

Het belang van het archiveren van websites

Niet alleen papier vergaat, ook websites verdwijnen nu al dagelijks in het niets.

De auteurs bepleiten een grondige aanpak van de archivering daarvan. Zij geven daarbij een voorbeeld van een Nederlands project van het DNPP.

HET GEHEUGEN VAN NEDERLAND verpulvert'. Zo luidde de noodkreet van ruim twintig grote bibliotheken in het voorjaar van 1998. In een advertentie in enkele landelijke dagbladen vroegen deze bibliotheken om honderden miljoenen gulden voor het restaureren, fotograferen en digitaliseren van boeken en kranten. Maar terwijl er reddingsacties op touw worden gezet om dit 'papieren geheugen' voor de toekomst te behouden, verdwijnt tegelijkertijd het 'digitale geheugen'.

Met het 'digitale geheugen' wordt hier niet bedoeld op digitale bestanden; voor de bewaring daarvan is meer aandacht gekomen, zoals de oprichting van het Nederlands Historisch Data Archief aan het einde van de jaren tachtig bewijst. Het 'digitale geheugen' heeft echter wel betrekking op de bouwstenen van het World Wide Web: de websites.

Begin jaren negentig deed het WWW (hier verder als 'web' aangeduid) zijn intrede, maar voorzover bekend worden nog nergens in de wereld websites systematisch gearchiveerd en ontsloten.¹ Wel is hier en daar een bescheiden

begin gemaakt met het archiveren van het web, maar dat neemt niet weg dat er reeds een substantieel deel van dit digitale erfgoed verloren is gegaan. En hieraan lijkt voorlopig nog geen einde te komen. "Internet dreigt een gat in de geschiedschrijving te veroorzaken", aldus Amerikaanse historici op een congres in 1998.² Voor toekomstig onderzoek naar de ontwikkeling van de 'virtuele' wereld van het web en zijn relatie met de 'werkelijke' samenleving, ontbreekt het bronnenmateriaal.

Het web breidt zich gigantisch snel uit. Volgens conservatieve schattingen zouden er elke maand meer dan 20 miljoen pagina's bijkomen. Het aantal sites zou maandelijks met ruim 150.000 toenemen en aan het begin van 2000 73 miljoen hebben bedragen.³ Deze fabelachtige groei voltrekt zich ondanks het feit dat ook veel sites weer verdwijnen: de gemiddelde levensduur van een site wordt geschat op 75 dagen. Zij worden vaak maar op één locatie op het web aangeboden; wanneer de aanbieder er om wat voor reden mee stopt, is de site voor altijd verloren. In dit

ke politici. Van deze laatste categorie zijn er overigens nog niet zoveel. Opvallend is wel dat nogal wat Europarlementariërs die relatief ver van huis opereren, al enige tijd de voordelen van een eigen website inzien.¹⁶

Het is overigens zeer de vraag of de in het digitale archief opgenomen websites en de verschillende versies via het web voor raadpleging en onderzoek kunnen worden aangeboden. Het archiveren van digitale bestanden impliceert per definitie kopiëren. Dit leidt automatisch tot het probleem van het auteursrecht. Een aantal politieke partijen heeft desgevraagd laten weten graag medewerking te verlenen aan dit archiveringsproject. Hoe belangrijk dit fiat ook is, toch zal moeten worden nagegaan in hoeverre er andere rechthebbenden zijn die toestemming moeten geven. Een website is immers een op een geheel eigen wijze vormgegeven verzameling digitale bestanden (bestaande uit tekst-, audio- of beeldmateriaal), waarop allerlei rechten kunnen liggen. Deze problematiek zal ertoe kunnen leiden dat de kopieën in het webarchief niet standaard online worden aangeboden.¹⁷

Archiveringsproject DNPP

Omdat partijsites vroeger of later in het niets dreigen te verdwijnen, is het Documentatiecentrum Nederlandse Politieke Partijen (DNPP) in 2000 samen met de Universiteitsbibliotheek van de Rijksuniversiteit Groningen begonnen met de voorbereiding van hun archivering. Dit project wordt mede gefinancierd door de Rijksuniversiteit Groningen en de stuurgroep Innovatie Wetenschappelijke Informatievoorziening (IWI). In tegenstelling tot de in het artikel genoemde Amerikaanse en Zweedse initiatieven richt het archiveringsproject van het DNPP zich op een specifieke, beperkte categorie websites.

Het gaat hier in de eerste plaats om de sites van de in de Staten-Generaal vertegenwoordigde partijen (en die van hun neveninstellingen – vooral de jongerenorganisaties), maar ook om de partijen die niet in het parlement zitting hebben. Onderzoek naar deze laatste categorie is interessant, omdat wel wordt beweerd dat het web door de lage kosten die aan een site verbonden zijn, de verschillen tussen de gevestigde partijen en nieuwkomers zou verkleinen.¹⁵ In een later stadium komen de sites van provinciale partijen in aanmerking, en die van landelijk



FOTO: ECON VIERRE

dynamische proces van opkomst en ondergang vormen de meeste bestaande sites geen statisch rustpunt. Zij veranderen immers voortdurend: een paar seconden na een bezoek aan een site kan die alweer gewijzigd zijn, doordat de aanbieder informatie heeft toegevoegd, of een bezoeker een bericht heeft achtergelaten.

Politieke partijen op het web

Panta rhei, zou men in navolging van de Griekse filosoof Heraclitus uit de zesde eeuw voor Christus kunnen zeggen. Ook op het web stroomt alles en is alles aan voortdurende verandering onderhevig. Dat geldt ook voor de websites van politieke partijen. De Nederlandse politieke partijen waren redelijk vroeg op het web te vinden. In januari 1994 was GroenLinks de eerste die met een website begon.⁴ In een beweging van links naar rechts volgden de andere partijen: de PvdA in november 1994, D66 en het CDA medio 1995, en de VVD in het voorjaar van 1997. Hekensluiter was de SGP in de herfst van 2000. Sinds het begin van hun aanwezigheid op het web hebben de meeste partijen al twee, drie keer hun site compleet gerestylet. Van de oudere versies is momenteel niets meer over. De eerste stappen van de Nederlandse politieke partijen op het web zijn dan ook niet meer digitaal te reconstrueren.⁵ Aan de vooravond van de Tweede-Kamerverkiezingen van mei 1998 hadden alle partijen – behalve de SGP – een site. Veel aanloop hadden zij niet tijdens de verkiezingscampagne. Naar schatting bezochten maximaal 100.000 personen de maand voor de stembusdag een partijsite.⁶ In de toekomst zullen de sites echter ongetwijfeld belangrijker worden voor partijen. Omdat leden en kiezers in toenemende mate toegang krijgen tot internet, zullen de sites een steeds grotere rol gaan spelen in de informatievoorziening aan de achterban. Dit blijkt nu bijvoorbeeld al uit de veranderende inhoud van de partijbladen. Deze ontwikkelden zich meer tot magazines, waarin niet langer de wat saaie

partijorganisatorische informatie wordt vermeld (spreekbeurten, agenda's van partijbijeenkomsten en dergelijke). Deze gegevens worden tegenwoordig op de website gezet. Hoewel de meeste sites heden ten dage nogal *top down* gericht zijn, dus vooral bedoeld om informatie te verspreiden, mag verwacht worden dat het interactieve aspect de komende tijd meer gewicht zal krijgen.⁷ In een tijd waarin partijen grote moeite hebben om op traditionele wijze leden aan zich te binden en deze te activeren (het ledental van de grote partijen – PvdA, CDA, VVD en D66 – loopt immers hard terug), zouden digitale vormen van participatie wel eens nieuwe mogelijkheden kunnen bieden (zonder dit nu meteen als panacee te willen zien). Het CDA bijvoorbeeld speelt hierop in bij de opstelling van het programma voor de Tweede-Kamerverkiezingen van 2002, door sitebezoekers in de gelegenheid te stellen hieraan een bijdrage te leveren. Er is dus alle reden om de digitale presentatie van partijen te documenteren of – zo men wil – te archiveren, net zoals dat bij de gedrukte publicaties gebeurt. Onderzoekers van velerlei discipline (historici, sociologen, politicologen, communicatiewetenschappers) en journalisten zullen hiermee hun voordeel kunnen doen.

Het Documentatiecentrum Nederlandse Politieke Partijen, DNPP (zie kader) is begonnen aan een project websites van politieke partijen te archiveren. In verband hiermee wordt eerst een globaal overzicht gegeven van min of meer vergelijkbare initiatieven elders.

Internationale initiatieven

In het midden van de jaren negentig begon men hier en daar na te denken over het downloaden en bewaren van websites. Het meest grootschalige project werd ontwikkeld door de Amerikaanse computerprogrammeur Brewster Kahle. Hij is de oprichter van het in San Francisco gevestigde, zogeheten "Internet Archive", dat officieel sinds de

zomer van 1996 het internet archiveert: van nieuwsgroepen tot homepages. Hiervoor wordt gebruikgemaakt van web crawling robots: programma's die via externe links van sites andere sites opzoeken en ze allemaal in hun geheel downloaden. Op deze wijze wordt een momentopname van het internet gemaakt. In september 2000 had het circa een miljard ongespecificeerde webpagina's verzameld en ongeveer 16 miljoen nieuwsberichten – overwegend html-bestanden overigens.⁸

In dezelfde tijd werd ook in Australië een eerste bescheiden aanzet gegeven tot het bewaren van websites, in het kader van het door de National Library opgezette project Pandora (Preserving and Accessing Networked Documentary Resources in Australia). Belangrijk geachte Australische online publicaties worden in het kader van Pandora gearcheveerd. Onder meer worden enkele websites van Australische politieke partijen bewaard.⁹

De Koninklijke Bibliotheek van Zweden begon in 1997 met het zogeheten "The Kulturarw3 Project". Doel is zoveel mogelijk te archiveren van het Zweedse deel van internet, dat wil zeggen alle url's met de extensie .se. Hier van zijn sindsdien enkele *snap-shots* gemaakt, momentopnamen waarbij in totaal ongeveer 56.000 websites zijn gearcheveerd. Het digitale archief is nog niet voor het publiek toegankelijk.¹⁰ Het Zweedse project heeft inmiddels bij de Nationale Bibliotheek in Finland navolging gevonden.

In Nederland richten twee instanties zich op de archivering van delen van het internet. De Koninklijke Bibliotheek is begonnen met de inrichting van het Depot van Nederlandse Elektronische Publicaties (DNEP). Hierin zullen naast *offline* digitale publicaties als cd-roms in de toekomst ook *online* publicaties als elektronische tijdschriften, boeken en artikelen worden opgenomen. Het Nederlands Uitgeversverbond heeft inmiddels ingestemd met een regeling waarbij deze publicaties bij het DNEP worden gedeponeerd. Verder kunnen ook bepaalde webdocumenten worden ondergebracht in het Depot; complete sites worden niet verzameld, noch archieven van discussielijsten.¹¹ Voor dat laatste heeft het Internationaal Instituut voor Sociale Geschiedenis (IISG) grote belangstelling.

Actiegroepen en sociale bewegingen ontdekten al vroeg het internet als een goedkope en effectieve manier om hun standpunten te verspreiden en steun te mobiliseren. Zoals het IISG vroeger de pamfletten en brochures van deze groeperingen verzamelde, collectioneert het Instituut nu hun berichten in nieuwsgroepen. In 1995 werd onder de naam "Occasio" begonnen met het archiveren van bijna duizend internetnieuwsgroepen van de Association for Progressive Communications. Het archief, dat onder meer veel materiaal bevat over de burgeroorlogen in het voormalige Joegoslavië en de opstand tegen Soeharto in Indonesië, telde begin 2000 ongeveer een miljoen berichten.¹² Hoewel er dus meer aandacht voor het probleem komt, is het duidelijk dat de archivering van websites nog in de kinderschoenen staat. De hierboven vermelde projecten verkeren alle feitelijk min of meer nog in een beginstadium. Bovendien richt geen van hen zich in het bijzonder op de archivering van websites, behalve het Zweedse. Dit project kent echter een aantal nadelen. Allereerst is het nogal grofmazig opgezet: het beoogt slechts één of twee keer per jaar zoveel mogelijk sites te archiveren. Hierdoor gaat er in de tussentijd alsnog veel informatie van belangrijke sites verloren. Daarnaast vindt er geen inhoudelijke ontsluiting plaats van het opgeslagen materiaal. Dit laatste is evenmin het geval met de door het Internet Archive gearcheveerde sites.

Moeilijkheden en mogelijkheden

De korte levensduur van soft- en hardware zal in de toekomst leiden tot problemen bij de opslag, het beheer en de beschikbaarstelling van de gearcheveerde sites. Ook wanneer de apparatuur, de programma's en de opslagformaten van tekst, audio- en videomateriaal verouderd zijn, moeten de gearcheveerde sites raadpleegbaar blijven. Dit betekent dat de digitale archieven periodiek moeten worden overgezet en geconverteerd naar een nieuwe generatie informatiedragers en soft- en hardwaresystemen. Sommigen achten emulatie een alternatief. Of dit alles kan gebeuren zonder dat de authenticiteit of integriteit van de digitale documenten (enigszins) aangetast wordt, is gezien de huidige stand van de techniek twijfelachtig.¹³

In 1996 verscheen in de Verenigde Staten het rapport *Preserving Digital Information* van de Commission on Preservation and Access en the Research Libraries Group. Hierin werd ervoor gepleit een decentraal netwerk op te zetten van digitale archieven, die de collectionering van digitale objecten als bijvoorbeeld websites, hun bewaring en beschikbaarstelling als taak hebben. Het rapport benadrukte het belang van een decentrale opzet, want "a distributed structure... places archival responsibility with those who presumably care most about and have the greatest understanding of the value of particular digital information objects".¹⁴

Het archiveringsproject van het DNPP past geheel in deze visie. Gezien zijn doelstelling en inhoudelijke kennis ligt het voor de hand dat het DNPP zich over de websites van de Nederlandse politieke partijen ontfermt. Het aandeel van het DNPP in het vastleggen en bewaren van het web voor de toekomst is hiermee zeer bescheiden. Wel levert het een grote bijdrage aan de archivering van de virtuele politieke cultuur van Nederland.

Noten

- 1 Zie C. Caey, 'The Cyberarchive: A Look at the Storage and Preservation of Web Sites', in: *College & Research Libraries*, juli 1998, 304-310.
- 2 *De Volkskrant*, 30 januari 1999.
- 3 Internet Software Consortium; <www.isc.org/ds/WWW-200001/report.html>; bezocht op 21 juli 2000.
- 4 S. Ward en G. Voerman, New media and new politics. Green Parties, Intra-party democracy and the Potential of the Internet (an Anglo-Dutch Comparison), in: *Jaarboek 1999 Documentatiecentrum Nederlandse Politieke Partijen*, Groningen, 2000, 192-215.
- 5 G. Voerman and J.D. de Graaf, 'De web-sites van de Nederlandse politieke partijen, 1994-1998', in: *Jaarboek 1997 DNPP*, Groningen, 1998, 238-269.
- 6 G. Voerman, 'Elektronisch folderen: de digitale campagne', in: Ph. van Praag en K. Brants, *Tussen beeld en inhoud. Politiek en media in de verkiezingen van 1998*, Amsterdam, 2000, 206.
- 7 G. Voerman, 'Dutch Political Parties on the Internet', in: *ECPR-News*, 10 (1998), nr. 1, 8-9.
- 8 Zie onder meer B. Kahle, 'Preserving the Internet', in: *Scientific American*, 1997, nr. 3; M. Cunningham, 'Brewster's millions', in: *The Irish Times* (webversie), 27 januari 1997. Zie voor de website van het Internet Archive <www.archive.org>.
- 9 'Electronic Publications and the Survival of Information'; zie <widow.nla.gov.au/nla/staffpaper/mjcwala.html>, 2 november 1998.
- 10 Zie A. Arvidson en F. Lettenström, 'The Kulturarw3 Project - the Swedish Royal Web Archive', in: *The Electronic Library*, 16 (1998), 2 (April), 105-108; zie ook <www.kulturarw3.kb.se/html/projectdescription.html>.
- 11 'Symposium "Het eeuwige leven"', in: *Historia & Informatica*, 5 (1998), 1 (apr.), 6-7; G.C. Nordermeer, 'Depot van Nederlandse elektronische publicaties', in: *HG-Nieuws*, 4 (1997), 1 (mrt.), 4-5; T. Noordermeer, 'Depot van Nederlandse Elektronische Publicaties', in: *Informatie Professional*, 1998, nr. 2, 22-24; zie ook *idem*, 3 (1999), nr. 3, 11.
- 12 J. Quast, 'OCCASIO Digital Social History Archive', in: *Historia & Informatica*, 5 (1998), nr. 2; zie ook *idem*, 'IISG: Occasio Archief nu toegankelijk via het Web', in: *idem*, 6 (1999), 3 (dec.), 1; en *idem*, 'Het internetarchief van het IISG', in: *Archievenblad*, 104 (2000), 7 (sep.), 16-17. Zie voor meer informatie: <www.iisg.nl/occasio>.
- 13 Zie bijvoorbeeld J.S. Mackenzie Owen, *Kennis in veelvoud*, Amsterdam, 1998, 24-25; G. Hodge and B.C. Carroll, *Digital Electronic Archiving: The State of the Art and the State of the Practice*. A Report Sponsored by: Inter-



national Council for Scientific and Technical Information, Information Policy Committee and CENDI, Oak Ridge, 1999, 7-8.

14 *Preserving Digital Information*. Report of the Task Force on Archiving of Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group, z.pl., 1996, 21; zie ook Hodge and Carroll, *Digital Electronic Archiving*, 8, 10.

15 M. Margolis, D. Resnick & Ch. Tu, 'Campaigning on the Internet: Parties and Candidates on the World Wide Web in the 1996 Primary Season', in: *Press/Politics*, 2 (1997), nr. 1, 59-78.

16 Het archiveren van de websites van de partijen past geheel in de Nederlandse traditie om de vele publieke uitingen van de politieke partijen te bewaren en toegankelijk te maken. De inbreng van hun vertegenwoordigers in de Eerste en Tweede Kamer is vastgelegd in de *Handelingen*. Beeld- en geluidsmateriaal van de partijen is gedeponeerd op het Nederlands Audiovisueel Archief. Het DNPP verzamelt, catalogiseert en ontsluit de vele publicaties van partijen (zoals beginsel- en verkiezingsprogramma's, brochures, ledenbladen, tijdschriften van wetenschappelijke bureaus) en over partijen (beschouwingen van journalisten, politicologen, historici, etcetera). Het webarchiveringsproject is een logisch vervolg op deze traditionele taak van het Documentatiecentrum.

17 Zie *Auteursrechtelijke aspecten van preservatie van elektronische publicaties*, Amsterdam, 1998; M. de Cock Buning, 'Europese Richtlijn Auteursrecht in de Informatiemaatschappij. Rechten versterkt & beperkingen beperkt', in: *Informatie Professional*, 3 (1999), nr. 7/8, 33-35, 37; Hodge and Carroll, *Digital Electronic Archiving*, 8.

Gerrit Voerman is hoofd van het Documentatiecentrum Nederlandse Partijen van de Rijksuniversiteit Groningen. André Keyzer is stafmedewerker elektronische dienstverlening van de Universiteitsbibliotheek Groningen. Frank den Hollander is voorlichtingsfunctionaris van de Groningse Universiteitsbibliotheek.

