



**Johan van der Knijff**

Onderzoeker Digitale Duurzaamheid  
bij de Koninklijke Bibliotheek

# WEBARCHEOLOGIE REDT EERSTE NEDERLANDSE WEBINDEX VAN CD-ROM

In de nieuwe rubriek 'KB Onderzoekskroniek' beschrijven medewerkers van de afdeling Onderzoek van de Koninklijke Bibliotheek hun resultaten, trends & vondsten.

De Koninklijke Bibliotheek is onlangs begonnen met webarcheologie, oftewel het reconstrueren van oude websites. Daarvoor worden cd-roms, harde schijven en floppy's met de inhoud van websites opgespoord.

**H**oe zag het Nederlandse web er vóór 2004 uit? Welke websites waren toen populair? Welke van die sites zijn intussen verdwenen en welke bestaan nog steeds? Net als de meeste andere nationale bibliotheken heeft de Koninklijke Bibliotheek (KB) een webarchief, waarin de belangrijkste websites van Nederland periodiek worden gearchiveerd. Omdat we pas in 2007 hiermee zijn begonnen met webarchivering, is het webarchief nu maar beperkt bruikbaar voor het beantwoorden van vragen over de vroege geschiedenis van het Nederlandse web.

## Webarcheologie

Daarom zijn we recent gestart met *webarcheologie*: het reconstrueren van oude websites door offline dragers zoals cd-roms, harde schijven of floppy's met de inhoud van de websites op te sporen. Door de gegevens op de dragers veilig te stellen, kunnen we de websites reconstrueren en alsnog opnemen in het webarchief.

Een mooi voorbeeld hiervan is *NL-menu*, de eerste Nederlandse webindex die in 1992 werd opgericht door SURFnet, universiteiten en de KB. Vanaf midden jaren negentig werd *NL-menu* door de KB beheerd, maar in 2004 stopte die hiermee, waarna de site in vergetelheid is geraakt.

## Reconstructie

Kort voor het einde in 2004 hebben de toenmalige beheerders de inhoud van *NL-menu* drie keer op een cd-rom gebrand. Met deze

schijfjes (die nog ergens in een stoffige kast bleken te staan) hebben we geprobeerd een reconstructie te maken van de site zoals die er in 2004 uitzag.

Hiervoor hebben we eerst de cd-roms uitgelezen. In eerste instantie leverden *alle* schijfjes hierbij leesfouten op. Geen verrassing, want zelfgebrande cd-roms zijn erg vergankelijk. Met speciale data-recoverysoftware (*ddrescue*) lukte het gelukkig om de data van één cd-rom volledig te redden. Door vervolgens de geredde data uit te serveren met een webserver konden we een lokaal werkende reconstructie van de site maken. Hier hebben we als laatste stap een 'webcrawler'-computerprogramma op losgelaten. Dit programma doorloopt alle onderdelen van de site en schrijft ze weg in een vorm die kan worden ingelezen in het webarchief.

## Unieke bron

Met deze geslaagde reddingspoging is *NL-menu* veiliggesteld, en hebben (web)historici ook in de toekomst toegang tot deze unieke bron van informatie over de (relatief) vroege geschiedenis van het Nederlandse web. Hoog tijd om onszelf een nieuwe competentie aan te leren: webarcheologie. <



## Verder lezen

Meer informatie over de reddingspoging van *NL-menu* is te vinden in onderstaande blogposts:

> Resurrecting the first Dutch web index: *NL-menu* revisited, [tinyurl.com/yc4yh423](http://tinyurl.com/yc4yh423)

> Crawling offline web content: the *NL-menu* case, [tinyurl.com/yaloasvf](http://tinyurl.com/yaloasvf).

Omdat het webarchief alleen toegankelijk is vanuit de leesalen van de KB, is ook een publiek toegankelijke versie van het gereconstrueerde *NL-menu* gemaakt: [www.kbresearch.nl/nl-menu/nl-menu/](http://www.kbresearch.nl/nl-menu/nl-menu/).