

Webarchivering in Nederland (3)

Webarchivering in de praktijk

Ervaringen uit Middelburg en Nijmegen

De eerste ervaringen in Middelburg en Nijmegen laten zien dat webarchivering behalve een technisch ingewikkelde kwestie, ook vooral een organisatorische uitdaging is. Hoe komen alle betrokkenen op één lijn? Hoe worden de juiste keuzes gemaakt in selectie? Hoe is de samenhang met andere archieven gewaarborgd? En hoe kan ervoor gezorgd worden dat de juiste kennis en expertise aanwezig is?



In een reeks van vijf artikelen wordt een beeld geschetst van webarchivering in Nederland. Het derde artikel in de reeks gaat over de praktijk in Middelburg en Nijmegen, waar twee archiefinstellingen de webarchiveringshandschoenen hebben opgepakt.

Webarchivering bij de gemeente Middelburg

TEKST FIEKE KRIKHAAR

De website www.middelburg.nl fungeert in toenemende mate als digitaal stadskantoor waar burgers gemeentelijke informatie halen en zaken kunnen regelen. Aan de informatie op websites kunnen burgers rechten ontlenuen. Websites vallen onder de Archiefwet en dienen duurzaam toegankelijk te worden gehouden. Webarchivering in archiefwettelijke zin betekent dat een website met een zorgvuldig vastgestelde frequentie wordt gearchiveerd en dat zowel de presentatie als de inhoud van de website in samenhang worden vastgelegd. Sinds 2011 archiveert de gemeente Middelburg de website één keer per week door het maken van *snapshots* (momentopnames) met het pakket Presurf van marktpartij Capsis. Daarnaast wordt één keer per dag een snapshot gemaakt van de wijzigingen. Het mag duidelijk zijn dat dit niet voldoet aan de archiefwettelijke eis van duurzame toegankelijkheid.



Overhandiging van het eindrapport van de pilot webarchivering. Foto: Zeeuws Archief/LH.

Aanpak en resultaten

De pilot* bestond uit drie onderdelen. In de eerste plaats zijn er snapshots van Capsis uit 2012 opgenomen in het e-depot, zowel volledige opnames als een aantal wijzigingen. Capsis maakt gebruik van WARC-bestanden, een van de voorkeursbestanden van het Zeeuws e-depot, dus opname was geen probleem. Raadpleging lukte echter niet, niet met de viewer in het e-depot zelf en niet met een generieke Web Archive Player. De oorzaak lag in de wijze waarop de WARC-bestanden zijn opgebouwd. Capsis gaat dat aanpassen waarna de snapshots raadpleegbaar zullen zijn, onafhankelijk van de applicatie van Capsis.

De website www.middelburg.nl fungeert in toenemende mate als digitaal stadskantoor

Het tweede onderdeel was de ingest (oftewel de 'opname' [red]) van de huidige 'dynamische' website van de gemeente in het Zeeuws e-depot. Zowel opname als raadpleging met de Web Archive Player verliepen voorspoedig. Om de live website in het e-depot ook zonder een extern hulpmiddel te kunnen bekijken, moet de reeds bestaande functie daarvoor in het e-depot nog worden doorontwikkeld.

En verder is ook gekeken naar de organisatie rond de website en archivering daarvan. Zijn de verantwoordelijkheden juist belegd en wordt er rekening gehouden met geldende wet- en regelgeving? Het antwoord op die vragen kon niet altijd worden gegeven. De gemeente overweegt om een handboek webarchivering op te stellen om het proces van webarchivering vanaf creatie tot overbrenging te kunnen borgen. Dat zal mede gebaseerd zijn op de handreiking die het Nationaal Archief nu opstelt.

Lessons learned en goede voornemens

Behalve de gemeente en het Zeeuws Archief waren ook het Nationaal Archief en Capsis bij de pilot betrokken. Vanzelfsprekend doen alle betrokkenen tijdens een dergelijke pilot veel ervaring en kennis op. Daarnaast is het ook zo dat mensen van verschillende organisaties elkaar beter leren kennen, en elkaars taal en insteek beter gaan begrijpen. Dat betaalt zich terug in het vervolg van de samenwerking bij aansluiting van deze en andere applicaties van de gemeente Middelburg op het Zeeuws e-depot.

Naast goede afspraken over de verantwoordelijkheid ontbrak ook een architectuurplaat van de website. Voor een website worden allerlei bronnen gebruikt en ook bevat deze vele links die meestal niet als onderdeel van de website worden gearcheveerd. Nadat de website is gearcheveerd, bestaat de interactieve functie van de website niet meer, maar de mogelijkheden die er waren moeten nog wel te zien zijn.

En als laatste bleek tijdens de metadatering dat TMLO minder geschikt is voor de archivering van websites. In principe dienen alle zoekmogelijkheden van een live website ook aanwezig te zijn in een gearcheveerde website. Als input voor het toekomstige handboek webarchivering zijn er tijdens de pilot een paar uitgangspunten voor toegankelijkheidseisen geformuleerd:

- goede beschrijving op hoofdlijnen met vermelding van de data van de gearcheveerde website in een boomstructuur;
- presentatie en inhoud in samenhang vastgelegd en bewaard waarbij het archivalistische principe van het handhaven van de oude orde in tact blijft.



Fieke Krikhaar

f.krikhaar@zeeuwsarchief.nl

Fieke Krikhaar is programma-coördinator van het e-depot Zeeuws Archief.

Aan het einde van de pilot heeft een presentatie en bespreking plaatsgevonden van deze pilot in relatie tot andere ervaringen en de landelijke ontwikkelingen. Bij deze bijeenkomst waren ook het Stadsarchief Rotterdam en het Regionaal Archief Dordrecht aanwezig die met dezelfde vraagstukken bezig zijn. Tot slot zijn de resultaten gepresenteerd tijdens een bijeenkomst van het Zeeuwse informatienetwerk op 27 juni 2017. ●

Noot:

* Het rapport van deze pilot is te vinden via: http://www.zeeuwsarchief.nl/media/uploads/e-depot/20170627_eindrapport_pilot_webarchivering_gemeentemiddelburg_zeeuwsarchief_versie_1.0.pdf

Webarchivering bij het Regionaal archief Nijmegen

TEKST HENK TRAPMAN

In 2016 heeft het Regionaal Archief Nijmegen (RAN) een begin gemaakt met het vanuit cultuurhistorisch belang archiveren van particuliere websites. Daarbij is het de bedoeling om uiteindelijk te komen tot een representatieve verzameling websites van Nijmeegse personen, instellingen en onderwerpen. De uitgekozen websites worden 'geharvest' met behulp van het zogeheten Web Archieven Dashboard (WAD) van webarchiveringsbedrijf GW Crossmedia, dat ook zorgt voor *hosting* van de gearcheveerde sites. Het WAD is bedoeld voor periodieke snapshots om de informatie en vormgeving van de website van dat moment te bewaren. De websites worden zo compleet mogelijk gearcheveerd, inclusief vormgeving, beeld- en geluidsmateriaal en links. Daardoor blijven de websites functioneel, en ogen en werken ze als echte websites. Door het archiveren periodiek te herhalen, bijvoorbeeld maandelijks of jaarlijks, ontstaan meerdere momentopnamen van dezelfde website.

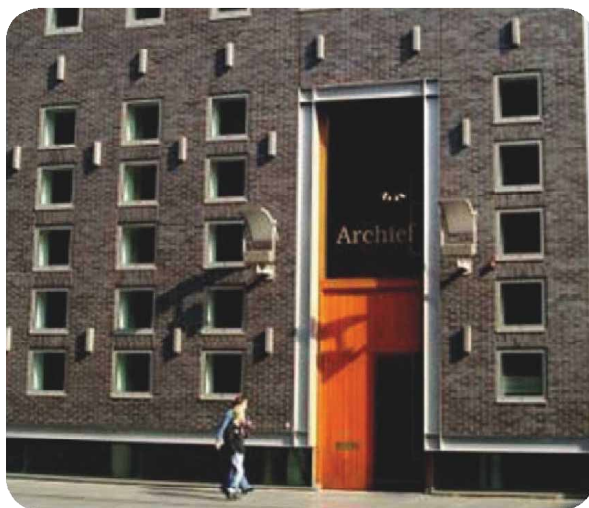
Aanpak

Het RAN wil de gearcheveerde websites in het collectiebeheersysteem Atlantis beschrijven, in de eigen studiezaal beschikbaar stellen en zo mogelijk ook online publiceren. Om hiermee een goed begin te maken is het RAN te werk gegaan volgens de *scrum*methode. Daarbij houdt een klein team van medewerkers gedurende een kort tijdsbestek (enkele weken) de agenda's vrij om zich vanuit de eigen disciplines (inventarisatie, dienstverlening) onder leiding van een scrummaster te buigen over een overkoepelend vraagstuk (beschikbaarstelling van gearcheveerde particuliere websites). De medewerkers rapporteren dagelijks aan de scrummaster en aan elkaar over de voortgang en de problemen die zij tegenkomen. Zo wordt toegewerkt naar het beoogde resultaat: de publicatie van de eerste gearcheveerde websites binnen de omgeving waar deze mogen worden bekeken.

Het scrumproject is begin 2017 succesvol afgerond en heeft geleid tot de online publicatie van twee gearcheveerde websites via de digitale studiezaal van het RAN. De rechthebbers gaven toestemming voor publicatie, waardoor het online beschikbaar stellen van de websites mogelijk werd.

Vraagstukken

De vraagstukken die tijdens het project naar voren kwamen waren veelal archivalistisch van aard. Om te beginnen kwam het verschil tussen de website als archiefdocument en de website als onderdeel van een collectie aan de orde. De beide online gepubliceerde websites maken deel uit van een archief. Websitearchivering kan als een mogelijk alternatief worden gezien voor de verwerving van met name kleinere, vaak bewerkelijke particuliere archieven. Zo is het denkbaar dat bij een amateurvoetbalclub kan worden volstaan met de archivering van de website waarop deze vereniging al zijn informatie (uitslagen, wedstrijdverslagen,



Het Regionaal Archief Nijmegen. Bron: CollectieGelderland

foto's etc.) heeft gepubliceerd. Gearchiveerde particuliere websites, waarbij niet wordt verwacht dat van de betreffende persoon of instelling nog meer archiefmateriaal zal worden geacquireerd, kunnen worden samengevoegd tot een collectie websites.

Het Regionaal Archief Nijmegen wil gearchiveerde websites zo mogelijk ook online publiceren

Een tweede vraagstuk betrof de beschikbaarstelling van de gearchiveerde websites binnen de eigen muren in de studiezaal of online via de website van het RAN. De gebruiksmogelijkheden van een gearchiveerde website zijn doorgaans beperkt vanwege de auteursrechten. Bij de twee gepubliceerde websites was er één contactpersoon die uit-

sluitsel kon geven over de rechten, waardoor het eenvoudig was om toestemming te krijgen voor online publicatie. Bij de meeste gearchiveerde sites blijkt dat bij de totstandkoming van deze websites niet of nauwelijks iets is vastgelegd met betrekking tot auteursrechten van de betrokken partijen. In die gevallen zullen de websites vooralsnog alleen binnen de muren van het RAN beschikbaar komen.

Er waren ook wensen ten aanzien van uitbreiding van de zoekmogelijkheden naar en in gearchiveerde websites op de publiekspagina van de website van het RAN. De bezoeker zou niet alleen moeten kunnen zoeken naar gearchiveerde websites, maar ook door de inhoud (*full-text*) van meerdere websites. Bij keuze voor de laatste optie zou een geavanceerd zoekscherm moeten verschijnen. Met de huidige technische middelen bleek een en ander echter nog niet realiseerbaar.

Het resultaat van het project is een opmaat voor het formuleren van verder beleid met betrekking tot websitearchivering door het RAN, waarbij er definitieve acquisitiecriteria moeten worden vastgesteld en de frequentie van archivering per gearchiveerde site moet worden bepaald. De focus van het project lag op de beschikbaarstelling van de gearchiveerde sites. Vanuit de website van het RAN wordt nu gelinkt naar de extern gehoste sites. Het is de wens van het RAN om de WARC-bestanden van de gearchiveerde websites op termijn zelf te beheren, duurzaam op te slaan in een e-depot en ze van daaruit beschikbaar te stellen. ●



Henk Trapman

h.trapman@nijmegen.nl

Henk Trapman is werkzaam als coördinator audiovisuele collecties bij het Regionaal Archief Nijmegen.