

Pagina 4 • CLARIN biedt taalonderzoekers infrastructuur • Het project CLARIN dat onlangs Europese erkenning kreeg gaat digitale toegang geven tot gesproken en geschreven taalmateriaal in heel Europa, tot gereedschap en tot expertise om ermee om te gaan.



Pagina 5 • Bijbelexegeet Talstra analyseert met computer • De Amsterdamse hoogleraar Eep Talstra ziet Bijbelteksten als data, en de analyse daarvan als het schiften tussen taalkundige en literaire verklaringen. 'Ik zeg altijd tegen mijn studenten: Begin nou eens met de data'.

Pagina 6 • Computers gekoppeld voor meer rekenkracht • Het BIGGRID project wil de in Nederland beschikbare opslag- en reken capaciteit vergroten voor het opslaan, archiveren, bewerken, koppelen en uitwisselen van steeds groter wordende wetenschappelijke datastromen.

EN VERDER

Agenda.....	2
Nieuws.....	3
Achtergrond.....	4
Gebeurtenissen.....	6
Column.....	8
Gelezen.....	8

SURF: 'Tekst en data koppelen'

De komende jaren gaat SURF onder meer aandacht besteden aan het toevoegen van data en visuele informatie aan publicaties, collaboratories en het beschikbaar maken van data. Dat bleek 25 januari op een bijeenkomst in Amsterdam ter afsluiting van het programma DARE.

Bij het beschikbaar maken van data gaat het erom, tekst en gegevens uit onderzoek aan elkaar te koppelen. 'Dat moet zo gebeuren dat de gebruiker er niets van merkt', lichte DARE-programmamanager Leo Waaijers toe op een ongeruste vraag van forumlid en taaltechnoloog Franciska de Jong. Waaijers: 'Tot nu toe hebben we in het kader van DARE gewerkt aan het opslaan en beschikbaar maken van rapportages. Dat is een flink succes geworden waar we internationaal mee voorop lopen. Ruim honderd-duizend publicaties zijn er nu via DAREnet te vinden, dat is tweemaal de jaarproductie. Geen land doet ons dat na. Voor de verduurzaming van die opgeslagen tekst hebben we een overeenkomst met het Depot van de Koninklijke Bibliotheek kunnen sluiten. Als het om data gaat is DANS daarvoor de natuurlijke partner. Met hen zullen we de komende jaren gaan kijken hoe het mogelijk is om het verslag en de data van een onderzoek tegelijk op een voor de gebruiker gemakkelijke manier beschikbaar te maken'.

Ook veel gasten waren goed te

spreken over de 'top record' van SURF, zoals voorzitter Sijbalt Noorda van de universiteitsclub VSNU het noemde. 'Op het gebied van samenwerking doen we het met SURF buitengewoon goed maar dat is eigenlijk jammer want het zou beter zijn als de situatie binnen de EU wat homogener zou zijn'.

Veel aandacht was er voor de remmende rol die aan uitgever wordt toegeschreven bij het nastreven van Open Access. Net als de Amerikaanse 'PR-pitbull' Eric Dezenhall volgens een ter plaatse uitgedeeld artikel in *Nature* waren ook verschillende sprekers van mening dat het systeem van peerreviewdaarbijeenbeslissenderol speelt. Dezenhall, die in opdracht van enkele grote uitgeverijrichting moet geven aan hun strijd tegen de Open Access beweging, benadrukt dat de traditionele uitgeefprocessen 'gelijk staan met' een kwalitatief hoogstaande peer review. Zonder een goed referee systeem blijft de betekenis van Open Access beperkt, was ook de constatering in Amsterdam maar dat leidde juist tot pleidooien voor het zoeken naar goede alternatieven,



MONIQUE KOOIJMANS

Waaijers: 'Geen land doet ons dit na...'

zoals door ex-Erasmusbaas prof. Jan van Bommel: 'Desnoods moeten we andere methoden van peer review gaan organiseren, bijvoorbeeld met de KNAW, als de uitgeverij het vertikken'. De pleiters werden op hun

wenken bediend want 'snellere en transparantere vormen van kwaliteitsbeoordeling' staan ook op de *to-dolist* die SURF onder de naam SURF-share presenteerde. (MdG)

NWO 'gematigd tevreden' met plannen van coalitie

De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) is 'gematigd tevreden' met de plannen van de nieuwe christelijk-sociale regering. Volgens directeur Communicatie Hein Meijer is het bedrag van 300 miljoen euro dat de coalitie structureel wil reserveren voor innovatie, kennis en onderzoek niet wat zijn organisatie wilde, maar toch positief. Uit het hele akkoord blijkt dat de nieuwe regering het belang van wetenschap en onderzoek onderkent, aldus Meijers die ook wijst op de voorgenomen heroprichting

van het Innovatieplatform en de erkenning van het belang van de z.g. tweede geldstroom – die loopt via instellingen zoals NWO. 'Jammer is wel dat de wens van NWO niet wordt gehonoreerd om structureel meer geld aan ons toe te kennen'. Het bedrag is daarvoor te klein en het moet over meer ontvangers worden verdeeld. Vorig jaar vroeg NWO om 433 miljoen extra per jaar voor excellent onderzoek, krachtenbundeling en betere terugkoppeling naar de samenleving. (MdG)

Cupido achter behang



Wie zich eenmaal door Cupido het juk heeft laten opleggen, komt daar niet meer zo makkelijk van af, aldus dit embleem van Daniël Heinsius, dat is gedigitaliseerd in het Emblem Project Utrecht (inzet). Eind 2006 ontving het project (zie e nr. 2/2006) een mail uit Frankrijk: de bewoon-

ster van een kasteel had een serie wandschilderingen aangetroffen van acht emblemen, waaronder het genoemde. Dankzij de bijschriften was ze via Google op het spoor van de site <http://emblems.let.uu.nl> gekomen.

Het gebruik van emblemen voor decoratie van privé-vertrekken is niet

uniek, maar goed bewaarde voorbeelden zijn zeldzaam. Inmiddels hebben medewerkers van het project in Frankrijk de wandschilderingen onderzocht, die sinds de vroege negentiende eeuw verborgen hebben gezeten achter het behang. (PB)

e-data: gratis voor wie het wil

Dit is het vierde nummer van e-data&research, het kwartaalblad over data en onderzoek in de alfa- en gammawetenschappen dat verschijnt sinds juni 2006. e-data&research heeft een aantal stakeholders die ervoor zorgen dat het blad kosteloos kan worden toegezonden aan onderzoekers en studenten in de betrokken disciplines. Dat zijn Data Archiving and Networked Services (DANS) en het Huygens Instituut in Den Haag, het Internationaal Instituut voor Sociale Geschiedenis (IISG) in Amsterdam en de Vereniging voor Ge-

schiedenis en Informatica (VGI). Leden van laatstgenoemde vereniging ontvingen vroeger het blad *Historia et Informatica*. e-data is daarvan een voortzetting. Wilt u e-data&research als onderzoeker of student voortaan ontvangen? Geef dan uw adresgegevens door aan het secretariaat: info@edata.dans.knaw.nl. Staat u wel op de verzendlijst maar stelt u geen prijs op toezending? Dezelfde procedure, met vermelding van uw wens. Adresgegevens voor de traditionele post staan in het colofon op pagina 7.

AGENDA

8 maart 2007

Den Haag, Ministerie van VROM
– Oranjehout

Workshop 'Wegwijs in WoOn'

Workshop van DANS en het ministerie van VROM om potentiële gebruikers wegwijs te maken in de gegevens van het Woon Onderzoek Nederland (WoON 2006), met informatie over de opzet, de structuur, de vragenlijsten, het veldwerk, en de veranderingen ten opzichte van de voorloper – het Woning Behoeft Onderzoek (WBO). Er wordt ingegaan op de dataprocessing en de valkuilen bij het gebruik van de gegevens. Ook komen onderzoekers aan het woord die ervaring hebben met het gebruik van WBO data. www.dans.knaw.nl/nl/dans_symposia/

14-17 maart 2007

New York, New York University

The 2007 Society for Textual Scholarship
14th Biennial International Interdisciplinary Conference

Topics: textual environments, textual cultures, textual ruins, textual arts, including the book arts, digital texts and editing projects. www.mith2.umd.edu/research/projects/sts/html/cprogram.php

14 maart, 28 maart, 18 april, 9 mei, 30 mei 2007

Amsterdam, Archiefschool

Cursus digitaal depot

Een plan van aanpak voor uw e-depot in vier dagen. De cursus is gericht op mensen die werkzaam zijn bij archiefbeherende instellingen en verantwoordelijk zijn voor de formulering van beleid, in het bijzonder het beleid ten aanzien van het beheer en beschikbaarstelling van digitaal erfgoed. www.divakoepel.nl en www.archiefschool.nl

10-12 april 2007

Manchester, University of Manchester,
Crawford House

Longitudinal Data Analysis

This course will provide a practical overview of the main classes of methods including population average and random effects models. www.ccsr.ac.uk/courses/external/2006-2007/index.html

20-21 april 2007

Frankfurt am Main, Deutsche Nationalbibliothek

The Challenge, long-term preservation, strategies and practice of European partnership

The conference aims to present and discuss the current technical and organizational status of measures for the long-term preservation of digital media in various countries in Europe, jointly identifying methods and strategies and sounding out the feasibility of a European alliance of the individual national alliances devoted to preventing digital data loss. <http://lza.ddb.de/eu2007/>

2-4 mei 2007

Baden-Baden

German e-Science Conference 2007

Organized by Helmholtz-Gemeinschaft, Max-Planck-Gesellschaft and German Rectors' Conference (HRK), this is the first event to demonstrate the results from four e-Science areas: Grid computing, knowledge networking, e-Learning and open access. The partners of the German e-Science community will present their results for the international e-Science community to get an overview of the German progress and for potential new users to meet the experts in the field of e-Science. www.ges2007.de/

16-18 mei 2007

Quebec

IASSIST Conference 2007 - Building Global Knowledge Communities with Open Data

Attention is focused on the globalization of knowledge and the importance of the open data concept in the development of knowledge communities. The inter-relationship of knowledge communities with open data will be explored. What is required to make data more 'open' and available; what are the outcomes from open data; and what is the role of the data community in helping this happen? www.iassistdata.org/

Web 2.0 technologie in erfgoedbeheer

'Dit project toont hoe men erfgoed kan ontsluiten met nieuwe media en web 2.0-technologie'. Dat zei Tijl Vereenoghe van de Vlaamse archeologiesite Archeonet bij de presentatie van het initiatief *Erfgoed.be* op een studiedag over digitale beeldcollecties op 8 december in Gent. Op de studiedag, die was georganiseerd door de Vereniging voor Geschiedenis en Informatica (VGI) werd een aantal beeldbankprojecten gepresenteerd door Vlaamse erfgoedbeheerders.

Erfgoed.be verzamelt foto's van beschermde monumenten, landschappen en dorpsgezichten op een interactieve kaart van Vlaanderen. Een klik op de kaart opent de foto met extra metadata. De bezoekers kunnen zelf een bijdrage leveren door foto's en metadata op te sturen, waarna een medewerker ze aan de kaart toevoegt. 'Tegenwoordig gebruikt de erfgoedconsument, net als iedereen, zoekmachines om informatie te vinden', aldus Vereen-

De Schuilplaats voor Werklieden te Antwerpen, op www.flickr.com geplaatst door erf-goed.be.

ooghe. 'Helaas is de *web presence* van Vlaamse monumenten niet altijd optimaal. Vaak blijkt het onmogelijk om op het net een geschikte rechten-vrije foto te vinden. Daarom besloten we zelf erfgoedfoto's te verzamelen en die onder een Creative Commons

Interactief erfgoed bij DEN

'Erfgoed interactief' was het thema van de drukbezochte Digitaal Erfgoed Conferentie, op 12 en 13 december 2006 georganiseerd door Digitaal Erfgoed Nederland in De Doelen in Rotterdam. Lezingen leverden nieuwe informatie, en workshops en debatten van hoog niveau boden de mogelijkheid om parallelle problemen vast te stellen en ervaringen uit te wisselen. Veel indruk maakte de *widget* van het Rijksmuseum: voor abonnees elke dag een ander kunstwerk op de desktop, gelinkt naar meer informatie over dat werk. Jeroen Loefen liet zien hoe de Stichting Rebeltoday software voor interactieve sites beschikbaar stelt aan scholen. Die kunnen daar hun eigen schoolkrant mee maken, waarin de leerlingen kunnen opnemen en aanpassen wat zij willen. Zijn ervaring was dat er weinig vervelende berichtjes op deze

sites terechtkomen: als je ook werkelijk alles plaatst, passen de leerlingen wel op om er gekke dingen neer te zetten. Een verfrissend bericht uit de praktijk bij alle zorgelijke blikken waar het vrije autorisatie betreft.

De informatica kwam aan bod in de sessie 'ICT-onderzoek en erfgoed: wat hebben we aan elkaar?'. Verschillende projecten uit het NWO-onderzoeksprogramma Continuous Access To Cultural Heritage (CATCH) werden kort gepresenteerd. Een daarvan was MuSeUM (*Multiple-collection Searching Using Metadata*), waarin informatici samenwerken met het Haags Gemeentemuseum aan het zoeken in een grote verscheidenheid aan digitaal materiaal. Het belang van traditionele ontsluiting lijkt met de nieuwe technische mogelijkheden aantoonbaar geringer te worden. (Karina van Dalen-Oskam)

@ www.den.nl/ www.rijksmuseum.nl/ www.rebeltoday.nl/

e-Humanities op e-Science 2006

Wetenschappelijk onderzoek is meer en meer een gezamenlijke onderneming van wetenschappers en onderzoeksgroepen. Daardoor groeit de behoefte aan een technische infrastructuur die digitaal samenwerken mogelijk maakt. Geen triviale zaak omdat vaak verschillen tussen computersystemen, softwarecomponenten, dataformaten en netwerken moeten worden overbrugd. De rol van *grid computing* bij de koppeling van heterogene computerbronnen werd nog eens belicht op de *2nd IEEE International Conference on e-Science and Grid Computing (e-Science 2006)*.

De 'harde' wetenschappen domineerden op de e-Science, maar gelukkig waren ook de geesteswetenschappen vertegenwoordigd met

licentie te verspreiden'. Door gebruik te maken van de mogelijkheden van Flickr, geotagging, Google Earth en Triggermap werd een succesvolle website gecreëerd: acht maanden na de lancering droegen tachtig geïnteresseerden samen reeds 1.400 foto's aan. Het toont volgens Vereenoghe aan dat er in Vlaanderen vraag was naar een dergelijk initiatief: 'Erfgoed.be bewijst dat de Vlamingen trots zijn op hun patrimonium'.

Nieuwe media bieden nieuwe mogelijkheden om erfgoed bij de

consument te brengen: 'Een open en dynamische benadering kenmerkt Erf-goed.be. Duurzaamheid van de site is weliswaar moeilijk te garanderen als je afhankelijk bent van web services, maar de voordelen wegen zwaarder. Zo is de bekendheid van een dienst als Flickr een enorm pluspunt. Momenteel concentreren wij ons op het bijeenbrengen van zo veel mogelijk foto's. Dit sluit echter niet uit dat we op termijn nieuwe functionaliteiten zullen toevoegen'. (Matthias Vandermaesen)

@ www.erf-goed.be/ www.archeonet.be/

Repositories voor kleintjes

Vele grote, nationale instellingen initiëren prestigieuze projecten voor digitale archivering met bijpassende budgetten. Kleinere spelers lopen het risico achterop te raken. Het Digital Curation Centre (DCC) en het Liverpool University Centre for Archive Studies (LUCAS) hielden op 30 november en 1 december 2006 een workshop in Liverpool met op de praktijk gerichte presentaties om die afstand te verkleinen.

De 56 deelnemers, grotendeels uit het Verenigd Koninkrijk, zagen presentaties van onder andere het Public Record Office of Northern Ireland

(strategieën om data vóór verwerving voor archiveren geschikt te maken), het Engelse Paradigm-project (verwerven van data van particulieren), en DCC (pilot audits on Trusted Digital Repositories). De Harvard University Library lichtte haar *knowing what you've got*-strategie toe, gekenmerkt door zware format identificatie, validatie en bewakingsprocedures tijdens de 'ingest' van materiaal. Daartegenover stond de benadering van de Old Dominion University: 'objects are responsible for their own preservation'. Hierbij worden technische gegevens van digitale objecten in laagjes XML-code aan het object toegevoegd. Bij het ter beschikking stellen van het object zorgen deze metadata voor de juiste afhandeling. Het Amerikaanse NARA demonstreerde het prototype 'Persistent Archives', een systeem dat documenten via een media-adaptor op het scherm kan tonen als txt-, pdf- of HTML-bestand, onafhankelijk van het oorspronkelijke format. Een onderzoeker kan vervolgens allerlei bewerkingen uitvoeren (vertalingen, annotaties) zonder aan de oorspronkelijke bitstream te komen.

Opvallend was hoe sterk het OAS-model ingeburgerd is geraakt, en hoezeer de nadruk meestal nog ligt op de processen van 'ingest' en 'preservation planning', en niet op 'access'. (MvdH)

@ www.dcc.ac.uk/events/archives-2006

@ www.mpi.nl/clarin/cvp-v-i-may-06_pewi.html
www.mpi.nl/clarin

Geheugen van Nederland en Multatuli

Recentelijk is een aantal nieuwe collecties van grote literaire waarde toegevoegd aan Het Geheugen van Nederland: twee met 'literaire hoogtepunten' van de Koninklijke Bibliotheek en het Nederlands Letterkundig Museum en Documentatiecentrum en één uit het Multatuli Museum. Ze bestaan alle drie grotendeels uit kwetsbaar materiaal uit de

periode 1840-1950, en waren eerder al op microfilm gezet in het kader van Metamorfoze, het landelijke programma voor papierbehoud.

In de Multatuli-collectie kunnen ruim zeshonderd publicaties van en over Multatuli worden bekeken. Te zien zijn onder meer afleveringen uit zijn beroemde *Ideeën*, en schermselingen op papier tussen de

schrijver en zijn criticasters. Al het materiaal komt uit de collectie van het Multatuli Museum in Amsterdam.

Het Geheugen van Nederland is het nationale programma voor digitalisering van cultureel erfgoed, dat wordt gesubsidieerd door het Ministerie van Onderwijs, Cultuur en Wetenschap. In de Koninklijke Bibliotheek is een bureau ingericht dat het programma coördineert en de website www.geheugenvanederland.nl in de lucht houdt, waar iedereen terecht kan die geïnteresseerd is in mooie afbeeldingen uit onze vaderlandse geschiedenis en de Nederlandse samenleving in het

algemeen. Erfgoedinstellingen kunnen via Het Geheugen van Nederland een subsidie krijgen van 70% voor het uitvoeren van digitaliseringsprojecten. In 2007 is een totaalbedrag van € 880.000 beschikbaar voor nieuwe projecten.

De website biedt op dit moment toegang tot gedigitaliseerde collecties van zo'n vijftig culturele instellingen, met ruim 700.000 afbeeldingen, 250 uur film en 100 uur audio. De collectie groeit gestaag. Ook biedt Het Geheugen kant en klare lessen voor gebruik in het voortgezet onderwijs. (Ingeborg Verheul)

Kaderprogramma: meer kans

Het nieuwe Europese Kaderprogramma voor Onderzoek (FP7), dat begin dit jaar van start ging, heeft anderhalf keer zoveel budget als het vorige programma (FP6) en een hogere vergoeding voor onderzoeksactiviteiten. Ook zijn er ruimere mogelijkheden om projectvoorstellen in de alfa- en gammawetenschappen in te dienen.

Het budget voor het zevende kaderprogramma, dat loopt tot 2013, werd eind vorig jaar vastgesteld op ruim vijftig miljard euro, en voor non-profit onderzoekinstellingen werd het vergoedingspercentage verhoogd van 60% naar 75%. Het thema 'Citizens and governance in a knowledge-based society' voor de alfa- en gammawetenschappen uit FP6 is vervangen door het aanzienlijk ruimere 'Socio-economic Sciences and Humanities (SSH)' in FP7.

Het thema SSH valt onder het programma *Coöperation*. Het kent acht activiteiten, die zijn onderverdeeld in gebieden, welke op hun beurt een aantal onderwerpen bevatten. In totaal zijn er binnen het thema SSH 32 onderwerpen benoemd, variërend van de gevolgen van demografische veranderingen, tot de bescherming van mensenrechten of de invloeden

van kunst en literatuur.

Daarnaast bieden de meer algemene programma's *Ideas*, *People* en *Capacities*, kansen voor alfa- en gamma-onderzoek. *Ideas* stimuleert excellent onderzoek. Nederlandse alfa- en gammawetenschappers hebben zich op dit punt de afgelopen jaren onderscheiden: in iedere ronde van het programma European Young Investigators Award wonnen ze tot nu toe een prijs. Bij *People* gaat het om beurzen voor training en mobiliteit, die niet gebonden zijn aan een onderzoeksgebied. *Capacities* ondersteunt onder meer onderzoeksinfrastructuur en onderzoek naar de rol van wetenschap in de maatschappij.

Voor alfa- en gamma-onderzoek zijn er dus genoeg kansen. Haast is echter geboden: de eerste sluitingsdata voor het indienen van projectvoorstellen zijn al in mei.

Meer informatie over FP7 is te vinden op de website van Cordis, de EU-portal voor research & development. Ook EG-Liaison, dat de benutting van het Europees onderzoeksgeld in Nederland bevordert, biedt informatie en advies. (Ellen Willemse)

www.cordis.europa.eu
www.egl.nl

www.multatuli-museum.nl

DANS begint project voor behoud van onderzoeksdata

In januari is bij DANS (Data Archiving and Networked Services) in Den Haag het tweejarige project MIXED begonnen. Het is een groot project om programmatuur te ontwikkelen die garandeert dat databases en spreadsheets in de verre toekomst bruikbaar blijven.

Voor de uitvoering van MIXED (Migration to Intermediate XML for Electronic Data) is als projectleider wiskundige dr. Dirk Roorda aangekomen. Naast Roorda werken verschillende medewerkers van DANS aan het project, terwijl ook externe programmeurs worden ingehuurd.

'We gaan een programma ontwikkelen', licht Roorda toe, 'dat in staat is om alle sporen te verwijderen van de toepassingen die een bepaalde dataset hebben voortgebracht, maar met behoud van de betekenisvolle kern van die data. Je haalt dus de franje eraf en je houdt iets over dat onder verschillende programma's te lezen en te gebruiken is'. De data moeten zo

worden opgeslagen dat hun waarden en hun onderlinge relaties behouden en leesbaar blijven. Daartoe worden ze omgezet in het duurzame bestandsformaat XML, dat gebruik maakt van de internationale schriftstandaard Unicode, die taalneutraal is en ook alle gangbare symbolen omvat. Unicode, dat nu rond 65.000 tekens omvat (ASCII: 255) kan nog tot een klein miljoen worden uitgebreid. MIXED moet beantwoorden aan een probleem dat in de hele wetenschappelijke wereld opgeld doet, namelijk van de duurzaamheid van bewaarde onderzoeksresultaten in een snel veranderende digitale omgeving.

Vervolg op pagina 6

Gratis toegang tot microbestanden CBS

Sinds 1 januari brengt DANS geen bijdrage meer in rekening voor het beschikbaar stellen van CBS-bestanden. Verder blijven de gebruikelijke voorwaarden gelden.

Het CBS stelt via zijn Centrum voor Beleidsstatistiek onder voorwaarden zoals geheimhouding statistische bestanden beschikbaar aan onderzoekers, planbureaus en ministeries. Deze microdata zijn *on site*- bij het CBS of via *Remote Execution*- en *Remote Access* te analyseren (zie ook e nr 2 en 3, 2006). Via DANS (Data Archiving and Networked Services) worden beveiligde versies beschikbaar gesteld voor extern gebruik.

Op dit moment zijn via DANS zo'n 35 belangrijke bestanden beschikbaar zoals de *Enquête Beroepsbevolking (EBB)*, het *Sociaal-economisch Panelonderzoek (SEP)*, en het *Permanent Onderzoek van de Leefsituatie (POLS)*. Regelmatig zal deze selectie uitgebreid worden met nieuwe bestanden. Een actueel overzicht van de beveiligde microbestanden is beschikbaar op de DANS website: www.dans.knaw.nl. Via info@dans.knaw.nl kunnen de bestanden bij DANS aangevraagd worden. Meer informatie over de microdataservices van het CBS, inclusief een catalogus van beschikbare microbestanden en bijbehorende documentatie, is te vinden op www.cbs.nl/microdata. (LS)

Geheugensite vernieuwd

Het afgelopen jaar is de techniek achter de website van het Geheugen van Nederland verbeterd: eenvoudiger zoeken, sneller en overzichtelijker bladeren door de vele afbeeldingen, nieuwe mogelijkheden om in te zoomen. Tegelijk wordt ook de huisstijl van Het Geheugen vernieuwd. De nieuwe versie van de website wordt op dit moment getest en is naar verwachting in mei 2007 beschikbaar. (Ingeborg Verheul)

www.geheugenvanederland.nl

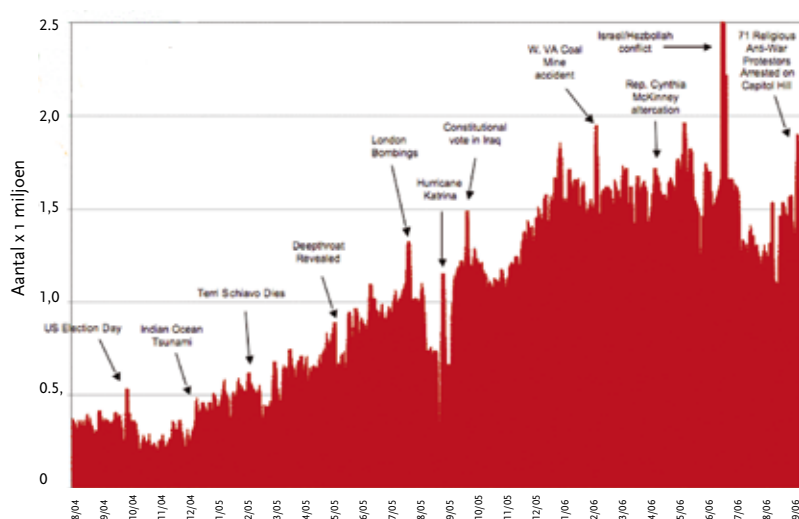
Petitie voor Open Access

Actieve steun voor Open Access en opnemen van de eis dat de rapportage in een open repository wordt opgenomen in de regels voor het financieel ondersteunen van onderzoek. Dat waren twee van de wensen in een petitie die op 15 februari is aangeboden aan EU-commissaris Potochnik (Science & Research).

Op het moment van aanbidding stonden er ruim 21 duizend handtekeningen onder de petitie, waaronder 500 namens instituten en organisaties voor wetenschappelijk onderzoek. De KNAW en de Deutsche Forschungsgemeinschaft tekenden ook.

De petitie verwees naar een rapport van de Commissie uit 2006 met aanbevelingen over duurzaam bewaren en openbaar maken van onderzoeksresultaten, en naar een advies van de onderzoeksadviesraad EURAB in december. Initiatiefnemers van de petitie waren onder meer SURF en het Engelse JISC.

Blogosfeer blijft groeien



De wereld van de weblogs – de 'blogosfeer' – blijft groeien. Het aantal weblogs verdubbelt in minder dan een jaar. Het bedrijf Technorati volgt deze ontwikkelingen op de voet. Het indexeert 60 miljoen weblogs en rapporteert elk kwartaal over de belangrijkste trends. De grafiek toont de groei van het aantal berichten per dag

en de relatie met de (Amerikaanse) actualiteit: het aantal berichten piekt bij verkiezingen, een oorlog, of een bomaanslag. Overigens is inmiddels het Japans in de blogosfeer bijna even belangrijk als het Engels. Het aantal Nederlandse weblog omvat minder dan één procent. (PB)

Bijbelexegeet Eep Talstra:

‘Begin nou eens niet met een mening maar gewoon met de data’

MARTIJN DE GROOT

Professor Eep Talstra voerde aan de Vrije Universiteit de studie van het Oude Testament met computerhulp tot grote hoogte. Hij ziet de teksten als data, en de analyse daarvan als het schiften tussen taalkundige en literaire verklaringen.

Stel voor: je fietst of wandelt langs de Waddenzeedijk, aan de zuidkant van het eiland Terschelling, over een verhard pad dat je moet delen met schapen en met andere toeristen. Je nadert een hek dat zich uitstrekt van de kruin tot de voet van de dijk, met een opening voor fietsers en wandelaars. Dan lees je de woorden die daarop zijn geschilderd: *Schapen op de dijk, honden aan de lijn.*

Veel mensen zouden het hek

‘Veel projecten stortten in omdat onze alfa’s de overstap van ponskaarten naar diskette en van mainframe naar pc niet konden maken’

gedachteloos zijn gepasseerd of hun hond hebben aangeliend, al of niet met aandrang tot een glimlach. Zo niet oudtestamenticus Eep Talstra. Hij proeft de literaire kwaliteit van het opschrift, betreurt het dat hij geen camera bij zich heeft, neemt de tekst dan maar in gedachten mee terug naar het vasteland en legt hem daarna regelmatig aan zijn studenten voor: ‘Hoe komt het nou dat wij snappen dat de ene zin een voorwaarde is voor de andere? En wat zou er met die relatie gebeuren als er achter had gestaan: *vogels in de lucht?*’

Het is het soort vragen dat Talstra dagelijks stelt, maar dan met het Oude Testament als onderwerp. Eep Talstra studeerde in de jaren zestig en zeventig aan de Vrije Universiteit: theologie en semitische talen. De eerste studie rondde hij af, in de laatste bracht hij het tot wat toen heette zijn kandidaats. Beide richtingen bleven zijn werk beheersen, samen met een grote belangstelling voor computers. Eenmaal afgestudeerd kreeg hij een halve aanstelling om de mogelijk-

Prof. dr E. Talstra (1946) is mede-maker van de op CD-ROM verschenen Stuttgart Electronic Study Bible (SESB), waarvan onlangs de tweede persing verscheen. Hij promoveerde in 1987 op een proefschrift over Het Gebed van Salomo en kreeg daarvoor de tienjaarlijkse Mallinckrodtprijs van de Universiteit van Groningen. In 1991 werd hij bijzonder hoogleraar aan de VU; in 2002 gewoon hoogleraar Oude Testament.

heden van de computer binnen de theologische faculteit te verkennen (‘We werkten alleen met ponskaarten in die tijd. Later zijn veel projecten ingestort omdat de alfa’s in de theologie niet over de kennis beschikten om de overstap van ponskaarten naar tapes en van mainframe naar pc te maken’). Die aanstelling veranderde in de loop van de tijd van aard en omvang en groeide tenslotte uit tot een gewoon hoogleraarschap Oude Testament.

Tegenwoordig huist Talstra met twee collega’s op een kamer op de veertiende verdieping van zo’n grijze VU-betonkolos in Amsterdam Buitenveldert, die nog helemaal de kale, ‘eerlijke’ bouwstijl ademt van de tijd waarin hij afstudeerde. Wie hem opzoekt voor een gesprek hoeft niet lang op de eerste Bijbeltekst te wachten. Het begint met Exodus 15:1 en dan heeft de interviewer nog maar net plaatsgenomen op de tot de draad versleten bureaustoel van een afwezige kamergenoot. Talstra: ‘Het klasieke voorbeeld van discussie over de uitleg is dat van het loflied van Mozes en Miriam, nadat de Israëlieten de Schelfzee hebben overgestoken. In de vroege rabbijnse commentaren was er al discussie over de beginwoorden: ‘Toen zong Mozes’, omdat je die net zo goed kunt lezen als: ‘Dan zal Mozes zingen’. Bevatten die woorden een verwijzing naar de opstanding of niet?’

Het is maar één voorbeeld. De geschiedenis van de Bijbel is er één van discussie, interpretatie, en van edities en vertalingen die tegelijk de

‘Onze aanpak beperkt de ruimte voor heel veel interpretaties’

oorspronkelijke tekst en de gedachten daarover van de vertalers en verspreiders lijken te weerspiegelen.

Talstra: ‘Lange tijd, vooral toen de computer voor het eerst een rol kreeg in de Bijbelstudie, was het streven van exegeten alleen maar om terug te komen tot ‘het echte’: wat heeft er nu oorspronkelijk gestaan? Alles wat er in de loop der tijd was bijgekomen wilde men er weer af schillen. Eigenlijk kon dat niet want we beschikken niet over voldoende eenduidige en betrouwbare bronnen om het te kunnen weten. Dat gaf weer ruimte aan een andere stroming in de tijdgeest van de jaren tachtig en negentig: alle



DIEDERIK VAN DER LAAN

toevoegingen hebben hun eigen zin en betekenis, alles vertegenwoordigt een belang van een latere bewerker en het heeft dus geen zin om naar de oorspronkelijke tekst te zoeken. Elke letter staat op z’n plaats waar hij nu staat.

Naar Talstra’s zin ontkent die laatste denkrichting weer teveel de eigen, inherente betekenis van de tekst: ‘In feite zegt die stroming: iedere interpretatie is goed’. Het heeft dus in zijn ogen wel degelijk zin om te zoeken naar de bedoelde betekenis, en dat is waar de bijzondere benadering van de Amsterdamse hoogleraar in beeld komt. Hij haalt de computer erbij en wel op een schaal die zijn vakbroeders soms doet fronsen. ‘Ik wil van die subjectiviteit af. Ik praat in de theologie, tot verbijstering van veel collega’s, over data – niet over gedachten of dromen. Tegen studenten zeg ik: begin nou eens niet met een mening maar gewoon met de feiten. Dat zijn dus de data’.

‘De teksten in de Bijbel hebben de gelaagdheid van verschillende bewerkingen door de tijd heen. Met een goede, ambachtelijke methode kan je de jacht op het oorspronkelijke

niet winnen, maar je kan wel die gelaagdheid ontdekken en benoemen. Computerbewerking is daarbij heel nuttig’. Opnieuw een voorbeeld. Jeremia is een gevangene in de door Babyloniërs belegerde stad Jeruzalem, want zijn profetieën zijn de heersers niet welkom. Koning Sedekia laat hem een boodschap sturen en vraagt: waarom profeteert jij: ‘jij wordt gevangen genomen en weggevoerd, en jullie zullen de strijd verliezen?’ Talstra:

‘Als je iets taalkundig kunt verklaren dan is daar niets diepzinnigs aan’.

‘Opeens gaat daar in de Hebreeuwse tekst de adressering over van *jij* naar *jullie*. En nu is de vraag: waarom die overgang? Is het een normale variant in de Hebreeuwse taal van die tijd of is het een literaire variant, dat wil zeggen met een door de schrijvers bedoelde bijzondere betekenis?’

Het onderscheid tussen taalkundige en literaire analyse komt regelmatig terug als Talstra de bedoeling van zijn computermatige benadering beschrijft. Die houdt te grondige, gelaagde ontleding van de Hebreeuwse

tekst in, ‘van de kleinste eenheden, de voor- en achtervoegsels, via woorden, woordgroepen en zinnen naar de tekstopbouw en het volgen van personen door de tekst heen’. Daarop volgt een structuuranalyse, die onderzoekt of een bepaalde constructie, in dit geval de overgang van enkelvoud naar meervoud, vaker voorkomt. Talstra: ‘Is dat het geval, dan hebben we waarschijnlijk te maken met een variant die uit de regelmaat, dus taalkundig verklaard kan worden. Daar is niets diepzinnigs aan. Is het niet zo, dan hebben we te maken met het bijzondere, waarvoor we dan naar een bedoelde verklaring kunnen gaan zoeken’. In het geval van Jeremia zou die eruit kunnen bestaan dat de tekst dateert van na de val van Jeruzalem, en dat de schrijver zich met de meervoudsvorm in feite richt tot de overgebleven bewoners van de stad of tot nog latere generaties.

De uitdaging van zijn type analyse is volgens Talstra: ‘Kan ik een databank formuleren waarbij ik een maximum aan taalkundige categorieën kan opslaan? Die machine moet zo ver komen dat hij rapporteert: er zit een merkwaardige tweede persoon meervoud in deze tekst die verder geen naam heeft’.

In de Bijbelexegese bestaat per traditie de neiging tot ‘maximaliseren van de betekenis’, zegt de Amsterdamse geleerde. Maar is zijn computermethode niet juist het omgekeerde? Wordt de Bijbel daardoor niet teruggerekend tot een verzameling formules zonder betekenis? Daarvoor is Talstra in het geheel niet bang: ‘Onze aanpak beperkt de ruimte voor heel veel interpretaties. Dus ik kan bijvoorbeeld van de vijf fantasievolle mogelijkheden bij een bepaalde passage er drie terzijde leggen. Maar er blijft genoeg ruimte over voor literaire analyse’.

Die ruimte kan sinds enkele jaren beter worden gedefinieerd met behulp van de Stuttgart Electronic Study Bible (SESB), waarvan enkele weken geleden een nieuwe versie verscheen. Voor dit product van samenwerking tussen onderzoekers van Duitse, Nederlandse en Amerikaanse universiteiten werd een beroep gedaan op de door Talstra en zijn collega’s ontwikkelde Hebreeuwse databank van het Oude Testament. Daaromheen zijn verschillende oude en nieuwe vertalingen van het Oude Testament opgenomen, woordenboeken ‘en van alles’, aldus Talstra. De bibliotheekfunctie van de computer, die lange tijd de enige was waar theologen raad mee wisten, gecombineerd met de analytische functie die immers veel meer te bieden heeft: ‘Een mooi pakket voor Bijbelbestudeerders’.

Is Talstra er zelf ook mee aan het werk? De professor glimlacht veelbetekenend: ‘Ik ben niet zo zeer een gebruiker maar meer een ontwikkelaar...’

'De vreugde van het publieke domein verspreiden'

Zolang er bibliotheken bestaan, en auto's, zullen er waarschijnlijk ook rijdende bibliotheken zijn geweest. Fotografisch bewijs is er in elk geval van een *Book Truck* die in de jaren twintig rondom het Amerikaanse Lexington de leeshonger van de platelandsbewoners hielp stillen (<http://ils.unc.edu/nclibs/davidson>). Het verschijnsel is sindsdien niet meer verdwenen, in leven gehouden door



De Book Truck in Lexington, 1928

de overtuiging van bibliothecarissen en subsidiegevers dat er altijd wel mensen aan het lezen gebracht kunnen worden die de bibliotheek niet weten te vinden. Afgelopen zomer waren de rijdende bibliotheken weer te zien op de stranden van Den Haag, Katwijk, Monster en Brouwersdam. In Amerika heet zo iets een *bookmobile*. Een heel nieuw dimensiewerd erdaar aan toegevoegd door het internet, dat



Vrouw met verse print van *The Wizard of Oz* in Washington DC

alles en ook dit overhoop haalde.

De mogelijkheden van het internet maken immers dat niet alleen uitgeverij boeken kunnen verspreiden, al of niet via de bibliotheek. In de wetenschappelijke sfeer woedt de discussie daarover dezer dagen nog in alle hevigheid.

Het werd dus mogelijk om met een Bookmobile de scholen en wijken te gaan opzoeken en daar geen boeken uit te lenen, maar ze ter plaatse van het internet te plukken, uit te printen, te binden en mee te geven aan leeshongerige kinderen. Boeken uit het auteursrechtvrije Public Domain vanzelfsprekend, want het was niet

de bedoeling om de wet te overtreden. Wel trouwens om die wet ter discussie te stellen. De eerste Internet Bookmobile, een initiatief van



De Bookmobile in San Fransisco

Internet Archive-baas Brewster Kahle zelf, legde een route af vanuit Californië dwars door de Verenigde Staten met als eindpunt Washington DC op 9 oktober 2002. Daar was op dat moment het spraakmakende proces *Eldred v. Ashcroft* bezig, met als inzet de duur van de periode waarin het auteursrecht op een boek beschermd is. 'Elke keer wordt die periode verlengd, waardoor er minder boeken in het publieke domein komen. Zonder dat soort boeken kan er ook geen digitale boekmobiel zijn' legt de website van het Internet Archive uit (www.archive.org/texts/bookmobile.php). En met aanstekelijk enthousiasme: 'Om te vieren dat we het Publieke Domein hebben komt de Internet Bookmobile binnenkort naar een plaats bij u in de buurt, met alle mogelijkheden om een van de bijna 20 duizend boeken in het publieke domein te downloaden, uit te printen en te binden'.

Na die eerste demonstratieve tocht volgt een periode van ups en downs. De ups gekenmerkt door groot enthousiasme tot bijna religieuze gedrevenheid, de downs waar-



Archive-baas Brewster Kahle zelf is vaak standwerker bij de Bookmobile

schijnlijk veroorzaakt door gebreken van vrijwilligers – en geld. Want ook boeken in het publieke domein moeten uiteindelijk echt gemaakt worden. De Bookmobile lijkt af en toe in de garage

te verdwijnen maar komt er steeds weer uit tevoorschijn, begeleid door aankondigingen op de Archive-site: 'Hij rijdt weer!'... 'Na een stille periode is het weer zo ver: de Bookmobile komt er aan!' Toch lijkt de Californische Bookmobile zich steeds vaker te beperken tot de directe omgeving. Het Golden Gate park is een populaire bestemming. Bookmobiliste Qirsten in een van haar laatste postings op 2 december 2005: 'Come and kick it with the Bookmobile! We'll be making books, doing jigs, and spreading general public domain merriment'.

Intussen zijn er initiatieven geweest om Bookmobiles de weg op te krijgen in gebieden in de wereld waar de kinderen minder zakgeld krijgen dan rond de San Fransisco Bay. Met geld van het Internet Archive en de Wereldbank start er een mobiel in het Egyptische Alexandrië – een plek van symbolische betekenis omdat in deze stad immers in de oude tijd



De bibliotheek van Alexandrië, naar verluidd uitvalsbasis van een Bookmobile

de beroemde Koninklijke Bibliotheek van Ptolemaeus had gestaan die het hele Middellandse-Zeegebied tot voorbeeld had gestrekt. Ook in India (<http://mobilelibrary.cdacnoida.com>) en Uganda wordt er een mobiel gelanceerd, maar het internet biedt anno 2007 geen bewijs dat die twee nog in bedrijf zijn. Op informatievragen per email komt in elk geval geen antwoord. Kahle weet het ook niet. Tegenover e-data@research antwoordt hij per mail het lot van de derde-wereldinitiatieven niet te kennen. De Open Content Alliance (www.opencontentalliance.org) is bezig, onder meer als antwoord op de ervaringen in Uganda, in hoog tempo boeken te scannen die door mobiles kunnen worden afgedrukt. Kahle: 'The idea is still in the air. We are now filling in the pieces we found missing so far'.

komen van leidende figuren in het veld, veelal geassocieerd met het Text Encoding Initiative; anderen hebben hun sporen verdiend in editieprojecten als die rond de Canterbury Tales (Peter Robinson) of Ludwig Wittgenstein (Claus Huitfeldt). Meegeleverde CD bevat elektronische versie van TEI-Guidelines.

<http://www.mla.org/store/>

Haan, J. de., R. Mast, M. Varekamp, S. Janssen: *Bezoek onze site, over de digitalisering van het culturele aanbod* (Den Haag, SCP) ISBN 978 90 377 02903

Steeds meer culturele instellingen maken hun culturele rijkdom door

Column

Gert van der Steen

Stof of vuur

In de computerkamer was het koud. Ik hing over een lage kast waarin een grote computerschijf ronddraaide. Lekker warm, met het knorrende geluid van de schijfarm die continu heen en weer schoot. Ik wist dat hij onze gebruikers probeerde bij te houden om ze te voorzien van een niet aflatende stroom van data.



Data afkomstig van romans, kranten teksten, corpora van allerlei slag, hettitische kleitabletten, muziek en historische boedelbeschrijvingen. Opeens voelde ik dat het niet zomaar bytes waren, maar vurige bits in ijzer gegrift. In staat om zich een weg te branden naar de toekomst.

Zijn onze dataverzamelingen het overgebleven stof van researchprojecten? Of zit er leven in? Laten we het laatste eens aannemen - dit is tenslotte een column. Dan mag je er vanuit gaan dat er in al die data een zweem van individualiteit zit. Die zich zal proberen in stand te houden en zich te ontwikkelen, met alle bijkomende problemen. Maar ook met alle potentie voor nieuwe ontwikkeling. Wat kun je dan met die metafoer doen?

Laten we zo'n individu eens Cor (pus) noemen. Cor redt het vege lijf door dagelijkse back-ups. Hij probeert zichzelf te ontwikkelen met annotaties voor structuur, morfologie, syntax en semantiek. Hem worden allerlei vragen gesteld, maar er zijn er meer dan hij kan beantwoorden. Cor probeert zich verder te ontwikkelen, maar de mensen kunnen hem niet veel meer leren. Ze maken fouten, de ene keer vertellen ze hem iets anders dan een andere keer, en op een kwade dag hebben ze geen tijd meer.

Maar dan komt er een dag waarop Cor zelf op weg mag gaan en kennismaken met collega's. Ze zien er anders uit en praten over andere zaken. Maar ze hebben een gemeenschappelijke taal waarin ze met elkaar kunnen uitwisselen. De mensen leggen Cor een plan voor om zich verder te ontwikkelen. En zo gaat Cor zelfstandig op bezoek bij vele collega's om zijn kennis aan te vullen.

Is dit een utopie? Een aantal technische hulpmiddelen is er. De wereld barst van zelfstandig met elkaar communicerende systemen. In mindere mate ook van losstaande woordenboeken en grammatica's, collega's van Cor. Als je Cor nu ook kleur, smaak en gevoel geeft dan krijg je dataverzamelingen die niet alleen de wetenschap dienen maar óók het grote publiek. Dat kan dan bijvoorbeeld beelden uit het verleden reconstrueren. Daarbij spelen wetenschappelijke beperkingen een kleinere rol. Leven gaat van nature om met onvolledigheid, met fouten en inconsistenties. Ik denk dat Cor zich meer gewaardeerd gaat voelen.

Gert van der Steen is consultant Documentmanagement en Taaltechnologie en oud-hoogleraar Uitwisselingstalen

digitalisering voor een breder publiek toegankelijk via het internet. Nederland neemt hierbij internationaal een vooraanstaande positie in. Dit rapport geeft een overzicht van de vele en diverse initiatieven in Nederland. Binnen de sectoren kunsten, cultureel erfgoed, bibliotheken en publieke omroep wordt steeds met perioden van

ongeveer vijf jaar achterom en vooruit gekeken. Het rapport biedt tevens inzicht in kansen en knelpunten. Met leerlingen als belangrijke doelgroep wordt verder apart aandacht besteed aan de digitalisering voor het onderwijs. <http://www.scp.nl>

SURF-platform ICT en Onderzoek: Jaarplan 2007 Platform ICT en Onderzoek

Met het DARE-programma 2003-2006 is de basis gelegd voor het SURFshare-programma 2007-2010. Waar SURF bij het vorige DARE-programma nog onderscheid maakte tussen het DARE-programma en overige activiteiten van het Platform ICT en Onderzoek, omvat SURFshare nu alle activiteiten van het Platform. Hiermee wordt gekozen voor meer samenhang en staan activiteiten van het Platform, zoals het ondersteunen van onderzoekers bij het uitoefenen van hun auteursrecht, in dienst van het gehele SURFshare-programma. <http://www.surf.nl/>



Gelezen

Neef, S., J. van Dijk, E. Ketelaar (eds.): *Sign here! Handwriting in the Age of New Media* (Amsterdam, AUP) ISBN 978 90 5356 8163

Zal het handschrift verdwijnen in het tijdperk van de digitale media? En wat gebeurt er dan met culturele en juridische concepten als originaliteit, imitatie, authenticiteit, reproductie, uniciteit en herhaling? Waar moeten we de schrijvende hand lokaliseren als het schrijven niet 'handmatig' wordt uitgevoerd maar door een tweede medium? En wat is de relatie van het handschrift tot dit tweede medium waarin het wordt afgebeeld, gereproduceerd of misschien zelfs verbeterd? De auteurs behandelen deze en an-

dere vragen over de toekomst van het handschrift. <http://www.aup.nl/>

Lou Burnard, Katherine O'Brien O'Keefe, John Unsworth (eds.): *Electronic Textual Editing: Modern Language Association of America; Bk&CD-ROM edition* ISBN 978-0873529716

Dit eerste handboek belooft een onmisbaar naslagwerk te worden voor de praktijk van de elektronische teksteditie. Het bevat de 'Guidelines for Editors of Scholarly Editions' van het MLA Committee on Scholarly Editions, plus een keur van artikelen over praktische en meer beschouwelijke aspecten van wetenschappelijke teksteditie. De meeste bijdragen