede is het binnen alle definities van het Nederlandse webdomein verstandig om periodiek te evalueren of de gekozen definitie en criteria doeltreffend en doelmatig zijn. Dit kan via reguliere verantwoordingscycli (zoals het jaarverslag van een uitvoerder) en/of via een uitgebreidere organisatie- of beleidsevaluatie.

Ten derde benoemen we nog de wijze waarop de definitie wordt vastgelegd. Dit is mede gerelateerd aan de gekozen juridische route (zie paragraaf 2.3). De gekozen juridische route bepaalt mede de wijze waarop de definitie vastgelegd *kan* worden. Zo sluit het *niet* wettelijk faciliteren van webharvesting uit dat definitie in wet- en regelgeving zal worden vastgelegd. Tegelijkertijd kan het vastleggen van (de kaders voor) de reikwijdte van de nationale domeincrawl en beredeneerde selecties in wet- en/of regelgeving bijdragen aan de voorzienbaarheid en transparantie van deze activiteiten.

Bijlage A Overzicht gesprekken

Hieronder volgt een overzicht van alle organisaties waar in het kader van dit onderzoek één of meerdere keren mee gesproken is.

Organisatie

Beeld en Geluid

Nationaal Archief

Koninklijke Bibliotheek

Koninklijke Vereniging Archiefsector Nederland

Ministerie van Justitie en Veiligheid, Directie Wetgeving en Juridische Zaken

Instituut voor Informatierecht (IViR)

Tresoar

Federatie Auteursrechtbelangen en VOICE

Groninger Archieven

Privacy First

Open State Foundation

Het Flevolands Archief

Autoriteit Persoonsgegevens

Kunstenbond

Groep Algemene Uitgevers

Bits of Freedom, het Internationaal Instituut voor Sociale Geschiedenis en het Huygens instituut zijn niet ingegaan op een verzoek voor een gesprek.

Bijlage B Deelnemers expertgroepen

Juridische expertgroep 26 augustus 2024

Koninklijke Bibliotheek

Nationaal Archief

KVAN

Beeld en Geluid

Technische/organisatorische/financiële expertgroep 27 augustus 2024

Koninklijke Bibliotheek

Nationaal Archief

KVAN

Beeld en Geluid

Juridische expertgroep 29 oktober 2024

Koninklijke Bibliotheek

Nationaal Archief

KVAN

Beeld en Geluid

VOICE

Federatie Auteursrechtbelangen

Kunstenbond

Groep Algemene Uitgevers

Ministerie van JenV

Ministerie van OCW

Juridische expertgroep 17 december 2024

Koninklijke Bibliotheek

Nationaal Archief

KVAN

Beeld en Geluid

Ministerie van OCW

Bijlage C Roadmaps voor uitvoering van webharvesting

We geven invulling aan de impactcategorie 'Uitvoering door de Rijksoverheid' in de vorm van verschillende roadmaps. De roadmaps beschrijven de acties die uitgevoerd moeten worden om een scenario daadwerkelijk in de praktijk te brengen. Het gaat daarbij niet om de uitvoering door erfgoedinstellingen, maar om de activiteiten die nodig zijn om te komen tot het (wettelijk) faciliteren van webharvesting.

We benoemen deze activiteiten hieronder en gaan daarbij in op het wetgevingstraject (indien van toepassing), het draagvlak en stakeholdermanagement, en de beheersing van risico's en onzekerheden. Voor het nulscenario beschrijven we geen roadmap, aangezien dit scenario de status quo betreft en er dus geen verdere acties nodig zijn om het in de praktijk te brengen. Voor scenario 1 en 2 beschrijven we een gedeelde roadmap, omdat de roadmap voor deze scenario's nagenoeg gelijk is, voornamelijk ten aanzien van het wetgevingstraject.

In het algemeen geldt dat de te nemen vervolgstappen enerzijds gaan over het maken van keuzes die in dit onderzoek niet zijn uitgewerkt en beoordeeld (bijvoorbeeld: de juridische keuze voor attribueren, delegeren en/of mandateren) en het nader onderzoek en uitwerken van sommige deelaspecten (bijvoorbeeld: de (mogelijke) wetten waarin een taak of bevoegdheid kan worden opgenomen).

In de onderstaande tabel staat een samenvatting per scenario over de te ondernemen acties:

Scenario	Wetgeving	Uitvoering
1. Centrale uitvoering	- Wetgeving voor de nationale domeincrawl (één wet/partij) - Wetgeving voor de loketfunctie (één of enkele wetten)	- Inrichting en beheer nationale domeincrawl - Loketfunctie inrichten, inclusief bijbehorende expertise, capaciteit en processen - Uitzoekpunten over werking loketfunctie ten aanzien van overdragen geharvest materiaal
2. Centrale en decentrale uitvoering	- Wetgeving voor de nationale domeincrawl (één wet/partij) - Wetgeving voor meerdere uitvoerende partijen (waarschijnlijk meerdere wetten, onzeker hoeveel en welke)	- Inrichting en beheer nationale domeincrawl - Inrichting en beheer beredeneerde selecties bij decentrale instellingen
3. Afspraken maken	Geen wetgeving	- Afspraken maken én beheren met belanghebbenden - Governance inrichten en regulier overleg (Beperkte betrokkenheid rijksoverheid bij deze acties)
4. Nulscenario	Geen wetgeving	Geen acties

Roadmap scenario 1 en 2

Wetgevingstraject

In scenario 1 en 2 is er sprake van een juridische route waarbij webharvesting wettelijk gefaciliteerd wordt. De belangrijkste stap is daarom het opstarten van een wetgevingstraject. De uitwerking en analyse van verschillende juridische routes in dit onderzoek vormen daarbij het startpunt, maar deze zullen in een wetgevingstraject nog verder geconcretiseerd moeten worden, waarbij er toegewerkt wordt naar een specifieke juridische oplossing.

Wij voorzien de volgende activiteiten in een wetgevingstraject:

- Start een wetgevingstraject om te komen tot een concrete juridische oplossingsrichting. Stel een startnotitie op met daarin de belangrijkste juridische routes.
 - Ga in de startnotitie in op de keuze voor attributie (van welke partij), eventueel in combinatie met mandatering/delegatie, en werk uit in welke relevante wetten de nieuwe publieke taak of bevoegdheid wordt toegevoegd. Mogelijke wetten zijn de Erfgoedwet en de Wet op het hoger onderwijs en wetenschappelijk onderzoek.
 - Onderbouw daarbij, mede op basis van dit onderzoek, welke combinatie van publieke taken en/of depotplicht nodig zijn.
 - Stel vast welke partij(en) webharvestingtaken kunnen uitvoeren, waarbij onderscheid gemaakt wordt tussen de nationale domeincrawl en beredeneerde selecties. Uit ons onderzoek is gebleken dat alleen de KB geïnteresseerd is in het uitvoeren van een nationale domeincrawl, en dat er al ongeveer elf partijen zijn die regelmatig een beredeneerde selectie uitvoeren (op basis van het netwerk digitaal erfgoed en hun register van webarchieven). Deze keuze hangt af van het gekozen scenario, waar we hieronder op ingaan bij de bespreking per scenario.
- Stem dit wetgevingstraject af met het Ministerie van Justitie en Veiligheid, specifiek op het vlak van de Auteurswet en de Uitvoeringswet AVG.
- Voer een Data protection impact assessment uit in het kader van het wetstraject ten behoeve van de bescherming van persoonsgegevens, in het bijzonder ten aanzien van de nationale domeincrawl.
- Voer gedurende het traject gesprekken met de Autoriteit persoonsgegevens en bereid een adviesaanvraag voor om de voorgenomen verwerking(en) aan hen voor te leggen.

Betrekken van belanghebbenden

Op basis van dit onderzoek zien we een belangrijke rol bij het betrekken van belanghebbenden in het wetgevingstraject, ook buiten de consultatiefase van een wetswijziging. De belanghebbenden zijn zowel de partijen die de urgentiebrief hebben gestuurd, erfgoedinstellingen die webharvesting willen uitvoeren, (auteurs)rechthebbenden en hun vertegenwoordigers, maatschappelijke organisaties en mogelijke toezichthouders. Dit kan parallel aan het wetgevingstraject worden gedaan en hoeft daar niet op te wachten.

- Organiseer een startbijeenkomst over het vervolgetraject, dat begint met een toelichting op de uitkomsten van dit onderzoek en de gekozen oplossingsrichting.
- Zorg voor reguliere herhaling van dergelijke bijeenkomsten in de opmaat tot uiteindelijke nieuwe wetgeving.

- Voorzie belanghebbenden zoveel mogelijk van informatie over de stand van zaken van het traject.

Beheersen van risico's en onzekerheden

Dit onderzoek heeft op een aantal vlakken nog risico's en onzekerheden in kaart gebracht ten aanzien van bijvoorbeeld de kosten of de uitvoerbaarheid. Het is van belang om deze risico's en onzekerheden te monitoren en waar mogelijk te reduceren door middel van verder onderzoek.

- Voer een uitvoeringstoets uit naar de uitvoerbaarheid van de voorgenomen juridische oplossing en het scenario voor uitvoering van webharvesting bij alle relevante instellingen.
- Voer een kostenraming uit wanneer duidelijker is welke partijen webharvesting gaan uitvoeren, en kijk daarbij in ieder geval naar de kosten voor opslag, aangezien dit onderzoek daar nog een bandbreedte voor heeft aangegeven die in de toekomst verder kan worden gespecificeerd.
- Maak afspraken over financiering van (opslag)kosten, ook gezien de voorziene stijging van deze kosten over de jaren heen.

Specifieke aspecten scenario 1

Scenario 1 bevat een aantal aspecten die specifiek uitgewerkt moeten worden:

- Kies een partij die de nationale domeincrawl gaat uitvoeren, omdat dit een specifieke (publieke) taak of bevoegdheid vraagt. De keuze voor wetgeving hangt af van de geselecteerde partij, maar zoals eerder gezegd lijkt de KB op basis van ons onderzoek de kandidaat hiervoor.
- Kies een partij die de loketfunctie gaat uitvoeren. Het is van belang dat deze partij voldoende expertise en uitvoeringscapaciteit heeft, in lijn met onze kostenraming voor dit scenario.
 - De eerdergenoemde uitvoeringstoets speelt hier ook een belangrijke rol bij, om vast te stellen onder welke condities het scenario uitvoerbaar is voor de loketorganisatie.
 - In dit scenario zijn er waarschijnlijk een beperkt aantal wetten die moeten worden aangepast; in het meest waarschijnlijke scenario enkel de wet voor de uitvoerende partij (die een nieuwe publieke taak of bevoegdheid krijgt).
- Selecteer partijen die beredeneerde selecties kunnen gaan uitvoeren. Ons advies is hierbij om aan te sluiten op de huidige uitvoeringspraktijk. De in dit onderzoek gesproken partijen hebben de suggestie gedaan om erfgoedinstellingen die vallen onder de archiefwet (regionale- en stadsarchieven) de mogelijkheid te geven om te harvesten, aangevuld met specifieke instellingen zoals Beeld en Geluid. Daarbij is nog niet gezegd dat deze partijen ook allemaal van deze mogelijkheid gebruik gaan maken, maar ze zouden wel een mogelijke behoefte hebben om dit te gaan doen.
- Werk de specifieke taken voor de loketfunctie uit, inclusief de in te richten processen en de eisen die aan deze processen gesteld worden (zoals reactie- en uitvoeringstermijnen).
 - Neem hierbij ook de verantwoordelijkheid/bevoegdheid van het loket om verzoeken tot harvesten te toetsen en mogelijk af te wijzen mee.
 - Een openstaand punt voor dit scenario is de mogelijkheid om materiaal over te dragen aan een verzoekende instelling, zodat deze dit op kan nemen in de collectie, kan preserven en kan presenteren via leesterminals. Dit scenario bevat namelijk een risico dat geharvest materiaal enkel op een centrale locatie kan worden ingezien nadat het geharvest is door het loket.

Specifieke aspecten scenario 2

Ook scenario 2 bevat een aantal aspecten die specifiek uitgewerkt moeten worden:

- Kies een partij die de nationale domeincrawl gaat uitvoeren, omdat dit een specifieke (publieke) taak vraagt. De keuze voor wetgeving hangt af van de geselecteerde partij, maar zoals eerder gezegd lijkt de KB op basis van ons onderzoek de kandidaat hiervoor.
- Selecteer partijen die beredeneerde selecties kunnen gaan uitvoeren. Ons advies is hierbij om aan te sluiten op de huidige uitvoeringspraktijk. De in dit onderzoek gesproken partijen hebben de suggestie gedaan om erfgoedinstellingen die vallen onder de archiefwet (regionale- en stadsarchieven) de mogelijkheid te geven om te harvesten, aangevuld met specifieke instellingen zoals Beeld en Geluid. Daarbij is nog niet gezegd dat deze partijen ook allemaal van deze mogelijkheid gebruik gaan maken, maar ze zouden wel een mogelijke behoefte hebben om dit te gaan doen.
- Het wetgevingstraject hangt af van de voorgaande stap: welke partijen webharvestingactiviteiten gaan uitvoeren bepaalt mede welke wetten er mogelijk moeten worden aangepast. In ons onderzoek zijn een aantal wetten genoemd die geraakt kunnen worden (hoewel de precieze impact moet blijken uit een wetgevingstraject):
 - De Wet op hoger onderwijs en wetenschappelijk onderzoek, waar onder andere de taak voor de KB is omschreven.
 - De Mediawet.
 - De Erfgoedwet, waarbij mogelijkheden kunnen zijn om een grotere groep aan (decentrale) instellingen een taak of bevoegdheid te geven.
 - De Archiefwet.
- In algemene zin is de verwachting dat in scenario 2 er meer wetten moeten worden aangepast dan in scenario 1.

Roadmap scenario 3

In het consensusmodel is er geen sprake van een wetgevingstraject. Wel is er noodzaak tot het maken van afspraken. Dit onderzoek doet geen uitspraken over de rol van de Rijksoverheid/OCW in het maken van deze afspraken. Desalniettemin hebben we een aantal acties geïdentificeerd die in dit scenario moeten worden uitgevoerd waarbij OCW een rol kan spelen.

Betrekken van belanghebbenden

Net zoals in de bovenstaande roadmap voorzien we op basis van dit onderzoek we een belangrijke rol bij het betrekken van belanghebbenden in het traject om te komen tot afspraken. De belanghebbenden zijn zowel de partijen die de urgentiebrief hebben gestuurd, erfgoedinstellingen die webharvesting willen uitvoeren, (auteurs)rechthebbenden en hun vertegenwoordigers, maatschappelijke organisaties en mogelijke toezichthouders.

- Organiseer een startbijeenkomst over het vervoltraject, dat begint met een toelichting op de uitkomsten van dit onderzoek en de gekozen oplossingsrichting.
- Zorg voor reguliere herhaling van dergelijke bijeenkomsten in de opmaat tot uiteindelijke nieuwe wetgeving.
- Voorzie belanghebbenden waar mogelijk van informatie over de stand van zaken van de gemaakte afspraken.

Maken en beheren van afspraken

De belangrijkste activiteit in dit scenario betreft het maken van algemene en specifieke afspraken. Algemene afspraken betreffen afspraken over webharvesting in het algemeen en specifieke afspraken gaan over specifieke sets van webdomeinen die worden geharvest. Bij het maken van beheren van afspraken is het de vraag of en zo ja welke rol OCW moet spelen; waarschijnlijk gaat het hier vooral om afspraken tussen de betrokken partijen (waarbij OCW geen partij is). De hieronder genoemde activiteiten moeten in dit licht gelezen worden.

- Bepaal welke partijen een rol hebben bij het vaststellen van afspraken met betrekking tot webharvesting en breng deze bij elkaar, in lijn met het voorgaande item.
- Stel een algemene set aan afspraken vast over het wat, hoe en wie (welke instellingen) van het uitvoeren van webharvesting en leg deze vast.
- Maak waar nodig specifieke afspraken over sets van webdomeinen, specifieke content en/of perioden met betrekking tot webharvesting en leg/stel deze vast zodat duidelijk en transparant is welke webharvestingactiviteiten worden uitgevoerd.
- Richt een notice-and-takedown procedure in voor rechthebbenden die bezwaar hebben tegen het harvesten van hun content.
- Zorg voor een periodieke verantwoording over webharvestingactiviteiten en publiceer deze.