

## Vernieuwde BNTL wordt web 2.0

In 2005 is op verzoek van de KNAW het vernieuwingsproject begonnen voor de Bibliografie van de Nederlandse Taal- en Literatuurwetenschap BNTL. Er komen meer mogelijkheden voor gebruikers, ook om zelf inhoud toe te voegen. De uitvoering is in handen van de afdeling e-Research van het Huygens Instituut.

Onderzoekers en docenten in de neerlandistiek kennen de BNTL, die is gestart in 1970. De bibliografie bevat Nederlandse, Vlaamse en buitenlandse titels van publicaties over Nederlandse taal- en letterkunde uit de periode 1940 tot heden: op dit moment ruim 230.000 elektronisch raadpleegbare titels.

In de eerste plaats betekent de vernieuwing een eigen webinterface, op basis van open standaarden. De BNTL kan daardoor overal en door iedereen geraadpleegd worden, onafhankelijk van toegangskanalen.

De gebruiker krijgt de mogelijkheid om gelijktijdig te zoeken op informatie uit de titels (publicatietype, auteur, jaar, titel), de toegevoegde informatie (trefwoorden, abstracts en andere informatie) en de volledige tekst, indien beschikbaar.

Het toevoegen van volledige teksten gebeurt voor een deel door de interne redactie, die ook links aanbrengt naar voor neerlandici interessante bestanden. Daarnaast worden auteurs, redacties en uitgevers van tijdschriften benaderd met de vraag hun teksten ter beschikking te stellen, al of niet op termijn.

De geregistreerde gebruiker krijgt ook zelf de mogelijkheid om titels en teksten toe te voegen.

Het invoersysteem wordt vereenvoudigd om het invoeren van titels en teksten minder arbeidsintensief te maken voor de interne redactie. Het zal naar verwachting ook de drempel voor de gebruikers om eigen titels

en teksten in te voeren verlagen. Ook op het web zal worden 'geooogst'. Tijdens het vernieuwingsproject wordt onderzoek gedaan naar een webcrawler en naar automatisch classificeren en indexerend. Het gehele titelbestand van de huidige BNTL wordt gemigreerd naar de nieuwe BNTL.

Als het project in juli 2007 is afgerond, wordt het 'format' ter beschikking gesteld aan de BFTL (de Bibliografie van de Friese taal- en literatuurwetenschap), die parallel aan de BNTL wordt samengesteld en gepubliceerd. Ook andere bibliografieën kunnen daar gebruik van maken. (Elly Kamp)

[www.huygensinstituut.knaw.nl](http://www.huygensinstituut.knaw.nl)

## Nieuwe data beschikbaar over primair onderwijs

Op verzoek van het Ministerie van OCW wordt sinds een aantal jaren het Cohortonderzoek Primair Onderwijs PRIMA uitgevoerd. Het wordt gedaan door het Instituut voor Toegepaste Sociale Wetenschappen ITS (RU Nijmegen), en het SCO-Kohnstamm Instituut (UvA) en moet een zo compleet mogelijk beeld van het primair onderwijs geven. De data van de zesde meting, in 2004/2005, zijn recentelijk bij DANS (Data Archiving and Networked Services) beschikbaar gekomen.

Bij deze zesde meting zijn op circa zeshonderd reguliere basisscholen gegevens verzameld bij leerlingen in groep twee, vier, zes en acht. Behalve bij leerlingen is ook informatie verzameld bij leerkrachten, schooldirecties en ouders. Aan ouders van leerlingen uit groep twee is bijvoorbeeld een uitgebreide schriftelijke vragenlijst voorgelegd. Daarin wordt inge-

gaan op sociale, culture en etnische achtergronden van het gezin van de leerling. Cohortonderzoek houdt in dat verschillende jaargroepen (cohorten) tegelijk worden onderzocht, en bovendien dat leerlingen worden gevolgd in hun onderwijsloopbaan. In totaal zijn van ongeveer 60.000 leerlingen gegevens verzameld. De onderzochte basisscholen komen uit een landelijke representatieve steekproef en een aanvullende steekproef van scholen met een hoge concentratie van kinderen (zowel allochtoon als autochtoon) uit de lagere sociaal-economische milieus.

Ook de data van de voorafgaande PRIMA metingen zijn beschikbaar via DANS. Geïnteresseerde onderzoekers kunnen de data, SPSS portable files met de bijbehorende documentatie, bestellen via [info@dans.knaw.nl](mailto:info@dans.knaw.nl). Aan levering zijn geen kosten verbonden.

## Techniek volstaat niet in informatica en letteren

De belangstelling voor de sociale aspecten van nieuwe media en technieken – in relatie met het traditionele onderzoek, met de samenleving en met andere digitale letterenprojecten – was het meest opvallende kenmerk van het congres CLiP (Computers, Literature and Philology) 2006, van 29 juni tot 1 juli in Londen.

CLiP is bedoeld als forum voor innoverend onderzoek op het raakvlak van informatica en letteren, en dan vooral wat de Romaanse talen betreft. Het bekende *Centre for Computing in the Humanities* van King's College London heeft dit jaar het initiatief genomen voor een herstart van de conferentiereeks en trad op als orga-

niserende instantie. De bijeenkomst heeft uitdrukkelijk de bedoeling om internationale contacten en samenwerking te stimuleren, en dat bleek ook uit het zeer divers samengestelde publiek: er waren vanzelfsprekend veel Engelsen, maar ook Spanjaarden, Italianen en een paar Fransen, en daarnaast Duitsers, Noren, en Esten. Die diversiteit maakte de discussies levendiger dan doorgaans het geval is op dergelijke congressen.

Uit de bijdragen bleek duidelijk dat XML (eXtensible Markup Language) algemeen aanvaard is als de basis voor gegevensopslag en -publica-

tie; het onderzoek richt zich nu vooral op het creëren, ontsluiten en presenteren van XML-gebaseerde corpora. Een nieuw aandachtspunt is het gezamenlijk gebruik van deze informatie, onder meer ingegeven door bezorgdheid over het langzame doorsijpelen van de 'traditionele' kennis en vaardigheden naar de nieuwe media, die in het maatschappelijk leven een steeds belangrijker plaats innemen. Meer dan eens werd gewezen op de noodzaak van meer continuïteit en samenwerking in de wereld van *humanities computing* – techniek alleen volstaat niet. (Bert van Elsacker)

<http://www.cch.kcl.ac.uk/clip2006/>

## Initiatieven voor wetenschappelijke datacitatie

Het lijkt erop dat de wetenschappelijke datacitatie de laatste tijd erg in de belangstelling staat. Op de IASSIST conferentie in Ann Arbor, Michigan was er een hele sessie aan dit onderwerp gewijd met vier verschillende sprekers. Ook tijdens de opening van DANS werd door methodologieprofessor Harry Ganzeboom nadrukke-

lijk stil gestaan bij het nut en belang van datacitatie.

De roep om datacitatie komt voort uit het idee dat een wetenschappelijk werk niet alleen voorzien moet zijn van literatuurverwijzingen, maar ook van goede documentatie van de gebruikte data. Dit heeft tot doel de empirische cyclus mogelijk te maken.

Zonder goede verwijzingen naar zowel literatuur en bronnen als de gebruikte data is het onmogelijk een wetenschappelijk artikel te beoordelen, laat staan om het onderzoek te dupliceren en de resultaten na te bootsen/na te rekenen. Daarnaast levert datacitatie ook *credits* op voor de producenten van data en is het mogelijk deze credits in wetenschappelijke beoordelingssystemen te verwerken. Producenten van data zouden daarvoor ook eerder geneigd zijn hun data te delen met andere gebruikers.

Wie op dit moment in zijn onderzoek naar data verwijst, volstaat meestal met vage aanduidingen of opmerkingen over de herkomst van de data: 'Bron: CBS', of een opmerking in het voorwoord waarin Pietje Puk bedankt wordt voor het leveren van de bestanden. Deze handelwijze wordt echter door steeds meer men-

sen als niet echt wetenschappelijk gezien. Statistics Canada (het Canadese CBS) is zich van het probleem bewust en heeft uitgebreide standaarden opgesteld voor het verwijzen naar haar data. Maar ook op Harvard University en bij het Amerikaanse *Interuniversity Consortium for Political and Social Research (ICPSR)* wordt hierover nagedacht en aan standaarden gewerkt. Inmiddels heeft de *International Association for Social Science Information Service and Technology (IASSIST)* het initiatief genomen tot een internationale werkgroep. Het probleem van deze initiatieven is dat ze afkomstig zijn uit de hoek van dataleveranciers en niet direct van datagebruikers. De huidige initiatieven hebben enkel kans tot slagen als ze in een vroeg stadium gebruikers van data weten te interesseren en betrekken bij de ontwikkeling. (LS)

<http://www.statcan.ca/bsolc/english/bsolc?catno=12-591-XWE>

<http://gking.harvard.edu/projects/repl.shtml>



CREDIT: DAGMAR STIEBRAL

Prof. Harry Ganzeboom onderstreepte bij de opening van DANS het belang van datacitatie

### VGI Innovatieprijs 2006

Ook in 2006 belooft de Vereniging voor Geschiedenis en Informatica (VGI) de meest vernieuwende toepassing op het terrein van de geschiedenis. Een deskundige jury zal uit de genomineerde inzendingen (zie [www.vgi-online.org/](http://www.vgi-online.org/)) drie Vlaamse of Nederlandse ICT-toepassingen selecteren die naar haar mening tot de meest vernieuwende van het jaar gerekend mogen worden. Deze drie zullen op een VGI studiedag met als thema 'Het gebruik van *social software* in historische context' op vrijdag 22 september worden gepresenteerd.

Winnaar 2005: Literair Gent, [www.literair.gent.be](http://www.literair.gent.be)

Winnaar 2004: Ename Expertisecentrum voor erfgoedontsluiting, [www.enamecenter.org](http://www.enamecenter.org)

Winnaar 2003: De schatkamer van het Gemeente-archief Amsterdam, <http://gemeentearchief.amsterdam.nl/schatkamer/>

### NWO MaGW vraagt datacontract

Het Gebiedsbestuur Maatschappijen en Gedragwetenschappen van NWO heeft DANS – *Data Archiving and Networked Services* – aangewezen om er op toe te zien dat onderzoeksdata uit projecten die door haar zijn gefinancierd, duurzaam worden gearchieveerd. Zo komen ze ter beschikking van verder onderzoek. NWO MaGW ziet het als haar verantwoordelijkheid, die verspreiding te stimuleren.

Ontvangers van subsidie zal daarom na toekenning van de subsidie verzocht worden contact op te nemen met DANS om een *datacontract* af te sluiten. Een kopie van dat contract wordt na ondertekening naar NWO gestuurd.

DANS, dat over de ervaring en deskundigheid beschikt om de archivering van data zowel formeel als praktisch goed te regelen, raadt aan om al voor het indienen van de subsidieaanvraag contact op te nemen. Dan kan in een vroeg stadium worden geadviseerd over data-formaten, metadata, beveiliging en digitale duurzaamheid.

[http://www.nwo.nl/nwohome.nsf/pages/NWOA\\_6PVA9j](http://www.nwo.nl/nwohome.nsf/pages/NWOA_6PVA9j)

### Social software

Op 22 september 2006 organiseert de Vereniging Geschiedenis en Informatica een studiemiddag over social software: internetdiensten die ontstaan en onderhoud van sociale netwerken faciliteren. De term wordt vaak genoemd in combinatie met web 2.0. Het 'oude web' wordt gezien als iets statisch, waar bezoekers informatie kunnen halen maar waar geen communicatie over en weer plaats vindt. Social software zorgt ervoor dat interactie kan plaatsvinden tussen de auteurs van de site en de bezoekers. Een voorbeeld is Wikipedia, een online encyclopedie waar iedereen aan bij kan dragen: <http://nl.wikipedia.org/wiki/weblog>. Tijdens de studiemiddag wordt gekeken wat de mogelijkheden van social software zijn voor het culturele erfgoed sector en de humaniora.

[www.vgi-online.org](http://www.vgi-online.org)