

Webarchivering bij de Koninklijke Bibliotheek

Nederland als internetpionier

Niet veel mensen realiseren zich dat Nederland een pionier in de geschiedenis van het internet is. Het .nl domein was het eerste nationale topleveldomein buiten de VS en werd vastgelegd in april 1986. De derde website van de wereld werd gepubliceerd in Nederland (NIKHEF) nadat de eerste website van het CERN in Zwitserland en die van Stanford in de VS online waren gegaan. Omdat twee website samen een lijn vormen en drie een web, kunnen we stellen dat het wereldwijde web (WWW) werd geboren op het moment dat de Nederlandse site online kwam. Ooit was één derde van het wereldwijde web Nederlands. Nu telt het .nl domein 5,8 miljoen sites en is daarmee nog steeds één van de grootste nationale domeinen van de wereld.

Deze rijke digitale historie is voor de Koninklijke Bibliotheek (de nationale bibliotheek van Nederland) aanleiding geweest om in 2007 te beginnen met het archiveren van het Nederlandse web door een webarchiveringsteam. De KB brengt mensen en informatie samen door de bibliotheekcollectie Nederland zichtbaar, houdbaar en bruikbaar te maken. Het webarchief is een belangrijk onderdeel van deze nationale bibliotheekcollectie. Het streven is een zo representatief mogelijk deel van het Nederlandse internet te bewaren. De missie is deze digitale collectie zo goed mogelijk samen te stellen, duurzaam op te slaan en zo open, compleet en authentiek mogelijk ter beschikking te stellen voor iedereen. Het belang van het behoud van websites als digitaal erfgoed is in 2003 wereldwijd erkend in het Unesco Charter on the Preservation of the Digital Heritage. Ook de KB signaleert dat websites als digitaal erfgoed verloren dreigen te gaan en dat het bewaren daarvan voor gebruik door huidige en toekomstige onderzoekers bijzonder urgent is.

Impact

De KB is de enige organisatie in de wereld die specifiek websites van en over Nederland selecteert en bewaart in haar collectie. De KB-webcollectie bevat daarom een unieke selectie van digitaal bronmateriaal over de Nederlandse internetcultuur die nergens anders ter wereld in deze vorm wordt verzameld en bewaard, of kan worden geraadpleegd. In totaal zijn 13.000 websites in de collectie opgenomen en de jaarlijkse groei is 1.000 sites. Daarmee is dit de grootste digitale webcollectie van Nederland: alle Nederlandse webarchieven samen bewaren ongeveer 16.000 sites. Door de innovatieve wijze van archiveren bewaart de KB veel sites die niet in het bekende Internet Archive te vinden zijn. De complete collectie omvat meer dan 31 TB: dit is groter dan de vergelijkbare collectie van de British Library (het Open UK webarchive). En dat terwijl het .uk domein twee keer zo groot is als het Nederlandse en men bij de British Library al in 2004 is begonnen met deze collectie.

Toekomstplannen

Het afgelopen jaar is de KB met drie innovatieve en baanbrekende projecten gestart om bedreigd digitaal erfgoed veilig te stellen voor de toekomst. Het eerste is het project Webarcheologie, waarbij “webincunabelen” of de “wiegedrukken van het web” (de vroegste websites op het Nederlandse web die nog online zijn) worden opgespoord en bewaard. Het tweede is het project Online Nieuws, waarbij nieuws- en krantenwebsites worden gearchiveerd. En ten derde loopt er een project “maatschappijkritische websites” waarbij Nederlandse sites die in de toekomst een bron of spiegel van ons post-truth tijdperk kunnen zijn, in samenwerking met wetenschappers en andere deskundigen worden geselecteerd en gearchiveerd. Deze collectie bevat onder meer unieke sites van een bont gezelschap van complotdenkers, ufo-onderzoekers, eindtijddenkers, maar ook complotbestrijders en wetenschappelijke sites over het fenomeen ‘nepnieuws’. Hierbij worden voor het eerst innovatieve technieken als online linkanalyse gebruikt om relevante en belangrijke websites op te sporen. Ook kan door deze techniek voor het eerst de “websfeer” rondom websites worden vastgelegd voor onderzoekers van de toekomst. Dit project is voor zover bekend uniek in de wereld: nergens anders is een dergelijk scherp tijdsbeeld in een webcollectie vastgelegd.

Webarchiveringsteam Koninklijke Bibliotheek

Het webarchiveringsteam van de KB bestaat uit drie professionals van de afdeling Collecties en de afdeling Digitale Duurzame Toegang met ieder hun eigen expertise. Peter de Bode is dienstcoördinator webarchivering en verantwoordelijk voor de selectie en invoer van de sites in de collectie. René Voorburg is als crawl-engineer verantwoordelijk voor de duurzame opslag van deze collectie voor de toekomst. Kees Teszelszky onderzoekt de verbetering van webarchivering en de webcollectie en zal vanaf 1 februari 2018 curator digitale collecties zijn.

Meer informatie over het webarchief en webarchivering in de Koninklijke Bibliotheek: <https://www.kb.nl/organisatie/onderzoek-expertise/e-depot-duurzame-opslag/webarchivering>