

Handreiking webarchivering  
overheidswebsites: deel 1,  
publiekscommunicatie

Versie 0.1

Datum      6 december 2017  
Status      Concept

# CONCEPT

## Colofon

Contact F +31-70-331 5460  
[contact@nationaalarchief.nl](mailto:contact@nationaalarchief.nl)  
Postbus 90520  
2509 LM Den Haag

E-Docs nr #1265833

<b>Versiegeschiedenis</b>			
<i>Versie</i>	<i>Datum</i>	<i>Auteur</i>	<i>Beschrijving</i>
0.1	06-12-2017	Erik Saaman, Suzi Szabó, Jeroen van Luin, Antal Posthumus, Subhaas Kisoensingh	Eerste versie. Ter review door de direct betrokken bij het opstellen en de eerste implementaties van de handreiking.

# CONCEPT

## Inhoud

1	Inleiding.....	4
2	Doel en doelgroep .....	5
3	Scope .....	6
4	Waarom webarchivering.....	7
5	Rolverdeling .....	8
6	Stappenplan .....	12
7	Bijlage: Begrippen .....	19
8	Bijlage: Checklist voor het beschrijven van een website .....	28
9	Bijlage: Kwaliteitseisen geharveste website .....	29
10	Bijlage: Over deze handreiking .....	33
11	Bijlage: Bronnen .....	35

## 1 Inleiding

Overheidswebsites vallen onder de werking van de Archiefwet. Overheidsorganisaties zijn daarmee verplicht om hun websites te archiveren, als deze van voldoende belang zijn. Deze wettelijke zorgplicht is duidelijk, maar hoe werkt dit in de praktijk? Waarom is het nodig, wanneer is een website van voldoende belang, en hoe archiveer je die website dan? Waar begin je, welke onderdelen van de website neem je mee, hoe vaak en wie doet dat dan?

Over webarchivering leven vele vragen. Wat er vaak toe leidt dat websites niet, of niet goed worden gearchiveerd. Deze handreiking geeft daarom antwoorden op die vragen. Zodat overheidsorganisaties over de kennis kunnen beschikken om hun websites goed te archiveren.

In dit document wordt als eerste ingegaan op het doel en de doelgroep van deze handreiking, en wordt de gekozen scope-afbakening toegelicht. Vervolgens worden de verschillende rollen beschreven die betrokken zijn bij de archivering van overheidswebsites. Daarna volgt een beschrijving van alle stappen in het proces van webarchivering, vanaf de voorbereiding tot de uiteindelijke overbrenging van het webarchief naar een archiefbewaarplaats. In de bijlagen wordt nader ingegaan op de gehanteerde begrippen en bronnen. Daarnaast zijn in de bijlage de kwaliteitseisen opgenomen waar een webarchief (geharveste website) aan moet voldoen. Tot slot wordt in de bijlage nadere informatie gegeven over de aanleiding en opdracht voor deze handreiking.

Onder 'webarchivering' wordt verstaan alle handelingen die nodig zijn om ervoor te zorgen dat alle overheidswebsites voldoen aan de Archiefwet. Dat is te omvangrijk om allemaal in deze eerste versie van de handreiking webarchivering te behandelen. Daarom beperkt de handreiking zich vooralsnog tot websites die gebruikt worden voor publiekscommunicatie. Zie voor een uitgebreidere toelichting op de afbakening het hoofdstuk 'scope'.

Deze handreiking is een zogenaamde best practice. Dat wil zeggen dat het inzichten en werkwijzen beschrijft die zich in de praktijk hebben bewezen. De toepassing hiervan gebeurt op vrijwillige basis. Overheidsorganisaties hoeven deze handreiking niet te volgen, maar zullen dan wel op een andere manier, naar eigen inzicht, invulling moeten geven aan de verplichtingen uit de Archiefwet. Met deze handreiking wordt hen echter een gemakkelijk te volgen best practice geboden.

De handreiking webarchivering is gemaakt en wordt beheerd door het Nationaal Archief, in opdracht van de minister van OCW. Daarbij is nauw samengewerkt met een groot aantal belanghebbenden en experts. Op deze conceptversie zal nog een openbare review plaatsvinden. Met als doel dat alle geïnteresseerden de mogelijkheid hebben onduidelijkheden aan te geven en verbeteringen voor te stellen. Op de zorgvuldigheid van het ontwikkelproces wordt toegezien door de Standaardisatieraad van het Nationaal Archief. Hierin zijn alle lagen van de overheid en verschillende archiefinstellingen vertegenwoordigd. De ontstaansgeschiedenis van de handreiking wordt nader toegelicht in de bijlage 'Over deze handreiking'.

Archivering van websites is nog geen standaardpraktijk. De toepassing van deze handreiking zal in de praktijk tegen vragen en problemen oplopen. Op basis van die praktijkervaringen zal de handreiking verbeterd en uitgebreid worden. Daarvoor werkt het Nationaal Archief mee aan een aantal praktijkcasussen van webarchivering. Organisaties die een praktijkcasus willen inbrengen, kunnen daarvoor contact opnemen met [contact@nationaalarchief.nl](mailto:contact@nationaalarchief.nl).

# CONCEPT

## 2 Doel en doelgroep

### 2.1 Doel

Het doel van de handreiking webarchivering is:

Overheidsorganisaties beschikken over de kennis die ze nodig hebben om de websites waar ze verantwoordelijk voor zijn te archiveren en passen die kennis toe. Zodat dat de inhoud van overheidswebsites toegankelijk is en blijft, voor iedereen die daar belang bij heeft en voor zolang als noodzakelijk. Zoals bedoeld in de Archiefwet.

Voor dit doel stelt Het Nationaal Archief deze handreiking beschikbaar. Maar de doelstelling vraagt ook andere activiteiten: zoals doorontwikkeling van de inhoud en het bekend maken van overheidsorganisaties met de handreiking om het gebruik ervan te bevorderen.

Deze activiteiten worden gerekend onder het beheer van de handreiking en behoren tot de taken van het Nationaal Archief.

### 2.2 Doelgroep

De handreiking is van toepassing op alle overheidsorganen die vallen onder de werking van de Archiefwet<sup>1</sup>. Zowel bij de centrale als decentrale overheid. Hieronder vallen ook particuliere instellingen die wettelijke taken uitvoeren, voor zover het deze taken betreft. Waar van toepassing wordt onderscheid gemaakt naar type overheidsorgaan.

De handreiking is bedoeld voor de archiefvormers. Dat zijn de overheidsorganisaties die de inhoud van de websites maken (of laten maken). Het zijn de archiefvormers die in eerste instantie verantwoordelijk zijn voor de archivering van de websites. De handreiking is niet bedoeld voor organisaties die websites van andere organisaties archiveren, zoals archiefinstellingen en collectievormers. Zij krijgen er mogelijk wel mee te maken in hun contacten met de archiefvormers.

Binnen de doelgroep van archiefvormers zijn twee specifieke doelgroepen te onderscheiden:

- **Ontwerpers**  
De handreiking is primair bedoeld voor de ontwerpers van de informatiesystemen die overheidsorganisaties gebruiken voor het publiceren van informatie op websites. Zowel voor eigen systemen als inkoop van systemen en diensten. Onder 'ontwerper' wordt verstaan elke functionaris die mede bedenkt hoe deze informatiesystemen ingericht worden. Zoals informatiemanagers, informatieanalisten, functioneel beheerders, architecten en (inkoop)adviseurs.
- **Opdrachtgevers**  
Secundair is deze handreiking bedoeld voor de opdrachtgevers van deze informatiesystemen. Zodat zij in staat zijn onderbouwde beslissingen te nemen over de voorstellen van de ontwerpers.

<sup>1</sup> Zie <https://www.erfgoedinspectie.nl/toezichtvelden/archieven/wet--en-regelgeving/archiefwetgeving/reikwijdte>

### 3 Scope

Het is ondoenlijk om alle aspecten van webarchivering en alle soorten websites in een eerste versie van deze handreiking te behandelen. De scope van deze handreiking is daarom beperkt en wordt daarna gefaseerd uitgebreid.

De eerste versie van de handreiking beperkt zich tot webarchivering gedefinieerd als: het binnenhalen, bewaren en beschikbaar stellen van websites die bedoeld zijn voor publiekscommunicatie. Voor deze beperkte scope gaat de handreiking uit van de bewezen en betaalbare techniek van webharvesting (binnenhalen en bewaren van webpagina's zoals ze aan de gebruiker gepresenteerd worden).

Hiermee vallen vooralsnog buiten de scope van deze handreiking:

- Niet openbare delen van websites. Zoals een intranet of een mijn-pagina met persoonlijke informatie. Het binnenhalen en beschikbaar stellen hiervan compliceert de harvesting. Bovendien spelen hier aspecten van privacy die de mogelijkheden voor bewaren en beschikbaar stellen beperken.
- Gedrag (dynamiek) van webpagina's waarvoor interactie met de webserver nodig is. Zoals een zoekfunctie of streaming media. Dit gedrag kan niet gearchiveerd worden doormiddel van harvesting omdat daarbij de interactie met de server niet wordt meegenomen. Gedrag die geheel binnen een webbrowser wordt afgehandeld valt wel binnen de scope.
- Communicatie van websitegebruikers naar de overheid. Zoals het invullen en versturen van een formulier. Deze interactie kan niet gearchiveerd worden door alleen de website te bewaren.
- Webpagina's waarvan de inhoud bepaald wordt door gegevens over de gebruiker van de pagina (zoals bijvoorbeeld vastgelegd in cookies).
- Eisen aan de websites zelf die nodig zijn om ze goed te kunnen archiveren. Zoals het gebruik van open standaarden en het toevoegen van metadata. Het realiseren van deze eisen vergt aanpassingen aan de website zelf.
- Gerelateerde informatie die niet in de website zelf is opgenomen. Zoals de notulen van de webredactie of het functioneel ontwerp van de website. Deze informatie moet mogelijk wel gearchiveerd worden, maar harvesting is daarvoor niet geschikt.
- Social media die geen website zijn maar een eigen applicatie gebruiken. Zoals WhatsApp. Hiervoor zijn andere technieken nodig, die vaak ook verschillen per social media platform. Openbare social media die via een website te benaderen zijn (zoals Twitter) vallen wel binnen de scope (voor zover het platform zelf geen beperkingen stelt).

Deze scopebeperkingen betekenen niet dat bovenstaande aspecten niet onder webarchivering vallen. Het betekent alleen dat deze handreiking daarvoor niet bedoeld is. Archiefvormers moeten zelf bepalen hoe ze hiermee omgaan, binnen de kaders die door de Archiefwet gesteld worden.

## 4 Waaron webarchivering

Overheden gebruiken websites om informatie te verstrekken aan burgers, bedrijven, instellingen en andere overheden. Verschillende partijen kunnen er belang bij hebben dat deze informatie nog toegankelijk is nadat ze verstrekt of ontvangen is (en misschien al op de website gewijzigd of verwijderd is). Deze belangen zijn zeer divers van aard:

- *Uitvoering van bedrijfsprocessen*  
Bij het uitvoeren van taken kan het nodig zijn om oude webpagina's te raadplegen. Bijvoorbeeld voor het afhandelen van vragen van burgers en het opstellen van rapportages. Ook kan de inhoud van oude websites soms hergebruikt worden.
- *Verantwoording over overheidshandelen*  
Websites zijn uitingen van overheidshandelen. Als een overheidsorganisatie zich achteraf moet verantwoorden, dan moet de informatie beschikbaar zijn zoals het destijds openbaar werd gesteld. Bijvoorbeeld voor het beantwoorden van WOB-verzoeken, in een debat, bij een parlementair- of raadsonderzoek, bij een audit of een rekenkameronderzoek.
- *Recht- en bewijszoekende burgers*  
Burgers en bedrijven kunnen rechten ontlenen aan bepaalde informatie op websites. Bijvoorbeeld bij het aanvragen van diensten, vergunningen, subsidies, of het starten van ondernemingen. Burgers en bedrijven baseren keuzes en handelingen op de informatie die de overheid op openbare websites aanbiedt. Die informatie kan daarom benodigd zijn als bewijs in het geval van beroep- en bezwaarprocedures.
- *Historisch onderzoek*  
Overheidswebsites bieden onderzoekers inzicht in hoe overheden hun taken uitvoerden en zichzelf presenteerden en zijn daarmee ook een bron voor geschiedschrijving en cultureel erfgoed.

Om deze belangen te borgen worden websites gearchiveerd. Welke websites moeten worden gearchiveerd en voor hoe lang, wordt bepaald door een afweging van de aard van de website, de belangen die daarmee gemoeid zijn en de kosten. Het maken van deze business case is onderdeel van het archiveringsproces.

Overheidsorganisaties zijn op grond van de Archiefwet verplicht om hun websites te archiveren omdat zij de makers zijn van de informatie op de website. Daarnaast zijn er ook andere wetten waarin elementen van toegang tot informatie terugkomen, zoals de WOB (Wet openbaarheid bestuur) of de WBP (Wet bescherming persoonsgegevens).

## 5 Rolverdeling

De Archiefwet maakt elke overheidsorganisatie verantwoordelijk voor het archiveren van de informatie die gemaakt of ontvangen wordt bij het uitvoeren van haar werkprocessen. Dat geldt daarmee ook voor de websites waarvan de inhoud (de content) door, of in opdracht van, een overheidsorganisatie is gemaakt. Ongeacht wie de eigenaar is van het platform waarop de website beheerd wordt.

Deze wettelijke verantwoordelijkheid wil niet zeggen dat een overheidsorganisatie ook zelf de webarchivering moet uitvoeren. Dat kan zij bij een andere overheidsorganisatie of marktpartij beleggen. Maar ook bij in- of uitbesteding blijft de overheidsorganisatie altijd zelf eindverantwoordelijk, en dus aanspreekbaar, voor de archivering.

Hieronder worden de verschillende personen of organisaties beschreven die een rol spelen bij webarchivering. Deze rollen kunnen anders verdeeld worden als dat beter uitkomt. Met uitzondering van de rol van zorgdrager, die wettelijk is vastgelegd. Daarbij kunnen meerdere rollen ook bij één partij belegd worden.

### 5.1 Zorgdrager

*Wie:* Degene die volgens de Archiefwet zorg draagt voor de archivering binnen een overheidsorganisatie. In de Archiefwet wordt per type organisatie bepaald wie de zorgdrager is. Meestal is dat de bestuurder, het dagelijks bestuur of het bestuur. Voor een ministerie is het bijvoorbeeld de minister en voor een gemeente het college van burgemeester en wethouders. De archiefzorg betreft, onder andere, het stellen van regels en het beschikbaar stellen van financiële, personele en materiële hulpmiddelen.

De taken van de zorgdrager zijn binnen een organisatie vaak gemandateerd. Bijvoorbeeld bij de CIO of het hoofd bedrijfsvoering.

In het dagelijks taalgebruik wordt het begrip 'zorgdrager' ook vaak gebruikt om de organisaties aan te duiden die de zorg dragen voor de archivering van bepaalde informatie. Maar strikt genomen klopt dat niet.

*Rol:* De zorgdrager draagt de formele verantwoordelijkheid voor de archivering van de websites. Dat betekent bijvoorbeeld dat bewaartermijnen zijn bepaald in een vastgestelde selectielijst, verantwoordelijkheden zijn belegd en voldoende mensen en middelen beschikbaar zijn. Een zorgdrager zal zelden rechtstreeks betrokken zijn bij webarchivering, maar op een hoger niveau de algemene archieffunctie van een organisatie regelen. De contenteigenaar en archiefbeheerder hebben de uitvoerende verantwoordelijkheid voor de webarchivering.

### 5.2 Contenteigenaar

*Wie:* Degene die de inhoud van een website bepaalt, als onderdeel van het uitvoeren van haar taken. De contenteigenaar is de zogenaamde archiefvormer van de website.

*Rol:* De contenteigenaar is de initiatiefnemer voor het archiveren van een website en ziet erop toe dat dat goed gebeurt. Hij kan dit zelf doen of het uitbesteden, bijvoorbeeld aan de platformbeheerder of de archiefbeheerder. De contenteigenaren bepalen daarmee welke websites gearchiveerd worden.



# CONCEPT

## 5.3 Contentbeheerder

*Wie:* Degene die de inhoud van een website maakt en onderhoudt. Dit kan de contenteigenaar zelf zijn of een opdrachtnemer van de contenteigenaar.

*Rol:* De contentbeheerder zorgt voor de inhoudelijke en procesmatige metagegevens. Zoals de beschrijving van de website als geheel (zie bijlage 'Checklist voor websitebeschrijving') en de metagegevens van de individuele webpagina's. [De metagegevens van individuele webpagina's vallen buiten de scope van deze versie van de handreiking.]

## 5.4 Platformbeheerder

*Wie:* Degene die de techniek beheert waarmee een website wordt gemaakt en onderhouden. Dit wordt webhosting genoemd.

*Rol:* De platformbeheerder verstrekt in ieder geval de technische metagegevens over de website. Hij kan daarnaast de harvesting (laten) uitvoeren, als de contenteigenaar daar opdracht toe geeft. Dat past goed bij het beheren van de techniek van de website. Maar het is niet noodzakelijk, de contenteigenaar kan de archivering ook door een andere partij laten uitvoeren.

*Binnen het Rijk:* De Dienst Publiek en Communicatie (DPC) van het ministerie van Algemene Zaken is de platformbeheerder voor de meeste van de grote websites van de ministeries. DPC verzorgt, via een onderaannemer, de harvesting van de door haar gehoste websites. DPC is voornemens om als centrale inkooporganisatie harvesting ook aan te bieden voor rijkswebsites waarvan het niet de platformbeheerder is. In de loop van 2018 komt hier meer duidelijkheid over.

## 5.5 Harvestingdienst

*Wie:* Degene die de harvesting van een website verzorgt.

*Rol:* Harvesting van websites en het ontsluiten van geharveste websites. In opdracht van de contenteigenaar (direct of indirect).

## 5.6 Archiefbeheerder

*Wie:* Degene die verantwoordelijk is voor het beheer van het archief van een organisatie, tot het moment van overbrenging naar een archiefbewaarplaats. Wordt ook wel aangeduid als Documentaire informatievoorziening (DIV)

*Rol:* Het beheer van de webarchieven (geharveste websites) die door de harvestingdienst gemaakt zijn. Vernietiging en overbrenging van, respectievelijk, tijdelijk en blijvend te bewaren webarchieven. Een archiefbeheerder kan als harvestingdienst of in koper daarvan optreden, maar dat hoeft niet.

## 5.7 Archiefbewaarplaats

*Wie:* De voorziening die bij of krachtens de Archiefwet voor blijvende bewaring van archiefbescheiden aangewezen is. De overheid heeft verschillende archiefbewaarplaatsen:

- *Het Nationaal Archief.* Beheert archieven van landelijk opererende rijksoverheden. Zoals ministeries en ZBO's.

# CONCEPT

- *Regionaal Historische Centra (RHC's)*. Deze bevinden zich in de provinciehoofdsteden en zijn samenwerkingsverbanden tussen het Rijk, provincies, gemeenten en waterschappen in verschillende samenstellingen. Hier worden de archieven van regionaal opererende rijksoverheden (bijvoorbeeld rechtbanken) en van de aangesloten provincies, gemeenten en waterschappen beheerd.
- *Provinciale archiefbewaarplaatsen*. Op dit moment fungeren de Regionaal Historische Centra als provinciale archiefbewaarplaats. Het Nationaal Archief is de provinciale archiefbewaarplaats voor de provincie Zuid-Holland.
- *Gemeentelijke archiefbewaarplaatsen*. Gemeenten dienen evenals de provincies een archiefbewaarplaats aan te wijzen. Sommige gemeenten hebben een eigen archiefbewaarplaats, zoals de gemeente Amsterdam. Andere gemeenten werken hierbij samen met andere gemeenten of binnen een samenwerkingsverband.
- *Waterschapsarchiefbewaarplaatsen*. Hiervoor geldt hetzelfde als voor de gemeentelijke archiefbewaarplaatsen.

*Rol:* Het beheer van overgebrachte webarchieven. Inclusief het ontsluiten daarvan voor het publiek. Daarnaast verzorgen archiefbewaarplaatsen richtlijnen, best practices, voorlichting en advies over (web)archivering. Een archiefbewaarplaats kan op verzoek als archiefbeheerder optreden voordat een webarchief is overgebracht. Dit wordt uitplaatsing genoemd.

## 5.8 Collectievormer

*Wie:* Een instelling die vanuit zijn eigen doelstellingen informatie van anderen archiveert. Typisch zijn dit culturele en wetenschappelijke instellingen. Zoals de Koninklijke Bibliotheek (KB) en het Nederlands Instituut voor Beeld en Geluid (NIBG)

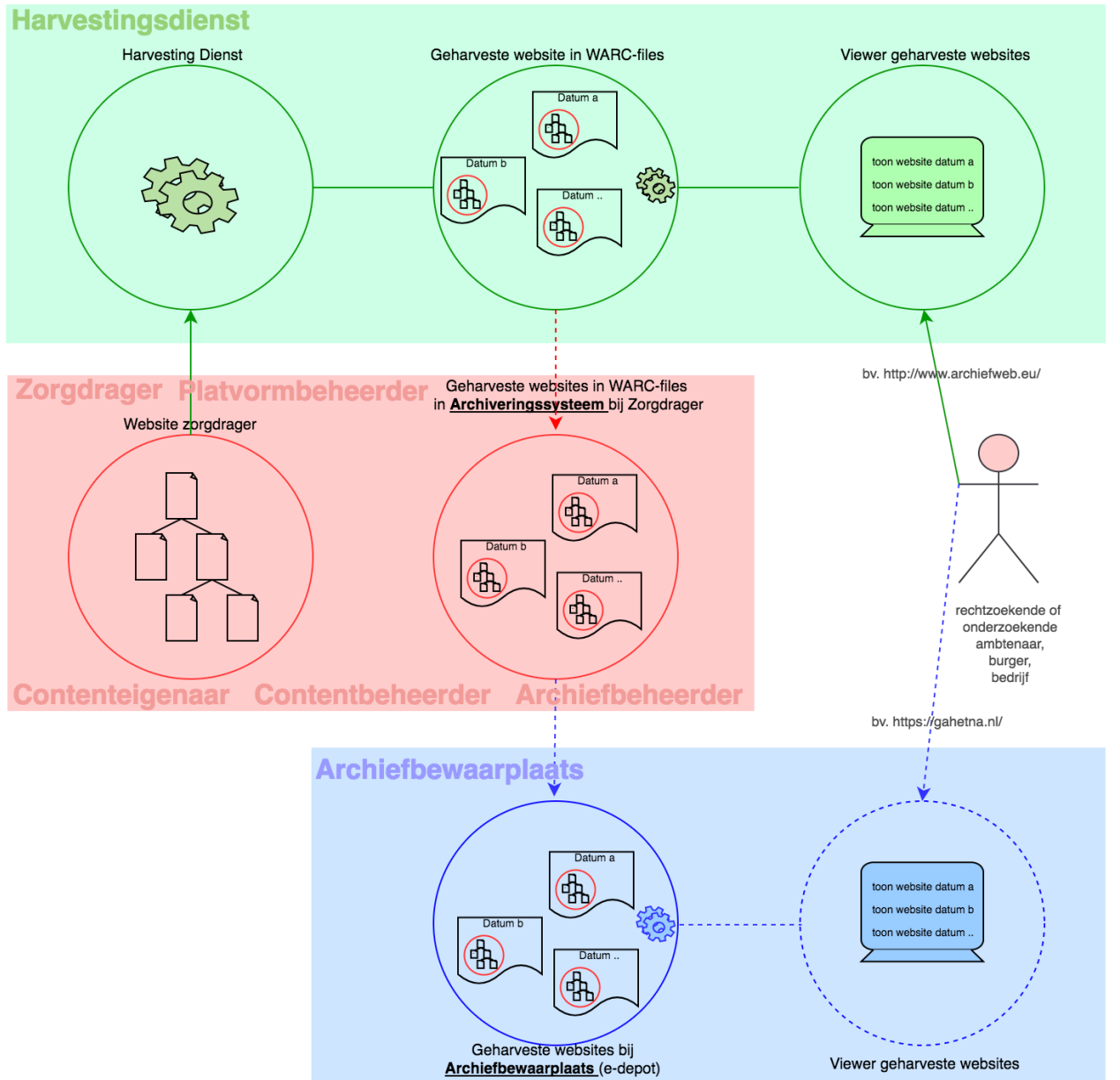
*Rol:* Collectievormers *kunnen* websites archiveren als dat in dienst staat van hun eigen doelstellingen. Zo bevat het KB-webarchief<sup>2</sup> een selectie Nederlandse websites – niet alleen van overheidsorganisaties- vanaf 2007. Deze websites worden één tot twee keer per jaar geharvest. Iedereen in het bezit van een KB-jaarpas kan het KB-webarchief in de leeszaal van de KB bekijken. NIBG archiveert websites van oproepen<sup>3</sup>.

Collectievormers archiveren niet voor de Archiefwet maar voor het invullen van hun eigen doelstellingen. Ze maken daarbij eigen afwegingen welke websites zij interessant en belangrijk vinden om te archiveren, met een zelfgekozen frequentie, selectie, bewaartermijn en kwaliteitskenmerken. Deze keuzes zullen zelden overeenkomen met de keuzes die voortvloeien uit de Archiefwet. Bovendien zijn de collectievormers geen verantwoording verschuldigd aan de zorgdragers. Dit alles maakt dat webarchivering door collectievormers niet in de plaats kan komen van de wettelijke webarchivering. Wel kunnen collectievormers hun soms jarenlange kennis en ervaring met webarchivering delen met overheidsorganisaties.

<sup>2</sup> Zie <https://www.kb.nl/bronnen-zoekwijzers/databanken-mede-gemaakt-door-de-kb/webarchief-kb>

<sup>3</sup> Zie <http://www.beeldengeluidwebarchief.nl>

# CONCEPT



## 6 Stappenplan

Onderstaand stappenplan is bedoeld als hulpmiddel voor een overheidsorganisatie om de archivering van haar websites goed in te richten. De stappen zijn bedoeld om in de aangegeven volgorde uit te voeren. Van de aangegeven invulling van de stappen kan worden afgeweken, mits op een nadere manier dezelfde doelen worden gerealiseerd.

### 6.1 Voorbereiding

*Actie:* Bepaal wie een rol spelen bij het realiseren van webarchivering. Informeer deze betrokkenen over het waarom, wat en hoe van webarchivering. Bijvoorbeeld in de vorm van een presentatie met discussie, op basis van de handreiking webarchivering. Behandel daarbij vragen en maak keuzes waar nodig.

*Doel:* Betrokkenen zijn voorbereid op de mogelijke impact van de archivering. Bijvoorbeeld op werkzaamheden, applicaties, rolverdeling, financiën en contracten. Zodat het stappenplan niet voor verrassingen zorgt die de uitvoering kunnen vertragen.

*Betrokkenen:* contenteigenaar, contentbeheerder, platformbeheerder, harvestingsdienst, archiefbeheerder, archiefbewaarplaats.

### 6.2 Inventarisatie

*Actie:* Maak een inventarisatie en beschrijving van alle websites (of subwebsites) waarvan de organisatie verantwoordelijk is voor de inhoud. Gebruik hiervoor de bijlage "Checklist voor het beschrijven van een website". Werk de inventarisatie periodiek bij.

Neem bij de inventarisatie ook de (sub)websites mee waarvan de overheidsorganisatie wel verantwoordelijk is voor de inhoud, maar niet voor het platform. Zoals de webpagina's van Twitter met tweets van de overheid.

*Doel:* Met de inventarisatie wordt de nodige informatie verzameld om te bepalen welke websites gearchiveerd moeten worden en wie daarbij betrokken moeten worden. De beschrijving bevat de informatie die de harvestingsdienst nodig heeft voor het uitvoeren van het archiveren. En die contenteigenaar nodig heeft om de kwaliteit van de webarchieven te (laten) controleren. Tegelijkertijd bevat de inventarisatie een beschrijving van de websites die in de toekomst helpt om de websites te vinden en te interpreteren.

*Betrokkenen:* contenteigenaar, contentbeheerder, platformbeheerder, archiefbeheerder, archiefbewaarplaats.

### 6.3 Bewaartermijn

*Actie:* Bepaal de bewaartermijn voor de websites in de inventarisatie en pas waar nodig de selectielijst aan op het moment dat deze periodiek geactualiseerd wordt.. Aangeraden wordt alle websites voor publiekscommunicatie<sup>4</sup> blijvend te bewaren.

<sup>4</sup> De scope van deze handreiking is vooralsnog beperkt tot publiekscommunicatie.

# CONCEPT

Dat geldt voor alle webpagina's en daarvan alle versies (bijvoorbeeld bij wijzigingen) en alle varianten (bijvoorbeeld voor verschillende talen en verschillende devices).

*Doel:* Duidelijkheid verschaffen welke webpagina's gearcheveerd worden en hoe lang de gearcheveerde webpagina's bewaard worden. De selectielijsten zijn hiervoor het wettelijke instrument.

*Toelichting:*

De keuze van de termijn hoe lang een webpagina bewaard wordt nadat deze binnen een website is verwijderd of gewijzigd<sup>5</sup>, bepaalt welke webpagina's gearcheveerd worden en hoe lang deze daarna bewaard worden. Deze bewaartermijn moet zodanig gekozen worden dat enerzijds de belangen bij het toegankelijk houden van webpagina's voldoende geborgd zijn, en anderzijds de inspanningen en kosten daarvoor acceptabel zijn.

Aangeraden wordt om websites voor publiekscommunicaties blijvend te bewaren (d.w.z. een oneindige bewaartermijn) omdat ze gelden als een informatieknoppunt zoals bedoeld in de nieuwe waarderingsmethodiek. Dat wil zeggen dat websites onderdeel zijn van de formele communicatiestructuur en tevens onderdeel van de manifestatie in de maatschappij van een overheidsorgaan. Daarnaast hebben websites beeldwaarde. Ze laten zien hoe de overheid zich door de jaren heen presenteert.

Aangeraden wordt om binnen één website alle webpagina (met bijlagen) en daarvan alle versies ven lang te bewaren. Ten eerste omdat differentiatie in bewaartermijnen onvolledig gearcheveerde websites geeft. Dit bemoeilijkt het gebruik in de toekomst.

Daarnaast is het een eenvoudig en duidelijk criterium. Meer verfijnde criteria voor bewaartermijnen leidt al snel tot misverstanden en daardoor de verkeerde toepassing. Onderscheid in bewaartermijnen kan extra werk met zich mee brengen bij het selecteren en vernietigen of er zelfs toe leiden dat er helemaal niets bewaard wordt omdat het als toepassing van het onderscheid als te ingewikkeld wordt gezien.

Tegen blijvend bewaren van alle webpagina's kan ingebracht worden dat het tot grote en daardoor mogelijk te kostbare webarchieven leidt. Vooralsnog zijn daar geen aanwijzingen voor, maar het is ook niet uit te sluiten. Het is daarom belangrijk de komende jaren de omvang en kosten van webarchieven goed in de gaten te houden. Mocht blijken dat de kosten te hoog worden, dan kan alsnog een differentiatie worden gemaakt in de bewaartermijnen. Tot die tijd wordt de veilige keuze gemaakt om alle webpagina's blijvend te bewaren.

De keuze voor blijvend bewaren sluit aan bij het 'Generiek Waarderingsmodel Rijksoverheid 2.0'<sup>6</sup> (GWR) en de 'Selectielijst gemeenten en intergemeentelijke organen 2017'<sup>7</sup> die ook niet differentiëren in de bewaartermijn. Gezamenlijk gelden de GWR en de gemeentelijke selectielijst als richtlijn voor het grootste groot deel van de overheid. Het GWR geeft aan dat "eindproducten van publieksvoorlichting" (websites vallen hieronder) blijvend bewaard moeten worden (categorie 5.1). De gemeentelijke selectielijst geeft aan dat "onlinecontent ten behoeve van externe communicatie" blijvend bewaard moet worden (nr. 20.1.2). Beide lijsten maken

<sup>5</sup> Merk op dat een wijziging van een webpagina doorgaans betekent dat de oude versie van de site wordt verwijderd. Daarom moet ook voor gewijzigde pagina's een bewaartermijn bepaald worden.

<sup>6</sup> Zie [https://www.earonline.nl/images/earpub/a/a7/Generiek\\_Waarderingsmodel\\_Rijksoverheid\\_2\\_0\\_def\\_pdf.pdf](https://www.earonline.nl/images/earpub/a/a7/Generiek_Waarderingsmodel_Rijksoverheid_2_0_def_pdf.pdf)

<sup>7</sup> Zie <https://vng.nl/files/vng/publicaties/2017/20170706-selectielijst-gemeenten-intergemeentelijke-organen-2017.pdf>

# CONCEPT

geen verder onderscheid in deze categorieën, daarom geldt het blijvend bewaren voor alle webpagina's en voor alle versies.

De bewaartermijnen van overheidsinformatie worden vastgelegd in een wettelijk verplichte selectielijst (zie toelichting op het begrip 'selectielijst'). Omdat archivering van websites nog geen standaardpraktijk is, zullen de meeste selectielijsten nog niet expliciet beschreven hebben wat de bewaartermijnen voor websites zijn. Neem daarom contact op met de beheerder van de selectielijst (meestal de afdeling Documentaire Informatievoorziening) om samen te bepalen wat de bewaartermijn is. Pas, waar nodig, de selectielijst aan op het moment dat deze periodiek geactualiseerd wordt.

Er kan een praktisch probleem optreden bij het archiveren van alle versies van een webpagina. Webpagina's kunnen namelijk op een willekeurig tijdstip veranderd of verwijderd worden. Zelfs meerdere keren op één dag. Periodieke harvesting van websites vindt daarentegen op vaste tijdstippen plaats, waarbij alleen de laatste wijziging tussen twee opeenvolgende harvests worden opgemerkt en vastgelegd. Aangeraden wordt daarom zo vaak als mogelijk te harvesten, maar minimaal één keer per dag. Nog beter is real time harvesting, waarbij elke wijziging in een website door het beheerplatform meteen wordt doorgegeven aan de harvestingdienst zodat deze de wijziging meteen kan vastleggen. Real time signalering vereist echter waarschijnlijk technische aanpassingen aan het beheerplatform, hetgeen buiten scope van deze handreiking valt.

*Betrokkenen:* zorgdrager, archiefbeheerder.

## 6.4 Contract harvestingdienst

*Actie:* Sluit een contract af met een leverancier van een harvestingdienst of richt er zelf een in. De harvestingdienst zorgt voor het harvesten van alle door de opdrachtgever aangemelde websites en de toegang tot de opgebouwde webarchieven.

*Doel:* De harvestingdienst voert het binnenhalen van de websites uit en zorgt ervoor dat de geharveste websites voor iedereen toegankelijk zijn. Een goed contract zorgt ervoor dat duidelijk is wat de rolverdeling is en aan welke eisen de dienst moet voldoen.

### *Toelichting*

Harvesting vindt plaats op basis van door de opdrachtgever bepaalde lijst met te harvesten websites. Waarin per website is vastgelegd:

- *De site-URL*  
Deze bepaalt welke webpagina's en bijlagen tot de website gerekend worden. Plus eventuele secundaire site-URL's voor webpagina's en bijlagen met een afwijkende URL, maar die wel tot de website gerekend worden.
- *De seed-URL*  
Deze bepaalt met welke webpagina de harvesting van een site begint.
- *De weespagina's*  
Een lijst van webadressen van webpagina's die tot de website behoren, maar die niet door middel van het volgen van hyperlinks vanaf de seed-URL bereikbaar zijn.
- *De blacklist*  
De lijst met webadressen van webpagina's die tot de website behoren, maar die niet geharvest moeten worden. Meestal om technische redenen.

# CONCEPT

De gecontracteerde harvestingdienst moet aan onderstaande vereisten voldoen:

1. De door opdrachtgever aangemelde websites worden periodiek geharvest. Minimaal één keer per jaar volledig, daarna dagelijks incrementeel.
2. Alle webpagina's en bijlagen worden gearcheveerd, voor zover:
  - a. ze vanaf de seed-URL of een weespagina bereikbaar zijn (ongeacht het aantal hops en de paddiepte);
  - b. ze tot de website gerekend worden (dat wil zeggen: het webadres begint met de site-URL of een van de secundaire site-URL's, er wordt geen "scope + 1" gehanteerd);
  - c. ze niet op de blacklist staan.
3. Alle bestanden die nodig zijn om een webpagina weer te kunnen geven worden gearcheveerd. Ook als deze door een andere webserver worden geleverd dan de webpagina zelf.
4. De opdrachtgever is en blijft volledig eigenaar van de geharveste websites. De leverancier kan hier geen rechten op uitoefenen.
5. Harvesting gaat niet ten koste gaan van het functioneren van de websites. De leverancier moet hiervoor vooraf kenbaar maken aan welke eisen de te harvesten websites moeten voldoen.
6. De geharveste sites worden aan de opdrachtgever beschikbaar gesteld in de vorm van WARC-bestanden die voldoen aan de kwaliteitseisen zoals beschreven in de bijlage 'Kwaliteitseisen webarchivering'.
7. De geharveste websites zijn (gedurende de contractperiode) voor iedereen online toegankelijk. Waarbij in ieder geval de volgende functies geboden worden (voor alle gangbare webbrowsers):
  - a. Full tekst-zoeken door alle geharveste webpagina's en bijlagen.
  - b. Van elke webpagina of bijlage kan de versie van een specifieke datum volledig worden weergegeven, zoals de originele website of bijlage op die datum zou worden weergegeven. Met uitzondering van server-side gedrag (deze kan niet achteraf worden gereconstrueerd uit een geharveste website).
  - c. Van zowel de website als elke individuele webpagina en bijlage kan metadata worden opgevraagd.
8. [Optioneel] Het op basis van aangeleverde WARC-files toegang geven tot eerder door een andere harvestingdienst geharveste websites.

Een contract met een harvestingdienst kan afgesloten worden voor een of meerdere websites. Hierbij kunnen overheidsorganisaties besluiten om samen in te kopen. Bijvoorbeeld door middel van een raamovereenkomst. In principe kunnen organisaties die al websites archiveren voor eigen doeleinden (zogenaamde collectievormers) de harvestingdienst leveren, mits ze daarbij voldoen aan de bovenstaande voorwaarden. Een overheidsorganisatie kan ook zelf een harvestingdienst inrichten, als het beschikt over de daarvoor benodigde mensen en middelen. De handreiking webarchivering geeft geen verdere aanwijzingen voor het inrichten van een harvestingdienst.

Het contract voor de harvestingdienst kan eventueel gecombineerd worden met de hostingdienst voor de website. Ook is het mogelijk de toegang tot de geharveste websites in een apart contract onder te brengen.

*Betrokkenen:* zorgdrager, contenteigenaar, harvestingdienst.

# CONCEPT

## 6.5 Start harvesting

*Actie:* Meld alle websites uit de inventarisatie aan bij de harvestingdienst, zodat harvesting start. Als er een website bij komt, meld deze dan meteen aan. Als een website wordt beëindigd, meld deze dan af.

Breng de functioneel beheerders van websites vooraf op de hoogte van de aanstaande harvesting. Zodat zij, waar nodig, maatregelen kunnen nemen om volledige harvesting mogelijk te maken zonder dat de website daar last van heeft.

Neem op de website een verwijzing op naar het archief, zodat bezoekers die voorgaande versies van de website willen bekijken deze gemakkelijk kunnen vinden.

*Doel:* websites worden bewaard en bewaarde sites worden beschikbaar gesteld. Voorkomen wordt dat de websites slecht functioneren door het harvesten of dat harvesting niet volledig kan worden uitgevoerd door configuratie van de website.

*Betrokkenen:* contenteigenaar, contentbeheerder, platformbeheerder, harvestingdienst.

## 6.6 Kwaliteitscontrole

*Actie:* Controleer de WARC-files periodiek op kwaliteit of laat deze door een kundige en onafhankelijke derde partij controleren. Gebruik hierbij de kwaliteitseisen die in deze handreiking zijn beschreven en de door de leverancier van de harvestingdienst opgeleverde lijst met niet-geharveste URL's. De leverancier moet geconstateerde onjuistheden meteen verhelpen en veranderingen doorvoeren om herhaling te voorkomen. Als de harvestingdienst blijvend niet aan de kwaliteitseisen voldoet, dan wordt het contract beëindigd en een nieuwe harvestingdienst gecontracteerd.

*Doel:* Zeker stellen dat de WARC-files voldoen aan de kwaliteitseisen. Zonder dat daar achteraf kostbare herstelacties voor nodig zijn of herstel zelfs onmogelijk is.

*Toelichting:* WARC-files die niet aan de kwaliteitseisen voldoen zijn mogelijk inhoudelijk onvolledig of onjuist, of technisch niet goed te beheren en ontsluiten. Fouten zijn achteraf vaak moeilijk of helemaal niet meer te herstellen. Daarom is tijdige signalering van groot belang. De leverancier van de archiveringsdienst is niet verantwoordelijk voor het valideren van alle documenten en bestanden die als bijlage in de website aanwezig zijn.

*Betrokkenen:* contenteigenaar, harvestingdienst, (archiefbeheerder).

## 6.7 Bewaar WARC-files

*Actie:* Direct na elke harvest worden de bijbehorende WARC-files (met bijbehorende metadata over de website en het harvestingproces) en eventuele lijst met niet-geharveste URL's beschikbaar gesteld aan archiefbeheerder van de contenteigenaar. De archiefbeheerder bewaart de WARC-files in één van zijn archiveringsystemen (bijvoorbeeld een DMS of e-depot). De WARC-files worden door de archiefbeheerder op dezelfde manier als andere digitale archieven beheerd.

*Doel:* De WARC-files worden beheerd volgens dezelfde normen als de andere digitale archieven van de zorgdrager.

*Toelichting*

Beheer van de WARC-files door de archiefbeheerder is vereist om te garanderen dat ze bewaard worden volgens de eisen die de zorgdrager stelt aan zijn archiefbeheer.



# CONCEPT

Maak vooraf goede afspraken tussen de harvestingsdienst en de archiefbeheerder. De opname van de WARC-files in het archiveringssysteem verloopt dan soepel.

Advies is om bij metadatering van de WARC-bestanden in het archiveringssysteem de Richtlijn Metagegevens Overheidsinformatie te volgen, zodat alle metadata aanwezig is om later de metadata volgens het Toepassingsprofiel Rijksoverheid of TMLO aan te kunnen leveren.

Het is mogelijk het bewaren en beheren van de WARC-files uit te besteden. Bijvoorbeeld aan de harvestingsdienst (want die beschikt immers toch al over de files) of aan de platformbeheerder. In dat geval moet de contenteigenaar aanvullende eisen aan de beheerder van de WARC-files stellen. Om ervoor te zorgen dat het beheer van de files plaatsvindt volgens de normen van de zorgdrager. Neem daarvoor contact op met de archiefbeheerder, die is bekend met deze normen. Aangeraden wordt om in dat geval in het archiveringssysteem wel een verwijzing (met metadata) naar de WARC-files op te nemen zodat het bestaan van de WARC-files bekend is in het archiveringssysteem. De eisen ten aanzien van het bewaren van WARC-files vallen buiten het bestek van deze handreiking. Omdat deze ervan uitgaat dat de archiefbeheerder de WARC-files bewaart, volgens normen die daarvoor al zijn vastgesteld.

*Betrokkenen:* harvestingsdienst, archiefbeheerder.

## 6.8 Beëindiging website

*Actie:* Zodra de website beëindigd wordt en de inhoud dus niet meer wijzigt, stop dan de harvestingsdienst. Het kan voorkomen dat een website voor harvestingsdienst aangemeld wordt terwijl hij al beëindigd is. In dat geval is slechts eenmalig een volledige harvestingsdienst noodzakelijk en kan incrementele harvestingsdienst achterwege blijven.

*Doel:* Voorkomen onnodige harvestingsdienst.

*Betrokkenen:* contenteigenaar, contentbeheerder, platformbeheerder, harvestingsdienst.

## 6.9 Overbrengen

*Actie:* Tenminste jaarlijks worden door de archiefvormer (de eigenaar van de inhoud van een website) voor de jaargang waarvan de overbrengingstermijn is verlopen, de betreffende archiefbestanden overgebracht naar een archiefbewaarplaats. Maak met de archiefbewaarplaats afspraken over de overbrengingstermijn.

*Doel:* Overbrenging zorgt ervoor dat het beheer van de blijvend te bewaren webarchieven veilig wordt gesteld en daarmee niet afhankelijk is van de continuïteit en expertise van de archiefvormer.

*Toelichting:* Zie paragraaf 7.12 voor een nadere uitleg over overbrengen

Bij overbrenging van de archiefbeheerder naar de archiefinstelling moet het webarchief worden aangeleverd volgens de eisen gesteld door de ontvangende archiefinstelling (voor het Nationaal Archief is dat het Voorwaarden Export-document). Neem al bij het inrichten en wijzigen van de webarchivering contact op met de betreffende archiefinstelling om zeker te stellen dat de webarchieven bij overbrenging op de juiste manier aangeleverd kunnen worden.

*Betrokkenen:* Zorgdrager, archiefbeheerder, archiefbewaarplaats.

# CONCEPT

## **6.10 Vernietigen**

Als er, in afwijking van deze handreiking, voor wordt gekozen een deel van de webarchieven niet blijvend te bewaren, dan moeten deze na afloop van de bewaartermijn vernietigd worden door de archiefbeheerder. Dat betekent dat het technisch onmogelijk gemaakt wordt om de betreffende webpagina's nog te raadplegen. De handreiking gaat daar verder niet op in, omdat ze ervan uitgaat dat alle publieke websites in zijn geheel blijvend bewaard worden. Het plannen en uitvoeren van de vernietiging van websites die niet blijvend bewaard zullen worden is een vervolgstap na de inventarisatie van wel en niet te archiveren websites door de archiefvormer.

## 7 Bijlage: Begrippen

### 7.1 Webpagina

Een webpagina is een document in de vorm van (X)HTML-code bedoeld om met een webbrowser te bekijken.

Een webpagina heeft een webadres in de vorm van een URL en wordt via het HTTP(S)-protocol beschikbaar gesteld door een webserver. Kenmerkend voor een webpagina is dat deze veelal hyperlinks (verwijzingen in de vorm van een URL) bevat naar bijlagen. Daarmee vormen alle webpagina's gezamenlijk een 'web' van informatie waarbinnen gemakkelijk genavigeerd kan worden.

Een webpagina wordt door de webserver beschikbaar gesteld in de vorm van een of meerdere bestanden. De basis daarvan is een bestand met (X)HTML-code. Daarnaast kunnen er aanvullende bestanden zijn voor de verschillende componenten van de webpagina. Zoals voor afbeeldingen, scriptcode en style sheets. Om een indruk te krijgen van alle bestanden waaruit een webpagina bestaat kan men de complete webpagina vanuit een webbrowser opslaan.

Binnen een webpagina worden de volgende aspecten onderscheiden:

- *Inhoud*  
De inhoud van een webpagina wordt gevormd door tekst en andere media zoals afbeeldingen, filmpjes, audio en animaties. Deze media kunnen in een afzonderlijk bestanden opgeslagen zijn (ieder met een eigen webadres).
- *Vormgeving*  
Hoe de inhoud door de browser wordt weergegeven wordt bepaald door de gebruikte HTML-elementen (bijvoorbeeld '<b>vette tekst</b>') voor vetgedrukte tekst) en de Cascading Style Sheets (CSS). In een CSS wordt de vormgeving van HTML-elementen vastgelegd. Zoals lettertype, lettergrootte, kleurgebruik, pagina-indeling. De CSS kan onderdeel zijn van een webpagina of als een losstaand bestand zijn opgeslagen (met een eigen webadres).
- *Gedrag*  
Met gedrag wordt bedoeld de manier waarop de weergave van een webpagina in een browser door de tijd verandert. Vaak onder invloed van interactie met de gebruiker. Bijvoorbeeld menu's, 'drag and drop', formulieren en interactieve kaarten. Er zijn twee manieren waarop het gedrag van een webpagina wordt vastgelegd:
  - *Client-side*: Het gedrag kan volledig door de browser worden uitgevoerd zonder dat deze aanvullende informatie van een webserver moet krijgen. Bijvoorbeeld door het uitvoeren van JavaScript of Dynamic HTML.
  - *Server-side*: Het gedrag is (deels) bepaald door aanvullende informatie die de browser bij een webserver moet ophalen. Bijvoorbeeld bij het tonen van de laatste tweet van de organisatie. Client-side gedrag kan gearchiveerd worden door alleen de webpagina te bewaren. Bij server-side gedrag kan dat niet.

Vaak zal een overheidswebpagina op een webserver van een overheidsorganisatie staan, maar dat hoeft niet. Het kan ook een webserver van een bedrijf of andere niet-overheidsorganisatie zijn. Al dan niet in opdracht van een overheidsorganisatie.

# CONCEPT

Dit doet niks af aan de wettelijke verantwoordelijkheid voor het archiveren van de webpagina. Als een overheidsorganisatie verantwoordelijk is voor de inhoud, is ze ook verantwoordelijk voor de archivering. Zo bevat [twitter.com/Rijksoverheid](https://twitter.com/Rijksoverheid) overheidsinformatie omdat de Tweets door de overheid zijn geplaatst. Daarmee valt (de inhoud van) [twitter.com/Rijksoverheid](https://twitter.com/Rijksoverheid) onder de Archiefwet. En zal de eigenaar van dit Twitter-account, zo nodig, deze tweets moeten archiveren.

## 7.2 Bijlage (bij een webpagina)

Een bijlage bij een webpagina is een bestand (anders dan een andere webpagina) waarnaar een hyperlink vanuit de webpagina verwijst. Zoals een Pdf- of videobestand.

## 7.3 URL (Uniform Resource Locator)

Een URL is het adres van een webpagina of bijlage.

Het eerste deel van een URL bevat de naam van het protocol dat gebruikt wordt om toegang te krijgen tot de locatie (zoals "http://" of "ftp://"). Het tweede deel bevat de naam van de locatie, het IP-adres of de Domeinnaam.

## 7.4 Webbrowser

Een webbrowser is een computerprogramma om webpagina's (en vaak ook hun bijlagen) te bekijken. Populaire browsers zijn Internet Explorer, Microsoft Edge, Mozilla Firefox, Google Chrome, Safari en Opera.

Merk op dat de weergave van webpagina's door een webbrowser niet gestandaardiseerd is. Dat betekent dat elke webbrowser een andere weergave kan geven van dezelfde webpagina. Dat kan zelfs verschillen per versie van die browser en per apparaat waarop de browser draait. Veelal zullen de verschillen niet groot zijn. Maar het is niet uitgesloten dat er essentiële verschillen zijn die van invloed zijn op de interpretatie van een webpagina door een gebruiker van de browser. Het is niet te voorzien hoe deze verschillen zich in toekomstige versies van de browsers ontwikkelen. Dat is een uitdaging voor vormen van webarchivering waarbij alleen de webpagina's bewaard worden en niet de browsers. Dan bestaat het risico dat de toekomstige weergave van een gearchiveerde webpagina op essentieel punten anders is dan bij een eerdere weergave.

## 7.5 Website

Een website is een verzameling samenhangende webpagina's met bijbehorende bijlagen.

Bij webarchivering is het van belang om exact te beschrijven welke webpagina's en bijlagen tot de website gerekend worden. Omdat alleen die worden meegenomen in het archief van die website. Meestal wordt de afbakening van een website bepaald door een site-URL (zoals "https://nos.nl/"). Alle webpagina's en bijlagen waarvan het webadres begint met de site-URL vallen dan onder de betreffende website.

Het komt voor dat het webadres van bijlagen van een website niet met de site-URL beginnen. Bijvoorbeeld omdat ze op een speciale server worden bewaard. In dat geval zijn er aanvullende URL's nodig om vast te leggen welke bijlagen tot de site gerekend worden.

## 7.6 **Webserver/Webclient**

Een webserver is een computer die via een netwerk aan andere computers webpagina's en bijlagen verschaft. Een webclient is een ontvangende computer.

Om een webpagina (of bijlage) op te vragen stuurt de client een request (verzoek) naar de webserver. De server reageert daarop met een response (reactie). Die response kan de opgevraagde webpagina bevatten of een melding waarom de webpagina niet geleverd wordt. Daarnaast bevat de response metagegevens over het afhandelen van het request.

## 7.7 **Content Management Systeem (CMS)**

Een CMS is de software waarmee webpagina's worden gemaakt en beheerd door een webredactie.

## 7.8 **Archiveren**

Archiveren betekent het duurzaam toegankelijk maken en houden van informatie. Zodat het gebruikt kan worden, zolang dat nodig is.

De Archiefwet kent geen expliciete definitie van het begrip archiveren. Maar deze zit impliciet verwerkt in Artikel 3:

De overheidsorganen zijn verplicht de onder hen berustende archiefbescheiden in goede, geordende en toegankelijke staat te brengen en te bewaren, alsmede zorg te dragen voor de vernietiging van de daarvoor in aanmerking komende archiefbescheiden.

Waarin archiefbescheiden alle informatie is die door een overheidsorgaan is gemaakt of ontvangen bij het uitvoeren van zijn taken, ongeacht de vorm van die informatie. Overheidswebsites vallen daar ook onder.

Het "in goede, geordende en toegankelijke staat te brengen en te bewaren" wordt doorgaans aangeduid met het *duurzaam toegankelijk* maken. Waarbij met toegankelijk wordt bedoeld:

- Vindbaar
- Beschikbaar
- Leesbaar
- Interpreteerbaar
- Betrouwbaar

En met duurzaam dat de toegankelijkheid van de informatie bestand is tegen veranderingen van elke aard. Zoals veranderingen in informatiedragers en formaten, maar ook veranderingen in organisaties of gebruikerswensen. Zie voor een nadere toelichting op 'duurzaam toegankelijk' de DUTO-wiki<sup>8</sup>.

## 7.9 **Archiveringssysteem**

Een archiveringssysteem is een informatiesysteem met als functie het archiveren van informatie. Een archiveringssysteem wordt ook wel archiefsysteem genoemd.

<sup>8</sup> Zie [https://wiki.nationaalarchief.nl/pagina/DUTO:Waarom\\_DUTO%3F](https://wiki.nationaalarchief.nl/pagina/DUTO:Waarom_DUTO%3F)

# CONCEPT

Onder informatiesysteem wordt verstaan het geheel van bestanden, procedures, apparaten en daarbij benodigde hulpmiddelen, ingericht door een persoon, groep personen of organisatie ten behoeve van de uitvoering van zijn of haar taken. Een archiveringssysteem omvat daarmee alles wat nodig is om archivering mogelijk te maken.

De applicaties die gebruikt worden in een digitaal archiveringssysteem kennen verschillende vormen. Zoals een DMS, RMA of een e-depot.

Merk op dat archivering mogelijk is in elk informatiesysteem. Mits deze voldoet aan de eisen van duurzame toegankelijkheid. Archivering is dan een nevenfunctie van het informatiesysteem. Met een archiveringssysteem wordt een informatiesysteem bedoeld waarvan archivering de hoofdtaak is.

## 7.10 Archiefbewaarplaats

Een archiefbewaarplaats is een bij of volgens de Archiefwet 1995 aangewezen bewaarplaats die het beheer van blijvend te bewaren overheidsinformatie als taak heeft en daarin gespecialiseerd is. Dat zijn onder andere het Nationaal Archief voor de centrale overheid, en voor regionale en lokale overheden de RHC's (Regionaal Historisch Centra) en de gemeentelijke archiefinstellingen. Na overbrenging van de WARC-files is de archiefbewaarplaats verantwoordelijk voor het beheer van de archiefbestanden, inclusief de toegang tot de archieven. Instructies voor de archiefbewaarplaatsen vallen buiten het bestek van deze handreiking, omdat deze alleen gericht is op de archiefvormers.

## 7.11 Selectielijst

Een selectielijst is een formeel document waarin een overheidsorganisatie vastlegt hoelang verschillende categorieën procesgebonden informatieobjecten bewaard worden dan wel voor vernietiging in aanmerking komen, voorafgegaan door een verantwoording.

In de handreiking Belangen in Balans is uitgebreid beschreven wat een selectielijst is en hoe hij gemaakt wordt. Voor de handreiking webarchivering volstaat te weten dat de Archiefwet geen bewaartermijnen vastlegt. De Archiefwet beschrijft alleen de procedure voor het opstellen en vaststellen van een selectielijst. Elke overheidsorganisatie moet de bewaartermijnen zelf bepalen door een afweging van belangen te maken, wat *waardering* genoemd wordt. Daarbij worden ze ondersteund door de archiefinstelling waaraan zij hun blijvend te bewaren informatie overbrengen.

Informatie die volgens de selectielijst tijdelijk bewaard wordt moet volgens de Archiefwet vernietigd worden na afloop van de bewaartermijn. Informatie die blijvend bewaard moet worden moet na uiterlijk twintig jaar overgebracht worden naar een archiefbewaarplaats.

## 7.12 Overbrengen

Overbrengen is een verplichte procedure uit de Archiefwet om overheidsinformatie die blijvend bewaard moet worden na uiterlijk twintig jaar aan een daarvoor aangewezen archiefbewaarplaats over te dragen.

Na overbrenging geldt dat de informatie in principe openbaar is en zonder kosten ingezien kan worden. Het verdere beheer wordt volledig door de archiefbewaarplaats verzorgd.

# CONCEPT

Overbrengen moet wettelijk plaatsvinden uiterlijk twintig jaar nadat de bewaartermijn gaat lopen (zoals aangegeven in de selectielijst). De archiefvormer kan met zijn archiefbewaarplaats afspreken om vervroegd over te brengen. Waardoor de archiefvormer minder lang verantwoordelijk is en het beheer over kan laten aan een specialist.

Overbrenging omvat meer dan alleen de overdracht van de archiefbestanden. Het betreft ook de formele overdracht van eigenaarschap, met eventuele afspraken die daarbij horen. De fysieke overdracht van de archiefbestanden zal waarschijnlijk plaatsvinden door de beheerder van de bestanden. Maar het is de archiefvormer die daarvoor verantwoordelijk is.

## 7.13 Webarchivering

Webarchivering betekent het duurzaam toegankelijk maken van webpagina's en hun bijlagen. Er zijn verschillende manieren om dat aan te pakken. Waarvan client-side archivering de meest gebruikte en bewezen manier is.

### 7.13.1 *Client-side archivering*

Het archiveren van een website kan uitgevoerd worden via de techniek die harvesten heet. Hierbij worden webpagina's (inclusief afbeeldingen, vormgeving, en downloadbare documenten) bij de webserver opgehaald met behulp van speciale software. Deze harvester benadert de website zoals een gebruiker en 'ziet' de website dus alleen aan de voorkant: aan de client-side.

Omdat er alleen contact wordt gemaakt met de client-side van de website en niet met het achterliggende CMS zelf kan client-side archivering onafhankelijk uitgevoerd worden van de onderliggende techniek. Dat is het voordeel van deze vorm van webarchivering.

Een belangrijk nadeel is dat de volledigheid en kwaliteit van de archivering afhangt van de toegang die de harvester krijgt tot de website. Een harvester zal bijvoorbeeld informatie die verstopt zit achter een zoekformulier of waarvoor ingelogd moet worden vaak niet of minder compleet kunnen binnenhalen. Ook kan een harvester server-side gedrag niet archiveren.

Voor een uitgebreidere beschrijving van varianten van harvesting-methoden die door het Nationaal Archief zijn getest zie "Ervaring met website archivering" (2016).

De meest gebruikte harvester is Heritrix. Deze is ontwikkeld door het Internet Archive in samenwerking met de Scandinavische nationale bibliotheken. Wereldwijd gebruiken diverse instellingen Heritrix voor het harvesten van webpagina's<sup>9</sup>.

### 7.13.2 *Server-side archivering*

Bij deze methode wordt de inhoud van een website bij het CMS opgehaald, of het CMS zorgt zelf voor de archivering. Deze vorm van webarchivering kan alleen uitgevoerd worden als de beheerder van het CMS daaraan meewerkt. Hiervoor moeten met ieder CMS afzonderlijk afspraken gemaakt worden en zijn mogelijk aanpassingen aan het CMS noodzakelijk.

Het voordeel van deze methode is dat alle informatie die in het CMS zit gearchiveerd kan worden, inclusief eventueel aanwezige geschiedenis-informatie, bijgehouden wijzigingen aan teksten en andere informatie die wel in het CMS zit maar die niet aan de voorkant (client-side) van de website zichtbaar is.

<sup>9</sup> Zie [https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives)

# CONCEPT

Een nadeel van deze methode is dat de uitvoering sterk afhankelijk is van de techniek van het CMS. Daarnaast moet de gearchiveerde inhoud van het CMS nog bewerkt worden om deze voor een gebruiker in de toekomst toegankelijk te maken als een navigeerbare website. Vooral bij dynamische websites, waarbij de inhoud van een database door middel van een op de server draaiend programma of script wordt omgezet in webpagina's, is het later weer tonen van de website een uitdaging.

### 7.13.3 *Transactie-archivering*

Bij transactie-archivering wordt tijdens het gebruik van een website de uitwisseling van data tussen de webbrowser en de webserver gearchiveerd. Daarmee wordt alle interactie tussen alle (of enkele) gebruikers en de website gearchiveerd.

Het voordeel van transactie-archivering is dat exact alle informatie wordt gearchiveerd die door de gebruikers is ontvangen en verzonden. Inclusief de inhoud die afhankelijk is van de gebruiker, de informatie die de gebruiker terugstuurt (bijvoorbeeld formulieren) en de interactie met de server voor het bepalen van het gedrag van een webpagina. Met transactie-archivering kan, in principe, het gehele websitebezoek van een gebruiker worden gereconstrueerd.

Transactie-archivering heeft een aantal zwaarwegende potentiële nadelen. De privacy van de gebruikers van de website is in het geding. Daarnaast levert het ook veel gegevens op.

Webrecording kan in principe gebruikt worden als een alternatief voor harvesting. Er wordt dan een sessie van één (anonieme) gebruiker opgenomen. Dit is echter onpraktisch omdat die gebruiker dan alle links handmatig moet aanklikken. En dit moet dan ook nog eens bij alle wijzigingen aan de website worden herhaald.

Transactie-archivering wordt (nog) weinig toegepast.

## 7.14 **Harvestingscope**

De harvestingscope is de afbakening welke webpagina's en bijlagen door een harvester worden binnengehaald. Deze wordt bepaald door een aantal parameters die aan de harvester als opdracht worden meegegeven.

Idealiter haalt een harvester alle webpagina's en bijlagen binnen die tot een website worden gerekend. Daarvoor zou het nodig zijn om vooraf een complete lijst te geven van alle webpagina's die tot de website gerekend worden. Dat is praktisch meestal onmogelijk. Omdat dit een hele lange lijst kan zijn, die ook nog eens voortdurend verandert. Daarom gaat een harvester anders te werk.

### 7.14.1 *Seed URL*

De harvester begint met het binnenhalen van een gegeven startpagina: de seed URL. Dit is vaak de startpagina van een website, bijvoorbeeld de startpagina van de website van de Rijksoverheid:

[www.rijksoverheid.nl](http://www.rijksoverheid.nl)

maar kan ook een subdomein zijn, zoals het Actorenregister:

[actorenregister.nationaalarchief.nl](http://actorenregister.nationaalarchief.nl)

of een sub-website, zoals de sub-website over typegoedkeuringen bij de RDW:

[www.rdw.nl/sites/tgk/Paginas/default.aspx](http://www.rdw.nl/sites/tgk/Paginas/default.aspx)



# CONCEPT

## 7.14.2 *Hyperlinks aflopen*

Op elke binnengehaalde webpagina worden alle hyperlinks afgelopen. Alle gelinkte webpagina's en bijlagen die binnen het domein van de website vallen (en nog niet binnengehaald zijn) worden ook binnengehaald. Dit proces herhaalt zich, totdat er geen nieuwe webpagina's meer gevonden worden.

## 7.14.3 *Blacklist*

Pagina's die om inhoudelijke of technische redenen worden uitgesloten van harvesting worden aangegeven op een blacklist. Deze pagina's worden door de harvester overgeslagen.

Een technische reden om delen van een website uit te sluiten zijn de *crawler traps*. Dit zijn delen van een website waar een harvester in een oneindige lus terecht komt. Een bekend voorbeeld van een crawler-trap is een agenda-element dat tot in het oneindige in het verleden en de toekomst kan lopen.

## 7.14.4 *Weespagina's*

Weespagina's zijn webpagina's die wel tot de website behoren, maar waarbij de harvester geen hyperlink-pad kan volgen vanaf de seed URL naar die pagina. Als bekend is wat de weespagina's zijn, worden deze op een lijst aangeleverd. Zodat de harvester weet dat hij deze pagina's ook mee moet nemen.

Een veel voorkomende oorzaak van weespagina's is het gebruik van JavaScript om dynamisch links te maken naar onderdelen van de website. Bij het harvesten wordt deze JavaScript-code niet uitgevoerd, en ziet de harvester deze links niet. Deze pagina's kunnen wel door bezoekers gevonden worden wanneer deze door zoekmachines geïndexeerd is, of wanneer men de rechtstreekse link naar de weespagina heeft.

Om te voorkomen dat een harvester weespagina's niet archiveert kan deze een lijst met URL's van weespagina's gegeven worden. Bijvoorbeeld in de vorm van een sitemap. Dit is een document waarin alle (of een deel) webadressen van webpagina's van een website zijn opgesomd.

## 7.14.5 *Scope van webpagina-componenten*

Een webpagina is meestal opgebouwd uit verschillende componenten: HTML, CSS, JavaScript, afbeeldingen, multimedia en links naar bijlagen zoals PDF-documenten. Voor elk van deze componenten geldt dat het op hetzelfde domein kan staan als de website zelf, of dat het op een externe locatie is gehost.

Bij het harvesten moet een keuze worden gemaakt welke van deze componenten onderdeel moeten zijn van het webarchief. Dit kan eenvoudigweg 'alles' zijn, of beperkt worden tot één of meer hosting-domeinen, bestandstypen of multimediaformaten.

## 7.14.6 *Hops en pad-diepte*

Voor sommige archieven kan een website te groot zijn. Een scope-afbakening die dan veel gebruikt wordt is het beperken van het aantal hops en van de pad-diepte:

- Het aantal hops is het kleinste aantal links dat je moet volgen om een pagina te bereiken vanaf de Seed URL.
- De pad-diepte is het aantal URL-elementen dat in de URL achter de domeinnaam staat. Zo heeft de homepage 'www.rijksoverheid.nl' een pad-diepte van 0, en heeft 'www.rijksoverheid.nl/actueel/agenda' een pad-diepte van 2.

# CONCEPT

Door het aantal hops en de pad-diepte te beperken zal het dieperliggende deel van de website niet in het webarchief wordt opgenomen.

## 7.14.7 *Scope + 1*

Het komt vaak voor dat webpagina's die binnen een website vallen een link hebben naar pagina's buiten de website, op een ander domein en mogelijk van een andere archiefvormer. Bij een 'scope + 1' worden deze externe pagina's ook binnengehaald bij het harvesten. De hyperlinks op die externe pagina worden echter niet gevolgd. Niet alle harvesters bieden deze optie.

Het voordeel van 'scope +1' is dat de gebruiker van het webarchief precies kan zien naar welke informatie de webpagina linkte. Het nadeel is dat het webarchief groter wordt, dat je als archiefvormer niet zeker bent van de rechten van de content op die externe pagina.

## 7.15 **Harvestingfrequentie**

Nadat is bepaald wat de harvestingscope is moet ook worden vastgesteld hoe vaak de archivering moet plaatsvinden. Daarbij is de keuze uit éénmalige harvesting, periodieke harvesting en real-time harvesting.

### 7.15.1 *Eenmalige harvesting*

Eenmalige harvesting wordt gebruikt wanneer een website op het punt staat om offline te gaan of vervangen te worden. Voorafgaand aan het offline-gaan of de vervanging kan de website dan worden gearchiveerd.

### 7.15.2 *Periodieke harvesting*

Periodieke volledige harvesting houdt in dat de harvesting met een bepaald tijdsinterval herhaald wordt uitgevoerd. De keuze voor het tijdsinterval bepaalt welke wijzigingen in het webarchief terug te vinden zijn. Een toevoeging of wijziging die weer verwijderd of gewijzigd wordt voordat er een nieuwe harvest heeft plaatsgevonden, zal niet in het webarchief zijn opgenomen.

Een variant hierop is incrementele harvesting. Hierbij worden alleen de webpagina's opgehaald en verwerkt die sinds de vorige harvest zijn gewijzigd of verwijderd.

### 7.15.3 *Real-time harvesting*

Bij real-time harvesting stuurt het CMS bij het toevoegen, wijzigen en verwijderen van een webpagina een signaal naar de harvester. Waarop de harvester die pagina (opnieuw) harvest. Niet alle CMS'en en harvesters ondersteunen real-time harvesting.

## 7.16 **WARC-bestand**

Het WARC-formaat (Web ARChive format) is een ISO-standaard (ISO 28500:2017) voor het opslaan van het resultaat van een harvest-actie. In een WARC-bestand zitten:

- De requests die de harvester heeft gestuurd naar de webserver.
- De responses die de webserver heeft teruggestuurd naar de harvester.
- Metadata over het harvestingproces.

Met speciale voor het WARC-formaat ontwikkelde afspeler-software (rendering) kan een in een WARC-bestand opgeslagen website bekeken worden. Bekende

# CONCEPT

voorbeelden van afspeel-software zijn de Wayback-machine (server-software) en de Webrecorder Player (stand-alone desktop-software).

Het archief van één website kan in meerdere WARC-bestanden zijn opgeslagen. Bijvoorbeeld omdat een website te groot kan zijn voor één bestand. In de metadata in een WARC-bestand kan daarvoor een verwijzing naar de andere bestanden zijn opgenomen. Zodat een in meerder bestanden gearchiveerde website goed afgespeeld kan worden.

## 8 Bijlage: Checklist voor het beschrijven van een website

Neem in de beschrijving van een website op:

- Algemeen:
  - Naam van de website
  - Site-URL
  - Secundaire site-URL's
  - Talen van de website
  - Korte beschrijving van de inhoud van de website (bijvoorbeeld uit "over deze site")
- De contenteigenaar die verantwoordelijk is voor de inhoud van de website:
  - Naam van de contenteigenaar (organisatie of organisatieonderdeel).
  - Link naar de beschrijving van de organisatie in het Actorenregister<sup>10</sup>. Voor zover deze daar al in is opgenomen.
- Beheergegevens:
  - Contentbeheerder
  - Platformbeheerder
  - Platform dat wordt gebruikt
  - Datum dat de website online ging
  - Datum dat website offline ging. Alleen in te vullen voor website die uitgefaseerd wordt.
- Parameters voor harvesting:
  - Seed URL
  - Weespagina's
  - Blacklist. Met daarbij toelichting waarom items op de blacklist staan.

<sup>10</sup> Zie <https://actorenregister.nationaalarchief.nl>

# CONCEPT

## 9 Bijlage: Kwaliteitseisen geharveste website

Onder de kwaliteit van een webarchief wordt verstaan de mate waarin de gearchiveerde versie van een website een betrouwbare weergave is van de website op het moment dat deze gearchiveerd werd. De kwaliteit van een gearchiveerde website wordt beïnvloed door de manier waarop de website gebouwd is, de technieken die daarbij zijn toegepast en door de gebruikte harvesting-software.

De kwaliteitseisen hieronder hebben betrekking op:

- Een gegeven harvestingscope.
- Een gegeven datum.
- Een verzameling bestanden die het resultaat zijn van een harvesting volgens de harvestingscope op de gegeven datum.

De kwaliteitseisen gaan zowel over de inhoudelijke kwaliteitseisen als de technische kwaliteitseisen.

#	Kwaliteitseisen webarchivering	Toelichting
1	De bestanden voldoen aan de WARC-standaard NEN-ISO 28500:2017	WARC is de de-facto standaard voor webarchivering.  Voor het valideren van WARC-bestanden kan bijvoorbeeld JHove (vanaf versie 1.14) gebruikt worden.
2	Elk WARC-bestand is maximaal 1GB groot.	Conform de aanbevelingen over bestandsgrootte in Annex C van de ISO 28500:2017 standaard.  Wanneer het totale webarchief groter is dan 1GB moet het worden opsplitsen in meerdere WARC-bestanden die elk maximaal 1GB groot zijn. Uit de naamgeving van de bestanden moet blijken welke WARC-bestanden bij elkaar één webarchief vormen. Zie hiervoor ook kwaliteitseis 3.  Alleen wanneer een individueel WARC-record van zichzelf al groter is dan 1GB mag het in een eigen, WARC-bestand zijn opgeslagen dat groter is dan 1GB.

# CONCEPT

3	<p>De bestandsnamen voldoen aan het volgende bestandsnaampatroon:</p> <p>&lt;ISIL-code van archiefvormer&gt;-&lt;datum en tijd start harvest&gt;-&lt;volgnummer&gt;-&lt;domeinnaam van seed&gt;.warc</p>	<p>Conform de aanbevelingen over bestandsnaamgeving in Annex C van de ISO 28500:2017 standaard</p> <p>Bijvoorbeeld: de eerste WARC van een webarchief van de GahetNA website van het Nationaal Archief die op 31 oktober 2017 om 10:15:00 gestart is:</p> <p>NL-HaNA-20171031101500-001-www.gahetna.nl.warc</p>
4	<p>De WARC-bestanden zijn ongecomprimeerd</p>	<p>De WARC-standaard staat 2 manieren van compressie toe, WARC in geheel gz-gecomprimeerd of alle individuele WARC-records los gz-gecomprimeerd.</p> <p>Uit duurzaamheidsoverwegingen kiezen wij nu voor volledig ongecomprimeerde aanlevering.</p>
5	<p>De WARC-bestanden zijn aangeleverd zonder index-bestanden</p>	<p>Index-bestanden worden door de rendersoftware in het e-depot zelf aangemaakt. In een SIP aangeleverde index-bestanden kunnen door het e-depot niet worden gebruikt</p>
6	<p>Elk WARC-bestand heeft een checksum in een gangbaar checksum-formaat.</p>	<p>Hoe vroeger in het bewaarproces een checksum berekend wordt, hoe beter de integriteit van de WARC-bestanden kan worden gecontroleerd.</p> <p>De door het Nationaal Archief gebruikte e-depotvoorziening ondersteunt in ieder geval MD5, SHA1, SHA256 en SHA512 als checksum-algoritmes.</p>
7	<p>Bij een volledige harvest zijn alle webpagina's met alle bijbehorende componenten en bijlagen die in de scope van de archiveringsopdracht vallen in het webarchief zijn opgenomen.</p> <p>Bij een incrementele harvest zijn alle webpagina's met alle bijbehorende componenten en bijlagen die sinds de vorige harvest zijn gewijzigd en die binnen de scope van de archiveringsopdracht vallen in het webarchief zijn opgenomen.</p>	<p>Zolang hier geen geautomatiseerde controle-tools voor zijn moet de controle hierop steekproefsgewijs visueel gebeuren.</p>

# CONCEPT

8	Pagina's die niet in scope van de archiveringsopdracht vallen of die op de blacklist staan, mogen niet in het webarchief zijn opgenomen.	
9	Alle componenten op een webpagina, uitgezonderd die componenten die op de blacklist staan, zijn geharvest, ongeacht de inhoud van robot.txt., ongeacht of de component lokaal of extern gehost is, tenzij bij het vaststellen van de scope is besloten om specifieke componenten uit te sluiten van archivering.	<p>Bij het vaststellen van de scope van een archivering moet de website worden gecontroleerd op het gebruik van crawler traps. Deze mogen worden uitgesloten van archivering door plaatsing op de blacklist. Alle andere componenten moeten worden opgenomen tenzij om inhoudelijke of technische redenen wordt gekozen om ze op de blacklist te plaatsen.</p> <p>Bij het achterhalen van crawler traps kan gebruik worden gemaakt van de inhoud van het robots.txt-bestand. Dit bestand wordt vaak gebruikt om indexeer-bots niet te laten verdwalen op een website.</p>
10	De WARC-bestanden bevatten alle verzoeken (requests) aan de webserver en alle antwoorden (responses) van de webserver ongewijzigd en integraal.	Er mag geen filtering worden toegepast, en geen delen van de pagina worden herschreven. URL-rewriting is niet toegestaan. Foutmeldingen, omleidingen en serverfouten worden ook allemaal opgenomen in het WARC-bestand.
11	De header van elk WARC-record bevat naast de verplichte velden ook tenminste de velden WARC-Block-Digest, Content-Type en WARC-Warcinfo-ID op, en bij de request- en response-blokken ook de WARC-Target-URI.	<p>Elk WARC-bestand bestaat uit één of meer WARC-records. Een WARC-record bestaat uit een header en een block, zoals beschreven in ISO 28500:2017.</p> <p>Met deze extra velden kunnen de authenticiteit, integriteit en onderlinge relaties van WARC-records beter worden gewaarborgd.</p> <p>Het gebruik van andere in de ISO 28500:2017 genoemde velden in de header van een WARC-record is toegestaan.</p>

# CONCEPT

12	Het Warcinfo-onderdeel van elk WARC-bestand bevat metagegevens aan over de robots-policy en over welke browser en welke versie daarvan gebruikt werd tijdens de harvest.	Gebruik in het Warcinfo-blok minimaal de in de ISO28500:2017, paragraaf 6.2 beschreven velden 'software', 'format' en 'robots'. Het gebruik van andere velden in het Warcinfo-blok is toegestaan
13	De WARC-bestanden kunnen gelezen worden in een gangbare, leverancieronafhankelijke viewer.	<p>Via een visuele controle controleert de archiefbeheerder zelf steekproefsgewijs de volledigheid en kwaliteit van de gearchiveerde website aan de hand van de afspraken die gemaakt zijn met de opdrachtnemer van de webarchivering.</p> <p>De eis voor een leverancieronafhankelijke viewer voorkomt een vendor lock-in.</p> <p>Voorbeelden van open-source en leverancier-onafhankelijke WARC-viewers zijn The Wayback Machine en de Webarchiveplayer.</p> <p>Wanneer een WARC-bestand wel voldoet aan alle eerdere eisen, maar bij visuele controle de inhoud niet toont zoals verwacht, kan dit veroorzaakt worden door fouten bij de harvest of door fouten bij het weer afspelen. De WARC zal dan samen met de harvestingsdienst nader moeten worden bekeken.</p>



## 10 Bijlage: Over deze handreiking

### 10.1 Aanleiding

Overheidswebsites vallen onder de werking van de Archiefwet 1995. Dit betekent dat websites van de overheid gearchiveerd moeten worden. Eind 2016 heeft de Erfgoedinspectie (EGI) onderzocht in hoeverre dit daadwerkelijk gebeurt bij de centrale overheid<sup>11</sup>. Conclusie van dit onderzoek is dat overheden hun websites en social media onvoldoende archiveren. Dit leidt tot verlies van informatie en is een (juridisch) risico. De belangrijkste redenen daarvoor zijn volgens de EGI:

- Het ontbreekt aan een beleidskader en aan specificaties. Mede daardoor hebben overheidsorganisaties onvoldoende handvatten;
- Er is onvoldoende kennis bij de archiefvormers;
- De taakverdeling tussen Nationaal Archief, Koninklijke Bibliotheek en de Dienst Publiek en Communicatie is onduidelijk.

Dit rapport leidde tot Kamervragen en daarmee tot een Kamerbrief<sup>12</sup> waarin de minister van OCW toezegt:

- De Algemene Rijksarchivaris<sup>13</sup> stelt een handreiking op en neemt die in beheer, gericht op de archivering van websites en uitingen op sociale media.
- In samenwerking met CIO-Rijk zal er tevens een voorstel komen ten behoeve van implementatie van een archiveringspraktijk voor deze media bij organisaties binnen de centrale overheid.

Dit document beschrijft de invulling van de eerste van deze twee toezeggingen: de ontwikkeling van de handreiking webarchivering door het Nationaal Archief. Het implementatieplan zal, samen met de CIO-Rijk, worden opgesteld zodra de eerste versie van de handreiking is opgeleverd.

### 10.2 Opdracht

De handreiking wordt ontwikkeld en beheerd in opdracht van de minister van OCW en wordt als kennisproduct onderdeel van de zakelijke dienstverlening van het NA. Sturing op de ontwikkeling en het beheer vindt plaats via de governance voor de zakelijke dienstverlening.

De interne opdrachtgever is de directeur Digitale Infrastructuur en Adviesdiensten (DIA) van het NA. De ontwikkeling en beheer van de handreiking wordt uitgevoerd door het beheerteam kennisproducten, volgens de daarbinnen gangbare werkafspraken.

De handreiking wordt ter vaststelling aangeboden aan de standaardisatieraad van het NA. De standaardisatieraad bestaat uit archiefvormers binnen alle bestuurslagen

<sup>11</sup> Zie het rapport 'Webarchivering bij de centrale overheid': <https://www.erfgoedinspectie.nl/publicaties/rapport/2016/12/8/rapport-webarchivering>.

<sup>12</sup> Zie <https://www.rijksoverheid.nl/documenten/kamerstukken/2016/12/14/kamerbrief-over-webarchivering-bij-de-centrale-overheid>

<sup>13</sup> De Algemene Rijksarchivaris staat aan het hoofd van het Nationaal Archief.

# CONCEPT

en archiefinstellingen op directieniveau. De raad vertegenwoordigt de belangen van organisaties die gebruik maken van de kennisproducten.

## **10.3 Verdere ontwikkeling**

De handreiking wordt gefaseerd ontwikkeld. Waarbij het uitgangspunt is dat bestaande kennis zo snel mogelijk beschikbaar komt, waarna deze direct in de praktijk getoetst wordt bij verschillende archiefvormers. Het NA is beschikbaar als adviseur voor deze praktijkcasussen. Zodat eventuele onduidelijkheden, onjuistheden en omissies in de handreiking ondervangen worden. Op basis van de praktijkervaringen en verder ontwikkelende kennis wordt de handreiking geregeld verbeterd en uitgebreid.

Als fasering voor de handreiking is voorzien:

- Eerste conceptversie (december 2017)
- Reviewversie (Q1 2018)
- Vastgestelde versie (na openbare review, Q1 2018)
- Aanpassingen op basis van praktijkervaringen (verwacht Q4 2018)
- Verdere uitbreiding van de scope (vanaf 2019)

De hier genoemde data zijn onder voorbehoud van de reacties op de verschillende versies van de handreiking. Die aard van deze reacties bepaalt hoe snel er een volgende versie wenselijk en mogelijk is.

## 11 Bijlage: Bronnen

Geraadpleegde bronnen (per 04-12-2017) die gebruikt zijn in deze handreiking:

Toelichting op overheidsorganen die vallen onder de Archiefwet:

<https://www.erfgoedinspectie.nl/toezichtvelden/archieven/wet--en-regelgeving/archiefwetgeving/reikwijdte>

Het KB-webarchief:

<https://www.kb.nl/bronnen-zoekwijzers/databanken-mede-gemaakt-door-de-kb/webarchief-kb>

Het NIBG-webarchief:

<http://www.beeldengeluidwebarchief.nl>

Het Generiek Waarderingsmodel Rijk:

[https://www.earonline.nl/images/earpub/a/a7/Generiek\\_Waarderingsmodel\\_Rijksoverheid\\_2\\_0\\_def.pdf](https://www.earonline.nl/images/earpub/a/a7/Generiek_Waarderingsmodel_Rijksoverheid_2_0_def.pdf)

De selectielijst gemeenten en intergemeentelijke organen 2017:

<https://vng.nl/files/vng/publicaties/2017/20170706-selectielijst-gemeenten-intergemeentelijke-organen-2017.pdf>

Toelichting op het begrip "duurzaam toegankelijk"

[https://wiki.nationaalarchief.nl/pagina/DUTO:Waarom\\_DUTO%3F](https://wiki.nationaalarchief.nl/pagina/DUTO:Waarom_DUTO%3F)

Toelichting op de harvester Heritrix:

[https://en.wikipedia.org/wiki/List\\_of\\_Web\\_archiving\\_initiatives](https://en.wikipedia.org/wiki/List_of_Web_archiving_initiatives)

Het actorenregister:

<https://actorenregister.nationaalarchief.nl>

Rapport Erfgoedinspectie "Webarchivering bij de centrale overheid"

<https://www.erfgoedinspectie.nl/publicaties/rapport/2016/12/8/rapport-webarchivering>.

Kamerbrief minister van Onderwijs, Cultuur en Wetenschap:

<https://www.rijksoverheid.nl/documenten/kamerstukken/2016/12/14/kamerbrief-over-webarchivering-bij-de-centrale-overheid>