



Nationaal Archief
Ministerie van Onderwijs, Cultuur en
Wetenschap

Richtlijn archiveren overheidswebsites

*Maak informatie
duurzaam toegankelijk*

Colofon

Contact: Telefoon: +31-70-331 5460
E-mail: contact@nationaalarchief.nl
Postbus 90520
2509 LM Den Haag

Versiegeschiedenis:

Versie	Datum	Beschrijving
0.1	06-12-2017	Eerste versie. Als concept gepubliceerd op KIA en voor een eerste consultatie besproken met de direct betrokken bij het opstellen en de eerste implementaties van de norm.
0.2	07-05-2018	Tweede versie waarin eerste opmerkingen van directbetrokkenen verwerkt zijn. Ter informatie voorgelegd aan CIO-Rijk, SG OCW en de Algemeen Rijksarchivaris.
0.3	14-05-2018	Publicatie reviewversie op KIA en start openbare review.
0.4	19-10-2018	Reviewcommentaar verwerkt. Versie voor tekstredactie en review door redactiegroep.
0.5	29-11-2018	Definitief concept. Ter vaststelling door de Standaardisatieraad Nationaal Archief op 20 december 2018. Alle wijzigingen n.a.v. review doorgevoerd en tekstredactie uitgevoerd.
1.0	20-12-2018	Definitief. Vastgesteld door de Standaardisatieraad Nationaal Archief op 20 december 2018.

Inhoud

1 Inleiding	3
2 Over dit document	5
3 Waarom websitearchivering?	8
4 Eisen aan websitearchivering	19
Bijlagen:	
5 Bijlage: Stappenplan implementatie webarchivering	19
6 Bijlage: Begrippen	25
7 Bijlage: Motivatie van gemaakte keuzes	35
8 Bijlage: Aanleiding en opdracht van deze norm	39

1 Inleiding

Websites zijn een van de belangrijkste vormen van communicatie door overheidsorganisaties met burgers, bedrijven, instellingen en medeoverheden. Daarom is het ook belangrijk om terug te kunnen halen welke informatie op een bepaald moment op een website stond. Dit gaat niet vanzelf. Als een webpagina wijzigt of verwijderd wordt, is de oude versie vaak niet meer beschikbaar voor inzage. Om voorgaande versies van webpagina's toegankelijk te houden zijn speciale maatregelen nodig. Dit wordt websitearchivering genoemd.

Omdat websites steeds veranderen is webarchivering een doorgaand proces. Dat moet zo snel mogelijk geregeld zijn nadat een website actief wordt. Het is te laat om pas maatregelen te nemen op het moment dat een organisatie een website niet meer bijhoudt, verwijdert of vervangt. Want dan zijn wijzigingen uit het verleden moeilijk of niet meer te achterhalen. Het beste kan websitearchivering al meegenomen worden op het moment dat een nieuwe website wordt ontwikkeld. Want dan is er financiële en technische ruimte om het goed te regelen.

Voor wie is dit document?

Maar hoe archiveer je een website? Waar begin je en welke onderdelen bewaar je? Wie moet dat doen en hoe vaak? Dit document geeft antwoorden op die vragen. Het is bedoeld voor iedereen die als manager, projectleider of adviseur betrokken is bij het beheer en de archivering van websites.

Erkende werkwijze

Dit document is een norm. Dat wil zeggen dat het een erkende werkwijze beschrijft. Deze norm is van toepassing op de openbare delen van bestaande en nieuwe overheidswebsites.

Elke overheidsorganisatie is verantwoordelijk voor de archivering van zijn eigen websites. Het toepassen van deze norm is daarbij vrijwillig maar niet vrijblijvend. Als een overheidsorganisatie deze norm niet toepast, dan zal ze op een andere manier moeten voldoen aan de verplichting van de Archiefwet 1995 om haar websites te archiveren.

Opbouw van deze norm

Deze norm is opgebouwd uit verschillende onderdelen en enkele bijlagen:

- **Vragen of suggesties?**

Het toepassen van deze norm kan in de praktijk nog vragen oproepen, problemen opleveren en tot nieuwe inzichten leiden. Met die praktijkervaringen wil het Nationaal Archief de norm verbeteren en uitbreiden. Organisaties die vragen of suggesties voor verbeteringen hebben of ervaringen uit de praktijk willen inbrengen, kunnen daarvoor een e-mail sturen naar contact@nationaalarchief.nl. Het Nationaal Archief zal deze norm updaten als de praktijkervaringen daar aanleiding toe geven.

- **Over dit document**

Toelichting op de aard, status, scope, doel en doelgroep

- **Waarom websitearchivering?**

Toelichting waarom de archivering van websites belangrijk is, en welke wettelijke verplichtingen daarvoor gelden.

- **Eisen aan websitearchivering**

Dit is de kern van dit document. Hij bestaat uit een lijst met eisen, met toelichting, op basis waarvan een overheidsorganisatie de archivering van haar websites kan inrichten.

Voor de leesbaarheid bevatten de eisen geen praktische aanwijzingen of uitleg van begrippen en gemaakte keuzes. Deze zijn in afzonderlijke bijlagen opgenomen:

- **Bijlage: Stappenplan implementatie webarchivering**

Een praktische handleiding voor het implementeren van webarchivering binnen een overheidsorganisatie.

- **Bijlage: Begrippen**

Definitie met toelichting van de belangrijkste begrippen die in deze norm gebruikt worden.

- **Bijlage: Motivatie van gemaakte keuzes**

Toelichting op de keuzes die gemaakt zijn bij het opstellen van de eisen.

- **Bijlage: Aanleiding en opdracht van deze norm**

Toelichting hoe dit document tot stand is gekomen.

Vragen of suggesties?

Het toepassen van deze norm kan in de praktijk nog vragen oproepen, problemen opleveren en tot nieuwe inzichten leiden. Met die praktijkervaringen wil het Nationaal Archief de norm verbeteren en uitbreiden. Organisaties die vragen of suggesties voor verbeteringen hebben of ervaringen uit de praktijk willen inbrengen, kunnen daarvoor een e-mail sturen naar contact@nationaalarchief.nl. Het Nationaal Archief zal deze norm updaten als de praktijkervaringen daar aanleiding toe geven.

2 Over dit document

2.1 Norm

Dit document is een norm. Een norm beschrijft een erkende werkwijze van een bepaald onderwerp en toepassingsgebied. Een norm komt via consensus tot stand en wordt door een erkende instelling ontwikkeld en beheerd. Een norm is bedoeld voor gemeenschappelijk gebruik en gericht op een eenduidige werkwijze door de gebruikers.

Het toepassen van een norm levert besparingen op doordat de gebruikers geen eigen werkwijze hoeven te ontwikkelen, gemakkelijker kunnen samenwerken en gebruik kunnen maken van producten en diensten die uitgaan van de norm.

2.2 Onderwerp

Deze norm gaat over het archiveren van websites.

Een website is een verzameling samenhangende webpagina's. Een webpagina is online informatie die via een url beschikbaar is en als eenheid getoond kan worden met een applicatie zoals een webbrowser of mediaspeler. Hieronder vallen html-pagina's, maar ook andere informatie die via een url beschikbaar is, zoals tekstdocumenten, afbeeldingen, filmpjes en geluidsfragmenten.

Archiveren is het geheel van activiteiten die nodig zijn om informatie (in dit geval websites) duurzaam toegankelijk te maken. Vanaf het moment van creatie en voor zolang als dat nodig is. Zo kunnen belanghebbenden de informatie raadplegen zoals deze op enig moment in het verleden ook raadpleegbaar was. Websitearchivering is dus meer dan alleen het bewaren van webpagina's.

De norm gaat alleen over het archiveren van websites zoals ze online gepubliceerd zijn. Ze bevat geen eisen aan de websites zelf, aan de archivering van informatie die via een website is ontvangen of aan de archivering van informatie die gerelateerd is aan de website.

2.3 Toepassingsgebied

Deze norm is van toepassing op overheidswebsites die online zijn of dat nog worden. En daarvan alleen de openbare en niet-gepersonaliseerde webpagina's. Voor websites die niet meer online zijn, is deze norm van toepassing voor zover dat technisch nog mogelijk is.

Een overheidswebsite is een website die beheerd wordt door, of in opdracht van een overheidsorganisatie. Onder een overheidsorganisatie wordt hier verstaan een overheidsorgaan zoals bedoeld in de [Archiefwet 1995](#). Daaronder vallen alle publiekrechtelijke organisaties binnen de centrale en decentrale overheid. Onder deze overheidsorganen vallen ook privaatrechtelijke organisaties en natuurlijke personen, voor zover deze een publieke taak uitvoeren. Deze norm is op hen alleen van toepassing als een website in zijn geheel voortkomt uit die publieke taak.

Openbaar betekent dat een webpagina via het internet beschikbaar is, dus niet via een besloten intranet. De bezoeker hoeft ook niet in te loggen om hem in te zien.

Niet-gepersonaliseerd wil zeggen dat de inhoud van een webpagina niet afhankelijk is van gegevens over de gebruiker die de webpagina opvraagt. Bijvoorbeeld opgeslagen in cookies of in een database bij de website.

2.4 Status

Het toepassen van deze norm is vrijwillig maar niet vrijblijvend. De norm toepassen is een manier om voor websites te voldoen aan de verplichtingen uit de archiefwetgeving.

De archiefwetgeving laat de manier van archiveren voor het grootste deel over aan de overheidsorganisaties zelf. Zij zullen dus zelf keuzes moeten maken over hoe ze dat doen. Voor informatie die blijvend bewaard moet worden (dus ook openbare websites) verplicht Artikel 16 uit de Archiefregeling overheidsorganisaties daarbij te voldoen aan toetsbare eisen. Deze norm is bedoeld als invulling van die toetsbare eisen. Als een overheidsorganisatie de norm niet volgt, dan zal ze zelf de toetsbare eisen moeten bepalen.

Websitearchieven van de Rijksoverheid die voldoen aan deze norm worden door het Nationaal Archief als [archiefbewaarplaats](#) geaccepteerd voor overbrenging. Andere archiefbewaarplaatsen kunnen deze norm hanteren als acceptatievoorwaarde voor overbrenging, maar zijn dat niet verplicht.

2.5 Archiefwet

De verplichting om websites te archiveren vloeit voort uit de Archiefwet 1995. Artikel 3 uit de Archiefwet 1995 verplicht overheidsorganisaties "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." De wet noemt daarbij niet expliciet websites. Maar onder archiefbescheiden wordt verstaan alle informatie die een overheidsorganisatie ontvangt of maakt bij het uitvoeren van haar taken. Ongeacht de vorm van de informatie. Daar vallen websites dus ook onder.

De overheidsorganisatie uit wiens taken overheidsinformatie voortkomt, is verantwoordelijk voor het archiveren van deze informatie. Voor websites is dat de [contenteigenaar](#). Dat is de overheidsorganisatie die eindverantwoordelijk is voor de [content](#) (inhoud) van de website. Het archiveren kan uitbesteed worden, maar de contenteigenaar blijft daar dan wel verantwoordelijk voor.

Sommige overheidsorganisaties werken samen aan één website. Er zijn dan meerdere contenteigenaren. Deze zijn allemaal verantwoordelijk voor het archiveren van hun eigen aandeel in de content. Het is onpraktisch dat allemaal afzonderlijk te doen. Daarom wordt aangeraden om samen te bepalen wie het archiveren op zich neemt. Vaak is er een coördinerende partij waar die rol goed bij past.

2.6 Doel en doelgroep

Het doel van deze norm is dat overheidsorganisaties de kennis hebben om hun websites te archiveren en deze kennis ook toepassen. Daardoor is en blijft de inhoud van overheidswebsites toegankelijk voor iedereen die daar belang bij heeft.

De norm is bedoeld voor overheidsorganisaties die hun eigen websites (laten) archiveren. Niet voor archief- en erfgoedinstellingen die websites van anderen archiveren als onderdeel van hun collectievorming.

Deze norm is bedoeld voor de volgende personen binnen een overheidsorganisatie:

- **Managers** die verantwoordelijk zijn voor websites of archivering. Bijvoorbeeld CIO, hoofd DIV, of hoofd communicatie. In onderling overleg moeten zij opdracht geven voor het archiveren van de websites. Door te eisen dat de archivering aan deze norm voldoet kunnen zij die opdracht specifiek maken.
- **Projectleiders** die de opdracht voor het archiveren van websites moeten uitvoeren. Het stappenplan in deze norm kan dienen als basis voor het plan van aanpak.
- **Adviseurs** die de opdrachten voor het archiveren van websites vertalen in concrete plannen. Voorbeelden van adviseurs: informatiemanagers en -beheerders, functioneel beheerders, architecten en (inkoop) adviseurs, en adviseurs digitale archivering. De eisen in deze norm kunnen dienen als basis voor de specificatie waaraan de websitearchivering moet voldoen. Bijvoorbeeld in de vorm van een inkoopbestek of een ontwerp voor een eigen voorziening.

De zogenaamde zorgdrager is bestuurlijk verantwoordelijk voor de archivering bij een overheidsorganisatie (zie 6.21). Meestal is dat de bestuurder of het dagelijks bestuur. Voor een ministerie is het bijvoorbeeld de minister en voor een gemeente het college van burgermeester en wethouders.

De zorgdrager zal zelf niet met de norm te maken krijgen. In de praktijk zijn een of meerdere managers daarvoor verantwoordelijk. Wie dat is of zijn kan per organisatie verschillen.

2.7 Ontwikkeling

Het Nationaal Archief heeft deze norm ontwikkeld als onderdeel van zijn taak als kenniscentrum op het gebied van archivering. In opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap ontwikkelt het Nationaal Archief normen om overheidsorganisaties te ondersteunen bij het implementeren van de Archiefwetgeving.

Deze norm is tot stand gekomen in nauw overleg met belanghebbenden en experts. Er heeft een openbare review plaatsgevonden op een conceptversie. Zo konden alle geïnteresseerden onduidelijkheden aangeven en verbeteringen voorstellen.

Nadat alle reviewcommentaren waren verwerkt, is deze norm vastgesteld door de Standaardisatieraad van het Nationaal Archief. In de Standaardisatieraad zijn alle bestuurslagen van de overheid en verschillende archiefinstellingen vertegenwoordigd. De Standaardisatieraad ziet erop toe dat de normen van het Nationaal Archief voor alle overheidsorganisaties inhoudelijk acceptabel en praktisch bruikbaar zijn. Daarnaast adviseert de Standaardisatieraad het Nationaal Archief bij het ontwikkelen van normen en andere kennisproducten.

3 Waarom websitearchivering?

Overheden gebruiken websites om informatie te geven aan burgers, bedrijven, instellingen en andere overheden. Die informatie kan in de loop van de tijd veranderen of van de site gehaald worden. Voor belanghebbenden kan het belangrijk zijn dat de informatie dan nog steeds toegankelijk is. De belangen daarbij zijn heel verschillend:

- **Uitvoeren van bedrijfsprocessen**
Bij het uitvoeren van taken kan het voor overheidsorganisaties nodig zijn om oude webpagina's te bekijken. Bijvoorbeeld voor het beantwoorden van vragen van burgers of het opstellen van rapportages. Ook kan de inhoud van oude websites soms hergebruikt worden. Veel ambtenaren gebruiken websites als naslagwerk bij hun dagelijks werk.
- **Verantwoording over overheidshandelen**
Overheidsorganisaties moeten zich kunnen verantwoorden over hun handelen. Websites zijn een belangrijke neerslag van dat handelen. De originele inhoud van de websites moet daarom toegankelijk zijn op het moment dat het handelen ter discussie staat. Dat kan nodig zijn voor het behandelen van Wobverzoeken of in een debat. Maar ook bij een parlementair- of raadsonderzoek, een audit of rekenkameronderzoek.
- **Recht- en bewijszoekende burgers**
Burgers en bedrijven kunnen rechten ontlenen aan informatie op websites. Bijvoorbeeld bij het aanvragen van diensten, vergunningen, subsidies, of het starten van een bedrijf. Burgers en bedrijven baseren keuzes en handelingen op de informatie van de overheid. Die informatie kan daarom nodig zijn als bewijs in beroeps- en bezwaarprocedures. Als een overheidsorganisatie een zaak verliest omdat ze niet beschikt over dit bewijsmateriaal, dan kan dat leiden tot hoge kosten en gezichtsverlies.
- **Onderzoek**
Overheidswebsites geven onderzoekers inzicht in hoe overheidsorganisaties hun taken uitvoerden en zichzelf presenteerden. Bijvoorbeeld als er onderzoek gedaan wordt naar de aanpak van een grote crisis, naar hoe bepaalde besluiten tot stand zijn gekomen, of naar de relatie tussen overheid en burgers.
- **Erfgoed**
Websites zijn ook een uiting van onze cultuur. Niet alleen de inhoud, maar bijvoorbeeld ook de vormgeving en het taalgebruik. Op het moment dat een website nog volop gebruikt wordt is het misschien moeilijk voor te stellen. Maar ooit zullen websites hetzelfde gevoel opwekken als het Polygoonjournaal en verkleurde jaren 70-foto's. Die willen we nu niet meer missen omdat ze ons herinneren aan het verleden en medebepalen hoe we ons zelf zien.

Door de combinatie van deze belangen dienen overheidsorganisaties hun websites te archiveren.

4 Eisen aan websitearchivering

Overheidsorganisaties die hun openbare websites archiveren volgens deze norm, moeten voldoen aan de volgende eisen.

4.1 Dagelijks harvesten

Eis: Elke website wordt dagelijks geharvest.

Een keer per jaar wordt er een volledige harvest gemaakt. Daartussen vindt dagelijkse incrementele harvesting plaats van alleen de nieuwe en gewijzigde webpagina's.

Bij een grote wijziging aan een website wordt vlak voor en direct na deze wijziging een volledige harvest uitgevoerd. Een grote wijziging is een wijziging waarbij meer dan de helft van de webpagina's wijzigen.

Toelichting: Harvesting is de basis voor de webarchivering. Het legt een website vast zodat iedereen die later nog kan raadplegen. Ook nadat die veranderd is.

Dagelijks harvesten zorgt ervoor dat de meeste wijzigingen gearchiveerd worden. Het kan voorkomen dat er meerdere wijzigingen plaatsvinden tussen twee opeenvolgende harvests. Daarvan wordt dus alleen de laatste wijziging vastgelegd.

Bij volledige harvesting worden alle webpagina's van een website bewaard. Bij een incrementele harvest alleen de nieuwe en gewijzigde webpagina's. Hierdoor is de impact van incrementele harvesting bij websites die weinig veranderen laag.

Het jaarlijks of bij grote wijzigingen herhalen van de volledige harvest geeft extra zekerheid voor als een eerder archief verloren is gegaan.

4.2 Volledig harvesten

Eis: Per website worden alle webpagina's die binnen de website vallen volledig geharvest.

Uitzondering is de content die is opgenomen in de lijst met niet-harvestbare content (zie eis 4.3).

Voor de harvestingscope geldt:

- Alle embedded content (ingesloten webbronnen) harvesten.
- Geen beperkingen aan het aantal hops.
- Geen beperkingen aan de padlengte.
- Geen 'scope+1'. Dat wil zeggen dat geen enkele webpagina die buiten de website valt wordt geharvest.
- Webpagina's waarvan de harvester niet kan bepalen of ze tot de site behoren staan op de lijst met weespagina's.
- Alle webpagina's in de blacklist zijn gedocumenteerd in de lijst met niet-harvestbare content.

Toelichting: Alle webpagina's dragen bij aan de betekenis van een website. Daarom worden ze allemaal geharvest, behalve webpagina's op de lijst met niet-harvestbare content.

Alle webpagina's wil zeggen de startpagina en alle webpagina's die van daaruit bereikbaar zijn met hyperlinks, voor zover het webadres begint met de site-url of een van de secundaire site-url's.

4.3 Niet-harvestbare content

Eis: Voor webcontent (een hele webpagina of een gedeelte daarvan) die niet geharvest kan of mag worden is een beschrijving beschikbaar.

De beschrijvingen zijn per website vastgelegd in een OASIS Open Document Format (Odf) bestand voor tabellen (.ods).

Per beschrijving staat in deze tabel één regel, waarvan de kolommen zijn:

- *Url:* De url van de webpagina waarvan content niet harvestbaar is.
- *Onderdeel:* Het onderdeel van de webpagina dat niet harvestbaar is. Bijvoorbeeld 'hele webpagina' of 'streaming video'.
- *Reden:* De reden waarom de content niet harvestbaar is. Deze kan technisch maar ook juridisch zijn.
- *Beschrijving:* Korte beschrijving van de niet-harvestbare content. Zodat achteraf bekend is wat er niet geharvest is.
- *Alternatieve representaties:* Beschrijving en verwijzing naar gearcheverde alternatieve representaties van de niet-harvestbare content (voor zover beschikbaar).

Toelichting: Met harvesting kunnen de meeste webpagina's goed gearcheverd worden. Maar er zijn situaties waarin webpagina's of onderdelen daarvan niet harvestbaar zijn. Ook al zijn dit uitzonderingen, ze moeten wel gedocumenteerd worden. Zo kan een raadpleger van het archief later zien welke onderdelen ontbreken en waarom.

Enkele voorbeelden van gevallen waarin content niet harvestbaar is:

- Bij sommige interactieve content wordt op basis van interactie met de gebruiker bepaald welke content bij een server wordt opgehaald. Dit kan tijdens het harvesten niet altijd worden vastgesteld.
- Bij een streaming video (zoals een YouTube-video) wordt pas tijdens het afspelen bepaald welke informatie bij de videosever wordt opgehaald. Dit hangt onder andere af van de bediening van de videospeler.
- Een opsporingsbericht van de Politie op de website bevat mogelijk persoonlijke informatie die niet bewaard mag worden.

De lijst met niet-harvestbare content mag niet gebruikt worden om naar eigen inzicht content uit te sluiten van archivering. Dit kan alleen als het technisch of juridisch niet mogelijk is om de content te harvesten. Vaak is er een alternatieve representatie van de niet-harvestbare content beschikbaar die wel goed te archiveren is. Zoals een avi-bestand bij een streaming video of een tabel met de achterliggende gegevens bij interactieve content. Als er een alternatieve representatie is, dan wordt deze gearcheverd en wordt hiernaar verwezen in de beschrijving.

Archivering van een alternatieve representatie kan bijvoorbeeld door het als bestand toe te voegen aan de website (het bestand wordt dan mee gearcheverd), of door het op te nemen in een DMS van de organisatie. Opname in de website heeft de voorkeur omdat het bestand dan integraal onderdeel zal uitmaken van de gearcheverde website. Maar als dat niet mogelijk is, dan kan archiveren op een andere plaats ook, zolang in de beschrijving duidelijk is waar het te vinden is.

4.4 Beperkt beschikbare webpagina's

Eis: Per website is er een actuele lijst van gearcheverde webpagina's waarvan de beschikbaarheid beperkt is op grond van wet- en regelgeving.

De lijst met beperkt beschikbare webpagina's is vastgelegd in een OASIS Open Document Format (Odf) bestand voor tabellen (.ods). Dit bestand is opgenomen bij de webarchiefbestanden (zie eis 4.6).

Per beperking staat in de tabel één regel, waarvan de kolommen zijn:

- *Url:* De url van de webpagina waarvoor de beperking geldt.
- *Onderdeel:* Het deel van de webpagina waarvoor de beperking geldt. Bijvoorbeeld 'hele webpagina' of 'opsporingsberichten'.
- *Reden:* De reden waarom er een beperking geldt, met daarbij een verwijzing naar de wettelijke grondslag en eventuele documentatie van het besluit om de beschikbaarheid te beperken.
- *Toegang:* Beschrijving van hoe en onder welke voorwaarden toegang tot de webpagina mogelijk is.
- *Begindatum:* Begin van de periode waarvan de webpagina beperkt beschikbaar is (als dat van toepassing is).
- *Einddatum:* Eind van de periode waarin de webpagina beperkt beschikbaar is (als dat van toepassing is).

Toelichting: Deze norm gaat over openbare websites. Daarom mogen de gearcheverde webpagina's in principe aan iedereen en op elke manier beschikbaar worden gesteld. Toch kunnen er redenen zijn om de beschikbaarheid van bepaalde webpagina's te beperken. Er gelden regels voor het beperken van de beschikbaarheid van informatie, zoals de Algemene verordening gegevensbescherming (AVG) en de Auteurswet. Deze regels dienen ook voor archieven gerespecteerd te worden. De lijst met beperkt beschikbare webpagina's maakt dit mogelijk. Bijvoorbeeld bij persoonlijke informatie die per ongeluk is geplaatst of bij beeldmateriaal waarop auteursrecht rust.

Gearcheverde webpagina's waarvoor de beschikbaarheid uit privacyoverwegingen is beperkt, worden niet vernietigd. Alleen de beschikbaarheid is beperkt, zolang de privacyoverwegingen gelden.

Alle websitearchieven die beperkt openbaar zijn of worden in de zin van de Archiefwet 1995, zijn ook beperkt beschikbaar. Maar omgekeerd hoeft dat niet zo te zijn. Een beperking die bijvoorbeeld geldt vanwege auteursrechten heeft niks met openbaarheid te maken. Bovendien is het begrip 'beperkt openbaar' alleen van toepassing op overgebracht archief. Ook archieven voor overbrenging kunnen beperkt beschikbaar zijn.

4.5 Webarchiefbestanden

Eis: De resultaten van een harvest zijn vastgelegd in een of meerdere bestanden (hierna aangeduid als 'webarchiefbestanden') die voldoen aan de volgende eisen.

4.5.1 WARC-standaard

Eis: Elk webarchiefbestand voldoet aan de standaard WARC NEN-ISO 28500. En daarvan een versie die geldig is drie jaar of korter voor het moment van harvesting.

Al bestaande webarchiefbestanden worden bij het verschijnen van een nieuwe versie van de WARC-standaard niet geconverteerd naar de nieuwe versie.

Toelichting: Het WARC-formaat is de internationale standaard voor webarchiefbestanden. Door deze standaard te gebruiken is duidelijk wat de inhoud van een webarchiefbestand betekent en zijn er hulpmiddelen om de bestanden te verwerken.

Bestaande webarchiefbestanden worden niet geconverteerd om te voorkomen dat daarbij informatie verloren gaat.

4.5.2 Grootte

Eis: Elk webarchiefbestand is maximaal 1GB groot.

Wanneer het totale webarchief groter is dan 1GB moet het worden opgesplitst in webarchiefbestanden die elk maximaal 1GB groot zijn. Uit de naamgeving van de bestanden moet blijken welke webarchiefbestanden bij elkaar één webarchief vormen. Alleen wanneer een individueel WARC-record van zichzelf al groter is dan 1GB, mag het in een eigen webarchiefbestand zijn opgeslagen dat groter is dan 1GB.

Toelichting: Dit is volgens de aanbevelingen over bestandsgrootte in Annex C van de ISO 28500:2017 standaard.

4.5.3 Bestandsnamen

Eis: De bestandsnamen voldoen aan het bestandsnaampatroon:

```
<naam of code van archiefvormer>-<datum en tijd start harvest>-  
<volgnummer>-<domeinnaam van seed>.warc.
```

Toelichting: Uniforme naamgeving helpt bij het herkennen en beheren van de webarchiefbestanden. Het hier gehanteerde formaat is volgens de aanbevelingen over bestandsnaamgeving in de WARC-standaard. Een voorbeeld: De eerste WARC van een webarchief van de GahetNA-website van het Nationaal Archief die op 31 oktober 2017 om 10:15uur startte heet 'NL-HaNA-20171031101500-001-www.gahetna.nl.warc'.

4.5.4 Compressie

Eis: Een webarchiefbestand is als geheel niet gecomprimeerd.

Toelichting: Dit voorkomt dat een webarchiefbestand in de toekomst in zijn geheel onleesbaar is als slechts een deel van het bestand beschadigd is.

Binnen een webarchiefbestand is wel compressie per record toegestaan.

4.5.5 Checksum

Eis: Van elk webarchiefbestand is direct na creatie een checksum berekend en vastgelegd.

Periodiek en voorafgaand aan overbrenging worden alle over te brengen webarchiefbestanden gevalideerd door de checksums te controleren. Als daarbij een corrupt bestand wordt aangetroffen, dan wordt dit bestand hersteld (als dat nog mogelijk is, bijvoorbeeld uit een back-up).

Toelichting: Met een checksum is op elk later tijdstip vast te stellen dat een bestand na creatie niet is aangepast. Dat vergroot het vertrouwen in de juistheid van de inhoud. Hoe vroeger in het bewaarproces een checksum toegevoegd wordt, hoe beter de betrouwbaarheid van de webarchiefbestanden kan worden gecontroleerd.

Bij overbrenging wordt de checksum meegegeven als onderdeel van de metagegevens (element 21.7).

4.5.6 Alle requests en responses

Eis: De webarchiefbestanden bevatten ongewijzigd en integraal alle requests aan de webserver en alle responses van de webserver die bij het harvesten zijn verzonden en ontvangen. Er is geen filtering toegepast en er zijn geen delen van de webpagina herschreven. Ook het herschrijven van url's (url-rewriting) is niet toegestaan. Foutmeldingen, omleidingen en serverfouten worden ook allemaal opgenomen in het webarchiefbestand.

Toelichting: Om de authenticiteit en betrouwbaarheid van de informatie te kunnen bepalen is volledig inzicht nodig in de communicatie tussen de webserver en de harvestingdienst tijdens het proces van harvesten. Zo kan bijvoorbeeld geregistreerd worden dat een webpagina niet meegeharvest is. Bijvoorbeeld omdat er op dat moment een serverfout was bij de webserver of omdat de server eruit lag voor onderhoud.

4.5.7 Verplichte velden

Eis: Naast de velden die al verplicht zijn volgens de WARC-standaard, bevat elk webarchiefbestand ook ten minste de volgende velden (zoals beschreven in ISO 28500):

- In het WARC-infoblok:
 - Software
 - Format
 - Robots
- In de header van elk WARC-bestand:
 - WARC-Block-Digest
 - Content-Type
 - WARC-WARCinfo-ID
- In de request en response-blokken:
 - WARC-Target-URI

Toelichting: Met deze extra velden kunnen de authenticiteit en onderlinge relaties van WARC-bestanden beter worden beschermd.

Het is toegestaan ook andere velden te gebruiken dan die verplicht zijn in de WARC-standaard.

4.6 Mappen

Eis: De gezamenlijke webarchiefbestanden en bijbehorende lijsten zijn als volgt samengevoegd in een hiërarchische mappenstructuur:

1. *Organisatie:* Alle websites van één organisatie. Daarbinnen een map per website.
2. *Website:* Alle bestanden van één website van de organisatie. Daarbinnen een map per periode.
3. *Periode:* Alle bestanden van de website voor een periode beginnend met een volledige harvest en alle daaropvolgende incrementele harvests. Met daarbij de lijst van beperkt beschikbare webpagina's (zie [4.4](#)) op de laatste dag van die periode. Hierbinnen een map per harvest.
4. *Harvest:* Alle webarchiefbestanden die gevormd zijn tijdens de harvest op één dag. Met daarbij de lijst van de niet-harvestbare content (zie [4.2](#)) van die dag.

Toelichting: Deze mappenstructuur is nodig voor een duidelijke vindbaarheid van de bestanden. Daarnaast maakt dit het mogelijk webarchiefbestanden per map van metagegevens te voorzien.

4.7 Metagegevens

Eis: Voor elk webarchiefbestand en de daarboven liggende mappen zijn metagegevens vastgelegd volgens het 'Toepassingsprofiel Metagegevens Rijksoverheid' of het 'Toepassingsprofiel metadatering lokale overheden'.

In aanvulling op de toepassingsprofielen voldoen de metagegevens aan de volgende eisen:

4.7.1 Elementen

Eis: Per map of webarchiefbestand zijn minimaal de volgende metagegevens vastgelegd.

Nr.	Elementnaam	Map organisatie	Map website	Map periode	Map harvest	Webarchiefbestand
2	Identificatiekenmerk	x	x	x	x	x
3	Aggregatieniveau	x	x	x	x	x
4	Naam	x	x	x	x	x
5	Classificatie		x			
6	Omschrijving		x		x	
9.1	Dekking in tijd		x	x	x	
11	Taal		x			
12	Eventgeschiedenis		x	x	x	x
15	Relatie	x	x	x		
19	Vorm		x			x
20	Integriteit		x	x	x	
21	Formaat					x

Toelichting: Door in de metagegevens deze contextinformatie mee te leveren, kunnen de webarchiefbestanden goed beheerd en ontsloten worden. Daarnaast helpen ze de toekomstige raadplegers om de websitearchieven te doorzoeken, te interpreteren en te beoordelen op hun betrouwbaarheid.

De nummering en elementnamen zijn gebaseerd op de 'Richtlijn metagegevens Overheidsinformatie: Beschrijving van entiteiten en elementen'. De beide toepassingsprofielen zijn gebaseerd op deze richtlijn.

De lijst met niet-harvestbare content (eis 4.2) en de lijst met beperkt beschikbare webpagina's (zie 4.4) zijn ook metagegevens. Deze metagegevens passen niet goed in de toepassingsprofielen. Daarom is ervoor gekozen deze lijsten afzonderlijk als tabellen op te nemen.

4.7.2 Waarden

Eis: Voor de waarden van de elementen geldt:

Nr.	Elementnaam	Bij map of bestand	Waarde
3	Aggregatieniveau	organisatie	Vaste waarde 'serie'.
3	Aggregatieniveau	website; periode	Vaste waarde 'dossier'.
3	Aggregatieniveau	harvest; webarchiefbestand	Vaste waarde 'archiefbestanddeel'.
5	Classificatie	website	De classificatie zoals deze binnen de organisatie geldt voor de betreffende website.
6	Omschrijving	website	Korte beschrijving van de website. Met daarbij: <ul style="list-style-type: none">• Site-url• Startpagina• Secundaire site-url's
6	Omschrijving	harvest	Beschrijving van de software met parameters die gebruikt zijn voor het maken van de harvest.
9.1	Dekking in tijd	website	Periode waarin de website online is (geweest).
9.1	Dekking in tijd	periode	Periode waarin de onderliggende harvests hebben plaatsgevonden.
9.1	Dekking in tijd	harvest	Datum van de harvest.
11	Taal	website	Taal van de website.
15	Relatie	organisatie	<u>Contenteigenaar</u> .
15	Relatie	website	Organisatieonderdeel verantwoordelijk voor de inhoud van deze website.
15	Relatie	harvest	Organisatie die de harvest in deze periode heeft uitgevoerd.
19	Vorm	website	Vaste waarde 'websitearchief'.
19	Vorm	webarchiefbestand	Vaste waarde 'webarchiefbestand'.

Toelichting: Deze beperkingen zijn bedoeld voor een eenduidige interpretatie van de velden die specifiek zijn voor websites.

4.7.3 Aanlevering metagegevens

Eis: Bij overbrenging van webarchiefbestanden worden de metagegevens aangeleverd in de technische vorm die de betreffende archiefbewaarpplaats bepaald heeft.

Toelichting: Er bestaat nog geen landelijke norm voor de technische vorm waarin metagegevens worden uitgewisseld. Elke archiefbewaarplaats bepaalt dat voor zichzelf. Het Nationaal Archief hanteert voor de aanlevering van metagegevens het XML-formaat ToPX.

Aangeraden wordt om al meteen bij harvesting de metagegevens vast te leggen in de vorm die voor overbrenging is voorgeschreven. Dat voorkomt extra inspanningen en mogelijke fouten bij het uitwisselen van de metagegevens.

4.8 Kwaliteitscontrole

Eis: Er wordt periodiek gecontroleerd of de websitearchivering voldoet aan de eisen volgens deze norm. Raadplegers van het webarchief kunnen online fouten melden.

Als er door controle of meldingen een afwijking van de eisen wordt geconstateerd, dan wordt de oorzaak hiervan onderzocht en worden maatregelen genomen om herhaling te voorkomen. Onderzocht wordt welke webpagina's hierdoor niet, onjuist of onvolledig zijn opgenomen in de al gevormde webarchiefbestanden. Dit wordt gedocumenteerd in de metagegevens van die webarchiefbestanden (veld 20 'integriteit'). De al gevormde webarchiefbestanden zelf worden niet aangepast.

Afwijkingen, de oorzaken daarvan en de genomen maatregelen, worden gerapporteerd aan de manager die bij de contenteigenaar verantwoordelijk is voor de websitearchivering. Deze manager bepaalt of er voldoende maatregelen zijn genomen.

Toelichting: Een websitearchief dat niet aan de eisen voldoet, kan voor een deel of helemaal onbruikbaar zijn. Het kan inhoudelijk onvolledig of onjuist zijn, technisch niet te beheren of onleesbaar. Achteraf zijn fouten vaak moeilijk of niet meer te herstellen. Daarom is het belangrijk fouten op tijd op te sporen.

Hoe de periodieke kwaliteitscontrole wordt uitgevoerd is niet gespecificeerd in deze norm omdat hiervoor nog geen best practices bekend zijn. Dat betekent dat elke overheidsorganisatie zelf kan bepalen hoe zij de kwaliteitscontrole uitvoert. Zolang dit maar voldoet aan gangbare normen voor kwaliteitsmanagement, zoals bijvoorbeeld vastgelegd in ISO 9001.

4.9 Beveiliging

Eis: De webarchiefbestanden en bijbehorende metagegevens worden bewaard in een informatiesysteem dat beveiligd is om verlies of veranderingen te voorkomen. Volgens de beveiligingsnorm NEN-ISO/IEC 27002 en de invulling van deze norm zoals die voor de organisatie geldt.

Toelichting: Dit garandeert dat gearchiveerde websites op het moment van harvesting precies hetzelfde zijn als op het moment dat iemand ze in de toekomst bekijkt. Overheidsorganisaties hanteren verschillende invullingen van NEN-ISO/IEC 27002. Voor de Rijksdienst is dat bijvoorbeeld de Baseline Informatiebeveiliging Rijksdienst (BIR). Voor andere overheden is dit naar verwachting vanaf 1 januari 2019 de Baseline Informatiebeveiliging Overheid (BIO).

4.10 Publicatierecht

Eis: De contenteigenaar beschikt blijvend over het recht zijn webpagina's te archiveren en de gearchiveerde webpagina's online te publiceren.

Als content wordt ingekocht of beheerwerkzaamheden worden uitbesteed, dan is dit recht contractueel vastgelegd.

Toelichting: Dit garandeert dat de overheidsorganisatie zelf kan bepalen hoe en aan wie ze toegang verleent tot de gearchiveerde websites.

Deze eis is vooral een aandachtspunt voor de auteursrechten van ingekochte content, zoals bij in opdracht gemaakte teksten, afbeeldingen en stockfoto's. Het gebruik van deze content mag contractueel niet beperkt zijn tot een bepaalde periode of server. Ook moet het onmogelijk zijn dat bijvoorbeeld de leverancier van een streaming videodienst contractueel de publicatierechten op de video's beperkt.

4.11 Toegang

Eis: De geharveste webpagina's zijn direct na harvesting voor iedereen online toegankelijk.

Minimaal zijn daarbij de volgende functies voor alle webbrowsers beschikbaar:

- Het is duidelijk aangegeven dat het om een gearchiveerde versie van de website gaat.
- Bij het aanklikken van een link naar een website die niet gearchiveerd is, wordt een duidelijke melding gegeven dat het websitearchief verlaten wordt.
- Mogelijkheid om bij elke webpagina in het archief de versie van een willekeurige datum in het verleden op te vragen. Bijvoorbeeld met een datumveld of een aanklikbare kalender.
- Full-tekst-zoeken door alle webpagina's in de webarchiefbestanden.
- Bij het opvragen van een webpagina die op de lijst met niet-harvestbare content staat (zie eis 4.2):
 - wordt een duidelijke melding gegeven dat de webpagina geheel of gedeeltelijk niet beschikbaar is in het websitearchief;
 - met daarbij de gegevens uit de lijst voor de betreffende pagina;
 - met daarbij een link naar de gedeeltelijke weergave zoals die in het websitearchief beschikbaar is (alleen als de webpagina gedeeltelijk niet harvestbaar is);
 - met daarbij een link om naar de actuele versie van de webpagina te gaan, buiten het websitearchief (als die nog aanwezig is).
- Mogelijkheid om de metagegevens te bekijken die bij het webarchief zijn vastgelegde (zie eis 4.7).

Voor gearchiveerde webpagina's op de lijst van beperkt beschikbare webpagina's geldt het volgende (zie eis 4.4):

- Deze zijn niet online vindbaar en opvraagbaar. Waar technisch mogelijk is dit alleen beperkt tot dat deel waarvoor de beperking geldt.
- Er wordt een duidelijke melding gegeven dat de webpagina, geheel of gedeeltelijk, niet online beschikbaar is. Met een uitleg waarom niet en hoe het eventueel wel mogelijk is om inzage te krijgen.

Toelichting: Het doel van deze eis is dat iedereen gemakkelijk gearchiveerde websites kan bekijken, voor zover dat mogelijk en toegestaan is.

Voor overbrenging van webarchiefbestanden is de contenteigenaar verantwoordelijk voor de online toegang. Na overbrenging is de archiefbewaarplaats daarvoor verantwoordelijk.

De online publicatie van een geharveste website wordt zelf niet geharvest. Dat zou dubbelop zijn.

4.12 Blijvend bewaren

Eis: De webarchiefbestanden worden blijvend bewaard. Dit geldt voor alle verschillende onderdelen en alle geharveste versies van de website.

Als een overheidsorganisatie in zijn selectielijst niet heeft opgenomen dat zijn websites blijvend bewaard worden, dan is dat een bespreekpunt met het Nationaal Archief bij de eerstvolgende actualisatie van de selectielijst. Tot die tijd worden de webarchiefbestanden nog niet vernietigd.

Toelichting: Openbare websites zijn een vorm van officiële overheidscommunicatie. Ze zijn een belangrijke bron van informatie over het handelen van de overheid. In het bijzonder over de relatie die de overheid onderhoudt met burgers en bedrijven. Websites laten ook zien hoe de overheid zich door de jaren heen presenteert aan de maatschappij. Ze hebben dus ook een culturele waarde. Daarom worden ze blijvend bewaard. Misschien zijn niet alle onderdelen van een website voor de lange termijn even belangrijk. Maar omdat het lastig is om een goed onderbouwd onderscheid te maken, wordt de hele website blijvend bewaard (zie ook 7.5).

Het is wettelijk verplicht om in overleg met het Nationaal Archief in een selectielijst vast te leggen hoe lang overheidsinformatie bewaard wordt. Maar in sommige selectielijsten is nog onvoldoende beschreven wat de vernietigingstermijn voor websites is.

Is in de selectielijst nog helemaal niets bepaald over websites, dan mogen ze niet vernietigd worden. Bij een complete actualisering van een bestaande selectielijst of opmaak van een nieuwe, kunnen dan alsnog bepalingen voor websites worden opgenomen.

Wordt in de selectielijst bepaald dat websites na verloop van tijd vernietigd moeten worden? Check dan eerst of de uitzonderingsgronden het mogelijk maken om websites tijdelijk uit te zonderen van vernietiging, tot de volgende actualisering van een bestaande selectielijst of opmaak van een nieuwe selectielijst. Bestaan die uitzonderingsgronden niet, neem dan contact op met het Nationaal Archief om mogelijke aanpassing van de selectielijst te bespreken.

4.13 Overbrenging

Eis: De webarchiefbestanden met de bijbehorende lijsten en metagegevens worden na uiterlijk twintig jaar overgebracht naar een archiefbewaarpplaats. Volgens de mappenstructuur van eis 4.6.

De overheidsorganisatie en de betreffende archiefbewaarpplaats bepalen samen na welke periode en hoe vaak overbrenging plaatsvindt.

Aanvullend op deze norm stelt elke archiefbewaarpplaats algemene eisen aan het overbrengen van een archief. Neem daarvoor contact op met de betreffende archiefbewaarpplaats.

Toelichting: Volgens artikel 12 van de Archiefwet 1995 moeten overheidsorganisaties archieven die blijvend bewaard moeten worden, overbrengen naar een archiefbewaarpplaats. Deze is gespecialiseerd in het blijvend bewaren van archieven. Zo blijven de webarchieven duurzaam toegankelijk.

Algemene eisen aan het overbrengen zijn bijvoorbeeld de procedure en de manier waarop bestanden en metagegevens worden aangeleverd. Dit betreft eisen die niet specifiek zijn voor webarchiefbestanden. Ze vallen daarom buiten deze norm.

5 Bijlage: Stappenplan implementatie webarchivering

Overheidsorganisaties die hun websites archiveren volgens deze norm, kunnen dat doen volgens het volgende stappenplan. Het is bedoeld als praktisch hulpmiddel, bijvoorbeeld als basis voor het opstellen van een projectplan.

5.1 Initiatief

Actie: Breng webarchivering onder de aandacht van de managers die binnen de organisatie verantwoordelijk zijn voor de websites en voor archivering. Zorg ervoor dat zij het belang van webarchivering inzien en vervolgens als opdrachtgevers het initiatief nemen om webarchivering in te voeren en daarop toezien.

Doel: Bereidheid van het management om webarchivering in te voeren. Dit is onmisbaar als er kosten gemaakt worden.

Betrokkenen: Opdrachtgevers van websites (typische organisatirollen hiervoor zijn CIO, hoofd communicatie, hoofd informatievoorziening), zorgdrager (of vertegenwoordiger daarvan), archiefbeheerder.

5.2 Alleen of samen

Actie: Bepaal op welke schaal webarchivering wordt ingevoerd. Per website, per organisatieonderdeel, voor de hele organisatie of samen met andere overheidsorganisaties. Wanneer gekozen wordt voor samenwerking, doorloop dan samen de volgende stappen.

Als er meerdere opdrachtgevers zijn, spreek dan vooraf af hoe die samen beslissingen nemen. Bijvoorbeeld door het onder te brengen bij een bestaand programma, project of overlegorgaan. Bij het gezamenlijk uitvoeren van webarchivering blijft elke individuele organisatie verantwoordelijk voor het archiveren van de eigen websites.

Doel: Omdat harvesting los van het CMS wordt geregeld, is het vaak goed mogelijk webarchivering voor verschillende websites gezamenlijk in te richten. Dit kan schaalvoordelen opleveren. Het inzetten van gedeelde kennis en samen inkopen van diensten kan bijvoorbeeld goedkoper zijn.

Betrokkenen: Contenteigenaar, contentbeheerder, functioneel beheerder CMS.

5.3 Projectleider

Actie: De betrokken managers wijzen een projectleider aan voor het invoeren van webarchivering.

Doel: Iemand wordt beschikbaar gesteld als aanspreekpunt, voor het (laten) uitvoeren van de benodigde vervolgacties en voor het inbedden in de organisatie na afloop van het project.

Betrokkenen: Opdrachtgevers webarchivering (zie bij 'initiatief').

5.4 Informeren

Actie: Bepaal wie een rol spelen bij webarchivering. Informeer deze betrokkenen over het waarom, wat en hoe van webarchivering. Bijvoorbeeld met een presentatie en discussie over de norm webarchivering. Behandel daarbij vragen en maak keuzes als dat nodig is.

Doel: Betrokkenen zijn voorbereid op de mogelijke gevolgen van het archiveren. Bijvoorbeeld op werkzaamheden, applicaties, rolverdeling, financiën en contracten.

Betrokkenen: Projectleider webarchivering, contenteigenaar, contentbeheerder, functioneel beheerder CMS, harvestingsdienstverlener (als die al bekend is), archiefbeheerder, archiefbewaarplaats.

5.5 Rollen verdelen

Actie: Bepaal wie welke rol op zich neemt bij webarchivering. Afhankelijk van hoe een overheidsorganisatie het archiveren regelt, kunnen de rollen anders verdeeld zijn.

Doel: Voor iedereen is duidelijk wat van hem of haar verwacht wordt.
Een typische rolverdeling is:

- *Contenteigenaar:* Eindverantwoordelijk voor het archiveren van de websites. Neemt het initiatief voor het archiveren en kijkt of dat goed gebeurt.
- *Contentbeheerder:* Zorgt voor de inhoudelijke en procesmatige informatie over de website (zie: 'Overzicht websites').
- *Functioneel beheerder CMS:* Verstreckt de technische gegevens over de website. Kan daarnaast de harvestingsdienst en toegangsdienst leveren of een externe leverancier daarvoor aansturen.
- *Harvestingsdienstverlener:* Zorgt voor de harvestingsdienst.
- *Toegangsdienstverlener:* Zorgt ervoor dat iedereen toegang heeft tot de geharveste websites. In de praktijk is de toegangsdienstverlener vaak dezelfde als de harvestingsdienstverlener.
- *Adviseur archiefbeheer:* Adviseert over alle algemene aspecten van archivering. En begeleidt de overbrenging van de webarchiefbestanden naar een archiefbewaarplaats.
- *Beheerder archiefbewaarplaats:* Beheert de overgebrachte webarchieven en ontsluit deze voor het publiek.

5.6 Plan en budget

Actie: Stel een plan van aanpak op. Beschrijf daarin minimaal doel, doelgroep, verwachte tijdsinvestering en budget. Als er nog geen budget is vastgesteld, is het belangrijk eerst duidelijk te krijgen hoeveel geld er beschikbaar is. Laat de opdrachtgevers het plan goedkeuren.

Doel: Voor de eigen organisatie duidelijk maken welke activiteiten nodig zijn en de benodigde middelen daarvoor reserveren zodat websitearchivering uitgevoerd kan worden.

Betrokkenen: Zorgdrager, contenteigenaar, contentbeheerder, functioneel beheerder CMS, harvestingsdienstverlener (als die al bekend is), archiefbeheerder, archiefbewaarplaats.

5.7 Maak overzicht websites

Actie: Zoek uit van welke websites de overheidsorganisatie contenteigenaar is en wie het directe aanspreekpunt is voor die websites. Vergeet daarbij niet de websites waarvan (een deel van) het beheer is uitbesteed. Ook die horen erbij.

Doel: De inventarisatie levert informatie op om te bepalen welke websites gearchiveerd moeten worden en wie daarbij een rol speelt. De harvestingdienstverlener heeft die informatie nodig voor het archiveren. En de opdrachtgever van de archivering om de kwaliteit van de webarchieven te laten controleren.

Betrokkenen: Contenteigenaar, contentbeheerder, functioneel beheerder CMS, archiefbeheerder, archiefbewaarpplaats.

5.8 Beheer overzicht websites

Actie: Maak afspraken over wie de inventarisatie bijwerkt als er grote wijzigingen in websites zijn. Bijvoorbeeld als er een nieuwe website bij komt, of als er een stopt. Zorg dat binnen de organisatie bekend is dat nieuwe en gewijzigde websites gemeld worden aan de beheerder van het overzicht. Zo kan die het overzicht voor de archivering bijwerken.

Doel: Het overzicht van de websites is altijd actueel, zodat de harvestingdienstverlener op de hoogte is van wijzigingen en daarop maatregelen kan nemen.

Betrokkenen: Contenteigenaar, contentbeheerder, functioneel beheerder CMS, harvestingdienstverlener.

5.9 Harvestingdienst

Actie: Maak afspraken over het uitvoeren van de harvesting van de websites (de harvestingdienst). Deze richtlijn gaat ervan uit dat hiervoor een contract met een externe leverancier wordt afgesloten. Harvesting vraagt specialistische kennis die niet standaard aanwezig is bij veel overheidsorganisaties, maar wel bij commerciële leveranciers. Een organisatie kan ook zelf een harvestingdienst inrichten maar dat vraagt vaak meer investering dan uitbesteden. De harvestingdienstverlener moet vooraf aangeven welke eisen voor websites gelden om ze te kunnen harvesten. Let op dat de harvesting niet ten koste gaat van de beschikbaarheid van de website voor gebruikers.

Leg de afspraken vast. Vooral hoe de opgebouwde webarchieven aan de opdrachtgever worden overgedragen als het contract met harvestingdienstverlener stopt (voorzien of onvoorzien). Leg ook de rechten van de opdrachtgever op de opgebouwde archieven vast zodat deze nog steeds kan beschikken over de webarchieven die de externe leverancier heeft gemaakt. De ARBIT en de CIBIT hebben voorbeelden voor het opstellen van escrow-regelingen die hierbij bruikbaar zijn.

Deze norm bevat geen informatie over het inrichten van een harvestingdienst. Raadpleeg daarvoor de documentatie en andere informatie over harvestingsoftware.

Doel: De harvestingdienstverlener maakt de webarchieven. En goede afspraken over een duidelijke rolverdeling en over eisen waaraan de harvestingdienst en de websites moeten voldoen.

Betrokkenen: Zorgdrager, contenteigenaar, harvestingdienstverlener.

5.10 Toegangsdienst

Actie: Maak afspraken over het beschikbaar stellen van de geharveste websites aan het publiek (de toegangsdienst). In de praktijk biedt de harvestingdienstverlener vaak de toegangsdienst aan, maar een organisatie kan hiervoor ook zelf een dienst inrichten. De toegangsdienst moet vooraf aangeven welke eisen gelden voor de websites die geharvest moeten worden om ze goed toegankelijk te maken. Wanneer je gearchiveerde websites op een eigen website publiceert, sluit deze dan uit van harvesting. Anders krijg je in één harvest ook alle eerder geharveste websites mee.

Doel: Vaak levert dezelfde leverancier de toegangsdienst en de harvestingdienst, maar dat hoeft niet. Goede afspraken maken duidelijk wat de rolverdeling is en aan welke eisen de toegangsdienst en de websites moeten voldoen. Deze norm bevat geen informatie over het inrichten van een eigen toegangsdienst. Raadpleeg daarvoor de documentatie en andere informatie over de betreffende software.

Betrokkenen: Zorgdrager, contenteigenaar, toegangsdienst.

5.11 Bewaarplaats

Actie: Bepaal waar de webarchiefbestanden direct na harvesting bewaard worden tot het moment van overbrengen. Dit kan bijvoorbeeld in een voorziening waar de zorgdrager ook zijn andere archieven bewaart (vaak aangeduid als DMS, RMA of e-depot). Of bij de leverancier van de harvesting- of toegangsdienst. Controleer of deze bewaarplaats voldoet aan dezelfde beveiligingseisen die de zorgdrager stelt aan de bewaarplaats van zijn andere archieven.

Doel: Beheren van webarchiefbestanden volgens dezelfde beveiligingsnormen als de andere digitale archieven van de zorgdrager.

Betrokkenen: Harvestingdienstverlener, toegangsdienstverlener, archiefbeheerder.

5.12 Beperkingen beschikbaarheid

Actie: Maak afspraken over het beheren van de lijst met webpagina's waarvoor beperkte beschikbaarheid geldt:

- Wie houdt de lijst met beperkingen bij?
- Waar en hoe kunnen verzoeken worden ingediend voor opname op de lijst?
- Wie bepaalt of en hoe het verzoek wordt gehonoreerd?
- Hoe wordt de leverancier van de toegangsdienst geïnformeerd over wijzigingen in de lijst?

Doel: Voorkomen dat uit het webarchief onrechtmatig informatie wordt verstrekt.

Betrokkenen: Contentbeheerder, beveiligingsbeambte, privacybeambte, functionaris voor de gegevensbescherming (FG), leverancier toegangsdienst.

5.13 Start harvesting

Actie: Meld alle websites uit de inventarisatie aan bij de harvestingsdienstverlener zodat harvesting kan starten. Lever de geharveste websites aan de toegangsdienstverlener zodat ze toegankelijk zijn.

Informeer vooraf de functioneel beheerders van de websites over de start van de jaarlijkse harvesting van de hele website. En over het dagelijks harvesten van de wijzigingen. Zo kunnen zij als dat nodig is maatregelen nemen om harvesting mogelijk te maken zonder dat de website daar last van heeft.

Doel: Websites bewaren en bewaarde sites beschikbaar stellen. Voorkomen dat de websites slecht werken door het harvesten of dat harvesting niet goed lukt.

Betrokkenen: Contenteigenaar, contentbeheerder, functioneel beheerder CMS, harvestingsdienstverlener.

5.14 Verwijzing naar gearchiveerde versie op live website

Actie: Zet op de websites verwijzingen naar de gearchiveerde versies van de websites, bijvoorbeeld bij de zoekfunctie.

Doel: Gebruikers van de website kunnen de gearchiveerde webpagina's gemakkelijk vinden als ze op zoek zijn naar oude versies van webpagina's.

Betrokkenen: Functioneel beheerders van websites.

5.15 Kwaliteitscontrole

Actie: Laat de kwaliteit van webarchiefbestanden regelmatig controleren. De leverancier moet uitleggen wat de oorzaak is van fouten en passende maatregelen nemen om herhaling te voorkomen.

Doel: Zorgen dat de webarchiefbestanden voldoen aan de kwaliteitseisen. Zonder dat daar achteraf dure reparaties voor nodig zijn of herstel zelfs onmogelijk is.

Betrokkenen: Contenteigenaar, harvestingsdienstverlener, archiefbeheerder.

5.16 Wijzigingen in websites

Actie: Neem in de beheerprocedures van de websites op dat de harvestingsdienstverlener op tijd ingelicht wordt als een website start, stopt of nooit meer wordt bijgewerkt. En ook bij een grote verandering waardoor veel webpagina's in één keer veranderen.

Stop de harvesting als de website stopt of de inhoud nooit meer wijzigt. Het kan gebeuren dat een website aangemeld wordt voor harvesting terwijl hij al niet meer bijgewerkt wordt. In dat geval is alleen een eenmalige volledige harvest nodig. Incrementele harvest heeft dan geen meerwaarde meer.

Doel: De harvestingsdienstverlener is voorbereid op grote wijzigingen in de websites. Voorkomen van onnodige harvesting.

Betrokkenen: Contenteigenaar, contentbeheerder, functioneel beheerder CMS, harvestingsdienstverlener.

5.17 Laat de selectielijst zo nodig aanpassen

Actie: Check of de selectielijst duidelijk aangeeft dat openbare websites blijvend bewaard (kunnen) worden en laat deze aanpassen wanneer dat niet het geval is. Zie ook eis [4.12](#).

Doel: Duidelijkheid over de vernietigingstermijn van websites.

Betrokkenen: Zorgdrager, archiefbeheerder, Nationaal Archief.

5.18 Overbrengen

Actie: Breng de webarchiefbestanden van jaargangen waarvan de overbrengingstermijn is aangebroken over naar de archiefbewaarpplaats van de organisatie.

De overbrengingstermijn is wettelijk maximaal twintig jaar. Een archiefbewaarpplaats kan toestaan de webarchiefbestanden al eerder over te brengen, maar is dat niet verplicht.

Welke organisatie de archiefbewaarpplaats verschilt per overheidsorganisatie (zie [6.1](#)). Ook de manier van overbrengen kan per archiefbewaarpplaats verschillen. Maak daarom al bij het inrichten van webarchivering afspraken met de archiefbewaarpplaats over wanneer, hoe vaak en hoe de bestanden worden overgebracht. Neem contact op met de eigen adviseur voor archivering. Die weet wat de archiefbewaarpplaats is en wat de afspraken zijn voor het overbrengen.

Doel: Overbrengen garandeert de toegankelijkheid van de websitearchieven voor de lange termijn zonder dat dit afhankelijk is van de contenteigenaar.

Betrokkenen: Zorgdrager, archiefbeheerder, archiefbewaarpplaats.

6 Bijlage: Begrippen

6.1 Archiefbewaarplaats

Een bij of krachtens de Archiefwet 1995 voor de blijvende bewaring van archiefbescheiden aangewezen bewaarplaats.

Bron: Archiefwet 1995.

Toelichting

Een archiefbewaarplaats heeft de taak om overheidsinformatie te beheren die voor altijd bewaard moet blijven. Nadat een overheidsorganisatie zijn archiefbescheiden (gearchiveerde overheidsinformatie) heeft overgedragen, is de archiefbewaarplaats verantwoordelijk voor de duurzame toegankelijkheid van de archiefbestanden.

Er zijn verschillende archiefbewaarplaatsen:

- *Nationaal Archief*. Beheert archieven van rijksoverheden die landelijk werken, zoals ministeries en zelfstandige bestuursorganen (ZBO's). Bijvoorbeeld de Kamers van Koophandel, Staatsbosbeheer en de Sociale Verzekeringsbank. Het Nationaal Archief is ook het Regionaal Historisch Centrum voor de provincie Zuid-Holland.
- *Regionaal Historisch Centrum (RHC)*. Beheert de archieven van de aangesloten provincies, gemeenten en waterschappen. En van regionaal werkende rijksoverheden zoals rechtbanken. RHC's zijn samenwerkingsverbanden tussen het Rijk, provincies, gemeenten en waterschappen. De RHC's bevinden zich in de provinciehoofdsteden.
- *Gemeentelijke archiefbewaarplaats*. Sommige gemeenten hebben een eigen archiefbewaarplaats. Andere gemeenten werken samen, bijvoorbeeld door deel te nemen aan een RHC of een ander samenwerkingsverband.
- *Waterschapsarchiefbewaarplaats*. Sommige waterschappen hebben een eigen archiefbewaarplaats, andere werken samen. Bijvoorbeeld door deelname aan een RHC of een ander samenwerkingsverband.

6.2 Archiveren

Het geheel van activiteiten die nodig zijn om informatie duurzaam toegankelijk te maken en te houden.

Bron: Eigen definitie. De Archiefwet 1995 bevat geen definitie van archiveren. Daarom wordt een eigen definitie gebruikt die aansluit bij de bepalingen over archiefbescheiden in de Archiefwet 1995.

Toelichting

Toegankelijk betekent vindbaar, beschikbaar, leesbaar, interpreteerbaar en betrouwbaar voor degenen die er recht op hebben, vanaf het moment van ontstaan en voor zolang als noodzakelijk. Duurzaam betekent dat de toegankelijkheid van de informatie bestand is tegen alle soorten veranderingen.

Overheidsorganisaties zijn bij wet verplicht te archiveren. Artikel 3 van de Archiefwet 1995 zegt er het volgende over:

“De overheidsorganisaties 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.”

Archiefbescheiden betekent alle informatie die een overheidsorganisatie heeft gemaakt of gekregen bij het uitvoeren van zijn taken. De vorm van die informatie maakt niet uit.

De duur van het archiveren wordt volgens de Archiefwet 1995 bepaald in selectielijsten.

Volgens deze definitie start archiveren al bij het ontstaan van de informatie en is het onafhankelijk van de plaats waar de informatie bewaard wordt. Hierin wijkt de definitie die gebaseerd is op de Archiefwet af van de veel gehanteerde betekenis dat archiveren betekent het opslaan van afgehandelde informatie in een speciaal daarvoor ingerichte archiefvoorziening.

6.3 Checksum

Een gegeven dat afgeleid is van een blok van digitale gegevens met het oog op het detecteren van fouten die zijn opgetreden tijdens het transport of opslag van het blok gegevens.

Bron: Engelse Wikipedia.

Toelichting

De checksum is meestal een korte reeks cijfers en letters. Het blok digitale gegevens is meestal een bestand of een deel van een bestand (zoals één record in een verzamelbestand). De essentie van een checksum is dat hij uniek is voor elk blok gegevens, maar veel kleiner dan het blok gegevens zelf. Je zou het kunnen zien als een digitale vingerafdruk van het blok gegevens.

Een checksum wordt met een algoritme berekend bij de creatie van het blok gegevens of net voor het versturen over een netwerk. Er bestaan verschillende checksumalgoritmes zoals MD5, SHA1, SHA256 of SHA512.

De checksum kan op ieder willekeurig tijdstip in de toekomst opnieuw berekend worden met het algoritme. Bijvoorbeeld na ontvangst van het digitale bestand. Als de tweede checksum overeenkomt met de eerste checksum, is vrijwel zeker dat het bestand niet gewijzigd of beschadigd is geraakt tussen de twee berekeningen. Als de checksums verschillen is dat een indicatie dat er iets mis is gegaan met het bestand.

6.4 Content / webcontent

Informatie op een website die aan de gebruiker doorgegeven wordt met een user agent, met inbegrip van code of opmaak die de structuur, presentatie en interacties van de content definieert.

Bron: [Web Content Accessibility Guidelines \(WCAG\) 2.0](#).

Toelichting

De content van een webpagina is de informatie die de gebruiker kan waarnemen als hij of zij de webpagina raadpleegt met een user agent.

6.5 Contentbeheer (van een website)

Maken en onderhouden van webcontent.

Bron: Eigen definitie.

6.6 Contenteigenaar (van een overheidswebsite)

De overheidsorganisatie uit wiens taken de inhoud van de website voortkomt.

Bron: Eigen definitie.

Toelichting

De contenteigenaar is verantwoordelijk voor de inhoud en daarmee ook voor de archivering van een overheidswebsite. Zowel het contentbeheer als de websitearchivering kan uitbesteed worden. Maar de contenteigenaar blijft daarvoor eindverantwoordelijk.

6.7 Content Management Systeem (CMS)

Een softwaretoepassing die het mogelijk maakt dat mensen eenvoudig, zonder veel technische kennis, documenten en gegevens op internet kunnen publiceren (contentmanagement). Meestal is een webapplicatie.

Bron: Nederlandse Wikipedia.

Toelichting

Een CMS is de software waarmee je contentbeheer kunt uitvoeren.

6.8 Embedded content (ingesloten webbron)

Webbron die onderdeel vormt van een webpagina.

Bron: Afgeleid van de definitie van webpagina (zie definitie [6.18](#)).

Toelichting

Embedded content is onderdeel van een webpagina die niet in de code van de webpagina zelf is opgenomen, maar die verwijst naar informatie op een server. Voorbeelden van embedded content zijn de plaatjes op een webpagina en de style sheets waarin de lettertypes van de webpagina zijn vastgelegd. Zie ook de definitie van webpagina.

6.9 Harvesting (van een website)

Met http ophalen van alle webpagina's van een website voor opslag of analyse.

Bron: bewerking van de Engelse Wikipedia.

Toelichting

Speciale software (de harvester) haalt webpagina's op bij de webserver. Inclusief afbeeldingen, vormgeving en bijlagen. De harvestingscope bepaalt welke webpagina's worden opgehaald (zie definitie [6.11](#)).

Harvesting wordt onder andere gebruikt voor webarchivering en voor het maken van zoekindexen voor zoekmachines.

6.10 Harvestingfrequentie

Hoe vaak harvesting van een website plaatsvindt.

Bron: Eigen definitie.

Toelichting

De harvestingfrequentie bepaalt voor een belangrijk deel de volledigheid van het harvesten. Er zijn verschillende mogelijkheden:

6.10.1 Eenmalige volledige harvesting

Harvesting vindt slechts op één moment plaats. Veranderingen die daarna plaatsvinden worden niet opgemerkt.

6.10.2 Periodieke volledige harvesting

Harvesting wordt met vaste tussentijden herhaald. Bijvoorbeeld dagelijks of wekelijks. Als hetzelfde item tussen twee opeenvolgende harvests vaker verandert, wordt alleen het resultaat van de laatste verandering vastgelegd.

6.10.3 Periodieke incrementele harvesting

Bij deze variant van harvesting gaat de harvester de hele website langs, maar slaat alleen de webpagina's op die sinds de vorige (incrementele) harvest zijn gewijzigd.

Het voordeel van incrementele harvesting is dat je opslag- en reken capaciteit bespaart voor webpagina's die niet zijn gewijzigd.

6.10.4 Real-time harvesting

Het CMS van de website stuurt bij het toevoegen, wijzigen en verwijderen van een webpagina een signaal naar de harvester. Die haalt de webpagina dan op.

Het voordeel van real-time harvesting is dat alle wijzigingen op een website worden opgemerkt. Bovendien wordt alleen maar opslag- en reken capaciteit gebruikt voor webpagina's die veranderen. Het nadeel van incrementele harvesting is dat het CMS zo ingericht moet zijn dat het een signaal verstuurt. Bij harvesting op vaste tijdstippen is de harvester niet afhankelijk van het CMS.

6.11 Harvestingscope

De afbakening van welke webpagina's bij harvesting van een website worden opgehaald en welke niet.

Bron: Eigen definitie.

Toelichting

Een harvester moet weten welke webpagina's bij een website horen. Theoretisch zou dat kunnen door vooraf een complete lijst te geven van alle webpagina's van de website. Maar in de praktijk is dat meestal onmogelijk omdat dit een hele lange lijst kan zijn, die ook nog eens steeds verandert. Daarom gaat een harvester uit van een harvestingscope die bepaald wordt door een aantal parameters.

De parameters van de harvestingscope kunnen per harvester verschillen. Voor de meeste harvesters zijn het varianten door de volgende parameters bepaald.

6.11.1 Startpagina (seed-url)

De harvester begint met het binnenhalen van de opgegeven startpagina. Vanuit de startpagina haalt de harvester vervolgens alle webpagina's binnen die via hyperlinks bereikbaar zijn.

De startpagina voor harvesting is meestal de hoofdpagina van de website (ook wel homepage genoemd). Bijvoorbeeld www.rijksoverheid.nl. Maar het kan ook een subdomein zijn, zoals het Actorenregister: actorenregister.nationaalarchief.nl. Of een sub-website, bijvoorbeeld www.overheidsorganisatie.nl/over-ons/wie-is-wie.

6.11.2 Weespagina's

Weespagina's zijn webpagina's binnen de website die de harvester niet kan ontdekken en daarom niet kan binnenhalen. Bijvoorbeeld doordat de hyperlink naar een weespagina doormiddel van een script pas berekend wordt op het moment dat een webpagina wordt weergegeven.

Weespagina's horen wel bij de website en moeten daarom wel geharvest worden. Als bekend is wat de weespagina's zijn, dan worden die op een lijst gezet. Zo weet de harvester dat hij die webpagina's ook moet archiveren.

6.11.3 Site-url

De hoofd-url waarmee alle webpagina's van een website beginnen. Zoals <http://www.overheid.nl>.

6.11.4 Secundaire site-url's

Soms kan het gebeuren dat een deel van een website met een andere url begint. Bijvoorbeeld bijlagen die op een andere server zijn opgeslagen. Als dat het geval is kunnen nog een aantal secundaire site-url's worden opgegeven.

6.11.5 Blacklist

Dit is een lijst met webpagina's die om technische redenen niet opgehaald kunnen worden. De harvester slaat deze webpagina's over.

Een voorbeeld van een technische reden om een webpagina niet te harvesten is een agenda op de site. Zo'n agenda kan tot in het oneindige doorlopen in het verleden en in de toekomst. Daardoor weet de harvester niet waar hij moet stoppen.

De blacklist is onderdeel van de lijst met niet-harvestbare content.

6.11.6 Scope voor embedded content (ingesloten webbronnen)

Afbakening welke embedded content van de webpagina's worden opgehaald. Dit kan alles zijn, of beperkt tot bepaalde hostingdomeinen, bestandstypen of multimediaformaten.

Een webpagina is meestal opgebouwd uit verschillende onderdelen: html, CSS, JavaScript, afbeeldingen, multimedia en links naar bijlagen zoals pdf-documenten. Het kan zijn dat sommige onderdelen op een andere server staan dan de site zelf.

6.11.7 Maximaal aantal hops

Het aantal hops is het aantal links dat de harvester volgt om een webpagina te bereiken vanaf de startpagina. Als het aantal hops beperkt is, dan wordt mogelijk niet de hele website geharvest. Deze norm staat een maximum aan het aantal hops niet toe.

6.11.8 Maximale paddiepte

De paddiepte is het aantal url-elementen dat in de url van een webpagina achter de domeinnaam staat. Een voorbeeld. De homepage www.rijksoverheid.nl heeft een paddiepte van 0. En www.rijksoverheid.nl/actueel/agenda heeft een paddiepte van 2 ('actueel' en 'agenda'). Als de paddiepte te beperkt is, dan wordt mogelijk niet de hele website geharvest. Deze norm staat een maximum aan de paddiepte niet toe.

6.11.9 Scope+1

Vaak bevat een webpagina links naar webpagina's buiten de website. Dat kan een andere site zijn van de eigen overheidsorganisatie, maar ook een website van een andere organisatie. Bij scope+1 haalt de harvester ook deze externe webpagina's binnen. De hyperlinks op die externe webpagina worden verder niet gevolgd. Zonder scope+1 haalt de harvester geen externe pagina's op. Deze norm staat het gebruik van 'scope+1' niet toe.

6.12 Overbrengen

Verplichte procedure uit de Archiefwet om overheidsinformatie die blijvend bewaard moet worden, na uiterlijk twintig jaar aan een daarvoor aangewezen archiefbewaarpplaats over te dragen voor beheer en beschikbaarstelling.

Bron: Eigen definitie.

Toelichting

Zie de [DUTO wiki](#) en op de [website van het Nationaal Archief](#).

6.13 Selectielijst

Een conform de Archiefwet vastgestelde systematische opsomming van categorieën informatieobjecten van een overheidsorgaan, waarin bij iedere categorie is aangegeven of de informatieobjecten blijvend bewaard worden of na welke termijn zij voor vernietiging in aanmerking komen, met bijhorende toelichting en verantwoording.

Bron: Ontleent aan [Artikel 5 Archiefwet 1995](#) en [Artikel 5 Archiefbesluit 1995](#).

Toelichting

Met selectielijsten maakt een overheidsorganisatie bekend hoe ze informatie selecteert voor vernietiging of blijvende bewaring. Dit heeft twee doelen. Ten eerste maakt het voor belanghebbenden duidelijk hoe lang informatie bewaard wordt. Daarom zijn selectielijsten officiële publicaties. Ten tweede is een selectielijst een praktisch instrument waarmee bepaald wordt of en wanneer een informatieobject vernietigd moet worden.

Informatie die volgens de selectielijst tijdelijk bewaard wordt, moet volgens de Archiefwet 1995 na die tijd vernietigd worden. Informatie die blijvend bewaard moet worden, moet de overheidsorganisatie overbrengen naar een archiefbewaarpplaats. Dat moet na uiterlijk twintig jaar.

In de [Handreiking waardering en selectie \(Belangen in Balans\)](#) staat wat een selectielijst is en hoe hij gemaakt wordt. De archiefwetgeving stelt geen vernietigingstermijnen vast voor webarchivering. De wet beschrijft alleen de procedure voor het opstellen en vaststellen van een selectielijst. Elke overheidsorganisa-

tie moet de vernietigingstermijnen zelf bepalen door belangen af te wegen. Dat noemen we waarderen. Archiefinstellingen waaraan zij hun blijvend te bewaren informatie overbrengen helpen daarbij.

6.14 Url (Uniform Resource Locator)

De unieke naam waarmee een webbron bij een webserver opgehaald kan worden.

Bron: [Bewerking van Wikipedia](#).

Toelichting

Bijvoorbeeld <https://en.wikipedia.org/wiki/url>. Zie verder de definitie op Wikipedia.

6.15 User agent

Alle software die webcontent voor gebruikers ophaalt en presenteert.

Bron: [Web Content Accessibility Guidelines \(WCAG\) 2.0](#).

Toelichting

Voorbeeld: Webrowsers, mediaspelers, plug-ins en andere programma's en hulptechnologieën die helpen bij het ophalen en weergeven van en het interacteren met webcontent.

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

6.16 WARC (Web ARChive format)

ISO-standaard (ISO 28500) voor het combineren van meerdere webbronnen in een samengesteld archiefbestand, samen met informatie gerelateerd aan de vorming van het archiefbestand.

Bron: [Bewerking van Wikipedia](#).

Toelichting

Typisch voor een WARC-bestand is dat dit webbronnen bevat die zijn opgehaald met harvesting (hoewel WARC ook voor andere methodes van archivering gebruikt kan worden).

Een WARC-bestand bevat:

- Requests die de harvester naar de webserver heeft gestuurd.
- Responses die de webserver heeft teruggestuurd naar de harvester.
- Metagegevens over het harvestingproces.

Zie voor verdere toelichting de ISO-standaard ISO 28500.

6.17 Webbron

Een informatieobject of service beschikbaar via een netwerk en geïdentificeerd door een url.

Bron: Vertaling van [W3C Glossary of Terms for Device Independence](#).

Toelichting

Een webpagina is bijvoorbeeld een webbron, maar ook de afbeeldingen op de webpagina zijn webbronnen.

6.18 Webpagina

Een verzameling informatie, bestaande uit een of meer webbronnen, geïdentificeerd door een enkele url en bedoeld om gezamenlijk weer te geven. Meer specifiek bestaat een webpagina uit een webbron met nul, een of meer vormen van embedded content die bedoeld zijn om te worden weergegeven als een eenheid en waarnaar wordt verwezen door de url van de ene bron die niet is ingesloten.

Bron: Vertaling van [W3C Glossary of Terms for Device Independence](#).

Opmerking: In de W3C-definitie wordt gesproken over URI in plaats van url. Het verschil is hier niet relevant. Hier is gekozen voor url omdat het gebruikelijk is een webpagina met een url aan te duiden.

Toelichting

Een webpagina is een online document met een webadres (url). Op basis van de url kan een user agent de webpagina opvragen bij een webserver en vervolgens weergeven.

Een webpagina kan een willekeurig formaat hebben. Veel webpagina's bestaan uit html-code, maar dat hoeft niet. Ook bijvoorbeeld een tekstbestand of een afbeelding is een webpagina als een user agent deze met een url kan opvragen bij een webserver.

Een webpagina bestaat meestal uit meerdere webbronnen voor verschillende onderdelen van de webpagina. Bijvoorbeeld de hoofdtekst, afbeeldingen, video's, scriptcode en style sheets (instructies voor de vormgeving). De hoofdbron wordt door de url aangeduid, de url's van de andere bronnen staan in de hoofdbron. Dit wordt de embedded content genoemd (ingesloten webbronnen). De embedded content wordt afzonderlijk opgehaald bij de webserver, maar vormt een integraal onderdeel van de webpagina. Waar er gesproken wordt over het archiveren van een webpagina wordt daar dus ook alle embedded content bedoeld.

Kenmerkend voor een webpagina is dat deze meestal hyperlinks bevat naar andere webpagina's. Daarmee vormen alle webpagina's samen een web van informatie waarin je gemakkelijk kunt rondkijken.

Vaak staat een overheidswebpagina op een server van een overheidsorganisatie, maar dat hoeft niet. Het kan ook een webserver zijn van een bedrijf of andere niet-overheidsorganisatie. Dit verandert niets 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.

6.19 Website

Een verzameling onderling verbonden webpagina's, inclusief een startpagina, met dezelfde begin-url (de site-url).

Bron: Vertaling van [Web Characterization Terminology & Definitions Sheet](#). Daarbij is "residing at the same network location" vertaald naar "met dezelfde begin-url". Zo sluit het beter aan bij de definitie van webpagina..

Toelichting

Een website is een samenhangende verzameling webpagina's die bij elkaar horen. Bijvoorbeeld omdat ze bij dezelfde organisatie horen of over hetzelfde onderwerp gaan.

De startpagina en de site-url bepalen welke webpagina's bij een website horen. Vaak is de url van de startpagina hetzelfde als de site-url. Bijvoorbeeld <https://nos.nl/> voor de website van de NOS. Onderling verbonden betekent dat elke webpagina van de website bereikbaar is door het volgen van een reeks van hyperlinks, te beginnen bij de startpagina via nul, één of meer webpagina's op dezelfde site.

Dezelfde begin-url betekent dat de url van elke webpagina van de website begint met de site-url. Bijvoorbeeld <https://nos.nl/nieuws/>.

Het kan voorkomen dat een deel van de webpagina's van een website begint met een andere url dan de site-url. Bijvoorbeeld bijlagen die op een andere server zijn opgeslagen. Als dat het geval kent een website ook een of meerdere secundaire site-url's. Webpagina's die bereikbaar zijn vanaf de startpagina maar beginnen met een secundaire site-url, worden dan ook tot de website gerekend.

De embedded content van een webpagina op een site hoort altijd bij die website, ongeacht wat de url van de embedded content is. Want de embedded content is een onlosmakelijk deel van een webpagina, en daarmee ook van de website.

6.20 Websitearchivering

Het archiveren van websites.

Bron: Eigen definitie.

Toelichting

Websitearchivering is het duurzaam toegankelijk maken van de webpagina's van een website. Soms wordt als definitie gebruikt 'het binnenhalen en opslaan van een website'. Maar dat is slecht één van de methoden voor websitearchivering.

Omdat webpagina's vaak veranderen en verdwijnen, gaat websitearchivering niet vanzelf. Hiervoor moet je aanvullende maatregelen nemen. Er zijn verschillende manieren om dat aan te pakken. Client-side archivering is de meest gebruikte.

6.20.1 Client-side archivering (harvesting)

Bij deze methode benadert een harvester de website zoals een gebruiker en 'ziet' de website dus alleen aan de voorkant: de client-side (zie ook definitie [6.9](#)).

Een voordeel is dat het niet ingrijpt in de techniek van de website. Er wordt immers alleen contact gemaakt zoals een willekeurige bezoeker van de website dat ook doet. Deze manier van harvesting kent daarmee echter ook een aantal beperkingen:

- Geharveste webpagina's zijn opgevraagd door één (anonieme) gebruiker. Webpagina's waarvan de inhoud afhankelijk is van kenmerken van de gebruiker (zoals een ingevuld formulier of de inhoud van cookies) kunnen daarom op deze manier niet gearchiveerd worden. Maar vooral kunnen webpagina's waarvoor de gebruiker moet inloggen niet geharvest worden.
- Tijdens het raadplegen van een geharveste webpagina is de achterliggende webserver niet meer beschikbaar in de toestand van tijdens het harvesten. Als voor de weergave van een webpagina interactie met de webserver nodig is, is het resultaat daarom niet gegarandeerd hetzelfde als op het moment van harvesting. In veel gevallen zal het zelfs een foutmelding kunnen geven. Zoals bij streaming video en interactieve kaartjes. Zolang de weergave in zijn geheel binnen een webbrowser wordt afgehandeld, bestaat deze beperking niet.
- Websites die niet meer online zijn kunnen niet geharvest worden. Dit kan soms opgelost worden door de website tijdelijk weer online te zetten.

6.20.2 Server-side archivering

Bij deze methode haal je de inhoud van een website op bij het CMS, of het CMS zorgt zelf voor het archiveren. Dit kan alleen als de beheerder van het CMS daaraan meewerkt. Hiervoor moet je voor ieder CMS afzonderlijk afspraken maken. En misschien moet het CMS aangepast worden.

Het voordeel van deze aanpak is dat je alle informatie in het CMS kunt archiveren. Inclusief eventuele geschiedenisinformatie, wijzigingen aan teksten en andere informatie die wel in het CMS zit maar niet aan de voorkant (client-side) zichtbaar is. Een nadeel van deze aanpak is dat die sterk afhankelijk is van de techniek van het CMS.

In paragraaf 7.3 staat waarom in deze norm niet voor deze vorm van archiveren is gekozen.

6.20.3 Transactiearchivering

Bij transactiearchivering archiveer je tijdens het gebruik van een website het uitwisselen van data tussen de user agent en de webserver. Daarmee archiveer je alle interactie tussen gebruikers en de website. Transactiearchivering wordt (nog) weinig toegepast.

Het voordeel van transactiearchivering is dat je precies alle informatie archiveert die de gebruikers kregen en verstuurden. Inclusief de inhoud die afhankelijk is van de gebruiker en de informatie die de gebruiker terugstuurt (bijvoorbeeld formulieren). Maar ook de interactie met de server voor het bepalen van het gedrag van een webpagina. Met transactiearchivering kun je het hele websitebezoek van een gebruiker terughalen.

Maar transactiearchivering kan ook een paar grote nadelen hebben. De privacy van de gebruikers van de website is namelijk in het geding. Daarnaast levert het ook veel gegevens en dat vraagt veel opslagruimte. Een alternatief voor harvesting kan zijn webrecording. Hierbij wordt een sessie van één (anonieme) gebruiker opgenomen. Maar dat is onpraktisch omdat die gebruiker dan alle links handmatig moet aanklikken. En dit moet dan ook nog eens worden herhaald bij alle wijzigingen aan de website.

6.21 Zorgdrager

Degene die bij of krachtens de Archiefwet 1995 belast is met de zorg voor de archiefbescheiden.

Bron: Archiefwet 1995.

De Archiefwet 1995 bepaalt per type organisatie wie de zorgdrager is. Meestal is dat de bestuurder of het dagelijks bestuur. Voor een ministerie is het bijvoorbeeld de minister en voor een gemeente het college van burgemeester en wethouders.

Het begrip zorgdrager wordt ook vaak gebruikt voor de organisatie waar het archief onder valt. Maar strikt genomen klopt dat niet.

7 Bijlage: Motivatie van gemaakte keuzes

Bij het schrijven van deze norm heeft het Nationaal Archief een aantal keuzes gemaakt, in afstemming met een aantal deskundigen. Sommige van deze keuzes hebben tot vragen geleid. Die keuzes zijn hieronder gemotiveerd.

7.1 Scopeafbakening

Op een aantal punten is de scope afgebakend omdat er op dit moment nog geen gangbare oplossingen bekend zijn. Op die punten kunnen toekomstige versies van deze norm uitgebreid worden.

7.1.1 Alleen openbare webpagina's

Voor niet-openbare delen van een website gelden beperkingen voor de toegankelijkheid en mogelijk technische beperkingen voor het harvesten (omdat de harvester toegang moet krijgen tot de niet-openbare delen). Daarbij is het de vraag of de niet-openbare delen van een website ook blijvend bewaard moeten worden. Dit maakt dat er extra inspanning nodig is om te bepalen hoe om te gaan met de niet-openbare delen van website.

7.1.2 Alleen niet-gepersonaliseerde webpagina's

Als de inhoud van een webpagina afhankelijk is van persoonlijke gegevens, dan moeten die voor de archivering bekend zijn. Dat brengt een aantal vraagstukken met zich mee die verder onderzocht moeten worden. Bijvoorbeeld: Hoe stel je vast welke gepersonaliseerde webpagina's zijn opgevraagd, hoe lang moeten deze gearchiveerd worden en hoe wordt daarbij de privacy gerespecteerd.

7.1.3 Geen sociale media

Zowel de rol als de techniek van sociale media zijn wezenlijk anders dan van websites. En die verschillen ook nog eens per type sociale media. Dat maakt dat de eisen voor het archiveren van websites niet zonder meer afdoende zijn voor het archiveren van sociale media. Op dit moment kan het Nationaal Archief geen standaard werkwijze adviseren voor het archiveren van uitingen op sociale media. Er is extra inspanning nodig om dat te bepalen.

7.1.4 Geen gerelateerde informatie

De norm gaat alleen over informatie in de vorm van een website. Niet over informatie die gerelateerd is aan de website Zoals het achterliggende communicatiebeleid, informatie over het CMS waarmee de website is gemaakt, de inhoud van de databases van het CMS, de informatie die een gebruiker via een website heeft verstuurd en de verwerking daarvan. Ook al deze informatie moet gearchiveerd worden, zolang als daar voldoende belang bij is. Dit betreft echter andere informatievormen dan websites en ze vallen daarom buiten het bestek van deze norm.

7.1.5 Beperking tot client-side harvesting

Deze eerste versie van de norm beperkt zich tot harvesting als archiveringstechniek. Omdat dat op dit moment de enige gangbare, eenvoudige en generieke techniek is. Bovendien heeft goed ingerichte harvesting weinig tot geen impact op de websites en de achterliggende processen en applicaties. Voor harvesting is open source software beschikbaar en zijn er meerdere commerciële dienstverleners. Dat maakt het ook een betaalbare techniek die breed ondersteund wordt.

7.2 Gebruik van WARC-standaard

De WARC-standaard is de gangbare standaard voor de opslag van geharveste webpagina's. De standaard staat niet ter discussie en wordt breed ondersteund.

7.3 Waarom niet alleen archivering van de onderliggende gegevens?

Deze norm vereist het harvesten (ophalen en bewaren) van de webpagina's zoals de gebruikers van een website ze ontvangen. Deze webpagina's worden gegenereerd door een CMS op basis van gegevens die in het CMS of ergens anders zijn vastgelegd. Een alternatieve strategie is het bewaren van alleen die onderliggende gegevens.

Alleen de gegevens archiveren is onvoldoende. Ook de verschijningsvorm van de webpagina moet achteraf bekend zijn. De achterliggende gegevens alleen geven onvoldoende inzicht in welke informatie de gebruikers van de website hebben ontvangen. De Archiefregeling zegt daarover: "De zorgdrager zorgt ervoor dat van elk van de archiefbescheiden te allen tijde kan worden vastgesteld: [...] de inhoud, structuur en verschijningsvorm bij het ontvangen of opmaken ervan door het overheidsorgaan."

Het is mogelijk om ook de functionaliteit van het CMS te behouden. Zo kunnen op elk later tijdstip webpagina's opnieuw gemaakt worden op basis van de oude achterliggende gegevens. Dit betekent meestal aanpassingen aan het CMS en levert extra werk op bij het aanpassen en vervangen van een het CMS. En dat voor alle verschillende types CMS. Daarbij moet je telkens controleren of het CMS de oude webpagina's nog op de juiste manier maakt. Harvesting is dan een eenvoudiger, goedkopere en beter te controleren manier van archiveren.

7.4 Archiveren embedded content van anderen

Sommige webpagina's bevatten embedded content die door een andere organisatie is gemaakt dan de contenteigenaar van de website. Een voorbeeld daarvan is een embedded Twitter tijdlijn. Archivering van deze embedded content is nodig om de webpagina's later nog volledig te kunnen geven zoals ze ooit waren.

Het zou voor de hand liggen dat de eigenaar van de embedded content deze archiveert. Maar de overheid kan er niet van op aan dat die eigenaar dat inderdaad doet. Daarom is het beter ook de embedded content van anderen mee te archiveren.

7.5 Blijvend bewaren van websites

Hoe lang overheidsinformatie gearchiveerd moet worden is vastgelegd als vernietigingstermijn (ook wel bewaartermijn genoemd) in de zogenaamde selectielijsten. Dit zijn per overheidsorganisatie of groep overheidsorganisaties (bijvoorbeeld alle gemeenten) vastgestelde formele besluiten. De vernietigingstermijnen voor websites zijn nog niet altijd ondubbelzinnig vastgelegd in de selectielijsten. Ook kunnen de termijnen verschillen per selectielijst.

Deze norm gaat ervan uit dat de selectielijsten bepalen dat openbare overheidswebsites blijvend bewaard moeten worden. Deze aanname is gebaseerd op het Generiek Waarderingsmodel Rijksoverheid (GWR) en de Selectielijst gemeenten en intergemeentelijke organen. Samen dekken deze het grootste deel van de overheid af. Het GWR 2.0 stelt dat “eindproducten van publieksvoorlichting” (waar websites onder vallen) blijvend bewaard moeten worden (categorie 5.1). De gemeentelijke selectielijst 2017 geeft aan dat “online-content ten behoeve van externe communicatie” blijvend bewaard moet worden (nr. 20.1.2).

Als een overheidsorganisatie in zijn selectielijst niet heeft opgenomen dat zijn websites blijvend bewaard worden, dan is dat een bespreekpunt met het Nationaal Archief voor de eerstvolgende actualisatie van de selectielijst.

7.6 Alle webpagina's even lang bewaren

Binnen één website worden alle webpagina's en daarvan alle wijzigingen even lang bewaard. Ook als van sommige webpagina's het belang kleiner is.

Verschillen in vernietigingstermijnen zou leiden tot onvolledig gearchiveerde websites waarin onderdelen niet beschikbaar zijn. Het is alsof je selectief pagina's uit een boek scheurt. Dit maakt het gebruik van de gearchiveerde website in de toekomst moeilijker. De ontbrekende onderdelen kunnen van belang zijn voor de interpretatie van de onderdelen die wel beschikbaar zijn.

Daarnaast is één vernietigingstermijn een duidelijke regel. Dat voorkomt misverstanden waardoor ten onrechte gearchiveerde webpagina's worden vernietigd.

De keuze om geen verschil te maken in vernietigingstermijnen sluit ook aan bij het Generiek Waarderingsmodel Rijksoverheid 2.0 (GWR) en de Selectielijst gemeenten en intergemeentelijke organen 2017. Deze maken geen onderscheid in vernietigingstermijnen voor onderdelen van websites. Er is een mogelijk argument tegen het even lang bewaren van alle webpagina's. De kosten voor het harvesten en beheren wegen misschien niet op tegen de meerwaarde van blijvend bewaren. Maar bij het vaststellen van deze norm waren daar geen aanwijzingen voor. Bovendien brengen verschillende vernietigingstermijnen binnen een website ook weer extra kosten met zich mee omdat je dan een selectie moet maken.

Het bewaren van alle wijzigingen aan webpagina's stuit soms op een praktisch probleem. Webpagina's kunnen op een willekeurig tijdstip veranderd of verwijderd worden. Zelfs meerdere keren op één dag. Maar periodieke harvesting van websites gebeurt op vaste tijdstippen. Daarom schrijft de norm één keer per dag harvesten voor. Nog beter is real-time harvesting. Daarbij geeft het CMS elke wijziging in een website meteen door aan de harvestingdienst. Deze kan de wijziging dan meteen vastleggen. Bij het opstellen van deze norm was real-time harvesting nog geen standaard optie bij CMS'en en harvestingsoftware. Daarom is het niet als eis opgenomen.

7.7 Waarom verzorgt het Nationaal Archief de webarchivering niet?

Het is geen wettelijke taak van het Nationaal Archief om archieven te vormen. Volgens de Archiefwet 1995 is elke overheidsorganisatie daar zelf verantwoordelijk voor. Wel worden blijvend te bewaren (web)archieven van het Rijk op termijn overgedragen aan het Nationaal Archief. Daarna neemt het Nationaal Archief het beheer op zich.

Daarnaast is het Nationaal Archief een kenniscentrum voor archivering en helpt daarmee overheidsorganisaties om hun (web)archivering in te richten. Deze norm komt daaruit voort.

7.8 De KB archiveert toch al overheidswebsites?

De Koninklijke Bibliotheek (KB) archiveert inderdaad overheidswebsites. Daarnaast zijn er meer instellingen die websites archiveren, zoals het Nederlands Instituut voor Beeld en Geluid (NIBG) en het Internet Archive. Het KB-webarchief archiveert een selectie van Nederlandse websites vanaf 2007. Deze sites worden één tot twee keer per jaar geharvest. Iedereen met een KB-jaarpas kan het KB-webarchief in de leeszaal van de KB bekijken. Het NIBG archiveert websites van omroepen.

Waarom moeten overheidsorganisaties zelf hun websites archiveren als andere instellingen dat al doen? De reden hiervoor is dat al deze instellingen niet archiveren voor de archiefwetgeving, 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, vernietigingstermijn en kwaliteitskenmerken. Deze keuzes zullen zelden overeenkomen met de keuzes die voortvloeien uit de archiefwetgeving.

Dit alles maakt dat webarchivering door culturele en wetenschappelijke instellingen de wettelijke webarchivering niet kan vervangen. Wel hebben deze instellingen soms jarenlange kennis en ervaring met webarchivering die zij graag delen met overheidsorganisaties.

7.9 Archiveren van webbrowsers buiten de scope

Webpagina's worden geraadpleegd met veel verschillende webbrowsers en apparaten die bovendien steeds veranderen. De weergave van webpagina's is niet gestandaardiseerd en is daarom anders voor elke combinatie van webbrowser en apparaat. Vaak zijn de verschillen klein. Maar relevante verschillen zijn niet uitgesloten, vooral niet tussen de weergave nu en in de verre toekomst.

Daarmee is de principiële vraag aan de orde wat er eigenlijk gearchiveerd wordt bij harvesting. Niet de eigenlijke webpagina zoals de gebruiker die waarneemt, maar slechts de code om een webpagina te maken. Moet dan niet ook de functionaliteit gearchiveerd worden waarmee uit de code de webpagina gemaakt wordt? Het lijkt misschien een oplossing om de code van de webbrowsers zelf ook te archiveren. Maar dan moet je ook de functionaliteit van de apparaten, operating systemen, enzovoort bewaren waarop die browsers draaien. Hier is nog geen betrouwbare en bruikbare oplossing voor.

8 Bijlage: Aanleiding en opdracht van deze norm

8.1 Aanleiding

Overheidswebsites vallen onder de werking van de Archiefwetgeving 1995. Dit betekent dat de overheid haar websites moet archiveren. Eind 2016 onderzocht de Erfgoedinspectie (EGI) in hoeverre dit daadwerkelijk gebeurt bij de centrale overheid. Conclusie is dat overheden hun websites en sociale media onvoldoende archiveren. Dit leidt tot verlies van informatie en is een (juridisch) risico. De EGI noemt als conclusies:

- De organisaties van de centrale overheid archiveren hun websites, een enkele uitzondering daargelaten, niet of onvoldoende.
- Hetzelfde geldt in nog sterkere mate voor officiële uitingen op sociale media.
- Dit leidt tot verlies van informatie en is een risico in het geval van een juridisch geschil.
- Voor de langere termijn dreigt digitale informatie met cultuurhistorisch, maatschappelijk en en/of wetenschappelijk belang te verdwijnen.
- Het ontbreekt aan een beleidskader en aan specificaties voor het archiveren van overheidswebsites en uitingen op sociale media.
- Mede daardoor hebben overheidsorganisaties onvoldoende handvatten voor het archiveren van hun websites en officiële uitingen op sociaalnetwerksites.
- Er is onvoldoende kennis over webarchivering bij de archiefvormende organisaties.
- De taakverdeling tussen Nationaal Archief, Koninklijke Bibliotheek en de Dienst Publiek en Communicatie is, voor wat betreft webarchivering, bij de archiefvormende organisaties onvoldoende duidelijk.

Het rapport leidde tot Kamervragen en daarmee tot een Kamerbrief waarin de minister van OCW het volgende toezegt:

- Ik zal de Algemene rijksarchivaris vragen een handreiking op te stellen en in beheer te nemen gericht op de archivering van websites en uitingen op sociale media.
- In samenwerking met CIO-Rijk zal hij tevens een voorstel doen ten behoeve van implementatie van een archiveringspraktijk voor deze media bij organisaties binnen de centrale overheid.

8.2 Uitwerking

In deze norm werkt het Nationaal Archief de eerste toezegging van de minister uit voor de hele overheid. De implementatie van de norm voor de centrale overheid is een afzonderlijk project. Dat wordt uitgevoerd in opdracht van de CIO-Rijk in samenwerking met het Nationaal Archief en DPC van het ministerie van Algemene Zaken. Deze norm gaat daar verder niet op in.

Het Nationaal Archief heeft ervoor gekozen om de handreiking vorm te geven als een norm met concrete eisen aan websitearchivering, met als bijlagen een stappenplan, een toelichting op de gebruikte begrippen en motivatie van de gemaakte keuzes. Op deze manier komt de norm tegemoet aan de behoefte aan zowel eenduidigheid (de eisen), praktische aanwijzingen (het stappenplan) als uitleg (de begrippen en de motivatie).

De definitie 'websitearchivering is het archiveren van websites' in deze norm wijkt af van de definitie in het EGI-rapport ('webarchivering is een containerbegrip voor het binnenhalen en bewaren van websites en van uitingen op sociale media.'). Het 'binnenhalen en bewaren van websites' is slecht één techniek van archiveren, meestal aangeduid als harvesting. Het omvat bovendien niet alle aspecten van archiveren, zoals de vindbaarheid en beschikbaarheid. Harvesting wordt in deze norm wel gehanteerd als de enige techniek voor websitearchivering. Daarmee heeft het verschil in definitie hier geen praktische gevolgen. De archivering van sociale media valt buiten scope van deze norm (zie paragraaf [7.1.2](#)).

8.3 Opdracht

Het Nationaal Archief ontwikkelt en beheert de norm in opdracht van de minister van OCW. De norm is als kennisproduct onderdeel van de zakelijke dienstverlening van het Nationaal Archief.

8.4 Standaardisatieraad

Deze norm wordt ter vaststelling voorgedragen aan de Standaardisatieraad van het Nationaal Archief. Dit overlegorgaan bestaat uit vertegenwoordigers uit alle overheidslagen en de archiefinstellingen en fungeert als inhoudelijke stuurgroep voor de kennisproducten van het Nationaal Archief. De Standaardisatieraad ziet erop toe dat de kennisproducten bruikbaar zijn voor overheidsorganisaties.

8.5 Verdere ontwikkeling

Het Nationaal Archief ontwikkelt de norm in stappen. Het uitgangspunt is dat bestaande kennis zo snel mogelijk beschikbaar komt. Daarna kunnen verschillende archiefvormers die direct toetsen in de praktijk. Het Nationaal Archief verwerkt deze praktijkervaringen in de norm.



Een uitgave van het Nationaal Archief
Postbus 90520
2509 LM Den Haag

Contact

Telefoon: +31-70-331 5460

E-mail: contact@nationaalarchief.nl

Website: <https://www.nationaalarchief.nl/archiveren>