

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.

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 internetkwesaties. 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 onderzoekers-gebruikers. 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 webarchiveringservice. 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/