E-DATA & RESEARCH

Jaargang 10 | nummer 2

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: CentERdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het RIVM.

INHOUD

Scan deze QR code met een smartphone om de website van E-data te bezoeken.
www.edata.nl

Hoe maak je gearchiveerde websites bruikbaar voor de wetenschap?

Nationale webarchief onderzocht door WebART

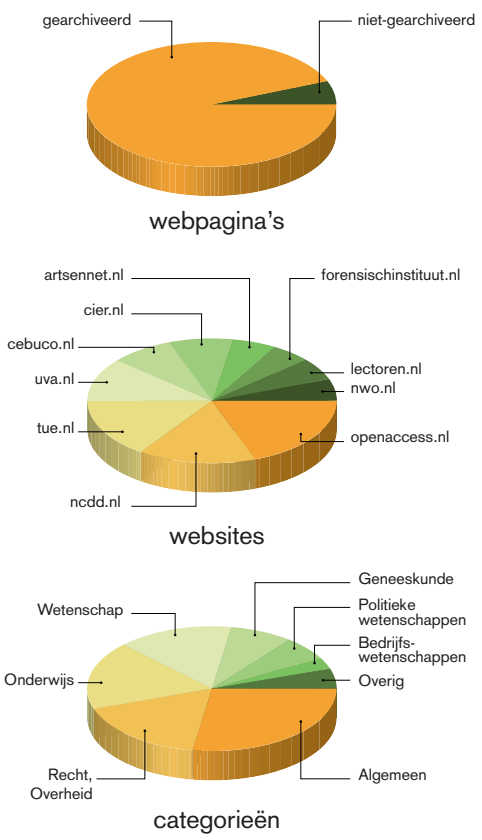
Het eerste grote onderzoeksproject in Nederland naar gebruik van gearchiveerde Nederlandse websites als primaire bron voor onderzoek sluit binnenkort de boeken. WebART-promovendus Hugo Huurdeman blikt terug.

Steven Claeyssens

Dit jaar ronden de laatste CATCH-projecten (Continuous Access to Cultural Heritage) hun werkzaamheden af en dus zet ook WebART (Web Archive Retrieval Tools) er een punt achter.
WebART was een samenwerking tussen de Universiteit van Amsterdam (UvA), het Centrum Wiskunde en Informatica (CWI) en de Koninklijke Bibliotheek (KB). Het WebART-team lichtte als eerste het Nederlandse nationale webarchief grondig door. Ze gingen daarbij na hoe zo'n heterogeen en omvangrijk *born-digital* archief voor onderzoeksdoeleinden bruikbaar kan zijn en bruikbaarder kan worden gemaakt.

10.000 websites

De KB archiveert sinds 2007 een immer groeiende selectie van Nederlandse websites. Op 1 januari van dit jaar stond de teller op 10.000 sites die met enige regelmaat worden *geharvest*. Het belang van dit *born-digital* archief voor onderzoek naar Nederlandse cultuur en samenleving zal naarmate de jaren verstrijken onvermijdelijk een steeds prominentere plaats opeisen. WebART onderschrijft dit belang en trok op onderzoek uit. Huurdeman: "In het WebART-project hebben we gekeken naar de onderzoeksvragen die wetenschappers aan webarchieven zouden willen stellen. Via een intensieve samenwerking met nieuwe media-onderzoekers hebben we vervolgens zoek- en onderzoekstools ontwikkeld die complexe onderzoekstaken kunnen ondersteunen. Denk bijvoorbeeld aan de initiële exploratie van het archief, het definiëren van een dataset en de analyse daarvan. Hiervoor was onderzoek nodig naar schaalbare extractie- en analysemethoden en naar bruikbare interfaces voor verschillende zoekstadia." Zo bouwde het team onder meer WebARTist, een interface waarmee onderzoekers op verschillende manieren het webarchief kunnen verkennen en bevragen.



Op basis van de zoekterm 'onderzoeksdata' toont WebARTist verschillende resultaten, waaronder deze grafieken. De bovenste grafiek laat de verhouding tussen de gearchiveerde en niet-gearchiveerde webpagina's zien, de middelste toont de belangrijkste websites voor deze zoekterm en de onderste grafiek vat de categorieën van de gevonden pagina's samen. De WebARTist-toolset biedt een veelheid aan mogelijkheden voor exploratie, analyse en visualisatie van de inhoud van het KB-webarchief. *credits WebART*

Ongearchiveerde websites

"Doordat webarchieven van nature incompleet zijn, vroegen wetenschappers ook om contextualisatie over wat er wel en niet in het archief zit. Dit heeft geleid tot verder onderzoek waarin we niet-gearchiveerde webinhoud hebben blootgelegd en gereconstrueerd." Zo slaagden Huurdeman en zijn mede-onderzoekers erin een fors aantal niet-gearchiveerde sites te identificeren op basis van verwijzingen in de vorm van URL's in het wel-gearchiveerde deel. Meer nog, door de afzonderlijke woorden uit deze URL's en de bijbehorende linkteksten te distilleren, maakten ze dit niet-gearchiveerde deel van het web tot op zekere hoogte toch vindbaar en daarmee ook onderzoekbaar.
"Deze informatie integreren we in de WebART-toolset. Helaas kan de toolset momenteel door auteursrechtelijke beperkingen nog niet volledig online worden aangeboden, maar de wens vanuit het projectteam om dit te bereiken, is er zeker."
webarchiving.nl

Principeakkoord open access VSNU en Elsevier

De Vereniging van Universiteiten (VSNU) en Elsevier hebben een principeakkoord bereikt waardoor Nederlandse wetenschappers toegang blijven houden tot de wetenschappelijke artikelen van Elsevier.

"Door deze overeenkomst," aldus prof. Gerard Meijer, hoofdonderhandelaar namens de VSNU en voorzitter van de Radboud Universiteit Nijmegen, "houden wetenschappers toegang tot Elseviertijdschriften en het biedt ze de mogelijkheid om in een selectie van die tijdschriften open access te publiceren. De universiteiten streven ernaar dat in 2018, het derde jaar van de overeenkomst, 30% van de Elsevierartikelen van Nederlandse auteurs open access beschikbaar is. Dit akkoord maakt dat mogelijk. Dit is echt geweldig nieuws en een 'big deal' voor open access." Philippe Terheggen, Managing Director Journals bij Elsevier: "Wij zijn content met deze overeenkomst, omdat blijvende subscriptietoegang tot onze hoogwaardige, 'peer-reviewed' wetenschappelijke artikelen essentieel is voor Nederland om zijn positie als één van de meest impactvolle onderzoekslanden te behouden. Daarnaast krijgen Nederlandse wetenschappers meer open access publicatiemogelijkheden om hun onderzoeksresultaten met de rest van de wereld te delen." De overeenkomst is in lijn met de ambitie van staatssecretaris Dekker (OCW), die wil dat artikelen van Nederlandse wetenschappers open access gepubliceerd worden. Blijf op de hoogte van deze en andere ontwikkelingen via de Open-Access nieuwsbrief van de VSNU, de Nederlandse universiteitsbibliotheken en de Koninklijke Bibliotheek. (VSNU)
vsnu.nl

OPROEP

Wint u de Nederlandse Dataprijs 2016?

Komend najaar wordt weer de Nederlandse Dataprijs uitgereikt. Een prijs voor een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor aanvullend of nieuw onderzoek.
De winnaars van de voorgaande edities zijn in ieder geval enthousiast: "De jury noemt onze database een grote aanwinst voor zowel het Nederlands academisch als cultureel erfgoed. Dat is een bevestiging dat we op het goede spoor zitten," aldus Martine de Bruin, Nederlandse Liederenbank, winnaar van de Dataprijs humaniora en sociale wetenschappen 2014.
"Door het winnen van de Dataprijs kunnen we nu ook een paar grotere, al langer gewenste verbeterslagen maken," aldus Johan Molenbroek en Marijke Dekker, DINED, winnaars van de Dataprijs exacte en technische wetenschappen 2014.
Naast de winnaars waren ook de bijna 50 andere inzendingen van hoog niveau. De jury sprak over 'allemaal mooie voorbeelden van het toegankelijk maken en delen van onderzoeksdata'.
De organisatie van de Nederlandse Dataprijs is in handen van Research Data Netherlands, een samenwerkingsverband tussen 3TU.Datacentrum, DANS en SURFsara. Binnenkort staat meer informatie over de Dataprijzen 2016 op de website van RDNL. (HB)
researchdata.nl