PETER BOOT

Beweging rond webarchivering

Webpagina's moeten bewaard blijven voor later onderzoek, vinden steeds meer mensen. *e-data&research* rapporteert over drie hoopgevende initiatieven op dat gebied.

In ruim een decennium is het World Wide Web uitgegroeid tot een medium dat elk aspect van onze levens beïnvloedt. Het web is informatiemedium, publicatieomgeving, winkel, manier van leven en huiskamer. Webpagina's zijn daarom een belangrijke bron voor: historici, letterkundigen, politicologen, en eigenlijk iedereen die zich bezighoudt met menselijk gedrag. Het is belangrijk dat webpagina's bewaard blijven voor latere onderzoekers. Er zijn op dat gebied in elk geval drie hoopgevende initiatieven: van de Koninklijke Bibliotheek (KB), Virtual Knowledge Studio (VKS) en het European Archive.

Het European Archive (EA) presenteerde zichzelf op het symposium 'Avoiding the Digital Memory Loss' op 27 september in Amsterdam. Het EA is een zusterinstelling van het (Amerikaanse) Internet Archive. De oprichter daarvan, Brewster Kahle, sprak op het symposium over zijn visioen van 'Universal access to human knowledge'. Het Internet Archive archiveert niet alleen webpagina's, maar ook boeken, films en muziek. De *Wayback Machine* van het Internet Archive kan laten zien hoe een website er in het verleden uitzag.

EUROPEAN ARCHIVE

Het archief van het European Archive bij xs4all

Het EA zal een eigen archief creëren, maar ziet zich ook als een katalysator voor ontwikkelingen elders. De directeur van het EA, Julien Masanès, zegt tegen *e-data&research* zowel een *competence center* als een archief te willen zijn. Voor het eigen webarchief zal de oppervlakte van het web worden doorzocht, zonder specifieke voorkeuren; samen met anderen zullen diepgaande deelarchieven worden gemaakt. Op het moment is het EA gevestigd in Nederland, onder meer vanwege het begrip voor internetkwesties. Als bewijs van dat begrip kan gelden dat web provider *xs4all* de technische behuizing voor het archief levert. De Koninklijke Bibliotheek is, vertelt Marcel Ras, aan een inhaalrace begonnen (*e-data&research* juni 2006). Ras is projectleider webarchivering bij de KB. De KB heeft veel tijd besteed aan het inrichten van een e-depot voor elektronische publicaties. In de toekomst zal het depot ook gebruikt worden voor het archiveren van internetsites – in eerste instantie selectief: zo'n duizend tot drieduizend sites in vooral de overheidssector. Later zullen ook volledige 'crawls' van het Nederlandse deel van het net worden gedaan.

De KB wil zorgvuldig omgaan met de juridische aspecten hiervan. Volgens een onderzoek van de Leidse universiteit maakt de Nederlandse wet het lastig om websites van derden te archiveren en toegankelijk te maken, bijvoorbeeld op grond van auteursrechten en bescherming van persoonsgegevens. 'Wat het Internet Archive doet is in Nederland strikt genomen misschien illegaal', bevestigt Ras. 'Maar als particuliere instelling kunnen zij natuurlijk anders opereren dan een overheidsinstelling zoals wij zijn'. De KB overweegt contracten af te sluiten met de te archiveren instellingen, of misschien een *opt-out* optie te gaan hanteren. Een andere mogelijkheid is de gearchiveerde data toegankelijk te maken binnen het KB-gebouw.

Een ander probleem bij het archiveren van websites zijn de vele bestandsformaten, voor afbeeldingen: Flash, pdf, Officebestanden en zo meer. Pdf is voor het e-depot goed onderzocht, maar voor andere bestandstypen moet nog veel gebeuren. Daarbij werkt de KB samen met andere instellingen. Ras ziet sowieso graag een nationaal samenwerkingsverband voor webarchivering tot stand komen. Sommige instellingen, zoals het Documentatiecentrum Nederlandse Politieke Partijen, archiveren al delen van het Nederlandse web. Ras werkt ook graag samen met de Virtual Knowledge Studio (VKS), als spreekbuis van de onderzoekersgebruikers. De VKS is een *e-research* onderzoeksgroep van de KNAW. Webarchivering is een VKS-speerpunt. De VKS is, uit hoofde van haar missie, geïnteresseerd in science studies, en heeft ervaring opgedaan met het archiveren van wetenschappelijke websites. Paul Wouters, hoofd van VKS, meldt aan *e-data&research* te willen werken aan de opzet van een breder toepasbare webarchiveringsservice. Onderzoekers zouden dan gebruik kunnen maken van een interface dat op specifieke onderzoeksvragen en onderzoeksstijl is aangepast. Wouters wil daarbij samenwerken met de KB, met DANS, en met het European Archive. Bij de workshop die de VKS in september organiseerde zijn verschillende bruikbare hulpmiddelen besproken.

Wat opvalt is dat alle partijen het belang van samenwerking benadrukken. Ook dat stemt hoopvol. Het gaat uiteindelijk om een klus die het vermogen van elke organisatie afzonderlijk te boven gaat.

www.kb.nl/hrd/dd/dd_projecten/projecten_webarchivering.html
www.europarchive.org/
www.virtualknowledgestudio.nl/

Beperkingen van vrijwillige onderzoekpanels

MARTIJN DE GROOT

'Vraag je eerst af wat je wil weten'

'Als je met een panel van vrijwilligers werkt, kijk dan of zo'n panel wel de goede weg is om je onderzoeksvragen te beantwoorden'. Dat zegt Ineke Stoop, hoofd van de afdeling Informatievoorziening en Automatisering van het Sociaal-Cultureel Planbureau.

Stoop geeft daarmee de belangrijkste uitkomst weer van het symposium *Access panels en Online Onderzoek* over internetpanels, op 16 september op initiatief van DANS gehouden in Amsterdam. Het moment van de bijeenkomst was goed getimed. Een maand eerder waren de resultaten gepresenteerd van het Nederlands Online Panel Vergelijkingsonderzoek (NOPVO) dat, georganiseerd vanuit de marktsector, tot kritische conclusies kwam over de kwaliteit van onderzoek met internetpanels. Ineke Stoop zelf leverde op het symposium een *keynote* af waarin ze benadrukte dat er veel verschillende soorten marktonderzoek, en daarbinnen panelonderzoek bestaan - ieder met hun eigen voor- en nadelen, ook wat betreft de betrouwbaarheid van de uitkomsten en het mogelijke effect van de non-respons. 'Voor de kritische conclusies van het NOPVO en van het hele symposium is niet de webbasis van de panels beslissend,' zegt ze dan ook terugkijkend, 'maar het vrijwillige karakter. Je kan ook panels organiseren met *face-to-face* interviews of schriftelijke enquêtes. Maar met een steekproef van vrijwilligers krijg je altijd een afwijking doordat het om een specifieke groep gaat. En daarvoor kan je niet controleren omdat je niet precies weet in welk opzicht die groep afwijkt'.

Dat was inderdaad een van de belangrijkste conclusies van het NOPVO. In marktonderzoek wordt veel belang gehecht aan de respons, waarmee dan het percentage beantwoorde vragenlijsten wordt bedoeld binnen het panel. De respons heeft echter wel invloed op de efficiency van het onderzoek maar niet op de uitkomsten, concludeerde NOPVO. Die worden sterk gekleurd door de selectie die voorafgaand aan het panelonderzoek heeft plaatsgevonden: wie neemt er wel en niet deel aan het panel. 'Het grootste deel van de populatie zit niet in de panels', aldus de NOPVO-presentatie. 'En die geven die àndere antwoorden'. Wie er wel in zitten zijn bovendien vaak mensen die aan meerdere panels deelnemen, niet zelden vanwege de interessante beloning. 'Ons vermoeden van een flinke overlap werd bevestigd door het onderzoek,' zegt achteraf onderzoeker Robert van Ossenbruggen (ProCression), die de NOPVO-presentatie in Amsterdam mede voor zijn rekening nam. Hij voerde ook het onderzoek uit, samen met Ted Vonk (Onderzoekpaleis) en Pieter Willems (Millward Brown).

Lastige conclusies, zo wordt algemeen ingezien in de marktsector, waar uit kostenoogpunt veel met de z.g. *volunteer panels* wordt gewerkt. Al maakt het natuurlijk verschil of je 'onderzoek doet naar toetjes en wasmiddelen dan wel naar technische producten zoals digitale camera's', relativeert Van Ossenbruggen. Maar het is ook een zelfkritische dus dappere conclusie, benadrukt Stoop die naar eigen zeggen 'naar adem zat te happen' toen ze de presentatie van het NOPVO een maand eerder bijwoonde.

Zijn er ook oplossingen? NOPVO komt tot een soortgelijke conclusie als Stoop: 'Vraag je af of je dat wat je wil weten nauwkeurig genoeg kan meten met volunteer panels'. En: 'Gebruik online panels vooral voor trendmetingen en *concept testing*'. Van Ossenbruggen pleit desgevraagd ook voor beter panelmanagement: 'Er moet een standaard komen. Daar wordt internationaal al aan gewerkt, maar onze studie laat weer zien dat het echt nodig is'.

Stoop wijst nog eens op het vrijwillige karakter als oorzaak van de problemen: 'Als je een *nette steekproef* trekt kan je daarmee via internet hele betrouwbare resultaten bereiken – ook in een panel'. Aan zo'n opzet wordt volgens de methodologe door sommige onderzoekers ook hard gewerkt. De steekproef voor CentERdata van de Universiteit Tilburg bijvoorbeeld, dat ook op het symposium werd gepresenteerd als een 'voor academische onderzoekers vrij toegankelijk platform' dat drijft op een panel van vijfduizend huishoudens. Die wordt 'op traditionele wijze' getrokken samen met het CBS, met extra aandacht voor moeilijke doelgroepen zoals ouderen en allochtonen die bijvoorbeeld apparatuur in bruikleen krijgen en toegang tot het internet. Een hele vooruitgang ten opzichte van de *volunteer* opzet, aldus Stoop. Maar toch: 'Als je ouderen aan de computer gaat zetten wéét je dat dat *een bepaald soort* ouderen is. En zijn allochtonen thuis op internet, dan zijn het hoog opgeleide allochtonen'.

www.onderzoekpaleis.nl/NOPVO presentaties.htm
www.dans.knaw.nl/nl/presentaties_access_panles/
ww.scp.nl/english/publications/books/9037702155.shtml