

Stappenplan voor de archivering van overheidswebsites

Versie 2

Nationaal Archief (contact@nationaalarchief.nl)

Versiegeschiedenis			
Versie	Datum	Auteur	Beschrijving
1	20-07-2017	Erik Saaman	Eerste concept
2	02-08-2017	Erik Saaman	Openbaar concept voor discussie op het kennisplatform webarchivering

1 Inleiding

Dit stappenplan geeft aan hoe overheidsorganisaties hun informatie die is opgenomen in websites¹ kunnen archiveren, conform de Archiefwet. Het stappenplan zal onderdeel vormen van de handreiking webarchivering die door het Nationaal Archief ontwikkeld wordt. Naast het stappenplan zal de handreiking bestaan uit een inleiding tot de achtergronden en concepten van (web)archivering. Zie voor verdere toelichting op de ontwikkeling van de handreiking de projectbeschrijving op het kennisplatform webarchivering.

Archivering van websites conform de Archiefwet is nog geen standaardpraktijk. Daardoor ontbreekt het aan de benodigde kennis en ervaring om op dit moment een volledig eenduidig stappenplan te maken. Het stappenplan moet daarom gezien worden als 'work in progress' waarin alle nu bekende best practices zijn verzameld. Het stappenplan zal op basis van toepassingen in de praktijk verbeterd worden. Om die praktijkervaringen op te doen zal het Nationaal Archief meewerken aan een aantal praktijkcasussen van webarchivering. Voor die praktijkcasussen stelt het Nationaal Archief een adviseur beschikbaar die de opgedane ervaringen zal documenteren en verwerken in volgende versies van de handreiking. Organisaties die een praktijkcasus willen inbrengen kunnen daarvoor contact opnemen met contact@nationaalarchief.nl.

2 Toepassingsgebied

Het stappenplan is bedoeld voor overheidsorganisaties. Deze zijn op grond van de Archiefwet verplicht hun websites te archiveren. De handreiking is *niet* bedoeld voor organisaties die op eigen initiatief websites van anderen archiveren (zoals de Koninklijke Bibliotheek en andere wetenschappelijke en culturele instellingen). Deze zogenoemde collectievormers archiveren ook een deel van de overheidswebsites. Deze archieven zijn echter niet gemaakt om te voldoen aan de Archiefwet en vormen daarom geen alternatief voor het stappenplan (maar kunnen daar mogelijk wel een rol bij spelen).

Het stappenplan bevat geen afspraken die gelden voor specifieke overheidsorganisaties. In aanvulling op de handreiking wordt, samen met de CIO-Rijk, een implementatieplan voor de rijksoverheid opgesteld. Hierin worden afspraken opgenomen die specifiek zijn voor de rijksoverheid.

Deze eerste versie van het stappenplan beperkt zich tot webarchivering op basis van harvesting². Omdat dit de meest gangbare vorm van webarchivering is, waarmee al een groot deel van alle overheidswebsites te archiveren is. Harvesting is echter niet geschikt voor de archivering van alle informatie die via het web gedeeld wordt. Daarom beperkt dit stappenplan zich tot:

¹ Vaak zal het gaan om websites die door of voor een overheidsorganisatie zijn gemaakt. Maar dat hoeft niet. Onder de Archiefwet valt alle informatie die gemaakt of ontvangen is door een overheidsorganisatie. Dat kan dus ook op een website zijn die niet van de overheid is. Zo bevat twitter.com/Rijksoverheid overheidsinformatie omdat de Tweets door de overheid zijn geplaatst. Daarmee valt (de inhoud van) twitter.com/Rijksoverheid onder de Archiefwet.

² Zie en.wikipedia.org/wiki/Web_archiving onder 'Remote harvesting'.

Het binnenhalen (harvesten), bewaren en beschikbaar stellen van websites die gebruikt worden voor de publicatie van openbare overheidsinformatie, en die voldoen aan de gangbare kwaliteitseisen voor websites³.

Buiten de scope van de eerste versie van de handreiking vallen (zie de projectbeschrijving voor nadere toelichting):

- Interactie tussen de overheid en de gebruikers van de overheidswebsites.
- Niet openbare delen van websites
- Dynamische content waarvoor interactie met de webserver nodig is (server site scripting).
- Social media die geen website zijn maar een eigen applicatie gebruiken (zoals WhatsApp).
- Verbeteringen aan de website zelf die nodig zijn om te voldoen aan de Archiefwet.
- Het bewaren van gerelateerde informatie die niet in de website zelf is opgenomen.

3 Stappen

Om haar websites (binnen het werkingsgebied van het stappenplan) te archiveren conform de Archiefwet wordt een overheidsorganisatie aangeraden⁴ de volgende stappen uit te voeren.

3.1 Voorbereiding

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

[Nog toe te voegen: Een lijst met meest voorkomende rollen van betrokkenen.]

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

3.2 Inventarisatie

Maak een inventarisatie van alle websites waarvan de organisatie verantwoordelijk is voor de inhoud (content), en werk de inventarisatie periodiek bij. Neem in de inventarisatie per website op:

- Naam van de website.
- URL van de startpagina.
- Korte beschrijving van de inhoud.
- Organisatieonderdeel dat verantwoordelijk is voor het beheer van de inhoud, met contactpersoon.
- Type platform dat gebruikt wordt.
- Organisatieonderdeel dat verantwoordelijk is voor het beheer van het platform, met contactpersoon.
- Valt de website binnen het werkingsgebied van het stappenplan.
- ... nog verder aan te vullen

Doel: Met de inventarisatie wordt de nodige informatie verzameld om te bepalen welke websites gearchiveerd moeten worden.

3.3 Bewaartermijnen

Bepaal op basis van de geldende selectielijsten van de organisatie per website de bewaartermijn. Neem daarvoor contact op met de beheerder van de selectielijsten. Deze kan helpen bij de interpretatie van een selectielijst.

³ Zoals het gebruik van open standaarden, eisen aan informatiebeveiliging, toegankelijkheidseisen, Wet Bescherming Persoonsgegevens/Algemene Verordening Gegevensbescherming, Wet Hergebruik Overheidsinformatie en de richtlijn bouwkwiteit.

⁴ Het stappenplan is niet verplicht. Een overheidsorganisatie mag een alternatieve aanpak gebruiken, mits deze voldoet aan de Archiefwet.

De bewaartermijn is de termijn dat een webpagina (of een bijlage waarna verwezen wordt) nog bewaard moet worden *nadat* hij in de website gewijzigd of verwijderd is. Een bewaartermijn van nul betekend dat een webpagina niet bewaard hoeft te worden na wijziging of verwijdering.

Elke website krijgt als geheel één bewaartermijn⁵. Als er op basis van de selectielijsten meerdere termijnen van toepassing zijn, dan geldt de langste termijn. Als de selectielijst geen duidelijke uitspraak doet over de bewaartermijn van (een deel van) een website, dan is deze oneindig.

Websites veranderen voortdurend. Dagelijks kunnen webpagina's (en bijlages) gemaakt, gewijzigd of verwijderd worden. Waarbij archiefwettelijk gezien verschillende versies van dezelfde webpagina verschillende informatieobjecten zijn. Alle versies van een websites moeten gearchiveerd worden, conform de selectielijst. In principe is het mogelijk om verschillende bewaartermijnen te hanteren voor verschillende versies van een website. Bijvoorbeeld, om de omvang van het archief te beperken, één versie per jaar blijvend te bewaren en alle tussenliggende versies slechts twee jaar (of zelfs nul). Dit kan alleen als de selectielijst dit expliciet aangeeft. Voor zover bekend zijn er geen selectielijsten waarbij dit nu het geval is. Gedifferentieerde bewaartermijnen worden afgeraden omdat hierdoor het maken en beheren van archieven ingewikkelder wordt.

Als de op basis van de selectielijst bepaalde bewaartermijnen geen recht doen aan bepaalde belangen (bijvoorbeeld te korte termijn of te hoge kosten), vraag dan de beheerder van de selectielijst om aanpassing van de selectielijst. Hierop is een wettelijke procedure van toepassing, waarvan de beheerder op de hoogte is.

Doel: Bewaartermijnen worden bepaald om vast te kunnen stellen hoe vaak een website geharvest wordt en hoe lang de geharveste site daarna bewaard wordt. Dit gebeurt op basis van officiële selectielijsten omdat bij het vaststellen van de selectielijsten de verschillende belangen op een zorgvuldige wijze zijn gewogen.

3.4 Harvestingdienst

Sluit een contract af met een leverancier van een harvestingdienst (of richt er zelf een in). De harvestingdienst moet voorzien in:

1. Het periodiek harvesten van alle door de opdrachtgever aangemelde websites:
 - a. Minimaal één keer per jaar volledig.
 - b. Daarna dagelijks incrementeel⁶ of real time⁷ bij wijzigingen.
2. Het aan de opdrachtgever beschikbaar stellen van de geharveste sites in de vorm van files die voldoen aan:
 - a. De WARC-specificatie (ISO 28500:2009).
 - b. Voorzien van metagegevens die voldoen aan het Toepassingsprofiel Metagegevens Rijksoverheid of het Toepassingsprofiel Metagegevens Lokale Overheden.
 - c. Eventuele aanvullende eisen die de archiefbewaarpplaats⁸ stelt aan over te brengen webarchief. Neem hiervoor contact op met de betreffende archiefbewaarpplaats.
3. Het toegankelijk maken van de geharveste websites. Waarbij in ieder geval de volgende functies geboden worden voor alle gangbare webbrowsers:
 - a. Full tekst-zoeken door alle geharveste webpagina's.
 - b. Van elke webpagina kan de versie van een specifieke datum worden opgevraagd.
 - c. Van zowel de website als elke webpagina kan metadata worden opgevraagd.

⁵ Hiervoor is gekozen omdat geharveste websites als één geheel worden opgeslagen. Het is lastig om daarbinnen verschillende bewaartermijnen te hanteren.

⁶ Incrementele harvesting wil zeggen dat alleen nieuwe en gewijzigde webpagina's (en bijlagen) worden opgehaald en opgeslagen.

⁷ Real time harvesting is mogelijk als er bij wijzigingen aan de website automatisch een signaal wordt verstuurd naar de harvestingdienst. Zodat deze weet welke pagina's te harvesten.

⁸ Voor elke overheidsorganisatie is er zogenaamde archiefbewaarpplaats aangewezen waaraan blijvend te bewaren archieven na maximaal twintig jaar worden overgebracht (zie Archiefwet).

4. Het toegankelijk maken van eerder door een andere harvestingsdienst geharveste sites (indien die er zijn). Op basis van aangeleverde WARC-files.
5. De opdrachtgever is en blijft volledig eigenaar van de geharveste websites. De leverancier kan hier geen rechten op uitoefenen.
6. Harvesting mag niet ten koste gaan van het functioneren van de websites. De leverancier moet hiervoor vooraf kenbaar maken aan welke eisen de te harvesten websites moeten voldoen.

[Nog te onderzoeken: aanvullende voorwaarden voor archiveren van mobiele versie van websites.]

Binnen de praktijkcasussen worden de eisen aan de harvestingsdienst verder uitgewerkt. De resultaten daarvan worden als standaard inkoopbestek opgenomen in een volgende versie van de handreiking.

Een lagere frequentie voor de incrementele harvesting is mogelijk als webpagina's heel weinig wijzigen. Zorg er daarbij voor dat de kans dat een pagina (of bijlage) tussen twee opeenvolgende harvests wijzigt te verwaarlozen is (kleiner is dan 1:10.000).

Een contract voor een harvestingsdienst kan afgesloten worden voor één of meerdere websites. Hierbij kunnen overheidsorganisaties besluiten om samen in te kopen. Bijvoorbeeld doormiddel van een raamovereenkomst (voor de rijksoverheid wordt deze mogelijkheid onderzocht). In principe kunnen collectievormers die al websites archiveren de harvestingsdienst leveren, mits ze daarbij voldoen aan de bovenstaande voorwaarden. Een overheidsorganisatie kan ook zelf een harvestingsdienst inrichten, als het beschikt over de daarvoor benodigde mensen en middelen. De handreiking webarchivering geeft geen verdere aanwijzingen voor het inrichten van een harvestingsdienst.

Het contract voor de harvestingsdienst kan eventueel gecombineerd worden met de hostingdienst voor de website. Ook is het mogelijk beschikbaar stellen van de geharveste websites in een apart contract onder te brengen.

Doel: De harvestingsdienst voert het daadwerkelijk binnenhalen en beschikbaar stellen van de websites uit. Een goed contract zorgt ervoor dat duidelijk is wat de rolverdeling is en aan welke eisen de dienst moet voldoen.

3.5 Start harvesting

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

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

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

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

3.6 Bewaar WARC-files

Direct na een harvest worden de bijbehorende WARC-files beschikbaar gesteld aan de eigenaar van de inhoud van de website (de zogenaamde archiefvormer). De archiefvormer bewaart de WARC-files in één van zijn bestaande archiveringssystemen (bijvoorbeeld een DMS of e-depot). Waarbij de files gemetadateerd worden zoals daar gebruikelijk. De WARC-files worden door de archiefvormer op dezelfde manier als andere digitale archieven beheerd.

Maak vooraf goede afspraken tussen de leverancier van de harvestingsdienst en de beheerder van het archiveringstelsel. Zodat opname van de WARC-files in het archiveringstelsel soepel verloopt.

Merk op dat de WARC-files bewaard en beheerd worden door de archiefvormer. Dat hoeft niet de beheerder van de website te zijn. Bijvoorbeeld de Dienst Publiek en Communicatie (DPC) beheert websites voor meerdere

rijksoverheidsorganisaties. Die organisaties zijn zelf verantwoordelijk voor het beheer van de WARC-files, niet de DPC.

Het is in principe mogelijk het bewaren en beheren van de WARC-files uit te besteden aan de leverancier van de harvestingdienst (want die beschikt immers toch al over de files) of de beheerder van de website. In dat geval worden aanvullende eisen aan de leverancier gesteld. Zodat het beheer van de files plaatsvindt volgens dezelfde normen als in het archiveringssysteem van de archiefvormer. Neem daarvoor contact op met de beheerder van het archiveringssysteem. Aangeraden wordt om in dat geval in het archiveringssysteem wel de metadata van de WARC-files op te nemen. Met daarin een verwijzing naar de bewaarplaats van de WARC-file. Zodat het bestaan van de WARC-files bekend is in het archiveringssysteem.

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

3.7 Kwaliteitscontrole

De archiefvormer laat de WARC-files periodiek op kwaliteit inspecteren door een kundige en onafhankelijke partij (op basis van de afspraken in het contract). De leverancier van de harvestingdienst moet, waar nodig, onjuistheden meteen verhelpen en veranderingen doorvoeren om herhaling te voorkomen. Als de harvestingdienst blijvend niet aan de kwaliteitseisen voldoet, dan wordt het contract beëindigd (en een nieuwe dienst gecontracteerd).

Binnen de praktijkcasussen wordt het inspectieproces verder uitgewerkt. De resultaten daarvan worden opgenomen in een volgende versie van de handreiking.

Doel: Zeker stellen dat de WARC-files voldoen aan de eisen in het contract. Zonder dat daar achteraf kostbare herstelacties voor nodig zijn.

3.8 Vernietigen en overbrengen

Jaarlijks wordt door de beheerder van het archiveringssysteem waar de WARC-files worden bewaard (na ontvangst van een nieuwe volledige harvest):

- De WARC-files van het jaar waarvoor de bewaartermijn verstreken is vernietigd.
- De WARC-files van het jaar waarvoor de overbrengingstermijn van twintig jaar verstreken is overgebracht naar de betreffende archiefbewaarplaats. Volgens de geldende processen voor overbrenging van digitale bestanden vanuit het archiveringssysteem.

Voor de harvestingdienst geldt:

- Na vernietiging van WARC-files mogen de betreffende pagina's niet meer beschikbaar worden gesteld. De harvestingdienst moet alle eventuele kopieën van de pagina's (als WARC-file of in elke andere vorm) verwijderen.
- Na overbrenging van de WARC-files is de archiefbewaarplaats verantwoordelijk voor het beschikbaar stellen van de betreffende pagina's. De beheerder van de archiefbewaarplaats kan besluiten dit uit te besteden aan de leverancier van de harvestingdienst.

Merk op dat de Archiefwet overbrenging eerder dan twintig jaar toestaat. Daarover moet de archiefvormer afspraken maken met de betreffende archiefbewaarplaats. Deze afspraken vallen buiten het bestek van de handreiking. Vervroegde overbrenging voor het rijk kan onderdeel zijn van het nog op te stellen implementatieplan voor de rijksoverheid.

Doel: Vernietiging zorgt ervoor dat archieven niet langer bewaard worden dan noodzakelijk. Overbrenging zorgt ervoor dat het beheer van blijvend te bewaren archieven veilig wordt gesteld en daarmee niet afhankelijk is van de continuïteit van de archiefvormer.