

Pagina 3 • NWO wil ook betere infrastructuur • Een betere data-infrastructuur voor de alfa- en gammawetenschappen hoort bij de NWO-strategie voor de komende jaren.

Pagina 4 • Letters, woorden en teksten • Het internet brengt efficiënte communicatie tussen taal- en literatuuronderzoekers en afstemming van onderzoeksinstrumenten binnen bereik.

Pagina 5 • Het verhaal bij de tabellen • Geen kwantitatieve aanpak zonder kwalitatieve dimensie, vindt historisch demograaf prof. Theo Engelen.



EN VERDER

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

KNAW: 'Betere toegang tot administratieve data'

Wetenschappelijk onderzoekers moeten betere toegang krijgen tot gegevens uit publieke en semi-publieke administraties. Als dat zou gebeuren zou dat niet alleen voordeel opleveren voor het onderzoek zelf maar ook voor bestuur en beleid. Dat blijkt uit een advies dat de Sociaal-Wetenschappelijke Raad van de Koninklijke Nederlandse Akademie van Wetenschappen eind juli publiceerde.

Het advies noemt met name de zes 'basisregistraties' met vitale gegevens van de overheid: gebouwen, adressen, kadaster, topografie, persoonsgegevens en handelsregister. Maar ook andere bronnen kunnen van belang zijn, zoals van de werknemersverzekeringen, de Belastingdienst en de InformatieBeheergroep.

Aard en omvang van de wetenschappelijke behoefte in kaart te brengen is een 'majeure uitzoekklus' volgens het advies, dat wijst op de landelijke data-agenda waaraan door DANS wordt gewerkt.

Het advies werd opgesteld op verzoek van het ministerie van Onderwijs, Cultuur en Wetenschap, dat aan een binnenkort te verschijnen nota over wetenschapinformatie werkt. Het ministerie zou over die data-agenda contact op moeten nemen met DANS, schrijft de Raad.

Digitale coalitie op komst

Er moet in Nederland een Nationale Coalitie voor Digitale Duurzaamheid komen. Daarover is een aantal grote partijen op het gebied van digitalisering en digitale archivering het eens geworden. De oprichtingsvergadering van de coalitie wordt verwacht in september of oktober van dit jaar.

Naast de initiatiefnemers Koninklijke Bibliotheek (KB) en DANS (Data Archiving and Networked Services) hebben organisaties als NWO, KNAW en SURF zich bij het initiatief aangesloten. Ook de drie technische universiteiten, die zich in een federatie hebben verenigd, en het KNMI doen mee. Met een aantal andere beoogde partners wordt nog gepraat.

De coalitie, waarvoor de Engelse *Digital Preservation Coalition* (DPC) model heeft gestaan, gaat streven naar het ontwikkelen en verspreiden van kennis, nationale en internationale samenwerking en profile-

ring van Nederland als expertiseland. Dat staat in een startnotitie die door de deelnemers is besproken. Het kan daarnaast een platform worden voor de uitwisseling van standaarden. Ook zou de coalitie het aanspreekpunt kunnen worden voor de overheid als het gaat om beleid op het gebied van digitale duurzaamheid.

Directeur Research en Development van de KB Hans Jansen bevestigt dat de voornemens wat hem betreft 'hard' zijn: 'Het staat nu ook zwart op wit in het beleidsplan van de KB'.

De coalitie moet een aantal 'zwa-

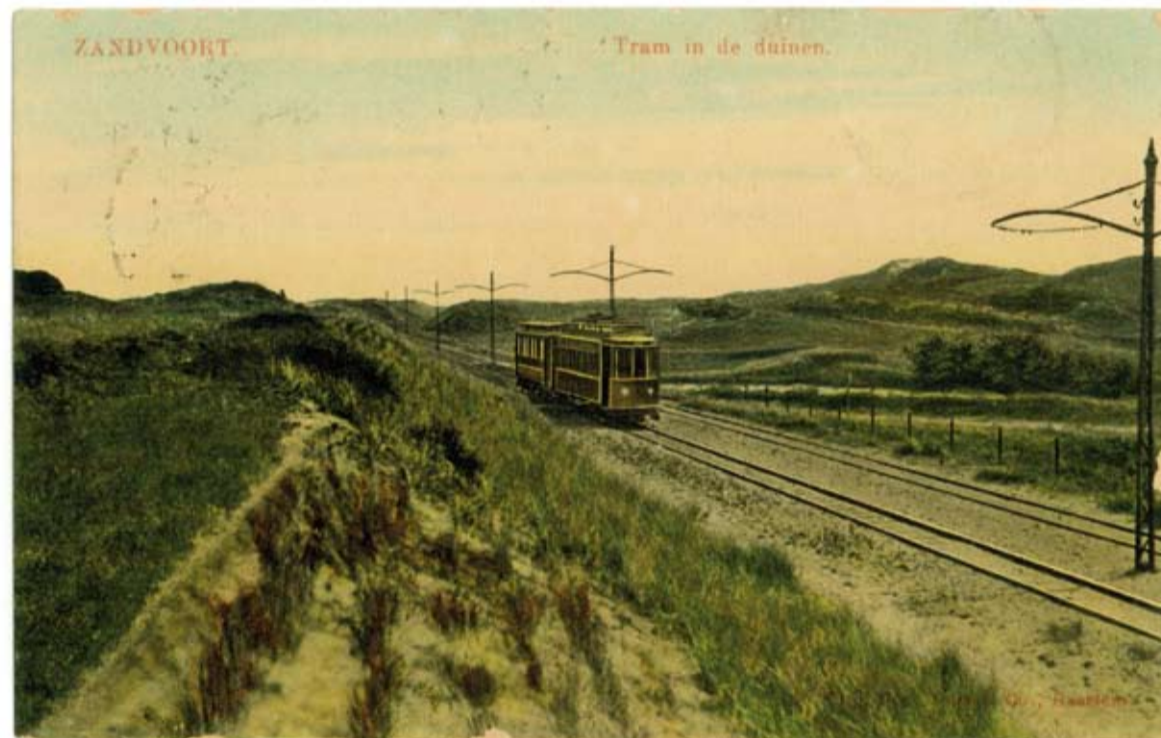
re jongens' samenbrengen, licht Jansen toe. 'Het gaat om organisaties of instellingen die ook echt zelf een belang hebben bij het duurzaam bewaren van hun spullen of die van anderen, en die in hun activiteiten en beleid op het gebied van digitaliseren en archiveren al een aantal stappen hebben gezet'. Die kunnen zich naar de overheid uitspreken over prioriteiten en ontwikkelingen die om financiering vragen. Zo wil de KB-directeur de overheid aanspreken om een onderzoek te laten doen naar 'wie wat bewaart in Nederland en hoe dat gaat. Er wordt op alle mogelijke plekken

informatie opgeslagen en beschikbaar gesteld; dat is nog onvoldoende uitgekristalliseerd'. Een bureau zou dat onderzoek uit kunnen voeren, de overheid moet het betalen en de coalitie zou als begeleidingscommissie kunnen fungeren, aldus Jansen. 'De DPC heeft een soortgelijk onderzoek laten doen, alleen nog breder. Daar waren ook data uit de industrie en van allerlei overheden bij betrokken. Ik bepleit alleen een studie naar de data die in opdracht en onder auspiciën van de overheid worden vervaardigd'.

De coalitie moet geen groot lichaam worden maar zal wel meer omvatten dan alleen een regelmatig overleg, als het aan Jansen ligt. 'Het bestuur zou uit vertegenwoordigers op directieniveau moeten bestaan. Daaronder komen werkgroepen te hangen die nadenken en afspraken maken over bijvoorbeeld technische standaarden en keuzes die daarin gemaakt moeten worden. Maar daar wordt ook kennis uitgewisseld: hoe doen jullie het, kan het ook efficiënter, kunnen we niet één systeem in gebruik nemen?'

Net als de DPC moet ook de Nederlandse coalitie een klein coördinerend bureau en een lidmaatschapsstructuur krijgen, met full members, associate members en individuele leden die naar rato van omvang en betrokkenheid betalen en beslissen. (MdG)

De eerste vakantie in Zandvoort



Leden van de vakbond van diamantbewerkers stuurden deze kaart in 1910 naar hun bestuur als dank voor de inspanningen die leidden tot hun eerste week betaald verlof. Bijna honderd jaar later biedt het Internationaal Instituut voor Sociale Geschiedenis aan bezoekers van zijn site de mogelijkheid om de kaart bij wijze van vakantiegroet als e-card te versturen. De Zandvoortse zomer was afgelopen maanden deel van een groter aanbod, dat elk jaar wat wordt aangepast.

Ⓜ <http://www.iisg.nl/collections/summerholidays.php>

Steun voor acht Kleine Dataprojecten

Data-instituut DANS heeft besloten om dit jaar acht Kleine Dataprojecten financieel te ondersteunen. Vier daarvan komen uit de sociale wetenschappen, de andere vier uit de geesteswetenschappen.

Het gaat om verkenningen, haalbaarheidsstudies en dataprojecten op het gebied van data-ontsluiting en archivering, waaraan DANS maximaal 10.000 euro bijdraagt. In de sociale wetenschappen zijn bijdragen toegezegd aan een archiveringsproject rond 'Jongeren en Cultuur' van de Vrije Universiteit, een project rond informatiesysteem voor opleidingsclassificaties van de Universiteit Maastricht, een Arbeidsvraagpanel

van de Universiteit van Tilburg en *Explanatory Notes for the Wageindicator* (Loonwijzer) data van de Universiteit van Amsterdam. In de geesteswetenschappen ging het om onderzoek naar historische beeldbronnen van de Radboud Universiteit, omeen project rond de ontsluiting van beeldmateriaal bij antropologisch onderzoek aan de Universiteit Leiden, een archeologisch project aan dezelfde universiteit en een haalbaarheidsstu-

die naar archivering en ontsluiting van geanalyseerde tekst van de Hebreeuwse Bijbel (Vrije Universiteit).

DANS noemt het Loonwijzerproject als een goed voorbeeld van een Klein Dataproject. In dit project van prof. Kea Tijdens van het Amsterdams Instituut voor Arbeidsstudies (AIAS), worden de 632 variabelen in de *Wageindicator* dataset meer gede-

tailleerd beschreven dan in de huidige codeboeken. Het gaat om gegevens uit een doorlopende survey, die sinds 2000 in Nederland en sinds 2004 in een groeiend aantal andere landen wordt gehouden. Eind 2006 zijn dat naar verwachting 17 landen. In totaal waren er eind 2005 ongeveer 250.000 waarnemingen. De *explanatory notes* bevatten uitleg over de inhoud en het belang van variabelen, de veranderingen in de tijd en de syntax waarin ze berekend worden. Omdat het project de toegankelijkheid van de dataset op termijn garandeert voldoet het aan het criterium 'duurzaamheid' van de subsidievoorwaarden. Volgend jaar komt er een nieuwe aanvraagronde Kleine Dataprojecten. (Heleen van Luijn)

Ⓜ www.dans.knaw.nl/nl/data_deponeren/kleine_projecten/

AGENDA

22 september

Studiemiddag "Het gebruik van 'social software' in historische context" en uitreiking VGI Innovatieprijs 2006
Vereniging voor Geschiedenis en Informatica
www.vgi-online.org

27-29 september 2006

Amsterdam, Internationaal Instituut voor Sociale Geschiedenis – IISG/KNAW
Workshop: Virtual Ethnography in Contemporary Social Science
Virtual Knowledge Studio, KNAW
This workshop focuses on challenges posed to ethnographic methods by the new digital and electronic media. It will be of interest to scholars engaged in ethnographic research which touches on digital media as well as other social scientists studying globalization and cultural change, digital communication, or cultures of digital media
<http://www.virtualknowledgestudio.nl/en/workshops/>

29 september 2006

Den Haag, NWO
Symposium Life Courses in Context – Digitale Nederlandse Volkstellingen 1795-2001
Centraal Bureau voor de Statistiek, DANS (Data Archiving and Networked Services)
Symposium ter gelegenheid van de afronding van het volkstellingen-gedeelte van het NWO-Groot digitaliseringsproject Life Courses in Context
www.volkstellingen.nl/nl/onderzoek_literatuur/collaboratory/studiedag_2006/

6 oktober 2006

Den Haag, Universiteit Leiden
Symposium: Hollanders op drift
Instituut voor Geschiedenis Universiteit Leiden, Historisch Tijdschrift Holland, Historische Steekproef Nederland – IISG/KNAW
Migratie en demografische ontwikkelingen in Holland in de 19e en 20e eeuw
<http://www.iisg.nl/~hsn/news/hollandersopdrift.php>

12 oktober 2006

Den Haag, NWO
Seminar: access panels en online onderzoek, panacee of slangenkuil?
Data Archiving and Networked Services, Sociaal en Cultureel Planbureau
Seminar om kwaliteit van en informatie over panels te bevorderen, en om meer wetenschappelijk onderzoek naar de kwaliteit van panels te genereren.
<http://www.dans.knaw.nl>

6 en 7 november 2006

Universiteit Utrecht
Congres: Learned Love, Dutch Love Emblems on the Internet
Onderzoeksinstituut voor Geschiedenis en Cultuur van de Universiteit Utrecht, Digitale Bibliotheek der Nederlandse Letteren, Data Archiving and Networked Services, Huygens Instituut
Congres ter afronding van het Emblem Project Utrecht (EPU) over de centrale thema's van het project: het Nederlandse liefdesembleem en de digitalisatie van embleembundels.
<http://emblems.let.uu.nl/emblems/congres.html>

15-17 november

Amsterdam, IISG – Internationaal Instituut voor Sociale Geschiedenis
Child Labour's Global Past 1500-2000
This Conference aims to trace and discuss the historical development of child labour from a global perspective, and brings together historians from all over the world who have been working on these issues for a particular area.
<http://www.iisg.nl/research/childlabour.php>

23-25 november 2006

London
Conference: Textual scholarship and the material book: comparative approaches
European Society for Textual Scholarship
Trends and developments in textual scholarship across the various national regions and historical periods.
www.textualscholarship.org/ests/conference06/index.html

Digital Humanities 2006

Aandacht verschuift van tekst naar omgeving

Kan de computer leren om sentimenteel of erotisch taalgebruik te herkennen? Dat is één van de uitdagingen die wordt aangegaan in het Amerikaanse NORA project. NORA maakt gebruik van *text mining*, een trend in de geautomatiseerde analyse van literaire teksten. Daarbij wordt de computer ingezet om aan grote aantallen teksten een score toe te kennen op een kenmerk als bijvoorbeeld sentimentaliteit. De computer leert dat kenmerk herkennen aan de hand van een aantal door de onderzoeker handmatig toegekende scores. Daarna kunnen verschillende vormen van visualisatie interessante verbanden suggereren tussen sentimentaliteit en bijvoorbeeld sekse of leeftijd van auteur, geografische achtergrond, sociale status, etc. De bedoeling is uitdrukkelijk om te exploreren, te suggereren, te provoceren – traditioneel, 'gedegen' onderzoek zal het daadwerkelijk bestaan van dergelijke verbanden vervolgens aannemelijk moeten maken.

Het NORA project kwam prominent naar voren tijdens de Digital Humanities conferentie, gehouden in Parijs van 5 tot 9 juli. De conferentie, voorheen bekend als ALLC/ACH,

is de belangrijkste verzamelplaats van wat bekend staat als 'computing humanists', geesteswetenschappers die computerhulpmiddelen ontwikkelen voor de analyse van teksten en andere cultuuruitingen. Traditioneel zijn de hulpmiddelen voor tekstanalyse (auteursherkenning, stilistische analyse) goed vertegenwoordigd, net als de projecten die teksten, afbeeldingen of ander materiaal via het web toegankelijk maken.

Voor de laatste soort projecten wordt het leven wat eenvoudiger dankzij de nieuwe versie van de veel gebruikte richtlijnen voor tekstcodering van het Text Encoding Initiative (TEI). Met het systeem Roma kunnen projecten de door het TEI aangeleverde basis aan de eigen wensen aanpassen. Het wordt mogelijk om documentatie in de eigen taal daaraan toe te voegen en deze direct oepbaar te maken bij de bewerking van documenten. Projecten die teksten ter beschikking stellen over het web houden het overigens steeds vaker niet bij tekst alleen. Het contextualiseren van informatie wordt steeds belangrijker, en vaak betekent dat integratie met kaartmateriaal, afbeeldingen en tijdbalken. Het Cervantes

project verzamelt zelfs sigarenbandjes die Don Quichot afbeelden. In eerste instantie wekt het verbazing, maar uiteindelijk zijn ook sigarenbandjes een aanwijzing voor de ma-

nier waarop een literair werk wordt gelezen. De aandacht in humanities computing verschuift zo steeds meer van de teksten naar hun omgeving. (PB)

<http://www.allc-ach2006.colloques.paris-sorbonne.fr/DHs.pdf>

IASSIST 2006: documentatie en gebruiksvriendelijkheid

Metadata en gebruiksvriendelijkheid waren belangrijke onderwerpen op de IASSIST conferentie van 23 tot 27 mei in Ann Arbor (VS). IASSIST is een jaarlijkse internationale bijeenkomst voor informatietechnologen in de sociale wetenschappen. Ditmaal was het thema *Data in a World of Networked Knowledge*.

Naast workshops op het gebied van het Data Document Initiative (DDI), Geographical Information Systems (GIS) en het opzetten van een on line data-archief, kon de bezoeker een breed scala aan presentaties bijwonen. Een onderwerp dat in verschillende sessies aan bod kwam, was het belang van metadata en de toepassing van DDI. Sprekers benadrukten dat metadata een essentiële rol spelen bij het lokaliseren, vinden en identificeren van data. Bij de selectie van geschikte metadata-elementen dient in het oog te worden gehouden welke elementen gebruikers helpen om succesvol naar data

te zoeken. Een goede oriëntatie op gebruikerswensen kan hierbij behulpzaam zijn. De nieuwste versie van DDI speelt onder andere hier op in.

In de sessie *Effective Design for Data-rich Websites* kwam duidelijk naar voren dat in een tijd waarin steeds meer instituten en organisaties hun data on line beschikbaar maken, vooral de gebruiksvriendelijkheid van websites goed in de gaten moet worden gehouden. Dit geldt in principe voor alle soorten websites, maar bij het ontwerpen van websites waarop grote hoeveelheden data worden ontsloten, gelden aanvullende eisen ten aanzien van het gebruik, de beveiliging en de toegankelijkheid. Tijdens drie leerzame presentaties kwamen deze eisen aan de orde. Met de beschreven *best practices* kunnen sites worden ontworpen die interactief en flexibel zijn, en tegelijkertijd eenvoudig en overzichtelijk. (Anouk de Rijk)

<http://www.iassistdata.org/conferences/2006/presentations>

Focus van aanbieders naar gebruikers van digitale tekst

'Het gebruik van elektronische teksten in de neerlandistiek' luidde de titel van een symposium dat op donderdag 8 juni in de Leidse universiteitsbibliotheek werd gehouden. De verwachting die door het woord 'gebruik' was gewekt, werd in het eerste deel niet helemaal waargemaakt. Daar hoorde het publiek namelijk vooral wat verschillende organisaties aan digitaal materiaal aanbieden.

Niets dan lof hiervoor natuurlijk. Het werk van de Digitale Bibliotheek voor de Nederlandse Letteren, vertegenwoordigd door René van Stipriaan, en het Woordenboek der Nederlandsche Taal op het Internet, waarvoor Tanneke Schoonheim sprak, is van onschatbare waarde. Maar hebben de aanbieders ook wel eens aan hun gebruikers gevraagd hoe die beschikbaarheid bevalt? Gebruiksanalyses beperken zich vaak, zo lijkt het, tot het noteren van aantallen hits. Maar in het digitale tijdperk verandert ook het gebruik van tekst.

Een voorbeeld daarvan liet Karina van Dalen-Oskam zien, met een becijfering van waar er in een middeleeuwse Nederlandse tekst sprake kan zijn geweest van een auteurs-

wisseling. Daarmee gaf ze een mooi voorbeeld van hoe digitale middelen nieuwe mogelijkheden scheppen om onderzoek te doen. Maar ook liet ze zien dat de aangeboden teksten veel voorbewerking vereisen voordat ze semi-automatisch geanalyseerd kunnen worden. Ook Adriaan van der Weel plaatste vraagtekens bij de wijze waarop digitaal materiaal wordt aangeboden. Tekst is – zeker in de digitale wereld – haast per definitie dynamisch, veranderlijk, bijna vloeibaar zelfs. Van der Weel hield een warm pleidooi om dit dynamische aspect niet te zien als een te bevechten nadeel, maar juist als een interessante nieuwe dimensie van tekst.

Maar zover zijn we nog niet. In zijn samenvattende bijdrage schetste Paul Rutten een scherp dilemma: in zijn rol als gebruiker hecht de onderzoeker vooral waarde aan de dynamische bruikbaarheid van tekst, maar in zijn rol als aanbieder worstelt hij met het probleem van de betrouwbaarheid en stabiliteit van tekst in de digitale context. Zo blijkt de problematiek onvermijdelijk, en blijft voor de onderzoeker altijd iets te wensen over. (Joris van Zundert)

ICPSR zomerprogramma in Michigan



Een rustige zomer in Ann Arbor

Elke zomer organiseert het Inter-university Consortium for Political and Social Research (ICPSR) een programma over kwantitatieve methoden in sociaalwetenschappelijk onderzoek, de *summer school* aan de University of Michigan. Deelname, door studenten, promovendi en onderzoekers uit alle windstreken, is bijzonder leerzaam en daarnaast leuk en nuttig. Dit jaar waren er maar liefst twaalf deelnemers uit Nederland.

Het programma bestaat uit workshops, waarbij ook studiepunten te behalen zijn, en lezingen. Voorbeelden van onderwerpen van lezingen zijn de programmeertaal R, ma-

trix-algebra en calculus. Workshops zijn doorgaans intensief en omvatten onderwerpen als longitudinale methoden, regressieanalyse, Bayesiaanse modellen en nog veel meer. Daarnaast zijn er kortere cursussen, lunchlezingen en informele bijeenkomsten voor deelnemers (zoals picknicks). Dit jaar zijn er in totaal ongeveer achthonderd deelnemers.

De meeste cursussen vinden plaats in Ann Arbor, een echt Amerikaans stadje met een rustieke uitstraling. In de zomer zijn de meeste studenten naar huis, waardoor het er rustig is. Accommodatie in Ann Arbor is doorgaans dan ook vrij eenvoudig te vinden. (Margit Averdijk)

<http://www.icpsr.umich.edu/training/summer/index.html>

NWO: 'Infrastructuur data hoort bij strategie 2007-2010'

Verbetering van de data-infrastructuur voor de alfa- en gammawetenschappen is integraal onderdeel van de strategie voor de komende jaren, die NWO in mei heeft gepresenteerd. Dat bevestigt een woordvoester van de organisatie.

'Wetenschap gewaardeerd!' was het motto van de strategie voor de periode tot en met 2010, die de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) op 22 mei aan minister van der Hoeven van OCW heeft aangeboden. De sterke positie die de Nederlandse wetenschap in de wereld heeft opgebouwd dreigt verloren te gaan als er geen forse extra investeringen worden gedaan, staat in de nota. Daartoe is volgens de organisatie een jaarlijks extra bedrag van

433 miljoen euro nodig, deels te gebruiken voor het voortzetten van de huidige incidentele financiering van toponderzoek (150 miljoen), deels om een 'sprong voorwaarts' te kunnen maken (283 miljoen).

Om de koppositie van de Nederlandse wetenschap te bereiken en verder te verstevigen heeft NWO in de strategie drie actielijnen uitgezet: 'ruimte voor onderzoekers', 'bundeling van krachten' en 'wetenschap voor de samenleving'. Het hoofd Voorlichting van NWO, Dominique de Vet, ziet verbetering van de data-infrastructuur vooral als onderdeel van de tweede actielijn: 'Die bundeling van krachten is op allerlei terreinen nodig. De data-organisatie DANS is daarvan een typisch voorbeeld. Ten eerste omdat DANS zelf een samenwerkingsproduct is van NWO met de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) en ten tweede omdat de opdracht van DANS, het verbeteren van de data-infrastructuur, een facilititeit is om de samenwerking tussen onderzoekers en disciplines te verbeteren'. Maar ook de eerste actielijn van de NWO-strategie, 'Ruimte voor onderzoekers', biedt aanknopingspunten als het om de data-infrastructuur gaat, verzekert De Vet: 'Daarbij gaat het erom, onderzoekers zo goed mogelijk te faciliteren bij het ontwerpen en uitvoeren van hun onderzoek.

De mogelijkheid om databestanden te kunnen vinden en hergebruiken hoort daar bij'.

DANS dong in december vorig jaar met een project om de data-infrastructuur te verbeteren vergeefs mee naar een eenmalige verdeling van honderd miljoen euro voor grootschalige onderzoekfaciliteiten.

Dat geld werd toen aan andere projecten gegund. Het bedrag – en nog iets meer – zou echter volgens de NWO-strategie elk jaar voor dat doel moeten worden besteed. 125 miljoen voor het ontwikkelen van grootschalige onderzoekfaciliteiten zijn onderdeel de het door NWO bepleite extra inspanning van 433 miljoen. (MdG)

'Betere toegang tot data door samenwerking DANS en CBS'

Er zijn goede mogelijkheden om door samenwerking tussen het data-instituut DANS en het CBS de toegang voor onderzoekers tot microdata van het CBS te verbeteren. Dat zegt drs. F.A. Hoeve, hoofd van het Centrum Beleidsstatistiek van het CBS.

Met zijn visie sluit Hoeve aan bij de uitlatingen van directeur Gosse van der Veen van het statistiekbureau. Die sprak bij de opening van DANS op 1 juni over 'een nieuw data-arrangement' tussen beide instituten, waarvoor voldoende ideeën aanwezig zouden zijn.

'Er liggen hier bijvoorbeeld historische datasets met microdata die door onderzoekers gebruikt kunnen worden voor tijdreeksen,' licht Hoeve toe. 'Het gaat onder andere om vele jaargangen van de Arbeidskrachtentellingen en om edities van het Woningbehoefteonderzoek. Maar om dat mogelijk te maken moeten ze geschikt gemaakt worden voor extern gebruik. Onderzoekers moeten van de variabelen en hun gecodeerde waarden weten wat ze betekenen en wat je er wel en niet mee kan doen'. Dat kan alleen met overzichtelijke hoeveelheden data voor een bepaald doel. 'Je moet daarvoor projecten definiëren en prioriteiten vaststellen. DANS lijkt ons daarvoor de uitgelezen partner'.

Daarnaast moet er veel gebeuren om de systemen waarmee data worden gedocumenteerd uniform te krijgen. Het CBS gebruikt nu een ander systeem voor deze metadata dan de universiteiten of DANS. Ook hier ziet het CBS kansen voor een vruchtbare samenwerking met het jonge data-instituut. Aan de sterke roep vanuit de academische wereld om de gegevens uit het Sociaal-Statistisch Bestand op

microniveau beschikbaar te maken wil het CBS ook graag tegemoet komen, aldus Hoeve. 'Maar het gaat daarbij om vele miljoenen records. Die kan je niet in één keer allemaal ontsluiten. De kunst zal zijn om een aantal bestanden te maken die beantwoorden aan de wensen van zo veel mogelijk onderzoekers. DANS zou kunnen optreden als vertolker van die wensen'.

Tenslotte werkt het bureau aan een *remote access facility* om onderzoekers op afstand gebruik te laten maken van haar data zonder dat die zelf worden overgedragen. Dit is een vervolg op het al bestaande onsite werken bij het CBS. Voor dit nog vrij nieuwe proces zijn speciale software en andere hulpmiddelen nodig en daarmee gaat het CBS op beperkte schaal beginnen. DANS zou kunnen helpen om vast te stellen welke onderzoeksinstituten het eerst voor aansluiting in aanmerking komen.

Op het gebied van de kosten voor het beschikbaar stellen van data zijn geen nieuwe ontwikkelingen te verwachten, zegt Hoeve desgevraagd. 'Wij geven onze data al enige tijd kosteloos af waar het om het materiaal zelf gaat; sinds kort ook de data die vroeger via het Sociaal Statistisch Agentschap werden geleverd. We rekenen wel kosten voor de faciliteiten om ze te prepareren, maar dat is reeel. Die kosten moeten nu eenmaal gemaakt worden'. (MdG)

Eerste e goed ontvangen



De verschijningsdatum van de eerste e-data&research, 1 juni, viel samen met de officiële start van DANS. Minister Maria van der Hoeven verrichtte dan ook in Den Haag niet alleen de officiële openingshandeling voor het data-instituut, maar nam ook het eerste exemplaar van het 'kwartaalblad over data en onderzoek' in ontvangst. De bewindsvrouw toonde zich ingenomen met het blad maar niet verrast. Dat kon ook niet, want ze was met een interview prominent aanwezig in dit openingsnummer.

Taakgroep voor infrastructuur

Een taakgroep met vertegenwoordigers van NWO, de KNAW, het CBS en het Sociaal-Cultureel Planbureau werkt sinds het voorjaar aan een landelijk actieplan voor een betere informatie-infrastructuur in het sociaal-wetenschappelijk onderzoek. De groep komt deze maand met een voorstel.

De taakgroep werkt vanuit DANS onder de naam AGENDA: 'Actieplan Gezamenlijke Effectiviteitsslag Nationale Dataverzamelaars Activiteiten'. Het is de bedoeling nationale partijen op het gebied van onderzoek en van overheidszijde in het proces te betrekken: universiteiten, SURF, onderzoeksscholen en instituten uit de eerste groep en RIVM, KNMI, TNO en anderen uit de tweede.

In het actieplan wordt uitgewerkt hoe het tot coördinatie en samenwerking kan komen op het gebied van onder meer het verzamelen, het beheer, de verspreiding en de archivering van data, het hanteren van kwaliteitsstandaarden, technische standaarden, metadata en procedures, onderzoek, methodenontwikkeling, onderwijs en scholing. (MdG)

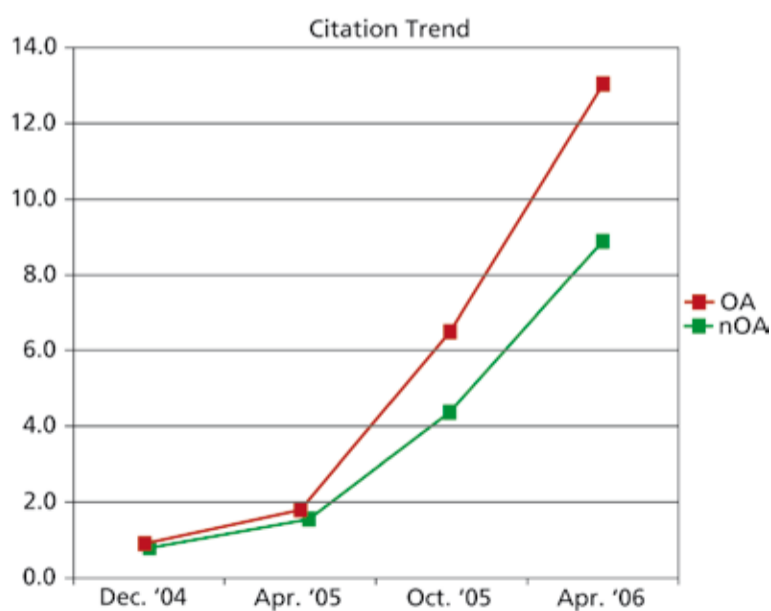
CBS werkt aan remote access

Een nieuwe techniek om microdata van het CBS van een afstand te kunnen benaderen is vrijwel praktijkrijp. Deze z.g. *remote access facility* biedt software die op de pc van een onderzoeker buiten het CBS wordt geïnstalleerd. Daarmee kunnen via een beveiligde verbinding gegevens worden bekeken en geanalyseerd, maar niet gekopieerd of geprint. De onderzoeker krijgt de resultaten op zijn beeldscherm maar kan de definitieve print alleen via het CBS krijgen. De beperkingen zijn nodig om de privacy te beschermen. Inloggen vereist ook nog een apparaatje om de vingerafdruk van de onderzoeker te checken.

100 miljoen onderweg

De honderd miljoen euro die vorig jaar december door NWO aan een vijftal projecten werden 'toegekend' zijn intussen daadwerkelijk onderweg naar de indieners. Dat werd half juli bekend. De toekenning door NWO was in feite een aanbeveling aan minister Van der Hoeven om het geld op de gewenste wijze te besteden. De bewindsvrouw heeft nu dus die aanbeveling overgenomen en uitgevoerd. Het ging om projecten voor het digitaliseren van kranten (13 miljoen), het opzetten van een e-science infrastructuur (29 miljoen), het verzamelen van data voor sociaalwetenschappelijk onderzoek (14 miljoen), voor materiaalonderzoek (26 miljoen) en voor systemen om afbeeldingen van de hersenen te maken (19 miljoen).

Open access meer geciteerd



Gemiddeld aantal citaties voor OA en niet-OA publicaties in Proceedings of the National Academy of Sciences, vanaf tweede helft 2004.

Voorstanders van Open Access (OA: onmiddellijke kostenloze en universele online toegankelijkheid na publicatie van wetenschappelijke artikelen) beweren sinds lang dat OA artikelen meer worden geciteerd. Tegenstanders schreven dat effect toe aan andere oorzaken. In een recente studie van Gunther Eysenbach (in PLOS Biology) wordt gecontroleerd

voor een serie andere variabelen, zoals aantal auteurs, aantal publicaties van de auteurs, land van herkomst, discipline, etc. De studie bevestigt het vermoeden van de voorstanders: OA artikelen worden tweemaal zo vaak geciteerd als niet-OA artikelen. Voor wetenschappers lijkt het dus verstandig te publiceren op basis van Open Access.



Het CBS-gebouw in Voorburg

Kabinetsnota 'Informatie op Orde'

MARTIJN DE GROOT

Van dementeren naar informeren

Alle ministeries gaan in de komende jaren hun digitale informatiehuishouding op orde brengen. Daarnaast moeten de papieren archieven versneld worden bijgewerkt. Dat staat in de nota 'Verbetering Informatie- en Archiefbeheer Rijksoverheid', die het kabinet op 29 juni naar de Tweede Kamer heeft gestuurd.

Het stuk, ook wel aangeduid als de nota *Informatie op Orde*, is nog afgeleverd door de bewindslieden Pechtold en Van der Laan uit het tweede kabinet-Balkenende, met als aanlei-

ding onder meer de nota 'Een dementerende overheid?' van de toenmalige Rijksarchiefinspectie. In het najaar zal het in de Tweede Kamer worden besproken.

DANS: 'Belangrijke wijziging in beleid'

Data-instituut DANS ziet de nota *Informatie op Orde* als een belangrijk signaal dat de overheid zich de kwetsbaarheid van digitale data op lange termijn realiseert. Voor het wetenschappelijk onderzoek is dat besef van groot belang omdat speciaal historisch, maar ook veel ander onderzoek vaak op informatie van de overheid berust. Heiko Tjalsma, senior expert research data van DANS ziet belangrijke wijzigingen ten opzichte van het bestaande beleid. 'Erkend wordt dat digitale documenten niet meer achteraf geselecteerd moeten worden, maar dat dit vooraf moet gebeuren, direct bij de vorming. Dat is wezenlijk want anders ben je te laat; na enige tijd kun-

nen de bestanden al onbruikbaar of onbegrijpelijk zijn geworden. Bovendien komt er bij de selectie nu meer aandacht voor de uitvoeringsdossiers en niet alleen voor het overheidsbeleid zelf. En juist in die uitvoeringsdossiers zit vaak een schat aan informatie voor historisch of sociaalwetenschappelijk onderzoek'.

Een ander punt is vooral voor DANS als nationale datacoach belangrijk: 'Er wordt nu ingezien dat het in het digitale tijdperk niet uitmaakt welke overheidsorganisatie de data beheert. Onderzoekbestanden van de overheid kunnen dus heel goed in een *trusted digital repository* worden bewaard, bij DANS of aan DANS gelieerd'.

De digitale informatiehuishouding van de overheid dreigt 'onbeheersbare proporties' aan te nemen en veel overheidsorganisaties kampen met een enorme achterstand in hun papieren archieven. Bij de eerste constatering speelt de 'digitale revolutie' een belangrijke rol; voor de tweede ontbreekt een rechtvaardiging. Woordvoerder Maarten Gresnigt van het ministerie van OCW wijst desgevraagd op de 'snelreinvahrt' waarmee de ambtenaren digitale uitingvormen tot hun beschikking hebben gekregen, 'terwijl het informatie- en archiefbeheer bij de departementen van oudsher nog een papieren proces is. Digitale en papieren processen hebben lange tijd naast elkaar bestaan. Daarbij heeft de papieren informatiehuishouding geleidelijk minder aandacht gekregen'.

De problemen moeten volgens de nota nu wel worden aangepakt want 'overheidsinformatie moet vindbaar en toegankelijk zijn maar ook betrouwbaar, authentiek en volledig'. Het inhalen van de achterstand in de papieren archieven lijkt nog het eenvoudigste voornemen. Een Rijksbreed project Wegwerken Archief-achterstanden t/m 1975 was al in

2005 van start gegaan en moet op 1 januari 2009 klaar zijn. Vanaf 2008 worden de 'werkvoorraden' van na 1975 onder handen genomen op basis van werkplannen die de ministeries dan klaar hebben.

Het aanpakken van de digitale archieven en vooral het aanpassen van de manier van werken opdat nieuwe stukken automatisch goed worden gearciveerd is veel moeilijker, blijkt uit de visienota. Voor dat laatste doel heeft het kabinet volgens de nota op 1 januari 2008 een set van minimale kwaliteitseisen gereed – een z.g. *baseline* – die moet worden gecontroleerd door de departementale auditdiensten. Ook moet er een nieuwe aanpak worden ontwikkeld van de selectie van documenten en met name metadata die ervoor kunnen zorgen dat cultuurhistorische selectie al bij de vorming van documenten mogelijk wordt. Die metadata worden onderdeel van de *baseline*.

Voor dit alles zijn nieuwe regels nodig. Een deel van die regels kan in de vorm van een convenant tussen de departementen worden afgesproken. Een ander deel zal bestaan uit het op elkaar afstemmen van bestaande wetten en regels zoals de Wet Openbaar-

Alleen Rijksoverheden

De nota Informatie op Orde geldt voor de Rijksoverheid: de departementen en de Hoge Colleges van Staat (zoals de Raad van State) en voor diensten zoals de Belastingdienst of Rijkswaterstaat. OCW-woordvoerder Gresnigt: 'Uitvoerende diensten vallen onder het beleid in zoverre ze onderdeel zijn van een ministerie. Over de verantwoordelijkheden en de uitvoering maken de ministeries afspraken met die diensten'.

De Zelfstandige Bestuurs Organen (zoals de Autoriteit Financiële Markten) vallen niet onder de nota, maar de ministeries waaronder ze vallen zullen die organen volgens de nota wel 'uitnodigen om de problemen op gelijke wijze aan te pakken en op te lossen'.

heid Bestuur, de privacywetgeving en de Archiefwet 1995. Hiernaar wil het kabinet vanaf het voorjaar van 2007 een onderzoek doen.

Voor informatie van wetenschappelijk belang heeft de nieuwe visie van het kabinet geen aparte gevolgen, verzekert het ministerie van OCW. Gresnigt: 'De nota richt zich op de kwaliteit van het informatie- en archiefbeheer van de rijksoverheid in het algemeen, ongeacht de vorm van de gegevens, en daarmee ook op een betere vindbaarheid en toegankelijkheid, ongeacht de gebruikersgroep'.

Letters, woorden en teksten: pijlers onder de data-infrastructuur

RENÉ VAN HORIK

Taal en literatuur worden op verschillende manieren wetenschappelijk bestudeerd. Zowel het materiaal als de instrumenten worden steeds vaker in elektronische vorm aangeboden. Het internet brengt efficiënte communicatie tussen onderzoekers en afstemming van onderzoeksinstrumenten binnen bereik. Enkele pijlers onder de benodigde data-infrastructuur komen hier voor het voetlicht.

De Digitale Bibliotheek voor de Nederlandse Letteren (DBNL: www.dbnl.org), opgericht vanuit de Maatschappij der Nederlandse Letterkunde, biedt een zeer goede website over de Nederlandse taal en literatuur. De site bevat literaire teksten, secundaire literatuur en aanvullende informatie zoals biografieën en portretten. Ook de digitale collecties van universiteitsbibliotheken en de Koninklijke Bibliotheek maken deel uit van de data-infrastructuur voor literair en taalkundig wetenschappelijk onderzoek. Het DBNL-project werkt samen met een groot aantal wetenschappers afkomstig van verschillende disciplines en vakgroepen.

Schatbewaarder

Naast het onderzoek dat aan de letterenfaculteiten van universiteiten wordt verricht bestaan er enkele toonaangevende instellingen in Nederland op het gebied van taalkundig en literair wetenschappelijk on-

derzoek: het Huygens Instituut, het Meertens Instituut, het Max Planck instituut voor psycholinguïstiek en het Nederlands Instituut voor Lexicologie (INL). Het Huygens Instituut (www.huygens.knaw.nl/) is geïnteresseerd in de toepassing van nieuwe methoden en technieken in

ontsluiting en tekstwetenschappelijk onderzoek. Op het gebied van de Nederlandse taal bestudeert het Meertens Instituut (www.meertens.knaw.nl) zowel de geografische als sociale variatie. Het Max Planck Instituut voor psycholinguïstiek (www.mpi.nl) richt zich op onderzoek naar de mentale processen gerelateerd aan het begrip en gebruik van taal. Het Instituut voor Nederlandse Lexicologie (www.inl.nl) beschouwt zich als de schatbewaarder van de Nederlandse taal door bijvoorbeeld de Taalbank bij te houden, waarin alle Nederlandse woorden voorkomen. Deze en andere instellingen zijn betrokken bij de creatie en het gebruik van een groot aantal digitale databronnen en onderzoeksinstrumenten, zoals *parsers* en *lemmatizers*.

Tekstcorpora zijn van groot belang

voor linguïstisch onderzoek en worden gebruikt om theorieën te toetsen. Er bestaat al een behoorlijk aantal corpora, maar deze zijn niet volledig qua type en tijdsperiode. Een corpus dient verrijkt te worden om de onderzoeksvragen te kunnen beantwoorden. Het Corpus Gesproken Nederlands (CGN) is bijvoorbeeld een belangrijk onderzoekscorpus. Daarnaast zijn er diverse taalbases, bijvoorbeeld op het gebied van dialecten, zoals de Syntactische Atlas van de Nederlandse Dialecten (SAND), te vinden via de website van het Meertens Instituut.

Scala van onderzoek

Databronnen en onderzoeksinstrumenten zoals deze maken een breed scala aan wetenschappelijk onderzoek mogelijk. Historici richten zich met hun onderzoek bijvoorbeeld op de dynamische relatie tussen taal en maatschappij. Taalveranderingen treden immers vaak op waar contact is geweest tussen verschillende talen of verschillende variaties van een taal. Ook kan het perspectief van het cultureel erfgoed centraal staan in onderzoek. Meer fundamenteel linguïstisch onderzoek kan zich bezig houden met de veranderingen van de syntaxis van het Nederlands, bijvoorbeeld op het gebied van het gebruik van persoonlijke voornaamwoorden.

In de literatuurstudie speelt de data-infrastructuur een belangrijke rol bij het vernieuwen van de bestaande

kennis. Daarbij moet rekening worden gehouden met de relatie tussen eigentijdse oordelen en de literaire geschiedenis. Hiervoor is een corpus van secundaire literaire bronnen nodig die gebruikt kunnen worden om literaire teksten te kunnen evalueren. Literaire tijdschriften zijn hierbij van het grootste belang, want zij vormen de 'barometer van de smaak'.

Goede fundering

Door de computerondersteunde auteurschapidentificatie wordt steeds vaker toegepast bij de analyse van gedigitaliseerde teksten. Het onderzoek richt zich dan op stijlkenmerken van auteurs door gebruik te maken van speciale algoritmen. Zo kan bijvoorbeeld vastgesteld worden wie de auteur is van een anoniem pamflet. De data-infrastructuur op het gebied van taalkunde en literatuurwetenschap heeft een goede fundering en maakt een breed scala aan wetenschappelijk onderzoek mogelijk. Meerjarige onderzoeksprogramma's, zoals *Stevin* (www.nwo.nl/stevin/) op het gebied van taal en spraaktechnologie en *Catch* (www.nwo.catch/) voor verbeterde toegang tot digitaal cultureel erfgoed, dragen bij aan verbeterde data-infrastructuur voor wetenschappelijk onderzoek. Uitwisseling van expertise, standaardisatie van datacreatie, dataopslag en data-uitwisseling zijn gebieden die meer aandacht vragen, waardoor de data-infrastructuur nog beter zou kunnen worden.



De Digitale Bibliotheek voor de Nederlandse Letteren biedt een zeer goede website over de Nederlandse taal en literatuur.

Nieuwe hoogleraar historische demografie Theo Engelen:

‘Uiteindelijk gaat het om het verhaal bij de tabellen’

Op 15 juni sprak Theo Engelen in Nijmegen de rede uit waarmee hij de leerstoel in de historische demografie aanvaardde die hij daar sinds 1 december vorig jaar bekleedt. Hij beschreef ‘een discussie over het kindertal in China en Europa’ waarmee ruim tweehonderd jaar geleden Thomas Malthus begon, schetste de botsende standpunten in die discussie en voegde daar tenslotte een eigen hypothese aan toe waarin de romantische liefde centraal stond.

e-data&research sprak met de man die het een voorrecht noemt zich te mogen bezighouden met de sporen van ‘liefde, dood en seksualiteit, de kern van het menselijk bestaan’.

‘Een revolutie?’ Theo Engelen is een man van bedachtzame, niet van hoogdravende formuleringen. De vraag wat nu eigenlijk de betekenis is geweest van de digitalisering voor zijn vak, de historische demografie, brengt zijn welsprekendheid echter onmiskenbaar in een hogere versnelling. ‘Het vak heeft absoluut een belangrijke kwalitatieve sprong gemaakt door de digitalisering. Kijk eens, als je je bezig houdt met liefde, dood en seksualiteit dan heb je in feite de kern van het menselijk bestaan te pakken. Dat mag je in ons vak doen en ik hoef dan niet te weten hoe in de negentiende eeuw de perceptie was van het Koningshuis of iets dergelijks. Ik ben tevreden als ik me mag bezighouden met die kernelementen van het menselijk leven. We weten nu eenmaal heel weinig over 99% van de bevolking. Maar we weten van een heel groot deel wel wanneer ze zijn geboren, getrouwd, wanneer ze kinderen kregen en hoeveel, en wanneer ze dood gingen. De digitalisering heeft het mogelijk gemaakt om die kennis op een veel grotere schaal te benutten en betrouwbaar statistisch te analyseren. En vooral, om je analyse te baseren op gegevens van het individuele niveau, waar we vroeger veel meer gewend waren om geaggregeerde data te gebruiken’.



GUY ACKERMANS

De mogelijkheid om grote hoeveelheden data te analyseren, en daarmee om verfijnde statistische methodes toe te passen, heeft echter ook zijn eigen gebruikers geschapen en dat roept bedenkingen op bij de juist gelanceerde hoogleraar. ‘Wat ik jammer vind is dat tegenwoordig de statistische technieken soms de boventoon gaan voeren en dan lijkt het tenslotte niet meer om concrete

mensen te gaan’. Er is een duidelijke verandering opgetreden sinds de jaren zeventig en tachtig, toen de historische demografie nog een richting was die door historici werd gedomineerd. Eind jaren tachtig en begin jaren negentig kwam er een nieuwe generatie van vakgenoten met een sociologische achtergrond en een pakket kwantitatieve technieken in hun opleiding. Dat legde de basis voor ‘een soort richtingstrijd’ aldus Engelen.

‘Je hebt de mensen die de nadruk leggen op het kwantitatieve, op de regressieanalyses enzo, en er zijn de collega’s die het in de ouderwetse beschrijvende aanpak zoeken. Ikzelf? Ik hoop dat ik er zo’n beetje tussenin sta. Ik wil heel graag die nieuwe kwantitatieve methodes gebruiken, maar ik wil ook dat het begrijpelijk is en dat het over mensen blijft gaan. Uiteindelijk gaat het om het verhaal. Je kunt wel een heleboel tabellen en

coëfficiënten presenteren, maar het gaat tenslotte toch om de interpretatie van de resultaten’.

Engelens inmiddels jarenlange studie van de demografie van China blijkt een goede illustratie van die visie. ‘De eerste serieuze demografische beschrijvingen’, afkomstig van de Engelse dominee Thomas Malthus, presenteerden China al als het

‘Je houdt je bezig met de kern van het menselijk bestaan’

spiegelbeeld van Europa. Het land was dus al heel lang onderwerp van demografische discussie, en Engelen’s groeiende belangstelling viel samen met een heftige ervaring van die discussie in de jaren negentig. ‘Malthus’ eigen basisidee dat China een hoge vruchtbaarheid kende omdat er niet zoals in Europa een maatschappelijke rem op het huwelijk bestond was intussen wel achterhaald geble-

ken. In de eerste helft van de vorige eeuw was namelijk aangetoond dat het kindertal in China niet hoger, maar juist lager was dan in Europa. Sindsdien verschoof de discussie naar de vraag hoe dat kon, terwijl Malthus waarneming van de verschillen in huwelijksleeftijd wel accuraat was’.

Is dit nu een van de belangrijkste discussies binnen de historische demografie? Theo Engelen is daarvan overtuigd: ‘Je moet natuurlijk altijd oppassen dat je niet het gebied waar je zelf mee bezig bent tot het centrum van de wereld verklaart. Maar als je nu een terrein zoekt waar de belangrijke vragen in dit vakgebied op het scherpst van de snede worden uitgevochten, dan is het toch wel het China-debat’.

Ook hier blijken de twee richtingen een rol te spelen. ‘Je hebt de echte kwantificeerders uit de Verenigde Staten, die de gedachte aan slechte economische omstandigheden en ondervoeding als oorzaak van lage geboortecijfers verwerpen. Zij houden bij hoog en bij laag vol dat Chinese echtparen heel bewust aan geboortebeperving doen en dat er dus sprake is van rationeel handelen. En je hebt de meer kwalitatieve benadering die het naast cijfers ook meer in bronnen zoekt, en die wijst op de rol van armoede, voedseltekorten en ondervoeding. We weten dat onder zulke omstandigheden de conceptiekans kleiner wordt’.

Het is een boeiende polemiek, die alleen wordt gehinderd door het feit dat de partijen niet meer serieus naar elkaar luisteren, aldus Engelen. Hijzelf koos in zijn oratie een derde positie: wat te denken van de romantische liefde als verklarende factor voor de huwelijksvruchtbaarheid? Europa kende immers ook in de negentiende eeuw deze romantische liefde als basis voor het huwelijk, terwijl die in China in het geheel geen rol leek te spelen.

‘Het is een hypothese waar nader onderzoek naar gedaan moet worden’ benadrukt de demograaf, maar bepaald niet het enige, zo blijkt al snel wanneer hem wordt gevraagd naar zijn aspiraties in zijn nieuwe hoedanigheid van hoogleraar.

Engelen: ‘Mijn benoeming is de eer-

ste in dit vakgebied. Dat is op z’n minst een opmerkelijk feit maar het is ook een stukje erkenning. Aan mij is daardoor echt niks veranderd maar je kunt gewoon vaststellen dat je meer mogelijkheden hebt en dat je anders wordt benaderd. Je hebt dus betere kansen om dingen binnen te halen en daar zijn we ook mee bezig’. Vergrijzing is een onderwerp waar onderzoekenergie in gestoken zou moeten worden. De politieke discussie daarover laait nu hoog op maar op historische inzichten in dat verschijnsel is decennia lang geen beroep gedaan. Sterfte onder meisjes en vrouwen is een ander voorbeeld: ‘We zijn vertrouwd met het verschijnsel van kindermoord – op meisjes – in China maar aan het eind van de negentiende eeuw bleek ook in Europa de levensverwachting van meisjes lager te zijn dan die van jongens. Hier had je dan weer het vondeling-

‘Voorlopig hebben we onze handen nog vol’

systeem dat in China niet bestond, en we weten dat de sterfte in vondelingentehuizen aanmerkelijk hoger was dan daarbuiten’. Dat moet uitgezocht worden, vindt Engelen, maar dan ‘op basis van goed gedocumenteerd historisch cijfermateriaal’.

Het beschikbaar komen van grote hoeveelheden cijfermateriaal ging in Nederland niet toevallig samen met het kwantitatiever worden van de historische demografie en met het tot stand komen van verschillende projecten van hoge kwaliteit, vindt Engelen. ‘Er is de Historische Steekproef Nederland, die veel belangrijk materiaal biedt op individueel niveau. Dan zijn er gegevens op gemeentelijk niveau, bijeengebracht door de Wageningse socioloog Hofstee en uitgebreid door Onno Boonstra. En er is de digitalisering van de Volkstellingen, een project dat hele rijke en nuttige bestanden oplevert. Allemaal waardevolle projecten, maar ze staan te los van elkaar en ze zijn niet makkelijk genoeg toegankelijk. Wat ik zou willen is dat er een interface gemaakt kon worden om al die gegevens met elkaar te verbinden en via één poort toegankelijk te maken voor alle onderzoekers die daar belang bij hebben. Gratis’.

Is er intussen genoeg gedigitaliseerd binnen het vakgebied?

Engelen: ‘Ik weet dat er nog schatten verborgen liggen in de gemeente- en provinciehuizen. Maar voorlopig hebben we onze handen nog vol aan de analyse en interpretatie van het materiaal dat we nu hebben’.

Theo Engelen (1950) studeerde geschiedenis aan de Radboud Universiteit, werd daar in 1976 tijdelijk medewerker, in 1982 universitair docent economische en sociale geschiedenis en later hoofddocent. Hij promoveerde in 1987 op een onderzoek naar de daling van de huwelijksvruchtbaarheid in Limburg. Naast zijn academische leven heeft Engelen zich sinds 1992 ontpopt als succesvol schrijver van jeugdboeken.

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

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. 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 ontwikkelingen. (LS)

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

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



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/) 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

Winnaar 2004: Ename Expertisecentrum voor erfgoedontsluiting, 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

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

Liefde en trouw gedigitaliseerd

Zo rond 1600 verschenen er op de Nederlandse boekenmarkt prachtig geïllustreerde emblembundels, met intrigerende teksten en afbeeldingen over thema's als liefdesverdriet, trouw en partnerkeuze. Elk embleem bevatte een les over de liefde, verborgen in puzzelstukjes die Nederlandse dichters aan klassieke, christelijke en eigentijdse bronnen ontleenden.

Kunsthistorici, neerlandici, latinisten en informatici werken binnen het Emblem Project Utrecht (EPU) aan de digitalisering van Nederlandse liefdesemblembundels, zo'n 25 in totaal. Doel van het onderzoek dat op basis van deze digitalisatie gebeurt, is onder andere de wisselwerking vast te stellen tussen religieuze en profane stromingen in de liefdesemblematiek.

Om de verknoppingen in kaart te brengen en doorzoekbaar te maken,

zijn de digitale edities die het EPU op het Internet publiceert gebaseerd op codering in XML (eXtensible Markup Language). Voor de tekstuele elementen van de emblemen wordt gebruik gemaakt van de *Guidelines* van het Text Encoding Initiative (TEI), en voor de pictoriale van Iconclass, een classificatiemethode die al voor het computertijdperk zijn waarde bewezen heeft. Uit de EPU-collecties worden dynamisch doorzoekbare edities gegenereerd met de open source software Cocoon.

Vanuit het Emblem Project Utrecht wordt ook deelgenomen aan het Open Emblem Portal, dat gebouwd is op dezelfde technische uitgangspunten en coderingsrichtlijnen. Via dit portal zijn inmiddels grote delen van het corpus Europese emblembundels te raadplegen en doorzoeken. (Els Stronks)



Zoals een fakkel door het vet waarmee hij gevoed wordt, zo kom ik om door wie ik leef en wordt gevoed en door degene op wie ik smoorverliefd ben. Uit: Daniel Heinsius, *Emblemata amatoria* (c. 1601), met dank aan de Universiteitsbibliotheek UvA.

<http://emblems.let.uu.nl>

<http://media.library.uiuc.edu/projects/oebp/>

VGI studiedag: nieuwe technologie in onderwijs en cultuur

De VGI organiseerde op 16 juni een drukbezochte studiedag over toepassingen van nieuwe technologie binnen onderwijs en cultuur. Met nieuwe technologie worden bijvoorbeeld bedoeld *location based services* (een informatiedienst die afhankelijk is van de locatie waar men zich bevindt) en *social bookmarking* (gereedschap waarmee de internetgemeenschap collectief informatiebronnen markeert en van kwaliteitsoordelen voorziet). Zogenaamde 'folksonomies' zijn hier nauw mee verbonden: het 'volk' bouwt *bottom up* een kennisstructuur die lijkt op een traditionele taxonomie. Ook werd de architectuur van de nieuwste versie van Surfnet (SURFnet6) gepresenteerd, waarbij snelle glasvezelverbindingen

en 'lichtpaden' hoogvolume datastromen mogelijk maken.

Netwerken van gebruikers, van mensen rondom kwalitatief goede inhoud bleken op de studiedag toch een eerste vereiste om te komen tot succesvolle toepassing van deze nieuwe technologie. Enkele bijdragen tijdens de studiedag presenteerden voorbeelden hiervan, zoals het themakanaal 'Geschiedenis' van de VPRO. Andere bijdragen waren van Dick Rijken (Hogeschool Den Haag) over 'De lerende samenleving'; Diederik van Wingerden (ZAPPwerk) met 'Cultuur in de buurt', Maurice Dohmen (Surfnet) over Surfnet 6 en Sybille Poortman (Fontys Hogeschool) met 'Folksonomy en social bookmarking'. (René van Horik)

Vlaams onderzoek naar ontsluiting van mondelinge bronnen

In Vlaanderen zal een studie worden gedaan naar de mogelijkheid om een flexibele applicatie te ontwikkelen voor het invoeren en ontsluiten van mondelinge bronnen. De coördinator van dit project is de Universiteit Gent. Partners zijn lokale erfgoedcellen in tien steden, enkele Vlaamse archiefinstellingen, het Geluidsarchief van de publieke omroep VRT, Volkskunde-Vlaanderen vzw en het Vlaams Centrum voor Volkscultuur (VCV), met de steun van de Vlaamse overheid.

Er bestaat momenteel in Vlaanderen een achterstand in de ontsluiting van mondelinge bronnen: onder andere (historische) opnames van interviews en getuigenissen, verhalen, dialecten of andere vormen van 'historisch klankarchief'. Er is een breed spectrum van 'producenten' van mondelinge bronnen, en vermoed wordt dat slechts een beperkt deel in bewaring gegeven wordt bij archiefinstellingen. In Vlaanderen bestaat geen instelling of organisatie die expliciet de opdracht heeft het ontsluiten, de opslag en bewaring van mondelinge bronnen te ondersteunen.

De voorbije jaren zijn mondelinge bronnen in Vlaanderen onder de aan-

dacht gekomen via de ontwikkeling van verhalenbanken in het erfgoedveld, zoals de *Verhalenbank Brugge* (www.verhalenbankbrugge.be). Diverse (erfgoed)organisaties ontwikkelen mondelinge geschiedenisprojecten en het Vlaams Centrum voor Volkscultuur (VCV) gaf in 2005 de DVD-handleiding *Van Horen Zeggen* uit.

Internationaal worden geavanceerde applicaties ontwikkeld voor het ontsluiten en tot waarde brengen van geluidscollecties via het internet, maar ook beheersapplicaties die op professionele wijze dit type bronnen inventariseren. Zo ook in Nederland waar in de context van het onder-

zoeksprogramma *Continuous Access To Cultural Heritage* (CATCH: www.nwo.nl/catch) projectlijnen zijn opgestart zoals *Mondelinge geschiedenis geïndexeerd* (CHORAL, <http://hmi.ewi.utwente.nl/project/Choral>). Ook het Nederlandse Instituut voor Beeld en Geluid ontwikkelt diverse interessante programma's. De UNESCO-conventie voor de bescherming van immaterieel cultureel erfgoed (2003) geeft nog een extra legitimatie en stimulans aan de zorg voor orale bronnen. (Zie: www.unesco.org)

De applicatie die nu in Vlaanderen zal worden verkend zou een open modulair systeem kunnen zijn met aandacht voor innovatieve zoek- en indexersystemen. Het systeem moet inspelen op een aantal van de behoeften op het vlak van productie, ontsluiting, opslag en bewaring van mondelinge bronnen. Wat realiseerbaar is hangt af van de financiële middelen; de studie brengt enkel de mogelijkheden in kaart. Tijdens de onderzoeksperiode zal een internationaal seminarie met experts worden georganiseerd en zal overleg gepleegd worden met de partners en de overheid. Het lijkt belangrijk te vermijden dat er op verschillende niveaus aan dure, parallelle ontwikkeling van systemen wordt gedaan. In dit stadium is er misschien een kans om instellingsoverschrijdend en schaalvergroterend te werken aan een gezamenlijk traject. Daarbij zal ook naar de ontwikkelingen over de grens gekeken worden, zoals in Nederland. (Jeroen Walterus)

www.vcv.be

TOPs komen van de grond

Met verschillende partners in de sociale- en geesteswetenschappen werkt het data-instituut DANS aan het opzetten van TOP's: Thematische Ontwikkel Programma's. Daarmee krijgt een van de belangrijke taken van DANS gestalte.

DANS heeft als taak om de data-infrastructuur binnen de sociale en geesteswetenschappen te verbeteren. De Thematische Ontwikkel Programma's (TOPs) zijn daarbij een belangrijk instrument. Ze stimuleren samenwerking met en tussen onderzoekers op een bepaald wetenschapsgebied en versterken zo de data-infrastructuur. Databestanden met nationale en/of internationale reikwijdte van de ene groep onderzoekers worden zo bijvoorbeeld beter toegankelijk en kunnen worden verrijkt door combinatie en harmonisatie met bestanden van een andere groep.

Momenteel werkt DANS binnen de sociale wetenschappen aan verschillende TOPs. Er zijn programma's in ontwikkeling rond *arbeid* (samen met de Organisatie voor Strategisch Arbeidsonderzoek OSA en het Amsterdam Instituut voor Arbeidsstudies AIAS), *familierelaties* (o.a. met Nederlands Demografisch Instituut NIDI en de Universiteit van Amsterdam, *sociale mobiliteit, kwalitatief onderzoek* (beide Vrije Universiteit), *opinions en attitudes* (o.a. met de Universiteit van Tilburg), en *epidemiologisch onderzoek naar psychiatrische stoornissen* (ZonMW). Ook binnen de geesteswetenschappen zijn diverse initiatieven in ontwikkeling.

In TOPs worden nieuwe terreinen van digitale archivering en ontsluiting van onderzoeksdata verkend, ze sluiten aan bij actuele wetenschappelijke ontwikkelingen en ze maken

nieuwe innovatieve onderzoeksvragen mogelijk. Onderzoekers uit het veld en medewerkers van DANS bemensen gezamenlijk een dergelijk programma. Financiering komt deels van de deelnemers, deels van externe (naast Nederlandse ook Europese of andere internationale) bronnen.

Contactpersonen bij DANS zijn dr. Heleen van Luijn (sociale wetenschappen), en dr. Rene van Horik (geesteswetenschappen), beide te vinden op www.dans.knaw.nl. (Heleen van Luijn)

COLOFON

e-data@research is het kwartaalblad over data en onderzoek in de alfa- en gammawetenschappen, verschijnend onder auspiciën van Data Archiving and Networked Services, het Huygens Instituut, het Internationaal Instituut voor Sociale Geschiedenis en de Vereniging voor Geschiedenis en Informatica.

e-data@research is ook een voortzetting van *Historia & Informatica* en van *Data News – Steinmetz Archive Newsletter*.

Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen. Oplage: 6500.

e-data@research is online te raadplegen op www.edata.nl en op de websites van de deelnemende instellingen.

Uitgever: Edita-KNAW, Postbus 19121, 1000 GC, Amsterdam

Redactieadres: Postbus 93067, 2509 AB Den Haag; Anna van Saksenlaan 51, 2593 HW Den Haag; T (070)3494450 F (070)3494451 E edata@dans.knaw.nl

Redactie: Peter Boot, Martijn de Groot (hoofd/eindredacteur a.i.), Marien van der Heijden, Jetske van der Schaaf, Luuk Schreven.

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Aan dit nummer werkten mee:

Margit Averdijk, Bert van Elsacker, René van Horik, Elly Kamp, Heleen van Luijn, Anouk de Rijk, Els Stronks, Jeroen Walterus, Joris van Zundert.

Pre-press, productie en vormgeving:

Edita-KNAW
Druk: Plantijn Casparie, Almere
ISSN: 1872-0374

Archiefbeheerder en historicus Jacques Dane:

'Psychologie in Nederland lijdt aan Peter Pan-syndroom'

De term Peter Pan-syndroom werd voor het eerst gebruikt door de Amerikaanse psycholoog Dan Kiley. Hij beschreef er in de jaren tachtig van de vorige eeuw het verschijnsel mee dat mannen bij het ouder worden soms geen afstand van hun jeugd willen nemen. Historicus dr. Jacques Dane stelde de Peter Pan-diagnose in een artikel in het *Archiefblad* van 2004 voor de Nederlandse psychologie, en hij kan nog steeds heel goed uitlegen waarom.

Dane (42) is coördinator en onderzoeker van het in Groningen gevestigde ADNP, het Archief en Documentatiecentrum Nederlandse Psychologie. Het is het enige sectorbrede archief dat de psychologen kennen. In 1989 opgericht als Stichting Historische Materialen Psychologie, was het oorspronkelijk bedoeld om allerlei instrumenten voor de vergankelijkheid te behoeden die psychologen in de loop der jaren hebben gebruikt. Later breidde de aandacht zich uit naar bronnen als boeken en tijdschriften, persoons- en instellingsarchieven, beeldmateriaal en een verzameling psychologische tests. Dat er in de psychologie gearhiveerd wordt weet Dane dus als geen ander. Maar hij weet ook hoe moeizaam dat meestal gaat: 'Psychologen archiveren niet, omdat ze er onbekend mee

zijn'. Archieven van gepensioneerde of overleden hoogleraren komen soms in ADNP-handen terwijl ze eigenlijk thuishoren aan de universiteit waaraan de geleerden waren verbonden. Dane vindt dat niet erg, maar wel tekenend: 'Er is een Archiefwet van 1995, en volgens die wet zijn onderzoekers en wetenschappelijke instellingen verplicht archieven fatsoenlijk te beheren en hun onderzoeksmateriaal te bewaren. Maar niemand lijkt dat te weten; laat staan hoe het dan moet en waarom en voor wie.'

Het ADNP heeft behalve voor papieren bronnen ook belangstelling voor elektronische, zoals de emailcorrespondentie van onderzoekers die maar al te vaak wordt gedeleet voordat een archivaris zich erover heeft kunnen ontfermen. En voor onderzoeksdata, maar daarin bestaat al helemaal geen traditie, aldus de coördinator. 'Onderzoekers gaan heel slordig om met hun data. Die worden op verschillende plaatsen bewaard als ze al bewaard worden. Laboratoriumboekjes raken zoek, floppies met data komen in dozen terecht die worden dan op zeker moment aan de portier gegeven. Die stapelt ze op in een fietshok en dat is dan het archief.'

Hoe dit kan in een respectabele wetenschappelijke discipline? Dane twijfelt. Misschien is het wel alge-

meen menselijk, oppert hij, om je achterland te verwaarlozen. Maar het blijft het jammer en, netjes uitgedrukt, niet verstandig. En dan komt Peter Pan aan de orde, de romanfiguur die niet volwassen wilde worden. 'Onderzoekers denken vaak alsof ze altijd jong willen blijven; alsof de psychologie pas vandaag begint. Ze willen dan niet erkennen dat er een verleden is. Ze denken dat wat hun voorgangers hebben gedaan niet belangrijk is, maar alleen wat vandaag gebeurt. Als je zo denkt, dan kan je dat gemakkelijk ook over je eigen werk gaan denken. Ik zie dat vaak. Dat onderzoekers niet meer willen weten waar ze tien of twintig jaar geleden mee bezig waren omdat ze denken 'ach, dat was toen'. Als ze wat ouder worden komt er vaak wel erkenning van het verleden, ziet Dane, 'maar de mensen die het nu voor het zeggen hebben denken meestal 'die oude rotzooi, wat moeten we ermee?'

En wat ook een rol kan spelen: het steeds meer overheersende motto *publish or perish*. Dane: 'Onderzoek gaat ten koste van publiceren' hoorde ik laatst iemand zeggen. Dat betekent dat nieuw onderzoek steeds sneller en kleinschaliger wordt opgezet om snel een publicatie op te leveren, maar dat de resultaten daardoor minder worden gezien als iets dat voor het nageslacht bewaard moet worden.'

Gelukkig duikt er nog genoeg materiaal op dat wel gearhiveerd moet en kan worden. Kostbaar materiaal waarmee het ADNP met zijn kleine personeelsbezetting de waarde van zijn collectie weer verder op kan voeren. Materiaal dat soms op een ouderwetse manier de gevaren van de digitale toekomst symboliseert. 'We krijgen hier wel eens prachtige instrumenten bezorgd waar we niets over weten. Dan denk je: wat zou hier mee gemeten zijn?' (MdG)

Column

Willem Kuiper

Leren van een schapenboer

De anekdote van de Australische schapenboer blijft mij achtervolgen. Ooit loofde Rolls Royce een gloednieuwe auto uit aan de eigenaar van de oudste, nog in normaal bedrijf zijnde Rolls. Dat bleek een schapenboer in Australië te zijn, die hem met veel tam-tam en fanfare in ontvangst nam. Een week later vroeg hij zijn oude Rolls terug. De nieuwe voldeed niet: hij kon zijn dode schapen niet op de treeplanken kwijt...



DIEDERIK V.D. LAAN

Voor mijn project *Repertorium van Eigennamen in Middel-nederlandse Literaire Teksten (REMLT)* maak ik intensief gebruik van digitale teksten. Toen dat project in 1993 van start ging, werd mijn voorstel om de te exciperen teksten eerst te digitaliseren verworpen. Ook een projectaanvraag van Ben Salemans (Neder-L) om een tekstencentrum op te richten werd afgewezen. Hoe zo'n tekstencentrum eruit zou kunnen zien demonstreerde Marc van Oostendorp met zijn Laurens Janszoon Coster-site.

Hierna werd Leiden de hoofdstad van de digitale neerlandistiek. Cees Klapwijk was de ijsberg onder water van de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL), Dirk Geirnaert van de CD-ROM *MiddeNederlands*. Hierop stond niet alleen een digitale versie van het monumentale *MiddeNederlandsch Woordenboek*, maar eraan toegevoegd waren driehonderd belangrijke *MiddeNederlandse versteksten* en tweehonderd prozateksten. Bovendien kon je de inhoud van die CD naar je harde schijf doen verhuizen en met een simpele muisbeweging een of al die vijfhonderd teksten naar je eigen tekstverwerker kopiëren. De CD die mijn leven veranderd heeft!

De DBNL is een website, een pracht van een website! Maar de teksten staan er niet 'los' op. Het is geen digitale boekenplank waar je een tekst naar keuze van af kunt pakken. Je kunt hem op het scherm lezen en bevragen, je kunt hem naar een printer sturen, maar je kunt hem niet met één muisbeweging naar je eigen tekstverwerker kopiëren. En laat dat nou precies zijn wat ik wil...

In de *Ferguut* wordt van Galiëne verteld dat moeder Natuur zoveel aandacht aan haar schoonheid besteedde dat honderd andere vrouwen eraan te kort kwamen. Dat gevoel bekruipt mij ook als ik weer eens zo'n subliem opgemaakte webtekst op het beeldscherm krijg. Dan moet ik denken aan al die goedkope facsimile-edities die je op Gallica kunt vinden, downloadbaar in pdf en tiff. Maar als ik die webtekst écht nodig heb, zit er niets anders op dan hem bij stukjes en beetjes te kopiëren en van alle codering te strippen totdat alleen de letters overblijven.

Ik wil geen zout in open (source) wonden strooien en niemand op zijn XML-hart trappen dat zijn TEI-ziel ervan kwekt. Maar met de *Australische schapenboer* in gedachten wil ik, met alle bewondering voor de extra functionaliteit die zo'n geavanceerde webtekst biedt, er toch voor pleiten teksten/bronnen ook als downloadbaar bestand aan te bieden in een algemeen geaccepteerd document formaat.

Dr. Willem Kuiper is als universitair docent verbonden aan de leerstoelgroep Historische Nederlandse Letterkunde UvA en als senior onderzoeker aan het Huygens Instituut.

Gelezen

H. Tjalsma (red.): Archiveren van Digitaal Academisch Erfgoed, een verslag als voorbeeld (Den Haag, DANS, 2006; Studies in Digital Archiving 2) DANS, Data Archiving and Networked Services, heeft een aanpak ontwikkeld om digitale bestanden achteraf te archiveren. Met behulp van deze z.g. ADA-methode, die speciaal is gericht op onderzoeksgroepen en wetenschappelijke instituten, kan DANS samen met de betrokken instelling een retro-archiveringsproject uitvoeren. Zo kunnen vaak oudere, digitale onderzoeksbestanden met wetenschappelijke waarde worden gered. Dit boek geeft inzicht in de mogelijkheden van deze vorm van retro-archivering.

Daarnaast biedt het een rapportage van het ADA-project dat van 2000 tot en met 2003 is uitgevoerd door het Nederlands Historisch Data Archief bij het Meertens Instituut. In dit project, het eerste in zijn soort, werden aanpak, mogelijkheden en kosten verkend van het archiveren van ouder, niet geordend digitaal wetenschappelijk materiaal. Gebaseerd op de ervaringen van dit project is de

ADA-methode ontwikkeld. Het ADA-project was dus een vingeroefening voor de nu door DANS ontwikkelde ADA-dienstverlening.

Ad van der Meer, Onno Boonstra: Repertorium van Nederlandse gemeenten 1812-2006. (Den Haag, DANS, 2006; Dans Data Guide 2)

De Nederlandse gemeenten van nu zijn niet meer die van 200 jaar gele-

den. Door opheffingen en samenvoegingen is het aantal gedaald van 1249 in 1820 tot 458 in 2006. Gemeenten die wel zijn blijven bestaan hebben vaak een enorme gebiedsuitbreiding ondergaan en een veelvoud aan inwoners gekregen. Ook zijn er honderden wijzigingen geweest in de namen die voor de gemeenten werden gebruikt. Vaak ging het om spellingvarianten, soms om officiële naamswijzigingen.

Historisch onderzoek wordt ernstig belemmerd door deze voortdurende veranderingen. Deze uitgave wil deze belemmeringen weg nemen door – voor het eerst – een volledige lijst te bieden van alle gemeenten die in Nederland vanaf 1812 tot 2006 hebben bestaan, inclusief alle bekende varianten van de gemeentenamen. Tevens laat het zien uit welke gemeenten een nieuwe is gevormd, of in welke gemeente een oude is opgegaan. De continuïteit wordt benadrukt door de introductie van een nieuwe gemeentecode – de Amsterdamse code – die, anders dan de codes die momenteel in gebruik zijn, onveranderd blijft bij grens- en naamswijzigingen.



Koninklijke Nederlandse Akademie van Wetenschappen: De toegankelijkheid van gegevens uit publieke en semi-publieke administraties voor wetenschappelijk onderzoek, KNAW-advies. (Amsterdam, KNAW, 2006)

Met de opkomst van informatisering en digitalisering worden in administraties van overheids- en semi-overheidsinstellingen op systematische wijze steeds meer gegevens gegenereerd die voor wetenschappelijk onderzoek van belang kunnen zijn. Het betreft veelal gegevens die in juridische zin wel voor wetenschappelijk onderzoek toegankelijk zijn, maar in de praktijk daarvoor niet of slechts met veel inspanning beschik-

baar komen. Vaak moet er door wetenschappelijke onderzoekers extra worden betaald voor gegevens die al met publieke middelen tot stand zijn gekomen, zoals de gemeenschappelijke basisadministratie persoonsgegevens (GBA). Bij andere gegevensbestanden zoals de belastingadministratie zijn er wettelijke beperkingen.

In dit advies gaat de de Sociaal-wetenschappelijke Raad in op de maatschappelijke baten van een betere toegankelijkheid van gegevens uit (semi-)publieke administraties voor wetenschappelijk onderzoek aan de hand van een aantal voorbeelden.