

# **Webarchivering, naar het volgende niveau.**

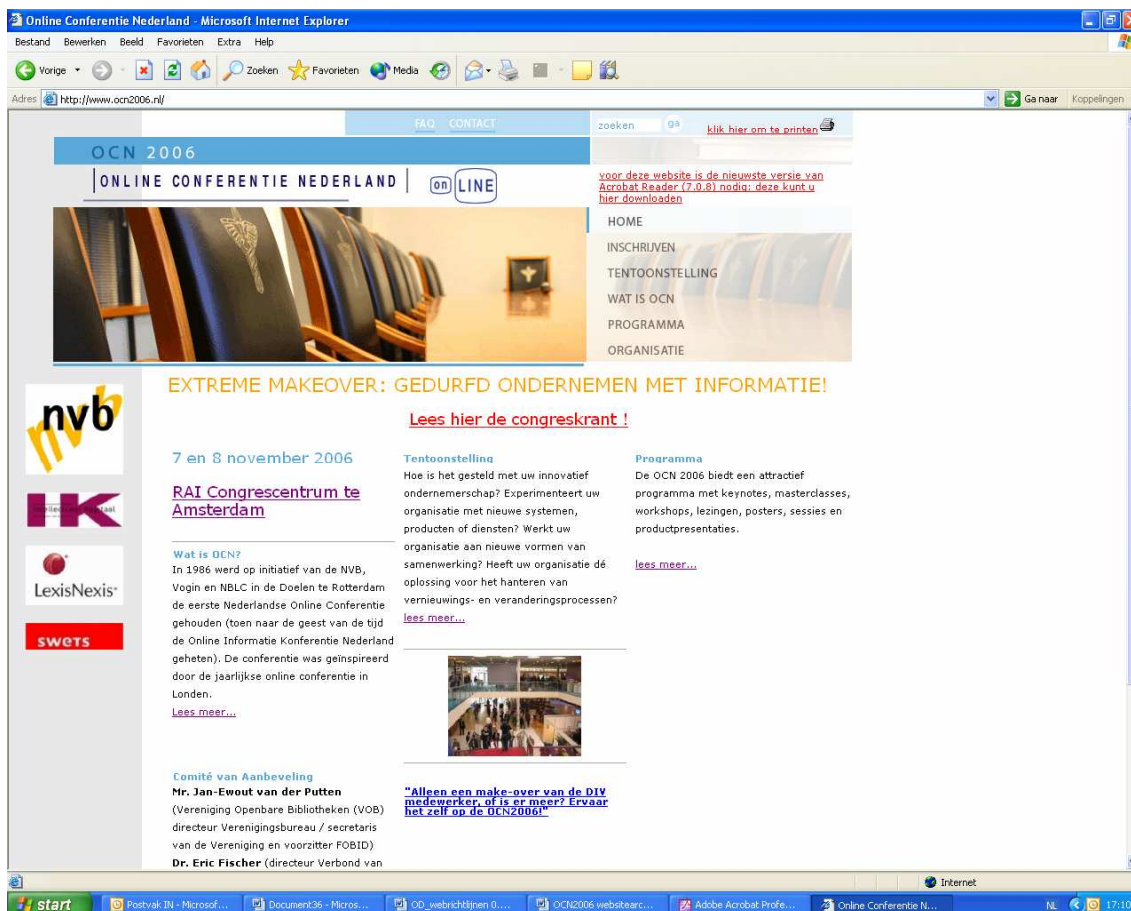
## **Theorie, beleid en praktijk**

Erika Hokke, Archiefschool,  
Hans Goutier, Ministerie van Verkeer en Waterstaat,  
Frans-Willem Duijnhouwer, Capsis B.V.

Versie 6 november 2006 (definitief).

## Inleiding

Op de OCN2006 website valt te lezen dat de eerste Nederlandse Online Conferentie gehouden werd in 1986. Ik was daar niet bij, maar ik vermoed dat er een paar dingen hetzelfde zijn gebleven: innovatieve onderwerpen, inspirerende sessies, een goede gelegenheid om collega's te ontmoeten. Maar ik kan me ook voorstellen dat er veel is veranderd, bijvoorbeeld de manier waarop u zich aangemeld of georiënteerd heeft op dit congres. Ik ga er vanuit dat u, net als ik, gekeken hebt op de website. U vond daar het programmaoverzicht, een routebeschrijving en het aanmeldformulier.



Figuur 1: homepage website OCN2006

Het is inmiddels de gebruikelijke manier om snel relevante en actuele informatie te vinden, om contact te leggen met organisaties en om transacties te doen, zoals de aanmelding voor dit congres. Het voorbeeld illustreert de snelle opmars van websites. Dit jaar zijn er al meer dan 101 miljoen websites, meldde het Amerikaanse bedrijf Netcraft deze maand.<sup>i</sup> Cijfers van Eurostat, het Europese bureau voor de statistiek, geven aan dat in Nederland 78% van de huishoudens een internetaansluiting heeft en dat 74% van de Nederlandse bevolking minstens één keer per week gebruik maakt van internet.<sup>ii</sup>

Actualiteit en verandersonnelheid hebben ook een schaduwzijde. Het maakt websites niet alleen tot relevante, maar ook tot kwetsbare informatiebronnen. Een al wat ouder onderzoek schatte in 2003 de levensduur van websites op ca. 75-100 dagen.<sup>iii</sup> Dagelijks worden nieuwe berichten, pagina's of hele sites gepubliceerd, maar ze verdwijnen ook dagelijks. Het ligt dan ook voor de hand dat dit type kwetsbare informatiebron een vorm van archivering vraagt. Inmiddels zijn er verschillende organisaties die zich hiermee bezig houden. Grote spelers zijn natuurlijk The Internet Archive en het Europese

internetarchief. Verder archiveren veel Nationale Bibliotheken en een aantal Nationale Archieven de websites in hun landelijk domein, zoals de bibliotheken van Zweden, Australië en The National Archives in Londen. Ook zijn er veel organisaties die heel gericht websites over een bepaald thema of onderwerp archiveren. Een bekend project in Nederland is Archipol, het archief met websites van politieke partijen.<sup>iv</sup>

Deze paper bespreekt vanuit theoretisch, beleidsmatig en praktijkperspectief i de volgende stap in webarchivering. Als eerste komt de vraag aan de orde waarom organisaties zouden willen archiveren. Vervolgens gaat de paper in op de organisatorische en praktische vraagstukken waar een organisatie die zijn eigen websites wil archiveren mee wordt geconfronteerd. De organisatie die daarbij centraal staat is het ministerie van Verkeer en Waterstaat.

## De theorie: de waarde van websites

### Begripsbepaling

Het theoretisch perspectief begint met een begripsbepaling. Wat is een website en wat is archivering? Het W3C consortium beschrijft<sup>v</sup> in een concept definitie uit 1999 een website als “a collection of interlinked Web pages, including a host page, residing at the same network location.” Een webpagina beschrijft het W3C vervolgens als “a collection of information, consisting of one or more web resources, intended to be rendered simultaneously and identified by a single URI”.

Volgend hierop heeft de Archiefschool in 2003 de volgende definitie van een website gegeven:

*Een website is het medium, waarmee via het Internet informatie gepresenteerd wordt op een statische of dynamische manier, vaak in combinatie met technologieën voor tweezijdige communicatie en transactie, zoals e-mail, nieuwsgroepen of discussielijsten.*<sup>vi</sup>

Deze definitie werd opgesteld om de vraag te kunnen beantwoorden of overheidswebsites archief zijn in de zin van de Archiefwet 1995. Het antwoord was bevestigend.

Archiveren is het zodanig beheren van documenten, dat wie ze nodig heeft ze snel kan vinden, zonder dat ze onnodig in de weg liggen. Onder archivering verstaan we een aantal activiteiten die een organisatie ontplooit om haar informatie toegankelijk te houden. In de moderne archivalische theorie<sup>vii</sup> wordt dit uitgewerkt als systeem. Onder het archiveringssysteem verstaan we daar:

*Het geheel van procedures, methoden, kennis, mensen, middelen en documenten, waarmee een persoon of samenwerkingsverband van personen (een organisatie) de archiveringsfunctie vorm geeft.*

Die archiveringsfunctie heeft als doel het beheer en beschikbaar stellen van de archieven die de betreffende persoon of organisatie nodig heeft, bijvoorbeeld voor de uitvoering van taken, als kennisbron of voor het afleggen van verantwoording.

Als we spreken over webarchivering, het archiveren van websites, bedoelen we dus alle activiteiten die een persoon of organisatie ontplooit met als doel om de eigen websites duurzaam toegankelijk te houden voor iedere mogelijke bezoeker. Deze invalshoek is dus breder dan de huidige webarchieven, die immers kopieën maken van websites van andere organisaties of personen. Het archiveren van websites heeft ook, of beter: juist, betrekking op de activiteiten die de ‘eigenaar’ van een website onderneemt om de gepubliceerde informatie langdurig toegankelijk te houden.

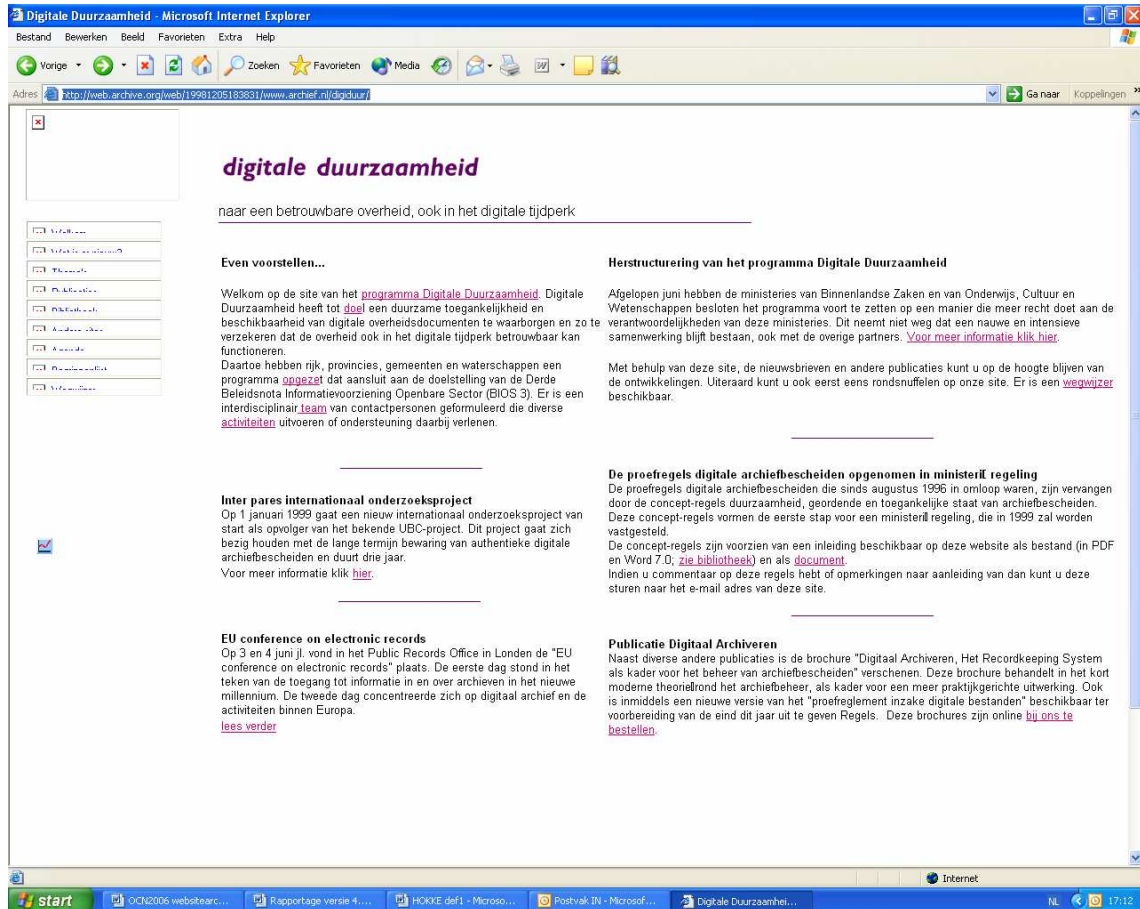
### Waarom archiveren?

De vraag waarom je informatie op websites duurzaam toegankelijk zou willen houden heeft te maken met de functie en de waarde van die informatie. Het gaat hier om de waarde voor de bezoeker van de website, maar ook om de waarde voor de persoon of instelling die de website publiceert. Er zijn verschillende waarden te benoemen:

#### 1. Emotionele waarde

Allereerst is er sprake van een emotionele waarde, vooral voor de ontwerper of eigenaar van de website. Deze waarde is natuurlijk sterk gekoppeld aan het ‘maken van iets moois’, vergelijkbaar met de publicatie van een boek of een kunstwerk, of iets anders waar tijd en creativiteit in is gaan zitten.

Als ik aan mensen vraag welke websites ze hebben bekeken, toen ze voor het eerst The Internet Archive bezochten, dan blijken dat bijna altijd websites te zijn waar men een band mee heeft. Bijvoorbeeld omdat het een eigen site is, of de oude site is van de organisatie waar men werkt. De eerste URL die ik invoerde in de Wayback-machine, was van de oude website van het Programmabureau Digitale Duurzaamheid, [www.archief.nl/digiduur](http://www.archief.nl/digiduur) mijn oude werkgever. Helaas is de oudste snapshot uit december 1998, toen ik al niet meer bij het programmabureau werkte.<sup>viii</sup>



**Figuur 2: snapshot [www.archief.nl/digiduur](http://www.archief.nl/digiduur) in The Internet Archive**

Websites kunnen ook om een andere reden een emotionele waarde hebben. Een voorbeeld hiervan is het archief met websites over de aanslagen op 11 september. Dit webarchief is direct na de aanslagen aangelegd door The Library of Congress, in samenwerking met onder andere The Internet Archive, WebArchivist.org en de Pew Internet & American Life Project. Het archief bevat meer dan 30.000 websites over de aanslagen. Het zijn websites van individuen, groepen, organisaties en de media, die gearcheveerd zijn tussen 11 september en 1 december 2001.<sup>ix</sup>



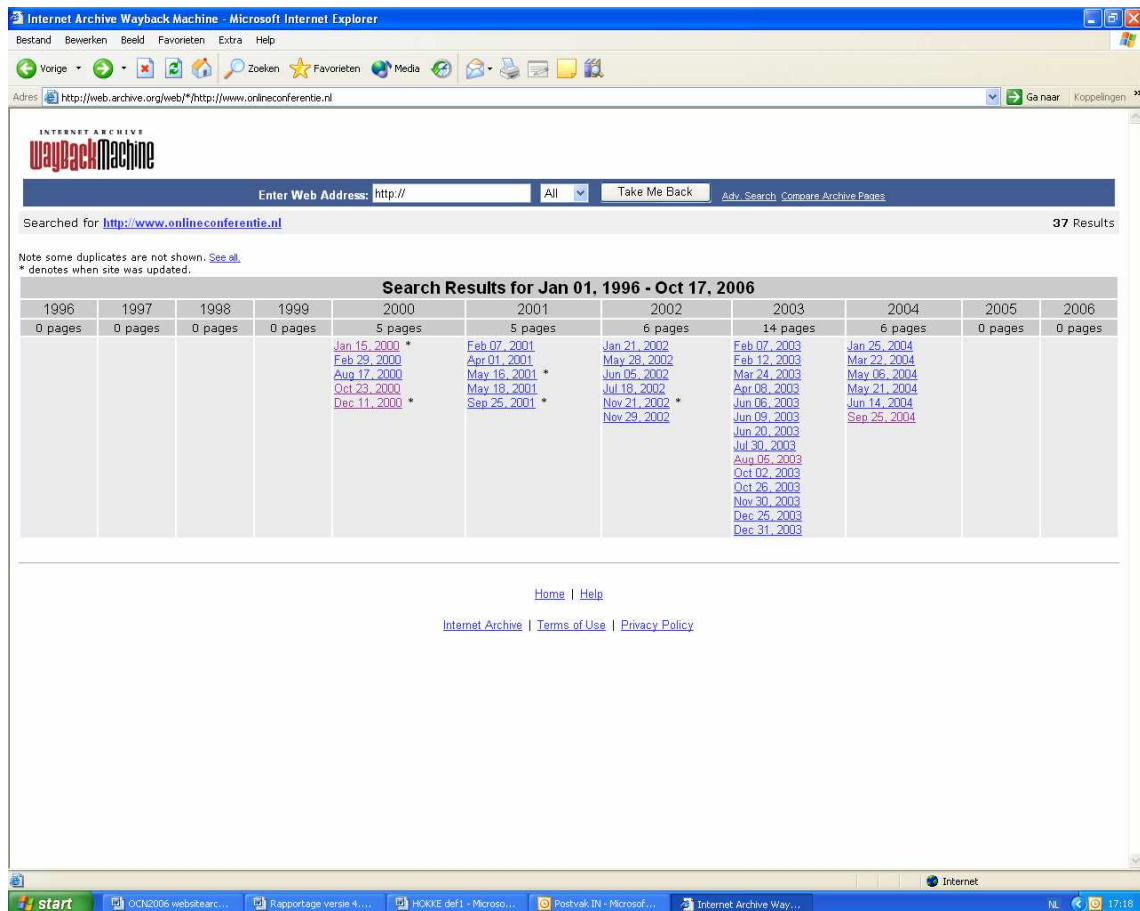
**Figuur 3: archief van websites over de aanslagen op het WTC 09-11-2001**

## 2. Informatie waarde

### 2.1. Actuele informatieve waarde

Websites hebben vooral informatieve waarde, of het nu om informatie over de Online Conferentie 2006 gaat of om sites over fietsen in China, websites zijn een belangrijke informatiebron. In de beginjaren van het World Wide Web was de website vaak nog een aanvulling op andere informatiebronnen en werd dezelfde informatie ook (of vooral) via traditionele papieren informatiebronnen verspreid. De huidige websites zijn vaak unieke informatiebronnen, waarvan de informatie niet in andere vorm beschikbaar is. Zo is de website van de vorige online conferentie, in 2004, met moeite terug te vinden.

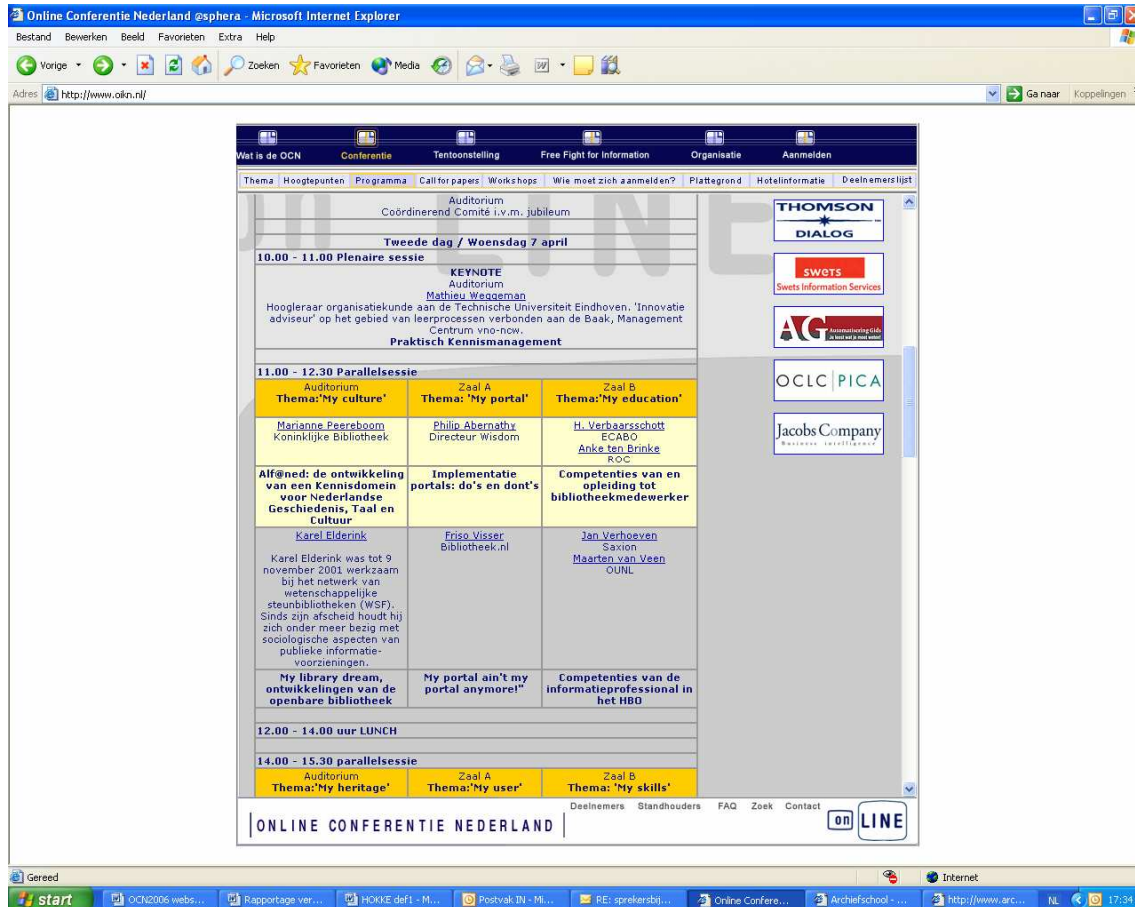
Mijn zoektocht in The Internet Archive leverde als resultaat 37 vermeldingen op van gearcheiverde versies van de website.<sup>x</sup>



**Figuur 4: The Internet Archive homepage, zoeken naar oude website van OCN**

Helaas kwam ik niet verder dan de vermeldingen, iedere poging om een gearcheiverde site te zien liep uit op een melding dat er geen connectie met de server gemaakt kon worden. Erg jammer, want ik was benieuwd wat het programma tijdens de vorige conferentie was.

Via een zoektocht in het world wide web vond ik wel een gearcheiverde versie, niet meer beschikbaar via de oorspronkelijke URL, maar via <http://www.oikn.nl>



**Figuur 5: website online conferentie 2004 via www.oikn.nl**

Deze site bevat wel het programma, maar op het moment dat ik de presentatie wil openen, word ik doorgelinkt naar de OCN2006 site. Helaas...

De verandingsnelheid van het World Wide Web bedreigt dus de beschikbaarheid van websites als bronnen, terwijl de informatieve waarde van die website blijft bestaan. Het probleem is bijvoorbeeld ook bekend uit de wereld van de academische publicaties, waar in publicaties verwezen wordt naar websites en andere digitale publicaties.

## 2.2. Onderzoekswaarde

Naast 'actuele' informatiebronnen met informatie over openingstijden, beleidsinformatie, vacatures, recente publicaties, vakantietips, etc. zijn websites relevante primaire bronnen geworden voor alle mogelijke onderzoeksdomeinen (historisch, sociologisch, cultureel-antropologisch, etc). Je ziet dan ook dat veel website-archiveringsinitiatieven<sup>xi</sup> gericht zijn op het behoud en documenteren van deze bronnen.

Een bekend Nederlands voorbeeld is natuurlijk Archipol, de verzameling van websites van de Nederlandse politieke partijen. Het Documentatiecentrum voor de Nederlandse Politieke Partijen is enige jaren geleden gestart met het verzamelen van kopieën van deze politieke sites, na de constatering dat de reguliere informatiebronnen (de schriftelijke publicaties) door politieke partijen werden vervangen door websites (die natuurlijk zo actueel mogelijk zijn) en digitale nieuwsbrieven. Nu kopieert het DNPP periodiek de verschillende websites en maakt het DNPP extra kopieën bij relevante ontwikkelingen in



de samenleving (zoals verkiezingen, maar ook direct na de moord op Pim Fortuyn). Deze websites zijn belangrijke bronnen voor politieke geschiedenis. Ook in Duitsland is ondertussen een dergelijk interessant archief met websites van politieke partijen ontstaan, het Politisches Internet-Archiv.<sup>xii</sup>

Een ander voorbeeld van het gebruik van gearchiveerde websites speelde de afgelopen jaren rondom een aantal websites in The Internet Archive. Gearchiveerde websites zijn gebruikt in een rechtszaak tussen twee Amerikaanse bedrijven. Het ene bedrijf, Healthcare Advocates, spande in 2003 een rechtszaak aan tegen Health Advocate Inc. De rechtszaak ging onder andere om merkenrecht (trademark). De advocaat van de gedaagde, kantoor Harding Earley Follmer & Frailey, heeft in de voorbereiding van de verdediging in The Internet Archive de websites van Healthcare gezocht en gedownload. Deze kopieën van de gearchiveerde sites hebben een zodanige rol gespeeld in het proces, dat de aanklager, Heathcare Advocate, de zaak heeft verloren.

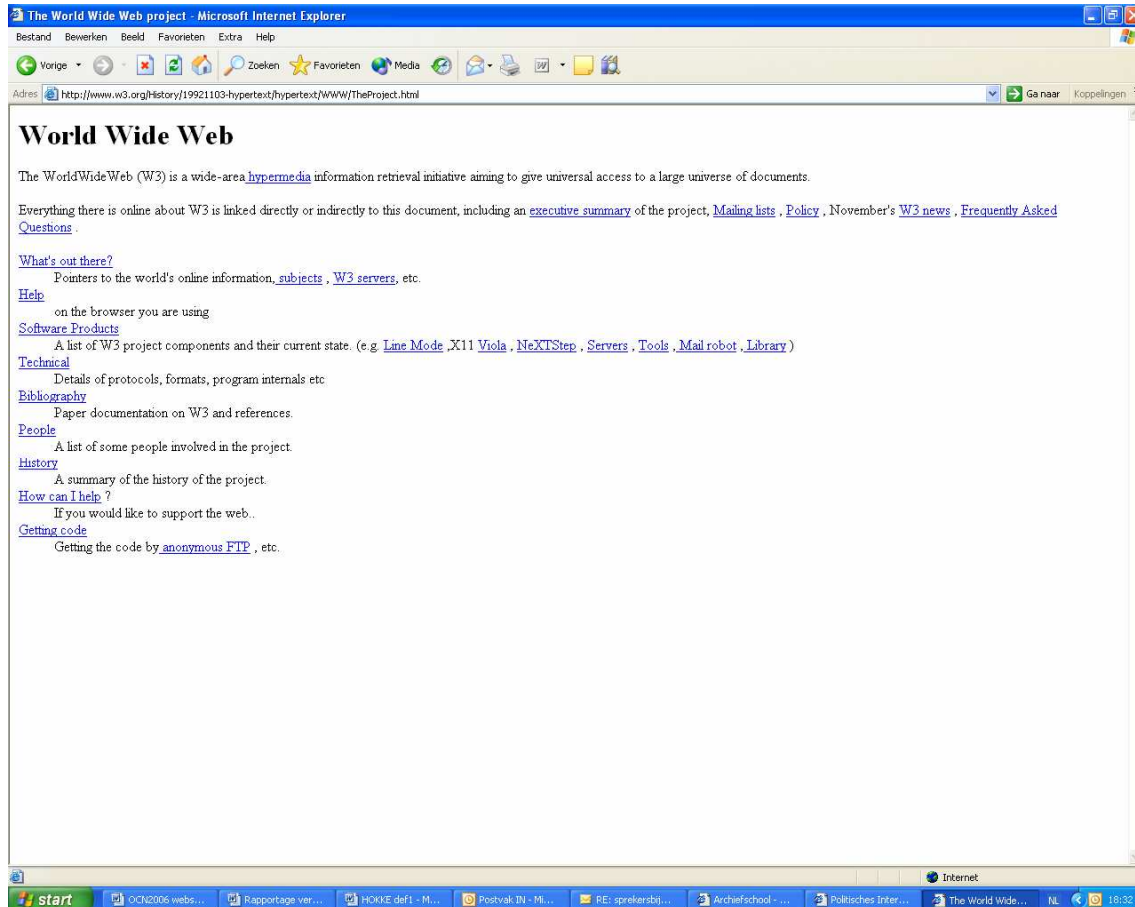
Deze rechtszaak had vervolgens nog een staartje in 2005, te weten een nieuwe rechtszaak. Deze rechtszaak werd door Healthcare Advocates Inc (de aanklager uit de eerste zaak) aangespannen tegen zowel het advocatenkantoor als tegen The Internet Archive. De aanklacht betrof overtreding van het copyright en van de Amerikaanse Digital Millenium Copyright Act. De aanklacht richtte zich primair op het advocatenkantoor, maar aangezien The Internet Archive aanbieder en beheerder van de gearchiveerde websites is, werd zij ook in de aanklacht betrokken. De zaak tegen The Internet Archive is in augustus 2006 geschikt, volgens een bericht op de website van The Internet Archive, maar de zaak tegen het advocatenkantoor loopt nog.<sup>xiii</sup>

### *3. Culturele waarde*

#### *3.1. Het verleden van het web*

De snelle ontwikkelingen in websites maken dat de websites ondertussen meer zijn geworden dan alleen een bron van informatie. Een website van tien jaar geleden is nauwelijks nog te vergelijken met een hedendaagse website: de verandering van een eenvoudige HTML-pagina met informatie over bijvoorbeeld openingstijden naar interactieve pagina's opgebouwd uit een CMS en ter ondersteuning van alle mogelijke werkprocessen in de organisatie. Ook het aantal websites en het gebruik van websites is enorm gegroeid en heeft z'n sporen nagelaten in de manier waarop mensen omgaan met informatie. De websites en het World Wide Web zelf zijn object van onderzoek geworden.

De allereerste website is bijvoorbeeld nog beschikbaar via het web. Tim Berners-Lee, de uitvinder van het world wide web, publiceerde de eerste website in 1991. De site geeft informatie over het world wide web project. De oer-versie van deze site is niet meer beschikbaar, maar een versie uit 1992 is te vinden bij het World Wide Web Consortium.<sup>xiv</sup>



**Figuur 6: afbeelding van de oudste website in het world wide web, gepubliceerd door Tim Berners-Lee**

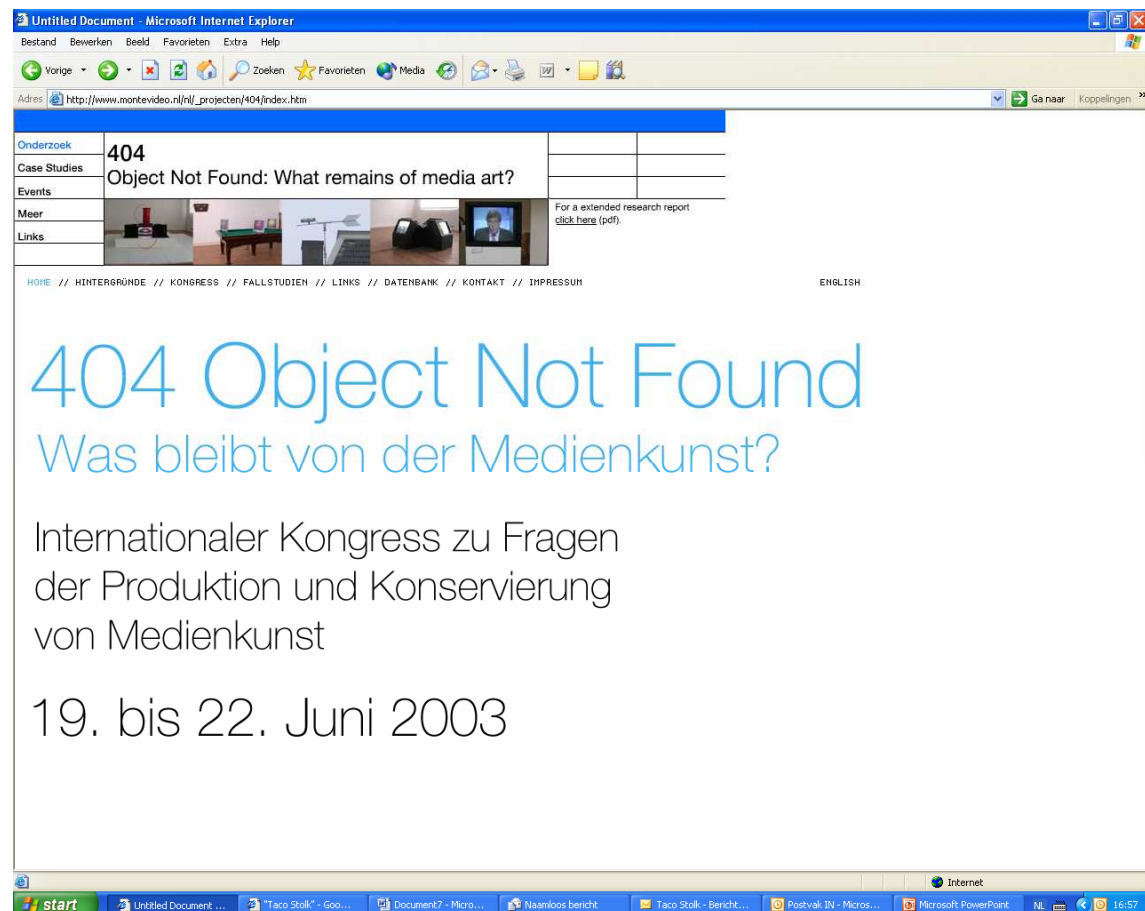
### 3.2. *Het culturele web*

Websites zijn in oorsprong informatiebronnen, maar inmiddels zijn ze een rol gaan spelen in het culturele domein; ze vormen de kern van e-Cultuur<sup>xv</sup>. Een sprekend voorbeeld hiervan is Droombeek<sup>xvi</sup>, een website over de door de vuurwerkcramp getroffen wijk Roombeek in Enschede. Via deze site worden verhalen verzameld over de oude wijk en wordt een relatie gelegd met de nieuwe wijk. De droombeek-site verbindt het heden en verleden van de wijk en de mensen die daar wonen en gewoond hebben, het is een virtuele ontmoetingsplaats. Een ander voorbeeld is de site van Thomas Schlijper. Schlijper publiceert op deze site al zes jaar iedere dag 1 of meerdere foto's die genomen zijn in Amsterdam.

Andere websites zijn zelf kunstobjecten.<sup>xvii</sup> Verschillende kunstenaars maken webkunst Bijvoorbeeld 'Making sense of it all', een internetproject van Blank & Jeron, of 'The Intruder' van Natalie Bookchin.<sup>xviii</sup> De open universiteit biedt zelfs een cursus over webkunst.<sup>xix</sup> In dit domein spelen de vraagstukken rondom de duurzaamheid van die websites. In 2003 werd onder de titel "404 Object nog found" een congres georganiseerd over de duurzaamheid van nieuwe mediakunst, zoals video maar ook webtechnologie.

Websites hebben dus een historische en een culturele waarde gekregen, ze horen tot het erfgoed van een gemeenschap. Ze worden in het Unesco 'Charter on the Preservation of Digital Heritage'<sup>xx</sup> art. 1 expliciet tot het digitaal erfgoed gerekend: Digital materials include texts, databases, still and moving

images, audio, graphics, software and web pages, among a wide and growing range of formats. They are frequently ephemeral, and require purposeful production, maintenance and management to be retained". Dit digitaal erfgoed dient toegankelijk te blijven voor toekomstige generaties en moet gevrijwaard blijven van verlies ten gevolge van technologische veroudering en andere risico's (art. 2 en 3).



**Figuur 7: wat blijft er over van e-Cultuur**

#### 4. Verantwoordingswaarde

De emotionele, informatieve en culturele waarde leveren belangrijke argumenten voor het duurzaam toegankelijk houden van websites. Voor overheidsorganisaties is een vierde overweging net zo belangrijk, of misschien zelfs van primair belang: verantwoording.

Voor overheidsorganisaties zijn websites belangrijke instrumenten in het vormgeven van de elektronische overheid. De burger krijgt via de website toegang tot alle voor hem relevante overheidsinformatie, hij kan gebruik maken van de diensten die daar aangeboden worden, vragen stellen, etc. Dat betekent dat de burger aan die website rechten kan ontleen, "*het stond immers op de website*". De overheid moet zich dus kunnen verantwoorden over de informatie en diensten die zij via haar websites aanbiedt. Dat betekent dat overheidsorganisaties er zorg voor moeten dragen dat de informatie die zij via websites verspreidt vindbaar en leesbaar is én blijft, en wel in dusdanige vorm (integer, betrouwbaar en authentiek) dat de burger de bewuste informatie ook vertrouwt.

Deze waarde van websites is recent benadrukt door de Algemene Rijksarchivaris<sup>xxi</sup>. Hij stelt dat websites, of delen daarvan, archief zijn: "een bundeling (verzameling) van archiefbescheiden gerelateerd aan verschillende bedrijfsprocessen". Dit betekent dat de Archiefwet 1995 van toepassing is voor websites van overheidsorganisaties.

***Webarchivering, het volgende niveau: behoud van waarde***

Aandacht voor de duurzame toegankelijkheid van websites kan dus vanuit verschillende invalshoeken worden gemotiveerd. Ze bevatten informatie die op korte of lange termijn van belang kan zijn voor de beantwoording van hele verschillende onderzoeksvragen. Aan die informatie kunnen rechten worden ontleend, wat dus betekent dat de organisatie die de informatie publiceert daarover verantwoording af moet kunnen leggen. En inmiddels horen ze tot het digitaal erfgoed, zowel op basis van inhoud als vormgeving als functie en gebruik. Ze zijn waardevol en tegelijkertijd bijzonder kwetsbaar.

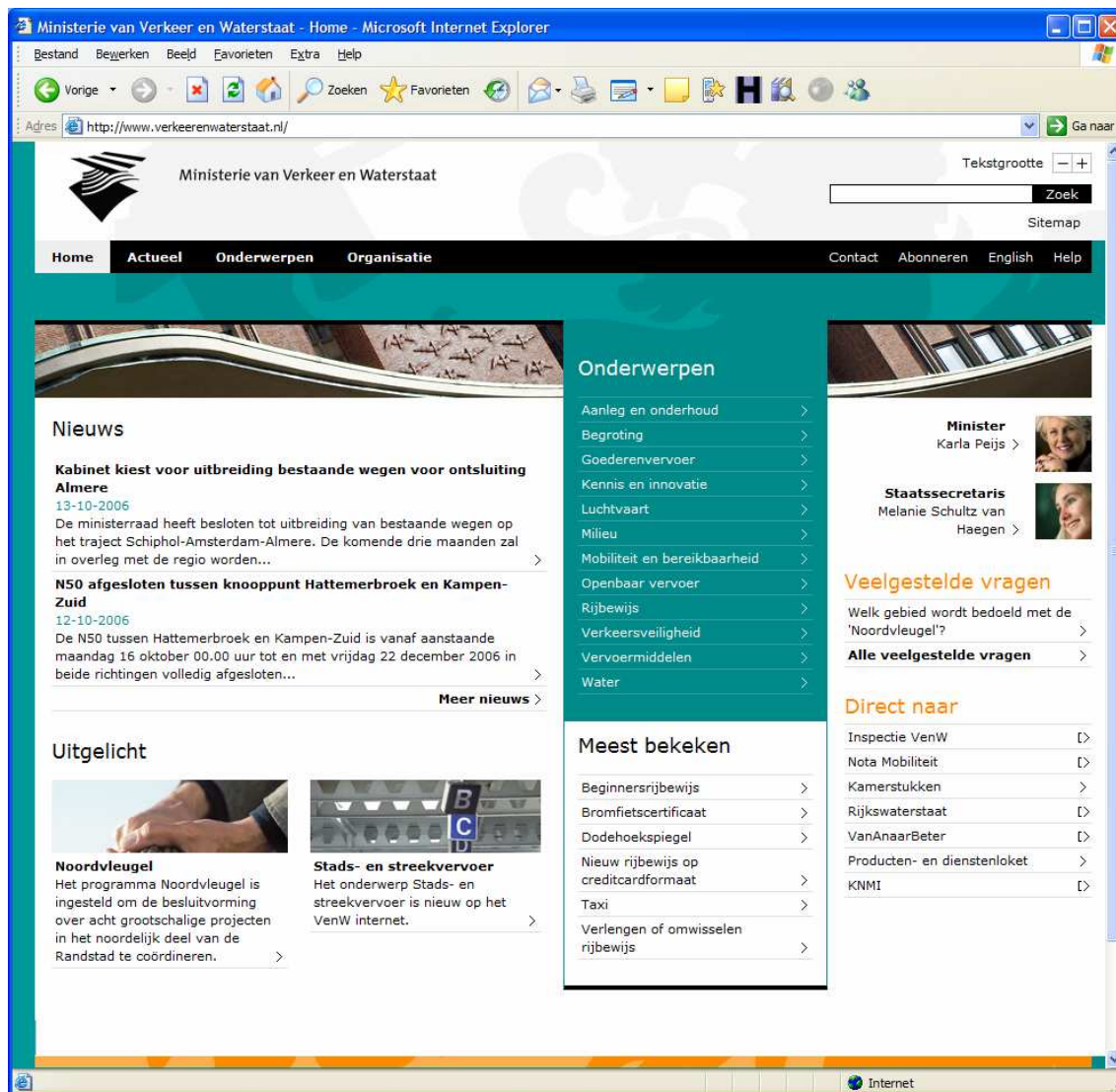
Archivering kan zorgen voor de duurzame toegankelijkheid van deze websites. Om deze archivering goed vorm te kunnen geven zijn vier randvoorwaarden van belang:

1. Allereerst bewustwording, namelijk dat websites meer zijn dan vluchtige publicaties. Het zijn mooie en waardevolle bronnen. Niet alleen om vandaag een boodschap uit te dragen of informatie te zoeken.
2. Het vraagt vervolgens om goed ontworpen sites. Dit zijn sites die toegankelijk zijn voor iedereen die ze wil raadplegen, die robuust zijn, die authentiek en betrouwbaar zijn. De Webrichtlijnen van de stichting ICTU geven een overzicht van eisen en richtlijnen waar goede sites aan moeten voldoen.
3. Een derde randvoorwaarde, vooral bedoeld voor organisaties, is een goed websitebeleid. Bestaande uit een integratie van het voorlichtingsbeleid, de visie op webdesign en archiveringsbeleid in een beleid voor duurzaam toegankelijke websites. Bij dit beleid hoort natuurlijk een toedeling van verantwoordelijkheden, niet alleen voor het actueel houden van de site, maar ook voor de toekomstvastheid. De ISO-NEN 15489:1 norm voor informatie- en archiefmanagement geeft hier richtlijnen voor. Door goede ontwerpen en goed beheer kunnen websites langdurig beschikbaar blijven. Pas bij grote wijzigingen, bijvoorbeeld als er een hele nieuwe site wordt ontworpen, zou de site verdwijnen van het contemporaine web.
4. De laatste randvoorwaarde sluit hierbij aan: er moet een voorziening zijn waar de website gearchiveerd kan worden. Men kan de eigen websites tegen betaling door andere organisaties te laten beheren. Bijvoorbeeld The Internet Archive heeft hier een omgeving voor ingericht. Het Nederlandse bedrijf Capsis levert software om zelf een website archief aan te leggen.

Hans Goutier zal nader ingaan op de overwegingen die voor het Ministerie van Verkeer en Waterstaat een rol spelen bij het archiveren van websites

## Webarchivering bij het ministerie van Verkeer en Waterstaat

Het ministerie van Verkeer en Waterstaat is een relatief groot ministerie, met vestigingen verspreid over heel Nederland. Verkeer en Waterstaat houdt zich bezig met beleid, uitvoering en inspectie. Bijna alle organisatieonderdelen van Verkeer en Waterstaat maken websites. Via deze websites communiceert het ministerie met burgers, bedrijven en andere overheidsorganisaties, worden transacties uitgevoerd en krijgen derden toegang tot de openbare informatie van het ministerie. Het gebruik van internet voor al deze doeleinden neemt hand over hand toe, hetgeen de vraag hoe er moet worden omgegaan met het archiveren ervan steeds prangender maakt.



Figuur 8: de huidige website van Verkeer en Waterstaat

Eén van de eerste websites die door het Ministerie ontwikkeld werd (ca. 1995/1996) was het Digitale Verkeersplein. Inmiddels is deze website verwijderd van het World Wide Web. Het is uitsluitend aan het initiatief van de toenmalige ontwerper te danken dat deze website behouden is gebleven. Naast het gevoel dat het “zonde” zou zijn geweest als deze (mogelijk oudste Nederlandse) overheidswebsite

zou zijn verdwenen, werd binnen het Ministerie geconstateerd dat het onduidelijk is of websites bewaard moeten worden. En, zo ja, hoe dat dan zou kunnen worden uitgevoerd.



**Figuur 9: de eerste versie van de website van Verkeer en Waterstaat (bron: The Internet Archive)**

Verkeer en Waterstaat heeft de Archiefschool daarom opdracht gegeven om te onderzoeken of websites aangemerkt kunnen worden als archief in de zin van de Archiefwet 1995 en hoe de archivering van websites binnen het ministerie vorm gegeven kan worden.

De achterliggende gedachte bij de opdracht aan de Archiefschool was, dat het in het geval van websites mogelijk zou gaam om archiefmateriaal. En voor het omgaan met archiefmateriaal gelden voor de Nederlandse overheid bijzondere regels, die met name zijn vastgelegd in de Archiefwet, het Archiefbesluit en enkele ministeriële regelingen.

Zouden de websites geen archiefmateriaal bevatten, dan zou het ministerie af kunnen zien van het archiveren ervan en zou het ministerie er b.v. van uit kunnen gaan dat andere partijen die websites wel zouden verzamelen, b.v. de Koninklijke Bibliotheek of, in het buitenland, The Internet Archive met hun Way Back Machine.

Maar zou het gaan om archiefbescheiden, dan zou dat niet kunnen. Dan zou het ministerie nadrukkelijk een eigen verantwoordelijkheid hebben voor het beheer van zijn eigen archief en dus niet blind mogen vertrouwen op de activiteiten die andere partijen op dit vlak verrichten.

De Archiefschool dit onderzoek hiernaar en kwam tot de conclusie dat websites inderdaad archiefmateriaal vormen, dan wel bevatten. Afgelopen voorjaar is deze conclusie, die hier en daar nog wel ter discussie werd gesteld, overigens bevestigd door de Algemene Rijksarchivaris (<http://www.advies.overheid.nl/attachment.db?5150>)

## **Wat betekent dit nu voor VenW?**

Op grond van de Archiefwet 1995 zijn overheidsorganen, verplicht hun archieven in goede, geordende en toegankelijke staat te brengen en te bewaren (AW1995 art. 3). Dit geldt voor archieven “ongeacht hun vorm”, dus ook voor digitaal archief en websites.

Deze formulering wil overigens niet zeggen dat alle websites voor eeuwig bewaard moeten worden. Zolang we ze hebben, moeten ze in goede, geordende en toegankelijke staat zijn. Dus we moeten in ieder geval in staat blijven om websites te presenteren zoals ze er nu uitzien. Om er voor te zorgen dat de websites worden geselecteerd die moeten worden bewaard dan wel vernietigd, is het nodig dat ze onder enig “archief regime” vallen, zodat de normale regels voor selectie en vernietiging kunnen worden toegepast. In het kort komen deze regels neer op het volgende: Hoe lang we archiefbescheiden moeten bewaren, hangt af van de rol die websites in het werkproces spelen.

Om dat te kunnen bepalen hebben we in Nederland de beleidsterreinen van de rijksoverheid uitvoerig beschreven in de zogenaamde Rapporten Institutioneel Onderzoek. Zo'n rapport beschrijft uitvoerig de doelen die de overheid op een bepaald werkterrein nastreeft en de partijen (in deze rapporten actoren genaamd) die op een bepaald terrein een rol spelen. Per actor is beschreven welke handelingen deze uitvoert. Per handeling wordt bepaald of het archiefmateriaal dat daar uit voortvloeit bewaard moet blijven of mag worden vernietigd en op welke termijn dat dan moet gebeuren. Als criteria worden daarbij b.v. gehanteerd dat materiaal over de voorbereiding, vaststelling en evaluatie van overheidsbeleid moet worden bewaard en dat het mogelijk moet zijn om het overheidshandelen op hoofdlijnen te reconstrueren op grond van archiefmateriaal. Besluiten over bewaren of vernietigen worden vastgelegd in een Basis Selectie Document (BSD), dat volgens een zware procedure, met inspraak van alle betrokken partijen, tot stand komt. Een BSD is dus de basis voor vernietiging of bewaren van websites.

Bij het toepassen van een BSD op websites lopen we al snel aan tegen het probleem dat BSD's zijn gericht op individuele handelingen, terwijl veel websites te maken hebben met een veelheid van handelingen. Het is dan vaak ondoenlijk om uit een website alleen de dingen te filteren die je er voor permanente bewaring uit wilt hebben. Bovendien, als je dit zou doen, raak je de context kwijt waarin de informatie ooit op het web is gezet. En ook deze context kan belangrijk zijn.

Tegen deze achtergrond heeft VenW de volgende stappen gezet. De Archiefschool heeft een onderzoek uitgevoerd en op grond daarvan een archiveringsstrategie en implementatieplan opgeleverd. Capsis heeft onderzocht wat je aan de voorkant, bij het bouwen van een website al kunt doen om het archiveren ervan te vergemakkelijken.

VenW en Capsis hebben gezamenlijk een pilot uitgevoerd om te onderzoeken of het mogelijk was om de websites van VenW met behulp van de “snapshot-methode” te archiveren. Alle internet- en intranetsites van VenW zijn offline beschikbaar gemaakt en hiervan is een eindrapportage opgesteld.

Verkeer en Waterstaat heeft verder besloten om het aantal websites drastisch te saneren. Ten tijde van de pilot met Capsis waren er meer dan 120 internetsites en bleek dat het intranet bijzonder onoverzichtelijk was. Op dit moment hebben beleid, uitvoering en inspectie ieder een eigen intra- en internetsite. Het veld is daardoor beter beheersbaar geworden.

Verder is er voor gezorgd dat de internetsites van VenW voldoen aan de Web-richtlijnen die door de ICTU zijn opgesteld. Hoewel deze allereerst zijn opgesteld om de toegankelijkheid van overheidswebsites te vergroten, hebben zij ook als effect dat zij de archiveerbaarheid bevorderen. Immers, als een gebruiker gemakkelijk bij een overheidswebsite kan komen, kan een robot die een snapshot moet maken dat ook. Overigens heeft de ICTU inmiddels aanbevelingen voor de archivering

van websites opgesteld (<http://www.advies.overheid.nl/webrichtlijnen/>), die ook voor het overgrote deel door VenW zijn overgenomen.

Op dit moment zijn we aan het onderzoeken of het mogelijk is om de recordsmanagement-applicatie die we bij Verkeer en Waterstaat hebben aangeschaft ook te gebruiken voor het archiveren van onze websites. Het grote voordeel daarvan zou zijn dat websites, die immers archiefbescheiden zijn in de zin van de Archiefwet, in samenhang met de andere archiefbescheiden van het ministerie op te slaan en te beheren.

Frans-Willem Duijnhouwer zal nu verder op ingaan op de ervaringen met het archiveren van websites, die Capsis onder andere bij Verkeer en Waterstaat heeft opgedaan.



## De praktijk

Nu duidelijk is waarom webarchivering uitgevoerd moet worden en waarom een organisatie kiest voor het archiveren van (eigen) websites, is het van belang te bekijken welke uitdagingen men tegenkomt bij het archiveren van een website. Aan de hand van de volgende onderwerpen wordt uitgelegd hoe men komt tot het goed kunnen archiveren van een website:

- Wat maakt webarchivering tot een uitdaging
- De werking van een website
- Het archiveren van een website
- Voorwaarden voor webarchivering
- De bouw van een website
- Afspraken
- Praktische voorbeelden en tips
- Concluderend

## Wat maakt webarchivering tot een uitdaging?

Webarchivering is vanuit het perspectief van archivering niet heel anders dan het archiveren van digitale documenten. Immers, selectiecriteria die ook gelden voor andere bronnen zijn toepasbaar, selectielijsten en bewaartermijnen zijn van toepassing en er dient gelet te worden op zaken als integriteit, authenticiteit en toegankelijkheid. Op basis van deze archivistische selectiecriteria valt dus te bepalen dat een website bewaard moet worden. De daadwerkelijke archivering roept vervolgens een aantal vraagstukken op, een website is tenslotte niet te vergelijken met een eenvoudig Word of PDF-document. Een website bestaat allereerst uit een grote verzameling objecten. Deze objecten vormen het document.

Een website heeft onduidelijk grenzen. Waar begint deze en waar eindigt de website? Een website met één eindresultaat – de publicatie – kan gevormd zijn uit een groot aantal verschillende bronnen. Daarnaast is de presentatie van een website niet overal en niet voor iedereen hetzelfde. Een website is vaak erg veranderlijk en interactief.



Figuur 10: één website, twee unieke URL's...

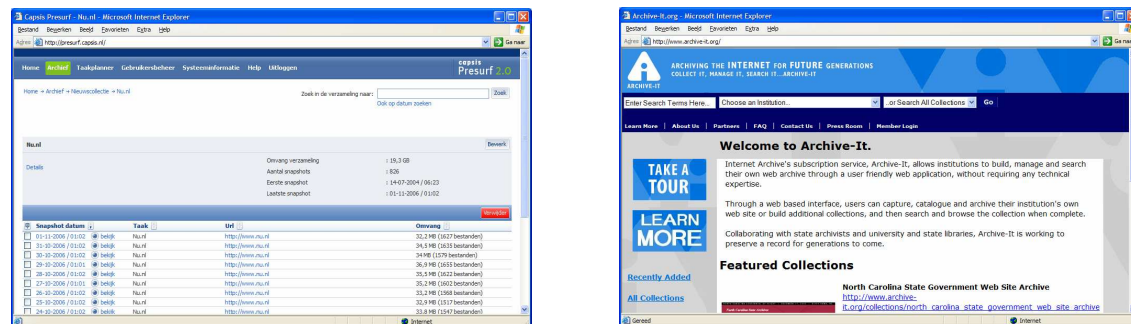
De objecten op een website die het document vormen, kunnen technisch gezien bestaan uit een grote hoeveelheid verschillende formaten. Vanuit verduurzamingoptiek is dit een extra uitdaging. Hoe kan gegarandeerd worden dat een website er later ook nog zo uitziet als oorspronkelijk? En hoe is het gesteld met de toegankelijkheid?

Het archiveren van een website kan volgens verschillende methoden gedaan worden. Elke methode kent een ander resultaat. Daarnaast speelt de manier waarop een website is gebouwd, een grote rol bij de archiveerbaarheid hiervan – welke methode ook gekozen wordt. Veel websites werken “live” al niet consistent in verschillende webbrowsers, dus hoe zag het document er oorspronkelijk uit?

Om te kunnen starten met webarchivering moeten beantwoord worden welke methode precies gebruikt kan en moet worden. Dit praktische onderdeel van de paper hoopt een start te kunnen bieden bij de beantwoording van de vragen die aan de basis liggen van webarchivering. Hierna kan gestart worden met het bepalen van een archiveringsstrategie, inbedding van het proces in de organisatie en eventueel een product om webarchivering uit te gaan voeren. Dit alles met het doel websites, bestaande uit een veelheid aan objecten, als één geheel zonder verlies van vorm, inhoud en interactie te bewaren en toegankelijk te houden voor later.

## De werking van een website

Om archieven van websites te kunnen maken is het van belang te weten hoe een website ongeveer werkt. Dit ondanks dat er tegenwoordig hulpmiddelen bestaan, zoals Capsis Presurf<sup>xxii</sup> of de diensten van Archive-It<sup>xxiii</sup>, die de technische aspecten van het webarchiveren afhandelen.



Figuur 11: Capsis Presurf en Archive-It

Een website bestaat uit twee zijden. Enerzijds is er het deel dat op een webserver actief is, anderzijds is er het deel dat door de webserver gepresenteerd wordt aan de bezoeker. Bij eenvoudige websites, ook wel statische websites genoemd, zijn beide delen gelijk. De webserver is in dit geval niets anders dan een onderdeel van het transportmedium.

Bij meer geavanceerde websites, die tegenwoordig het grootste deel van het internet vormen, zijn deze delen verschillend. Beide zijden interacteren vaak zeer actief met elkaar op basis van gebeurtenissen aan de zijde van de bezoeker of de webserver. Websites zijn tegenwoordig meer te vergelijken met min of meer ingewikkelde programmatuur dan alleen inhoud aangevuld met eenvoudige meta-informatie voor de opmaak.

Aan de bronzijde is programmacode te vinden. Deze kan van alles doen. Bijvoorbeeld het presenteren van een website in de juiste taal, het formuleren van een reactie op een klik van een gebruiker aan de presentatiezijde. Op dynamische websites wordt aan de bronzijde vaak tekstuele inhoud en de opmaak voor presentatie gescheiden opgeslagen. Wanneer een pagina wordt opgevraagd, wordt door de programmacode de juiste opmaak verweven met de juiste tekstuele inhoud. Het resultaat hiervan wordt aangeboden aan de bezoeker. Op deze manier wordt de beheerbaarheid van een website sterk vergroot. Een hierin gespecialiseerde partij maakt de opmaak van de website, terwijl de eigenaar van de website zelf tekstuele inhoud kan toevoegen. Een "Content Management Systeem" verzorgt deze mogelijkheid, variërend van zeer eenvoudig, tot zeer complex waarbij de gebruiker op allerlei manieren wordt ondersteund bij zijn werkzaamheden. Vaak wordt alle inhoud, behalve externe documenten en afbeeldingen voor opmaak van de website, opgeslagen in een database.

De programmacode van een website kan in een groot aantal programmeertalen gemaakt worden. Verschillende webservers (die ook weer deels op verschillende platformen werken) ondersteunen verschillende talen.

Wanneer een "pagina" naar de presentatiezijde verstuurd wordt, bevat deze naast eenvoudige broncode vaak allerhande kleine programmaatjes die uitgevoerd worden door de webbrowser. Ze worden gebruikt om bijvoorbeeld mooie bewegende menustructuren te maken. Iedere webbrowser heeft zo zijn eigen kracht en belangrijk om te weten is dat webbrowsers soms ook verschillend omgaan met dezelfde scripts. Zo kan het zijn dat script wel werkt in Microsoft Internet Explorer, maar niet in Mozilla Firefox – en omgekeerd. Dit heeft te maken met de termijnen waarbinnen een standaardisatieproces zich afspeelt en de snelheid van softwarebouwers. Zo bieden commerciële leveranciers vaak extra mogelijkheden buiten de bestaande vastgestelde specificaties om, terwijl bijvoorbeeld open source software vaak heel precies de van toepassing zijnde standaarden volgt. Wanneer programmeurs gewend raken aan een bepaalde browser – en dit is in veel gevallen zo – zullen beschreven problemen ontstaan.

Voor archivering van websites is het belangrijk om de volgende informatie goed te onthouden:

- Webarchivering is anders dan het archiveren van andere digitale documenten omdat er geen heldere begrenzing is in een website. Onderdelen op het internet zijn met elkaar verweven.
- De presentatie van een website is niet overal gelijk. Welke presentatie dient gearchiveerd te worden?
- Een website bestaat uit veel samengestelde technieken wat de duurzaamheid en de authenticiteit in gevaar brengt.
- Standaarden zijn vaak "vervuild" met extra mogelijkheden die toegevoegd worden door commerciële partijen.

## Het archiveren van een website

Een website is dus een zeer complex document. De begrenzing is niet helder, de presentatie kan verschillen, een website is vaak zeer veranderlijk en een site bestaat uit verschillende technologieën.

Wanneer men begint met webarchivering is het dus van belang rekening te houden met deze kenmerken. Webarchivering begint met bepalen 'wat' gearchiveerd moet worden en 'hoe' dat gearchiveerd gaat worden.

Voor de keuze van wat gearchiveerd moet worden geldt op dit moment een pragmatische insteek. Voorlopig is het verstandig een gehele website te archiveren. Omdat de begrenzing niet vooraf vastgesteld kan worden en nog niet bepaald is hoe om te gaan met archivalistische selectie van websites, is het verstandig om de stelling in te nemen dat deze later bepaald wordt door de gebruiker van het archief. Eventuele archivalistische selectiecriteria die in de toekomst volgen, kunnen achteraf toegepast worden op het gearchiveerde exemplaar van de website. Ruime selectie geeft op dit moment de grootste kans op het vormen van een in de toekomst bruikbaar archief.

Na het kiezen van de te archiveren bronnen, volgt de keuze voor de te gebruiken techniek. Het archiveren van een website kan, analoog aan de werking van een website, vanaf twee zijden uitgevoerd worden. Vanaf de bronzijde en vanaf de presentatiezijde. Beide methoden hebben voor- en nadelen.

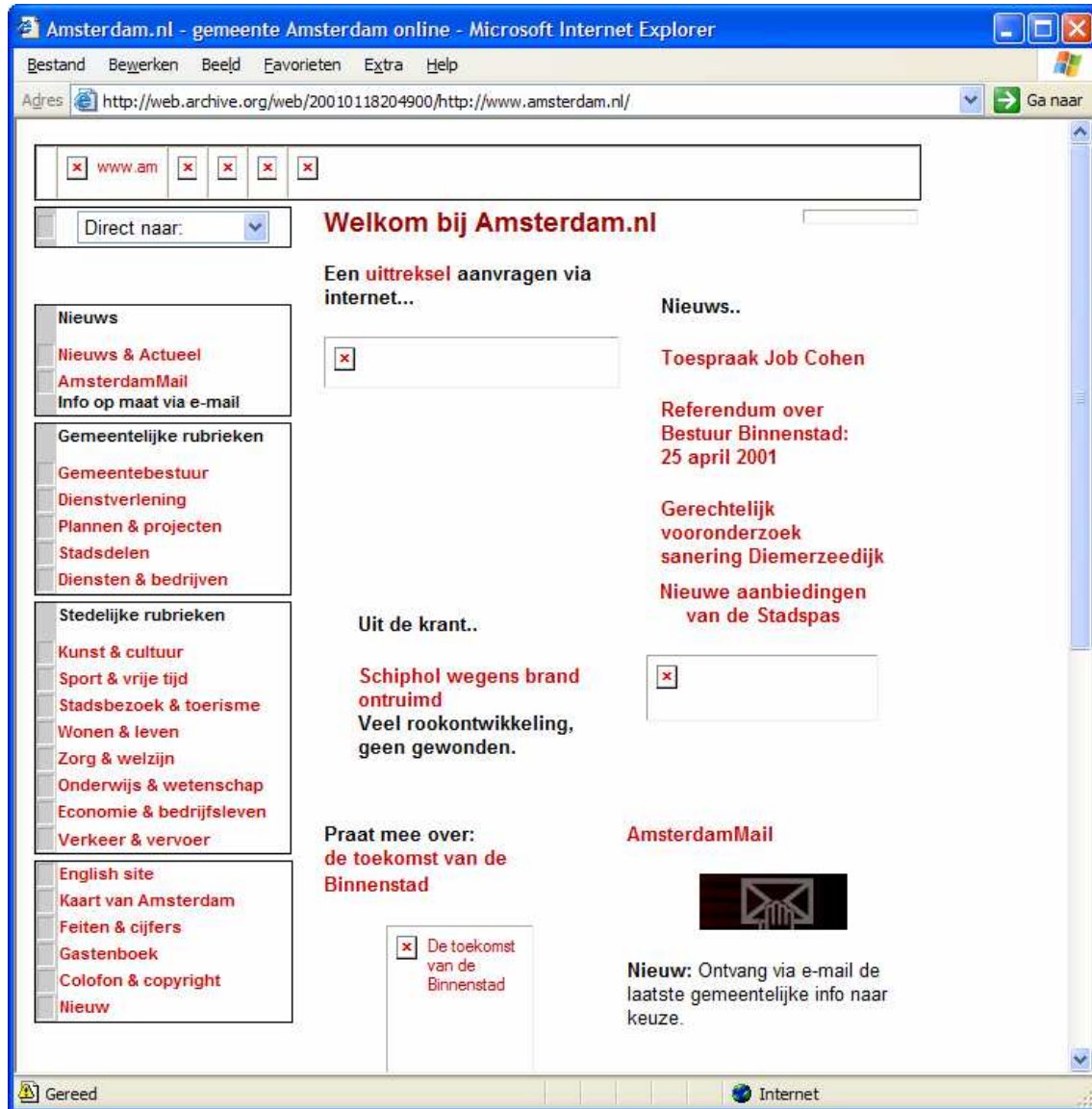
Archiveren vanaf de serverzijde moet plaatsvinden door alle gebruikte technieken en bronnen te archiveren. Een archiefkopie die via deze methode gemaakt is moet ook alle software bevatten inclusief de broncode, en mogelijk moet ook het hardwareplatform – als het gaat om de langere termijn – gepreserveerd worden. De programmatuur is immers afhankelijk van het geheel aan gebruikte technieken. Een emulator van deze hard- en software kan mogelijk op langere termijn een oplossing bieden. De complexiteit van dit soort oplossingen is feitelijk een nadeel van dit type archivering. Een voordeel van deze methode is dat een 100% kopie mogelijk is.

De tweede methode is het maken van een archiefkopie vanaf de presentatiezijde. Dit wordt ook wel de snapshotmethode genoemd. Er wordt door een zogenaamde crawler een kopie gemaakt van datgene wat een bezoeker ziet als deze de website bezoekt. Een soort foto van de website op een bepaald moment. Een verschil met foto's is wel dat de website wel als website benaderbaar blijft. Een voordeel van de snapshotmethode is dat hierdoor alle afhankelijkheden van de technologie aan de serverzijde verdwenen zijn. Alle programmacode is immers al uitgevoerd bij het opvragen van de pagina's. Een nadeel is dat een 100% kopie niet haalbaar is. Alle gebieden waar interactie tussen de bezoeker en de webservice nodig is, wordt niet in het archief opgenomen. Zo zal een formulier wel worden opgeslagen, maar niet de reactie van de programmatuur na het verzenden van het formulier. De programmatuur die door de webbrowser uitgevoerd moet worden wordt ook in het archief opgenomen dus afhankelijkheid van een webbrowser blijft bestaan.

Een andere manier van archiveren die vanaf de voorkant werkt, is "archiving by browsing". Hiervoor geldt dat het archief opgebouwd wordt terwijl over een website wordt gebladerd door een gebruiker. Voordeel is dat ook de hyperlinks die door een crawler niet worden benut, door de gebruikers kunnen worden aangeklikt. Het resultaat is een vollediger archief. Dit archief is in sommige gevallen wel weer afhankelijk van een specifieke webbrowser.

De meest gebruikte methode is de snapshotmethode. Deze methode is goed te automatiseren zodat het archiveren van websites niet te arbeidsintensief wordt. Daarnaast is het vanuit optiek duurzaamheid goed dat de afhankelijkheden aan de serverkant alvast verdwenen zijn.

Een lastig punt is dat de organisatie moet bepalen hoe lang de website bewaard moet worden. Overheidsorganisaties hebben hiervoor bewaartermijnen gedefinieerd in selectielijsten. De bewaartermijn hangt af van het primaire proces waaruit de website voortkomt. Vaak zal een website uit meerdere processen komen waarvoor verschillende bewaartermijnen gelden. Voor de beschikbaarstelling van het archief is net als voor alle andere bestandsformaten een softwareapplicatie benodigd om de formaten te openen. Voorlopig volstaat een webbrowser. In de meeste gevallen is het dan nog steeds ingewikkeld om een archiefkopie te bekijken, omdat niet duidelijk is welke delen van de website direct vanaf internet komen en welke delen in het archief zitten. Ook zullen veel hyperlinks in de pagina's de bezoeker direct naar het internet brengen. De authenticiteit van het archief is in gevaar. Daarom maakt bijvoorbeeld de applicatie Presurf van Capsis gebruik van een speciale tussenlaag die gebruikt wordt om dit soort problemen op te lossen. Deze tussenlaag kan in de toekomst worden gebruikt om de webbrowser na te bootsen zodat ook de afhankelijkheid van de webbrowser verdwijnt.



Figuur 12: de website van gemeente Amsterdam op The Internet Archive – hoe authentiek is dit?

## Voorwaarden voor webarchivering

Kennis van de werking en de opbouw van een website biedt de mogelijkheid om te kunnen bepalen of een website archiverbaar is en waar eventuele problemen schuilen. Maar alleen kennis is niet voldoende. Om eerder beschreven problemen op te lossen is het ook vaak nodig om bij de bouw van een website alvast rekening te houden met het archiveren ervan. Gewoon als extra onderdeel van de ontwerpactiviteiten. Om deze extra activiteit succesvol te laten zijn, is het verstandig om een archiefmedewerker te betrekken bij de bouw van de website. Deze kan bepalen welke onderdelen in het bijzonder archiverbaar moeten zijn en adviseren over te gebruiken standaarden en testmethoden. En wanneer een onderdeel goed archiverbaar is, is de toegankelijkheid ook vaak een stuk beter. Een programmeur is niet vaak bezig met toegankelijkheid en toekomstvastheid – bijvoorbeeld van gebruikte bestandsformaten – van wat deze maakt. Daarom is het van belang deze

verantwoordelijkheid ergens anders te beleggen. Ook het testen op minder gangbare systemen is een dergelijke activiteit en altijd geldt dat de toegankelijkheid vaak alleen maar vergroot wordt. Verder is het van belang goede software te gebruiken voor het toegankelijk maken van het archief. Software die garanties biedt over integriteit, beschikbaarheid, authenticiteit, benaderbaarheid en toegankelijkheid.

## Afspraken

Voor de meeste succesvolle webarchivering zijn afspraken nodig. Dit zijn praktische afspraken die te maken hebben met het ontwerp en de bouw van een website. De bestaande producten voor webarchivering bieden ook technologie om veelvoorkomende problemen op te lossen, maar wanneer er geen rekening gehouden is met archivering bij de bouw van een website, zal het nooit optimaal gaan.

Praktische afspraken zijn opgenomen in de richtlijnen voor toegankelijkheid en duurzaamheid van overheidswebsites van Overheid.nl<sup>xxiv,xxv</sup>. Daarnaast heeft Overheid.nl aanvullende richtlijnen opgesteld die voor webarchivering gelden<sup>xxvi</sup>. Ook René Voorburg van Uselab heeft in de pilot met het Ministerie van Verkeer en Waterstaat richtlijnen opgesteld die gebruikt kunnen worden bij webarchivering, de Webontwerp richtlijnen voor archivering<sup>xxvii</sup>. Door zoveel mogelijk van deze afspraken na te leven wordt archiverbaarheid van een website makkelijker, sneller en kwalitatief beter.

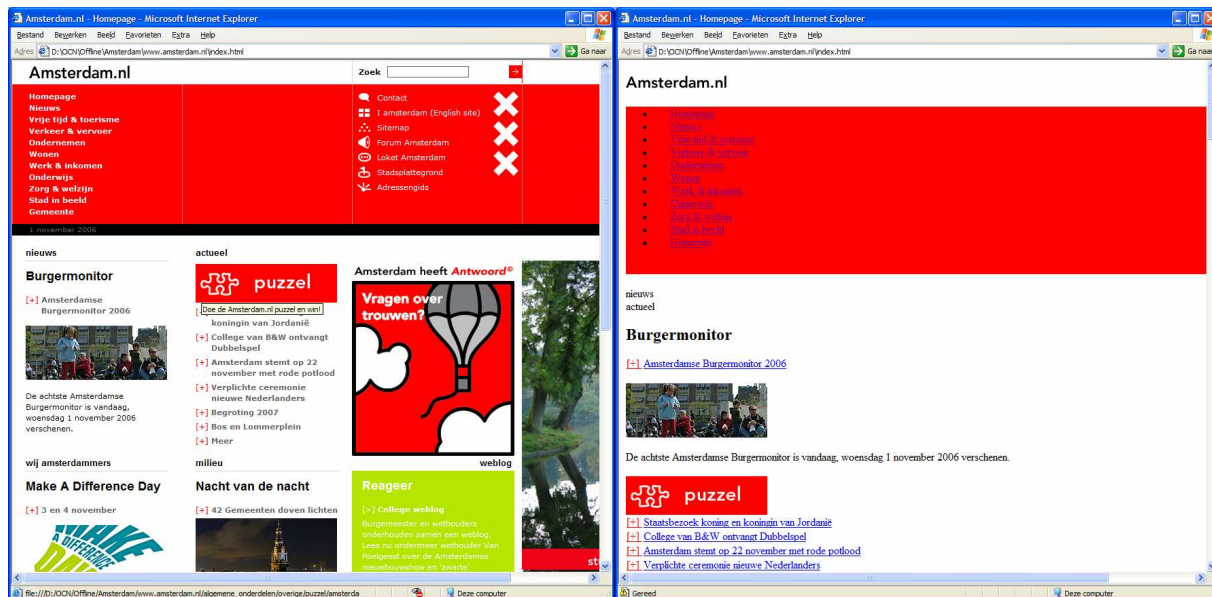
## Praktische voorbeelden en tips

Over de bouw van een website zijn inmiddels voldoende mogelijkheden gegeven om de archiverbaarheid naar het volgende niveau te brengen. Voor het archiveren van websites zijn ook nog praktische tips en voorbeelden te geven. In deze paragraaf is hier meer aandacht voor.

### *Basis*

Aan de basis van het archiveren met behulp van de snapshotmethode staat het binnenhalen van de website. Hiervoor wordt een crawler gebruikt. Voorbeelden hiervan zijn HTTrack<sup>xxviii</sup> en Heritrix<sup>xxix</sup>. Beide producten komen uit het open source domein. HTTrack is een offline browser. Deze maakt kopieën van een website op de manier waarop een website is opgebouwd. Heritrix is aanzienlijk ingewikkelder. De door deze tool gemaaktearchieven zijn niet direct bruikbaar om te bekijken. Hier zijn andere tools voor nodig, ook te vinden op de website van Heritrix of via het Internet Preservation Consortium<sup>xxx</sup>. Voor het gebruik van beide tools is een degelijke kennis van websites en de crawlers zelf een noodzaak.

Wanneer men wat ervaring wil opdoen met het archiveren van een website, is het aan te raden HTTrack te proberen. Hierbij moet wel de opmerking gemaakt worden dat dearchieven er vaak erg goed uitzien totdat de internetverbinding op de archiefmachine wordt weggenomen. Vaak blijkt dat grote delen uit de archiefkopie toch van internet worden opgevraagd.



**Figuur 13: één offline versie van de website van gemeente Amsterdam – Met en zonder internetverbinding**

Het is voor het gebruik van HTTrack raadzaam om een korte cursus te volgen, aangezien het een tool is “door techneuten, voor techneuten”. Wanneer een echt archief van websites opgebouwd moet gaan worden, is het van belang speciale aandacht aan de volgende onderwerpen te besteden.

Voor het maken van een offline kopie bestaan ook een aantal hulpmiddelen die op Windows werken. Een voorbeeld hiervan is Teleport Pro. Dit soort tools maken vaak hele goede offline kopieën van websites, maar veranderen hiervoor een en ander in de broncode van de website waardoor deze kopie niet meer één op één overeenkomt met de werkelijke website. Het “archief” is geheel afhankelijk van deze tools en daarom niet toekomstvast. Deze tools zijn veelal gemaakt onder “closed source” dus het is ook niet precies duidelijk wat ze wijzigen en waarom.

Het gebruiken van een offline browser voor het maken van een archiefkopie van een website zorgt ervoor dat er op de harddisk een kopie van de geselecteerde website staat. Maar vanuit optiek van archiveren is dat niet voldoende. Hoe wordt het geregeld met de integriteit van het archief? Hoe kan het archief weer beschikbaar gesteld worden aan bijvoorbeeld onderzoekers? En hoe wordt omgegaan met duurzaamheid in de zin van beheer op het archief?



## Integriteit

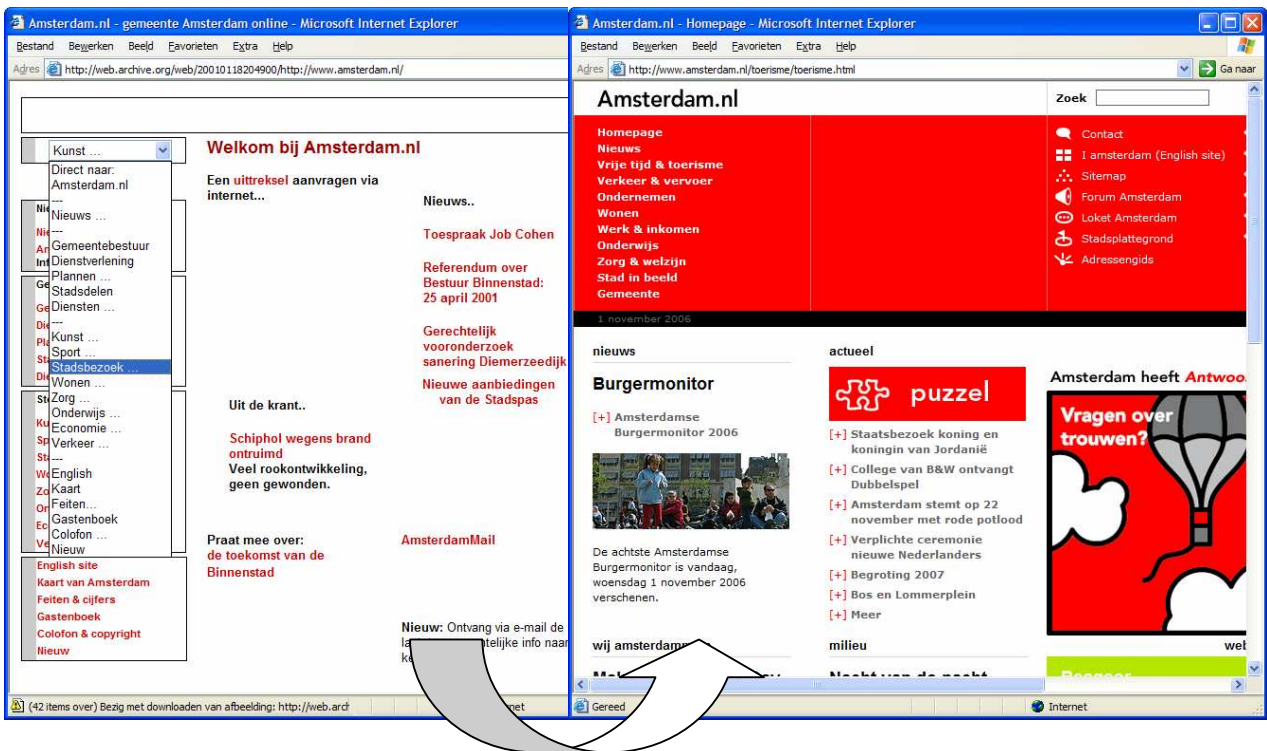
### Correctheid, geldigheid, onweerlegbaarheid, controleerbaarheid

Voor een archief is het van het grootste belang de integriteit hiervan te waarborgen. Is het archief nog steeds het archief zoals dit destijds is opgebouwd? Er zijn technische ingrepen noodzakelijk om een integriteitcontrole voor gearchiveerde websites in te bouwen. Een bezoeker dient te allen tijde te kunnen vertrouwen op de weergave uit het archief. Een archivaris heeft als taak integriteit te bewaken. Voor websites geldt dat dit zoveel mogelijk automatisch uitgevoerd dient te worden gezien de grote hoeveelheid bestanden waaruit een website bestaat.

## Authenticiteit

### Volledigheid, presentatie, context

Ziet de website er nog zo uit als in het verleden? Is de weergave in het archief hetzelfde als de oorspronkelijke. Controleer bij ieder archiefexemplaar welk deel van de website daadwerkelijk in het archief aanwezig is en hoe dit eruit ziet. Vaak resulteert een klik op een hyperlink in een bezoek aan de huidige website op hetzelfde adres. Bijvoorbeeld na het geven van een zoekopdracht wordt verder gesurft op het echte internet. Ter controle is het handig om de netwerkverbinding te verwijderen om zodoende te bepalen wat in het archief zit, en wat niet. Bekijk ook of hyperlinks u binnen het archief houden, of dat u tijdens het bekijken van het archief naar het internet leiden.



Figuur 14: één willekeurige klik brengt de bezoeker naar de huidige website van gemeente Amsterdam.

## *Beschikbaarheid*

### Leesbaarheid

Als het archief gevormd is, moet bepaald worden welke bestandsformaten aanwezig zijn en welke mogelijk niet duurzaam zijn. Van belang is te bepalen welke bestandsformaten in het webarchief mogen voorkomen en welke niet. Een verstandige keuze is alleen bestandsformaten te kiezen welke aan bijvoorbeeld ISO standaarden voldoen. Dus geen Word bestanden, maar PDF bestanden (van de juiste versie!)<sup>xxxi</sup>. Natuurlijk heeft dit in beginsel te maken met het plaatsingsbeleid op de website, maar in het kader van het maken van een duurzaam archief is het mogelijk een verstandige keuze de niet duurzame bestanden te vertalen naar duurzame formaten. Hiervoor is het noodzakelijk een afgeleide versie van het archiefexemplaar te creëren waarin koppelingen naar gewijzigde bestanden hersteld worden.

## *Toegankelijkheid*

### Vindbaarheid gegevens

Een archief is geen goed archief als er geen informatie te vinden is. Het is van belang een heldere structuur in de afzonderlijke snapshots aan te brengen om zodoende eenvoudig terugvinden van gegevens in het archief mogelijk te maken. Bijvoorbeeld door middel van het toevoegen van metadata voor ontsluiting van de gegevens, liefst in een daarvoor bestemd systeem (een Records Management Applicatie). Daarnaast is een webarchief na verloop van tijd zo'n grote collectie van gegevens dat een zoekfunctie niet mag ontbreken. Zowel zoeken in de metadata als full text zoeken door het archief zijn functionaliteiten die geboden zouden moeten worden.

## *Benaderbaarheid*

### Beveiliging, toegang, raadpleegbaarheid

Ten laatste moet een webarchief benaderbaar zijn voor bezoekers. Het is mooi als er een geslaagde archiefkopie van een website is gemaakt, maar niemand heeft er wat aan als deze in een la of kluis belandt. Gegevens in het archief moeten voor bezoekers toegankelijk zijn. Misschien lijkt het nu onbelangrijk, maar over 20 jaar is het zeker interessant om dit soort archieven tot onze beschikking te hebben. Omdat sommige websites (denk aan intranetten, beveiligde zones) niet voor iedereen toegankelijk moeten zijn, is een vorm van beveiliging op de archieven ook van belang.

## **Concluderend**

Webarchivering is in theorie niet anders dan de archivering van andere soorten archief-bescheiden. In de praktijk is het echter moeilijker. De reden daarvoor is dat niet altijd even duidelijk is wat precies de website is. Bovendien is een website vaak een dynamisch geheel, dat moeilijk is vast te leggen.

Maar hebben we eenmaal een website-archief, dan is de behandeling hiervan analoog aan die van andere archiefstukken. Ook vanuit duurzaamheidsoptiek is de omgang met gearchiveerde websites niet echt anders. Het feit dat datgene wat op websites staat soms niet in eigen beheer is gemaakt, kan het noodzakelijk maken deze websites te verduurzamen.

Webarchivering is ook uitdagend omdat websites veel verschillende vormen kunnen aannemen, met veel verschillende achterliggende technologieën. Bij archivering vanaf de achterkant moeten al deze technologieën worden bewaard bij het archiefexemplaar. Wanneer in de toekomst dezelfde soorten systemen beschikbaar blijven als de huidige, dan is deze methode 100% dekkend. De kans dat in de toekomst dezelfde soorten software beschikbaar blijven als tegenwoordig is echter maar klein.

Daarom is een andere methode, archiveren vanaf de voorkant, misschien beter. Er blijft een afhankelijkheid bestaan met technologie vanaf de voorzijde, maar deze is aanzienlijk minder

gecompliceerd en makkelijker te reproduceren. Een nadeel van deze methode is dat er grenzen zijn van de te archiveren bestanddelen van een website. Vooral de interactieve delen worden niet tot nauwelijks gearchiveerd.

Bij het opbouwen van een webarchief, is het van belang na te denken over integriteit, authenticiteit, beschikbaarheid, toegankelijkheid en benaderbaarheid, net zoals bij andere archieven. Het is van belang er over na te denken hoe deze aspecten in het archief kunnen worden gewaarborgd en getest.

---

<sup>i</sup> NRC Handelsblad d.d. 2 november 2006.

<sup>ii</sup> Zie het persbericht van Eurostat d.d. 6 april 2006 via:

[http://epp.eurostat.cec.eu.int/pls/portal/docs/PAGE/PGP\\_PRD\\_CAT\\_PREREL/PGE\\_CAT\\_PREREL\\_YEAR\\_2006/PGE\\_CAT\\_P\\_REREL\\_YEAR\\_2006\\_MONTH\\_04/4-06042006-EN-AP.PDF](http://epp.eurostat.cec.eu.int/pls/portal/docs/PAGE/PGP_PRD_CAT_PREREL/PGE_CAT_PREREL_YEAR_2006/PGE_CAT_P_REREL_YEAR_2006_MONTH_04/4-06042006-EN-AP.PDF) 16-10-2006

<sup>iii</sup> Zie bijvoorbeeld de uitspraak van Brewster Kahle via The Washington Post (2003) via:

<http://www.washingtonpost.com/ac2/wp-dyn?pagename=article&node=&contentId=A8730-2003Nov23&notFound=true>

<sup>iv</sup> De initiatieven vinden vooral plaats in Europa, de VS, Canada, Australië en Nieuw Zeeland. Een uitgebreid overzicht van archiveringsprojecten is te vinden in de bibliografie websitearchivering: H.A. Hokke *Eerste Hulp bij Websitearchivering* (Archiefschool, februari 2006) via: <http://www.archiefschool.nl/onderzoek/aw.htm> 16-10-2006.

<sup>v</sup> Brian Lavoie, Henrik Frystyk Nielsen *Web Characterization Terminology & Definitions Sheet W3C Working Draft 24-May-1999* via [www.w3.org/1999/05/WCA-terms/01](http://www.w3.org/1999/05/WCA-terms/01) 16-10-2006

<sup>vi</sup> H.A. Hokke *Blijvend Beschikbaar, onderzoek naar de archivering van websites* (Amsterdam, 2003) p. 5

<sup>vii</sup> Zie hiervoor P.J. Horsman Msc *Archiveren. Een inleiding* ('s-Gravenhage, 2004) p. 37-43

<sup>viii</sup> Zie <http://web.archive.org/web/19981205183831/www.archief.nl/digiduur/> 16-10-2006

<sup>ix</sup> Zie <http://lcweb4.loc.gov/911/index.html> 16-10-2006

<sup>x</sup> Zie [http://web.archive.org/web/\\*/http://www.onlineconferentie.nl](http://web.archive.org/web/*/http://www.onlineconferentie.nl) 16-10-2006

<sup>xi</sup> Voor een overzicht van deze initiatieven, zie de bibliografie over websitearchivering H.A. Hokke *Eerste Hulp bij Websitearchivering* (Amsterdam, 2006) via <http://www.archiefschool.nl/onderzoek/aw.htm> 16-10-2006

<sup>xii</sup> Zie <http://www.fes.de/archiv/spiegelung/default.htm> 16-10-2006

<sup>xiii</sup> Internet archive announcement: <http://www.archive.org/iathreads/post-view.php?id=67826> 16-10-2006 Voor de aanklacht zie: [http://www.geocities.com/bledrydudenet/Healthcare\\_Advocates\\_v.\\_Harding\\_Complaint\\_FINAL.pdf.pdf](http://www.geocities.com/bledrydudenet/Healthcare_Advocates_v._Harding_Complaint_FINAL.pdf.pdf) 16-10-2006

Inhoudelijke informatie over deze zaken is onder andere te vinden via de New York Times:

<http://www.nytimes.com/2005/07/13/technology/13suit.html?ex=1278907200&en=377b4f3f0d459300&ei=5090&partner=rssuserland&emc=rss> 16-10-2006

<sup>xiv</sup> Time: *80 days that changed the world, Aug. 6, 1991* zie <http://www.time.com/time/80days/910806.html> 16-10-2006 en

*Welcome to info.cern.ch, the website of the world's first-ever web server* zie <http://info.cern.ch/> 16-10-2006 De website zelf is beschikbaar via de website van W3C <http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html> 16-10-2006

<sup>xv</sup> Onder e-Cultuur worden de nieuwe culturele uitingen verstaan die beïnvloed zijn door digitalisering. Zie hiervoor ook het advies van de Raad voor Cultuur over e-Cultuur: *eCultuur: van e naar i: advies over de digitalisering van cultuur en de implicaties voor cultuurbeleid*. (Raad voor Cultuur, juni 2003) via <http://www.cultuur.nl/files/pdf/advies/200505271117440.rc-2002.4251.3.pdf> 09-06-2006

<sup>xvi</sup> Droombeek, zie [www.droombeek.nl](http://www.droombeek.nl), 16-10-2006

<sup>xvii</sup> Ik wil Marieke de Haan en Adelheid Feryn hartelijk danken voor de verschillende tips en verwijzingen over dit onderwerp kunstwebsites.

<sup>xviii</sup> Zie bijvoorbeeld: "In Deaths Kingdom" via [http://www.mecad.org/hm/prod\\_i/frm\\_ivan.htm](http://www.mecad.org/hm/prod_i/frm_ivan.htm) ; "Making sense of it all" via <http://www.makingsenseofitall.net/> .

Andere sites: <http://www.medienkunstnetz.de/kunstprojekte/#projekt2> ; <http://www.hartware-projekte.de/links.htm> .

<sup>xix</sup> Open Universiteit cursus Webcultuur: creatieve technologie voor internet. Via: <http://www.ou.nl/eCahce/DEF/76/143.html> d.d. 02-11-2006

---

<sup>xx</sup> Unesco *Charter on the Preservation of Digital Heritage* (oktober 2003) Via: [http://portal.unesco.org/en/ev.php-URL\\_ID=17721&URL\\_DO=DO\\_TOPIC&URL\\_SECTION=201.html](http://portal.unesco.org/en/ev.php-URL_ID=17721&URL_DO=DO_TOPIC&URL_SECTION=201.html)

<sup>xxi</sup> Zie hiervoor de brief van de Algemene Rijksarchivaris aan de Programmamanager AdviesOverheid.nl d.d. 3 april 2006, kenmerk C/DD/06/656 via: via <http://www.advies.overheid.nl/webrichtlijnen/> 17-08-2006

<sup>xxii</sup> Capsis, Capsis Presurf , <http://www.capsis.nl>

<sup>xxiii</sup> Archive-It, <http://www.archive-it.org/>

<sup>xxiv</sup> Overheid.nl, Webrichtlijnen voor de overheid, <http://webrichtlijnen.overheid.nl/>

<sup>xxv</sup> Overheid.nl Webrichtlijnen, <http://www.advies.overheid.nl/webrichtlijnen>

<sup>xxvi</sup> Overheid.nl, Archivering van websites, <http://www.advies.overheid.nl/5149/>

<sup>xxvii</sup> Uselab, René Voorburg, Webontwerp richtlijnen voor archivering, [http://www.webarchivering.nl/files/externe\\_docs/Webontwerp\\_Richtlijnen\\_voor\\_Archivering\\_-\\_Uselab.pdf](http://www.webarchivering.nl/files/externe_docs/Webontwerp_Richtlijnen_voor_Archivering_-_Uselab.pdf)

<sup>xxviii</sup> HTTPTrack web site copier, Xavier Roche, <http://www.httrack.com>

<sup>xxix</sup> Heritrix, Internet Archive, <http://crawler.archive.org/>

<sup>xxx</sup> Internet Preservation Consortium, <http://netpreserve.org/about/index.php>

<sup>xxxi</sup> PDF/A, [http://www.adobe.com/nl/aboutadobe/pressroom/pr/sep2005/PDF\\_A\\_200905.pdf](http://www.adobe.com/nl/aboutadobe/pressroom/pr/sep2005/PDF_A_200905.pdf)