

**Pagina 4 • Schrijfsters in een databank**

Drie jaar is er gewerkt aan een databank met 17 duizend records over de ontvangst van schrijfsters in Nederland in de 18<sup>e</sup> en 19<sup>e</sup> eeuw. Nu kan het onderzoek echt beginnen.

**Pagina 4 • Handgeschreven tabellen**

Bij het digitaliseren van de Volktellingen werden handgeschreven tabellen verwerkt. In een pilotproject is onderzocht of dat ook met optische tekenerkenning kon.



WILLIAM HOOGTEYLING

**Pagina 5 • Passie voor data**

Een onderzoeker met een passie voor cijfers: historicus en econoom Jan Luiten van Zanden houdt van 'tellen vanuit een theoretisch en conceptueel kader'. Dat is een cumulatief proces, zegt hij in een interview met e-data& research, waarbij onderzoekers op elkaars kennis kunnen voortbouwen.

**Pagina 6 • Een nieuwe start voor het TEI**

De nieuwe Guidelines maken het Text Encoding Initiative in zijn 21<sup>ste</sup> levensjaar opener.

**Pagina 7 • Panelstudie Jeugd en Cultuur**

De gegevens van deze multi-actor panelstudie over de periode 1998-2005 zijn vrij beschikbaar via DANS

**EN VERDER**

Agenda.....	2
Nieuws.....	3
Achtergrond.....	4
Ontwikkelingen.....	7
Column.....	8
Gelezen.....	8

## Huygens Instituut werkt aan Digitaal Wetenschapshistorisch Centrum

Het Huygens Instituut in Den Haag werkt sinds vorig najaar aan de ontwikkeling van een 'Digitaal Wetenschapshistorisch Centrum' (DWC). Het moet de digitale toegang worden voor onderzoekers en andere geïnteresseerden tot de Nederlandse wetenschapsgeschiedenis.

Het centrum gaat via een website nieuws, recente ontwikkelingen en relevante links uit de professie aanbieden, maar ook toegang tot een bibliografisch apparaat over het thema 'Geleerdheid & Wetenschapsbeoefening in de Lage Landen'. De voornaamste ingang zal worden geboden via een biografische database. Biografieën in deze database moeten worden ontsloten met behulp van high-tech zoekfuncties. Via de biografieën wordt toegang geboden tot bijvoorbeeld fulltext bestanden en audiovisuele presentaties. In de verzameling van biografieën zullen ook minder bekende wetenschapsbeoefenaren worden opgenomen.

Wetenschapsgeschiedenis behandelt de vraag hoe het wetenschappelijk bedrijf in verschillende contexten werkt, hoe methoden en technieken

## Datakeurmerk gaat eerste fase in

Zeventien richtlijnen bevat de eerste versie van het Datakeurmerk die in februari aan de openbaarheid is prijsgegeven. Ze zijn te downloaden vanaf [www.datakeurmerk.nl](http://www.datakeurmerk.nl). De makers bij DANS verwachten nu een brede discussie die tot verdere verbetering kan leiden.

Het datakeurmerk bevat de voorwaarden waaraan dataproducent, -consument en -archief moeten voldoen om het etiket 'toekomstbestendig' te verdienen. Het tot stand brengen van zo'n keurmerk heeft DANS (Data Archiving and Networked Services) als opdracht meegekregen van haar beide moederorganisaties KNAW en NWO. 'Maar het is niet de bedoeling om er een monopolie op te gaan koesteren,' verzekert adjunct-

directeur Henk Harmsen van DANS die het keurmerk in december binnen de KNAW en half februari op een congres in Zuid-Afrika presenteerde. 'Wij zijn er juist op uit om betrokkenheid te organiseren van grote dataproducenten en -beheerders, zodat het een gedeelde verantwoordelijkheid wordt. We streven naar een hele reeks relevante logo's op de home page van de website'

DANS heeft bij het formuleren

van de eerste *draft* gebruik gemaakt van het werk dat eerder door anderen is gedaan. Harmsen: 'Wij hebben veel gehad aan de richtlijnen die in documenten als TRAC, DRAMBORA en in de *Kriterienkatalog* van het Duitse NESTOR zijn neergelegd. Het verschil is dat wij minder diep zijn gegaan op het gebied van de data zelf, en meer aandacht besteden aan de verantwoordelijkheid van de partijen die aan het archiveren te pas

komen. TRAC kent bijvoorbeeld 75 richtlijnen terwijl het keurmerk er maar 17 heeft. Maar wij hebben wel weer richtlijnen voor de dataproducent en de dataconsument'.

In de komende tijd wil DANS bekendheid geven aan het keurmerk en discussie losmaken. Nationaal wordt het gesprek gezocht met belangrijke repositories en datapartners en met de Nationale Coalitie voor Digitale Duurzaamheid, internationaal met projecten als Planets en Driver II. NWO zal het keurmerk gaan hanteren als een van de eisen die worden gesteld bij het toekennen van nieuwe subsidies. (MdG)

### Ook internationale belangstelling

## Subsidie voor boomtjidskunde

'Nu ons project kan worden uitgevoerd gaat het mes aan twee kanten snijden', reageert prof.dr. Esther Jansma op het besluit van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) om er geld voor beschikbaar te stellen. Zij krijgt 565.000 euro voor de ontwikkeling van een digitale databank op het terrein van de dendrochronologie.

Jansma, bijzonder hoogleraar aan de Universiteit van Utrecht en onderzoeker aan de Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten (RACM), verwacht veel van de uitstraling van het project: 'We kunnen bouwen aan een grote database die heel veel nieuw onderzoek mogelijk maakt. En er ontstaat een norm voor de manier waarop data moeten worden opgeslagen. Dat is nodig, want op dit moment bestaat er eigenlijk vanuit de opdrachtgevers voor dendro-

nologisch onderzoek geen aandrang in die richting'. De projectleider verwacht dat van haar project ook internationale invloed zal uitgaan in de richting van standaardisering.

Bij het project, getiteld 'Digital Collaboratory for Cultural-Historical Dendrochronology in the Low Countries (DCCD)', gaat het om het aanleggen van een databank op het gebied van de boomtjidskunde, waarin alle materiaal wordt samengebracht en doorzoekbaar gemaakt dat nu verspreid is over universiteiten en laboratoria in Nederland en België. Boomtjidskunde of dendrochronologie is de wetenschap die onder meer het bepalen van de leeftijd van bomen en hout mogelijk maakt op basis van de binnenkant van dat hout, die variaties verradert in de jaarlijkse groei. Maar Jansma wil graag toelichten waarom het belang van de wetenschap, en van haar project, veel verder reikt. 'Van hout in cultureel erfgoed kan niet alleen de absolute leeftijd, maar ook de herkomst worden bepaald. In de Nederlandse bodem zijn bijvoorbeeld heel wat Romeinse rivierschepen gevonden, waarvan men steeds dacht dat deze hoger stroomopwaarts langs de Rijn in tegenwoordig Duitsland waren



RACM

### De Romeinse rivierpraam 'De Meern 1' (Utrecht, Leidsche Rijn) werd gebouwd rond 148 n.Chr. uit enkele reusachtige eiken die groeiden in tegenwoordig Nederland, blijkt uit de jaarringpatronen

gebouwd. Maar de vergelijking met de stamresten van oude bossen die in onze bodem bewaard zijn gebleven, heeft laten zien dat ze gemaakt zijn uit in ons land gegroeide eiken, en dus in tegenwoordig Nederland. Dit type onderzoek wordt door het DCCD op grotere ruimtelijke schaal en ook voor andere perioden mogelijk. Ik verwacht door de toenemende kennis over de houterkomst ook spin-off op het terrein van de vegetatie- en klimaatreconstructie'.

Het project wordt uitgevoerd door de Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten (RACM), in samenwerking met het

data-instituut DANS (Data Archiving and Networked Services) en de Faculteit der Geowetenschappen van de Universiteit Utrecht (UU). De RACM en de laboratoria die over de betrokken dataverzamelingen beschikken, investeren zelf gezamenlijk voor € 300.000 capaciteit in het project. 'Veel van het door NWO toegezegde geld gaat naar de instituten die de data hebben', licht Jansma toe. 'Dat deel wordt gebruikt voor het inkloppen maar vooral ook upgraden van de gegevens. Een groot deel gaat verder naar de soft- en hardware die nodig zijn voor de infrastructuur rond de nieuwe database'. (MdG)



ARBEIDERSPERS

Esther Jansma



## AGENDA

14 maart

Den Haag, NWO Java-gebouw, DANS i.s.m. het Veteranen Instituut en UT.

Studiedag 'Interviews uit de kast'

Centrale vragen: welke mogelijkheden zijn er om interviewmateriaal duurzaam te bewaren en ontsluiten, hoe kan het uniform worden genoteerd, hoe blijft het na gebruik beschikbaar voor andere onderzoekers? Ook is er aandacht voor verschillen tussen de disciplines die interviews als bronnenmateriaal gebruiken. Onderzoekers vertellen hoe zij bestaande interviewdata gebruiken. Ook is er aandacht voor beheer en ontsluiting van interviewcollecties, methodologische aspecten en mogelijkheden van ICT bij analyse van interviewdata. [www.dans.knaw.nl/interviewsuitdekast](http://www.dans.knaw.nl/interviewsuitdekast)

Maart/april 2008

VGI: Workshop Social Software; wat levert het in de praktijk op?

Verschillende mogelijkheden van social software worden getoond en besproken. Afhankelijk van de wensen en het kennisniveau van de groep, wordt er gewerkt met sociaal bookmarken, het uitlezen en bijhouden van RSS-feeds, het maken van een sociaal netwerk of het onderhouden van een wiki. Eventueel kan er ook een blog, een fotopagina of een Twitter-pagina gemaakt worden. Zie ook pagina 7. [www.vgi-online.nl](http://www.vgi-online.nl)

9 april 2008

Amsterdam, Trippenhuis, DANS, NPSO Symposium 'Wat is goed survey-onderzoek? Doel en middelen'

Bij de eerste bijeenkomst van het Nederlands Platform voor Survey Onderzoek (NPSO) in oprichting. Sprekers uit de drie 'bloedgroepen' (overheid, academia en bedrijfsleven) over wat goed survey-onderzoek is. Ook zal worden gesproken over het vormgeven van de activiteiten van het platform. [www.npsos.net](http://www.npsos.net)

11 april

Amsterdam, IISG

Levenslopen in de 19e en 20e eeuw; afsluitend congres van het NWO-groot programma Life Courses in Context

In het programma *Life Courses in Context* is de vijf jaar gewerkt aan de reconstructie van de levensloop van 40000 Nederlanders geboren tussen 1863 en 1922. Deze database werd aangevuld met macrodata uit de beroeps- en volkstellingen van 1859 tot 1947. Mogelijkheden van deze nieuwe database worden hier gepresenteerd door historici, demografen en sociologen. [www.iisg.nl/~hsn/news/2008congres.php](http://www.iisg.nl/~hsn/news/2008congres.php)

21-22 april

Amsterdam, OB, NUV, Vereniging van Letterkundigen, FOBI, Stichting Auteursrechtmanifestaties

Symposium 'The Book in the Internet Era: Copyright and the Future for Authors, Publishers and Libraries'

In afzonderlijke sessies voor elk van de drie betrokken partijen werpt dit symposium licht op het problematische gebied van auteursrechten en internet: de auteur, de uitgever en de bibliotheek. Zo wordt geprobeerd de drie onderling te verbinden en tegelijk de publieke belangstelling voor het onderwerp te stimuleren. [www.amsterdamworldbookcapital.com/index.cfm?page=Copyright%20symposium](http://www.amsterdamworldbookcapital.com/index.cfm?page=Copyright%20symposium)

16 mei

Nijmegen, Huize Heyendaal, Huizinga Instituut/TS. Tijdschrift voor tijdschriftstudies

Conferentie Bits, Bytes en Bladen.

De stand van zaken op het gebied van tijdschriften en digitalisering wordt besproken vanuit zoveel mogelijk verschillende praktische en theoretische invalshoeken. Het is de bedoeling dat verschillende tijdschriftgenres aan de orde komen. Aanmelden en informatie: [m.koffeman@let.ru.nl](mailto:m.koffeman@let.ru.nl)

27-30 mei

Stanford University Palo Alto – California, USA, IASSIST 2008 Conference

Technology of Data: Collection, Communication, Access and Preservation

There have been vast changes in compilation and dissemination methods; increasing awareness about access and associated licensing and privacy issues; and growing concern about the safeguarding and protection of valuable data resources. The role of technology – past, present, and future – in all of these arenas is discussed here. [www.iassistdata.org/conferences/](http://www.iassistdata.org/conferences/)

## Presentatie Volkstellingenbundel

## 'Grijp kansen, pas op voor verkeerd gebruik'

Enthousiasme over de grote mogelijkheden voor onderzoek; behoedzaamheid voor de gevolgen van onoordeelkundig gebruik. Dat waren hoofdstromingen in de reacties toen op 15 november bij het CBS in Voorburg het boek werd gepresenteerd waarmee de digitalisering van de Volkstellingen werd afgerond.

Het CBS en drie KNAW-instituten werkten sinds 1997 samen aan het project. De resultaten zijn te vinden op de website [www.volkstellingen.nl](http://www.volkstellingen.nl). Die site heeft sinds de openstelling in 2004 al meer dan honderdduizend bezoekers uit binnen- en buitenland gescoord, zo werd bij de presentatie bekendgemaakt. De publicatie 'Twee eeuwen geteld; Onderzoek met de digitale volks-, beroeps- en woningtellingen, 1795-2001' toont aan de hand van een reeks analyses op de vrijgekomen gegevens de vele mogelijkheden die het Volkstellingenbestand biedt. Er staan artikelen in over bevolking, gezin, arbeid, wonen, infrastructuur en veel andere onderwerpen.

## De domme maar vruchtbare vragen van de computerwetenschapper

Wat kan de computerwetenschap betekenen voor de humaniora? Dat was de vraag die Willard McCarty, hoogleraar *humanities computing* in Londen, probeerde te beantwoorden in zijn keynote lezing op het Catch 'midterm event', 30 november. Catch is het NWO-programma voor computerwetenschappelijk onderzoek ten behoeve van de toegankelijkheid van cultureel erfgoed. Bij het event deden de verschillende projecten verslag van hun vorderingen.

McCarty is de auteur van het invloedrijke boek *Humanities Computing*, waarin hij de rol en plaats van *humanities computing* ('alfa-informatica', 'historische informatiekunde', 'computer en letteren') tussen de andere disciplines onderzoekt. Een van zijn centrale stellingen is dat het er niet om gaat problemen op te lossen, maar om ze op te sporen. De computer helpt de geesteswetenschapper zijn terrein te modelleren, en dwingt af dat dat volgens formele regels verloopt. De wetenschapper kan zo het model met de computer manipuleren en de gaten tussen model en werkelijkheid zichtbaar maken. Meer dan het model zelf is dit proces van iteratieve verbetering van het model de essentie van *humanities computing*. Het voorbeeld dat McCarty in zijn boek gebruikt, is zijn eigen langdurige studie van het verschijnsel personificatie in Ovidius' *Metamorfosen*.

De invalshoek voor deze keynote was de kennelijke aversie van de humaniora tegen een expliciete methode, die al snel mechanisch en triviaal lijkt. De methoden van

toonde zich, nadat hij een van beide eerste exemplaren van het boek in ontvangst had genomen, onder de indruk van de inhoud en dat was te meer opmerkelijk in het licht van zijn aanvankelijke 'sceptis over blinde digitalisering'. De vraag 'wie hier nu echt mee werkt' was met de bundel voor hem afdoende beantwoord. Wel vroeg hij zich af 'of er niet een bijsluiter bij levering van de gegevens moet worden toegevoegd om ondeskundig gebruik te voorkomen?'. Gebeurt dat niet, dan zouden veel media op onverantwoorde wijze met de cijfers aan de haal gaan, veronderstelde hij. CBS-directeur Gosse van der Veen, die het andere 'eerste' exemplaar had aangepakt, voegde toe dat ook het CBS vaak een rapport meeleverd 'om uit te leggen wat je er wel en niet mee kan'.

De Nijmeegse historicus prof. Paul Klep, naast socioloog Pieter

Hooimeijer een van de commentatoren op de pas verschenen bundel, haalde een paar heikle periodes uit de geschiedenis van de volkstellingen naar voren en concludeerde dat de politiek er goed aan had gedaan

## Zoeken is veel meer dan Google

Onder de titel 'Zoeken is veel meer dan Google' organiseerde de Vereniging voor Geschiedenis en Informatica eind 2007 een studiedag in Leiden. De drukbezochte dag stond in het teken van zoeken en vinden op het web. Commercieel gezien beheerst Google de markt van de zoekmachines, in Nederland nog meer (90% marktaandeel) dan in de VS (ca. 66%). 'Is Google wel geschikt voor de wetenschappelijke wereld?' was de rode draad van de dag. Het was interessant om te zien hoe verschillende organisaties eigen zoekstrategieën en -machines ontwikkelen. Jacqueline Limpens van

om het instituut toch steeds weer in stand te houden. Het boek demonstreert het wetenschappelijk gebruik en belang van de volkstellingen en de website [www.volkstellingen.nl](http://www.volkstellingen.nl) vond hij 'werkelijk schitterend'. (MdG)

het Academisch Medisch Centrum van de UvA nam in haar presentatie de toehoorders mee op een bijna interactieve zoektocht naar medische informatie. Eric Sieverts liet zien hoe men bij de Universiteitsbibliotheek van de UU het succes van Google niet kon negeren bij de ontwikkeling van het eigen Omega zoekstelsel. En het ingenieuze zoekstelsel Open Boek, gepresenteerd door Hans Paijmans van de Rijksdienst voor Archeologie, Cultuurlandschap en Monumenten, is gespecialiseerd in het zoeken van jaartallen en chronologische referenties in Nederlandse teksten.

In deze voorbeelden gaat het om wetenschappelijke zoekmachines en -strategieën waarbij geen gebruik wordt gemaakt van Google. Maar gebruikt dan geen enkele wetenschapper Google? Voor Peter Verbist van de KU Leuven staat vast dat Google veel meer is dan alleen maar zoeken en dat de services die Google biedt zich steeds verder uitbreiden. Verbist was de enige die in zijn presentatie het standpunt innam dat Google's services ook wetenschappers prima van dienst kunnen zijn. Hij richtte daarbij zijn blik vooral op een toekomst waarin die services nog verder uitgebreid en toegespitst kunnen worden. Daarmee zal Google volgens hem ook wetenschappers uitstekend kunnen bedienen zonder dat instellingen ingewikkelde eigen zoekmachines en -strategieën hoeven te ontwikkelen.

In de afsluitende presentatie belichtten Arjan van Hessen (Universiteit Twente) en Peter van der Maas (NIOD) het 'Radio Oranje'-project. Invalshoek daarbij was de zeer interessante toepassing van een full text zoekstrategie gekoppeld aan gesproken woord. Door OCR-technieken te combineren met spraakherkenning zijn toespraken van Koningin Wilhelmina voor Radio Oranje toegankelijk gemaakt en kan er 'op zijn Googles' gezocht worden. De presentatie van de zoekresultaten waarbij beeld, geluid en tekst worden gecombineerd was indrukwekkend.

De studiedag leverde geen definitief antwoord op de vraag of Google wel geschikt is voor de wetenschappelijke wereld. Maar wel bleek opnieuw dat Google in de commerciële markt een zeer belangrijke plek inneemt en dat bijna iedere zoekmachine, ook de wetenschappelijke, met Google wordt vergeleken. (LS)



Bernini, Apollo en Daphne. Daphne verandert in een boom om te ontsnappen aan Apollo. Naar aanleiding van Ovidius' *Metamorfosen*.

de geesteswetenschapper blijven vaak onuitgesproken. Maar juist daarom kan de ontmoeting met de computerwetenschapper vruchtbaar zijn. 'In setting out to make things, computer scientists have to ask *dumb questions*, that is, questions which have been kept voiceless because the keepers of the domain of those questions have not thought to ask them. Giving these questions a voice – which is to say, *problematising* that which provokes them – is a

core function of the humanities. So in asking these *dumb questions* computer scientists enter into the game of the humanities on the common ground of *humanities computing*.'

McCarty's filosofische en erudiete stijl van schrijven maakt zijn teksten lezenswaardig maar soms ook moeilijk toegankelijk. Het was een goed idee van de Catch-organisatie hem uit te nodigen voor een mondelinge uiteenzetting. (PB)



# Project Getuigen Verhalen van start

In januari zijn de eerste voorstellen gehonoreerd in het kader van het project Getuigen Verhalen. In dat project worden persoonlijke getuigenissen uit de Tweede Wereldoorlog vastgelegd. Het is onderdeel van het programma Erfgoed van de Oorlog, waarvoor in totaal 21,6 miljoen euro beschikbaar is.

Met Erfgoed van de Oorlog wil het ministerie van VWS belangrijk materiaal uit en over de Tweede Wereldoorlog behouden en het gebruik ervan stimuleren. Het project Getuigen Verhalen hanteert een 'pluriform perspectief': bij het interviewen van ooggetuigen van de Tweede Wereldoorlog krijgen vooral nieuwe of onderbelichte thema's, invalshoeken en groeperingen aandacht.

De projectleiders verwachten dat het Getuigen Verhalen project ongeveer vierhonderd interviews zal opleveren, die voor iedereen toegankelijk zullen zijn. Het project schrijft voor het afnemen van de interviews het

gebruik van hoogwaardige opname-apparatuur voor een elk interview dient goed gedocumenteerd te worden. Transcripties maken de interviews inhoudelijk toegankelijk. Enkele belangrijke interviewcollecties zijn al laagdrempelig en open toegankelijk, maar voor veel andere geldt dat niet – ook los van privacybescherming en auteursrechten. Voor het project Getuigen Verhalen is de openbare toegang tot de interviews een belangrijke doelstelling. Daarom zal dat project een presentatie verzorgen op de studiedag Interviews uit de Kast die op 14 maart georganiseerd wordt (zie Agenda, p.2). (Rene van Horik)



Razzia op het Amsterdamse Paul Krügerplein, 20 juni 1943.

[www.dans.knaw.nl/interviewsuitdekast](http://www.dans.knaw.nl/interviewsuitdekast)  
[www.minvws.nl/dossiers/erfgoed-van-de-oorlog](http://www.minvws.nl/dossiers/erfgoed-van-de-oorlog)

## Evaluatie van bestandsformaten voor duurzame opslag en toegang

Hoe kunnen we bepalen in welke formaten objecten het best kunnen worden aangeleverd bij het e-Depot? Wat zijn de risico's van formaten voor de langdurige opslag en de toegankelijkheid in de toekomst? En welke strategie moet er worden toegepast om de documenten toegankelijk te houden? Om deze

vragen te beantwoorden heeft de Koninklijke Bibliotheek (KB) een methode ontwikkeld die bestandsformaten evalueert op een zevental criteria op deze gebieden.

Sinds 2003 deponeren nationale en internationale uitgevers digitale publicaties in het e-Depot van de KB. Tot voor kort werd het

overgrote deel van deze publicaties aangeleverd in het Portable Document Format (PDF). De komende tijd zal de inhoud van het e-Depot echter steeds diverser worden door de aanlevering van materiaal uit projecten als webarchivering en massadigitalisering.

Elk van de zeven criteria is onderverdeeld in één of meerdere karakteristieken waarop het formaat een score ontvangt. Elke karakteristiek ontvangt een weging. De weging ligt niet vast, maar is afhankelijk van het beleid van de specifieke instelling

die de methode toepast. Uiteindelijk ontvangt het formaat een score op een schaal van 0-100. Deze score wordt vergeleken met de scores van andere formaten in dezelfde categorie (bijvoorbeeld 'opgemaakte tekst documenten', 'afbeeldingen', 'geluid'). De uitkomst kan worden gebruikt in het opstellen van een strategie voor digitale duurzaamheid. De methode is gepresenteerd op de IPRES 2007 conferentie en zal binnenkort worden gepubliceerd op de KB website. (Judith Rog, Caroline van Wijk)

[http://www.kb.nl/hrd/dd/dd\\_links\\_en\\_publicaties/links\\_en\\_publicaties\\_intro.html](http://www.kb.nl/hrd/dd/dd_links_en_publicaties/links_en_publicaties_intro.html)

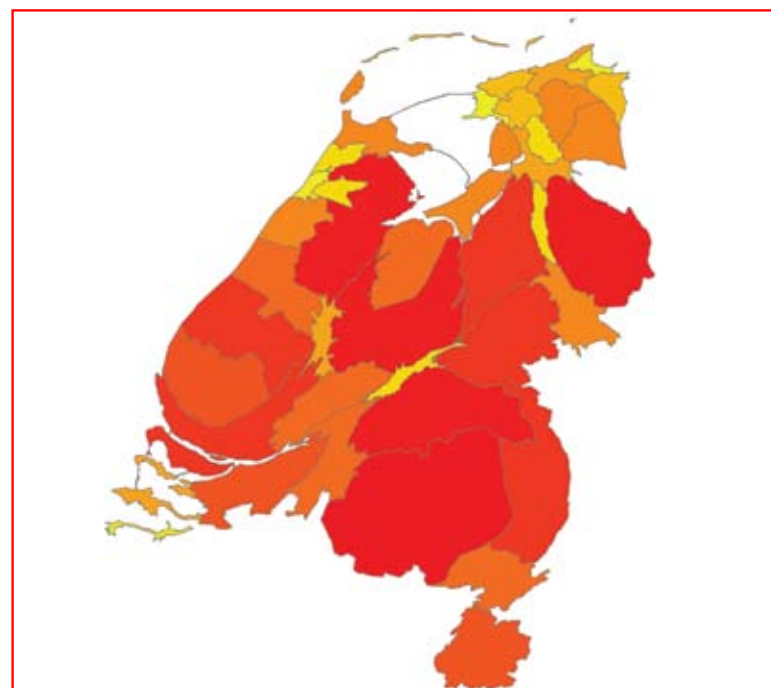
## Anamorphose van high-tech werkgelegenheid in Nederland

Anamorphoses zijn kaarten waarop de grootte van geografische eenheden (regio, gemeenten of landen) geschaald is naar een bepaalde indicator, bijvoorbeeld de omvang van de bevolking of het inkomen. In de figuur zijn de veertig COROP-regio's in Nederland geschaald met de omvang van de werkgelegenheid in high-tech sectoren. Duidelijk blijkt dat de regio Eindhoven de grootste concentratie high-tech werkgelegenheid kent.

Niet iedere indicator kan gebruikt worden om er een anamorphose van te maken. De indicator mag niet afhankelijk zijn van het schaalniveau waarop wordt gemeten. Zo moet de optelsom van alle indicatorwaarden van gemeenten in een provincie

gelijk zijn aan de waarde van deze indicator in die provincie. Ook moet een juiste minimumgrootte moet worden gekozen waarmee de verhoudingen van alle regio's onderling worden bepaald.

Een probleem bij het maken van anamorphoses is om rationele grenzen tussen de gebieden te behouden, zodat de lezer de kaart in één oogopslag met een werkelijke geografische kaart kan vergelijken. Gelukkig zijn hiervoor enkele gratis via internet beschikbare programma's te krijgen. Hier is gebruik gemaakt van het programma Cartogram, ontwikkeld door Michael Gastner. Het maakt gebruik van shapefiles; basiskennis van arcinfo is noodzakelijk. (Mark Thissen en Frank van Oort)



## Alliantie wil gevaar keren voor wetenschappelijke informatie

Op initiatief van de Koninklijke Bibliotheek richtten vooraanstaande organisaties uit wetenschap en de wetenschappelijke informatievoorziening de *European Alliance for Permanent Access* op.

Net als andere sectoren is ook de wetenschap afhankelijk geworden van digitale informatie. Het behoud en de toegankelijkheid van die informatie kunnen op de lange termijn moeilijk worden gegarandeerd. Wanneer er digitale onderzoeksgegevens verloren gaan, loopt het wetenschappelijke proces direct gevaar. De alliantie wil er voor zorgen dat in Europa een organisatorische infrastructuur wordt opgezet die de duurzame toegankelijkheid tot digitale wetenschappelijke informatie waarborgt. Er zijn tot op heden zeventien organisaties lid van, uiteenlopend van de European Space Agency tot het Max Planck Geselschaft en van de European Science Foundation tot de organisatie voor kernonderzoek CERN.

Onderzoek naar en ontwikkeling van methoden en technieken om digitale informatie voor de lange termijn bruikbaar te houden staan nog in de kinderschoenen. De kosten voor hoogwaardige bewaring

en duurzame toegankelijkheid zijn grotendeels onbekend. De verantwoordelijkheden voor een duurzaam beheer van de digitale *records of science* zijn volgens de leden van de Alliantie in Europa nog onvoldoende bij betrouwbare organisaties belegd. De beoogde Europese infrastructuur is een virtuele infrastructuur die de alliantie stap voor stap wil opbouwen in overleg met relevante wetenschappelijke partijen, de Europese Commissie en nationale overheden. Als aftrap voor dit overleg heeft de Alliantie zich eind vorig jaar met een strategische conferentie in Brussel aan vertegenwoordigers van grote (Europese) onderzoeksinstituten, de Europese Commissie en nationale overheden gepresenteerd. Ze wil met deze partijen praten over het belang van een infrastructuur die permanente toegang tot wetenschappelijke informatie voor Europa garandeert. (Ingrid Dillo)

[www.alliancepermanentaccess.eu](http://www.alliancepermanentaccess.eu)

### DANS neemt deel in DPE

Het Nederlandse data-instituut DANS is associate partner geworden van het Europese netwerk Digital Preservation Europe (DPE). Dit netwerk bundelt in Europa kennis op het terrein van de bewaring van digitale objecten voor de lange termijn. DPE heeft in het verleden al instrumenten opgeleverd voor audit en certificering van digitale depots, een checklist en richtlijnen voor het opzetten en beheren van digitale depots, een systeem voor unieke identificatie van digitale objecten en een onderzoekagenda. DANS wil workshops en seminars organiseren over de digitale bewaring en 'briefing papers' produceren op het gebied van data, vooral over de uitkomsten van research and development-projecten. Deelnemers in DPE zijn universiteiten, bibliotheken, cultureel erfgoed instellingen en archiefdiensten.

[www.digitalpreservationeurope.eu](http://www.digitalpreservationeurope.eu)

### COST-subsidie InterEdition

Het Europese programma om internationale wetenschappelijke samenwerking te bevorderen, COST-ESF, heeft eind vorig jaar subsidie toegekend aan het voorstel 'An interoperable supranational infrastructure for digital editions (InterEdition)'. Het voorstel (zie ook *e-data* nr. 1-4, maart 2007) was door het Huygens Instituut ingediend namens tien Europese onderzoeksinstituten, waaronder ook DANS. Het project moet een internationaal gedragen protocol opleveren en een voorbeeldimplementatie voor een infrastructuur voor gedistribueerde digitale tekstedities. Het uit te keren bedrag kan oplopen tot 400.000 euro. (Joris van Zundert)

### Nieuw datacentrum in VS

In de Verenigde Staten is vorig najaar het International Data Resource Center (IDRC) opgericht om internationale data aan onderzoekers over de hele wereld te bieden. Initiatiefnemer is het Amerikaanse Inter-university Consortium for Political and Social Research (ICPSR). Aanleiding was volgens het IDRC de groeiende behoefte aan zulke data: 'Toegang tot internationale data is een kritische factor geworden voor onderzoekers over de hele wereld'. Via de website van het centrum zijn data te verkrijgen die worden beheerd door oprichter ICPSR. Studenten en medewerkers van de meeste universiteiten in Nederland kunnen ze downloaden via de service ICPSR Direct. Gedetailleerde informatie daarover biedt de website van DANS.

[www.dans.knaw.nl](http://www.dans.knaw.nl)  
[www.icpsr.umich.edu/IDRC/](http://www.icpsr.umich.edu/IDRC/)

### Platform survey-onderzoek

Onderzoekers van het Centraal Bureau voor de Statistiek, het Sociaal-Cultureel Planbureau en DANS (Data Archiving and Networked Services) willen een Nederlandstalig Platform voor Survey-Onderzoek oprichten. Er verandert veel in de surveymethodologie, zeggen ze, en er is behoefte aan een overzicht van standaarden, *best practices*, collegiaal advies en ondersteuning. Internationaal zijn zulke mogelijkheden er wel maar in het Nederlandse taalgebied niet, aldus de oprichters op hun website [www.npsa.net](http://www.npsa.net). In april organiseert het Platform in oprichting een eerste symposium in Amsterdam (zie pagina 2, Agenda)



# Schrijfsters in een databank

SUZAN VAN DIJK

Op 21 september 2007 vond in het Huygens Instituut de afsluiting plaats van het NWO-digitaliseringsproject *The International Reception of Women's Writing*, waaraan van 2004 tot 2007 met vier assistenten was gewerkt. Opgeleverd werden 17 duizend records met informatie over de ontvangst in Nederland van Nederlandse en buitenlandse schrijfsters uit de achttiende en negentiende eeuw. In het verlengde van de grote bloemlezing uit 1997 over Nederlandstalige schrijfsters uit de periode 1550-1850, *Met en zonder lauwerkrans*, kan nu het onderzoek naar het aandeel van vrouwen in de literaire wereld grootschalig worden voortgezet.

De records zijn onderdeel van de database *Women Writers*. Deze bevat informatie over de ontvangst door tijdgenoten en vroege literatuurhistorici (tot ca. 1900) van het werk van vrouwelijke schrijvers, ontleend aan verschillende grootschalige bronnen zoals tijdschriften, bibliotheekcatalogi en privé-correspondentie. De records verwijzen en linken soms door naar deze bronnen; een enkele keer is de tekst van bijvoorbeeld een

artikel of brief integraal overgenomen. De informatie is afkomstig uit zowel online toegankelijke als papieren bronnen.

Idealiter zullen ooit databanken zoals *Women Writers* fungeren als distributieapparaten, waar steeds kan worden doorgelinkt naar achterliggende bronnen. Het is nog niet zo ver, maar dankzij de DBNL (Digitale Bibliotheek der Nederlandse Letteren) was het mogelijk om voor enkele

tijdschriften (bijv. *De Gids*, *De Nieuwe Gids*) een directe toegang te maken naar artikelen over schrijfsters die nu veelal onbekend zijn, en waarover belangrijke negentiende-eeuwse critici als Conrad Busken Huet uitgebreid schreven – in positieve of negatieve zin. Met subsidie vanuit het academische repositoryproject DARE is samen met het Huygens Instituut het tijdschrift *Vaderlandsche Letteroefeningen* online gezet. Vanuit *Women-*

*Writers* zijn hyperlinks gemaakt naar de daar gevonden artikelen.

De database biedt dus onderzoekers, studenten en belangstellenden informatie uit bronnen die die informatie, indien gezocht op individueel niveau, niet gemakkelijk prijsgeven. Bovendien is het informatie waarnaar onderzoekers niet gemakkelijk op zoek gaan: *buitenlandse* onderzoekers beseffen niet altijd dat kleine landen in receptie-historisch opzicht interessant zijn, en *Nederlanders* zijn zich er niet per se van bewust dat meer dan zeventienhonderd Nederlandse vrouwen actief zijn geweest als vertaalster/journaliste/schrijfster, én dat hun werk enigermate de aandacht heeft getrokken.

Er is zo een voorraad aan gegevens verzameld die eerder nauwelijks betrokken werden bij het onderzoek

naar het vrouwelijk auteurschap. Nadere bestudering en interpretatie ervan zullen nieuw licht werpen op de rol van vrouwelijke auteurs in het Nederlandse literaire veld tot circa 1900, bezien in de Europese context. In beoogde vervolprojecten zullen ook gegevens uit Portugal, Rusland, België en andere landen worden ingevoerd. Vervolgens zal aan de hand van *Women Writers* in een Europees samenwerkingsproject een 'internationale vrouwenliteratuurgeschiedenis' worden voorbereid.

Daarbij laten wij ons onder andere inspireren door de ideeën van de literatuur- en cultuurhistorici Linda Hutcheon en Mario Valdés, die in *Rethinking Literary History* (2002) uitgaan van de 'literary text as a historical event of production and subsequently of reception' (p. 67). De benadering van het literaire veld als continu verwickeld in dialoog tussen auteur en lezers vormde ook het uitgangspunt voor de database *Women Writers*, met dien verstande dat de aandacht zich richt op de plaats die vrouwen zich wisten te verwerven, bij een vrouwelijk zowel als een mannelijk contemporair publiek. Zoals Margaret Ezell enige tijd geleden al betoogde in *Writing Women's Literary History* (1993), was deze plaats in de eeuwen vóór de negentiende dikwijls veel groter dan de in die tijd ontstane canon deed vermoeden.

 [www.womenwriters.nl](http://www.womenwriters.nl)  
[www.databasewomenwriters.nl](http://www.databasewomenwriters.nl)



MARIE-PIERRE LEGRAND

Leden van de projectgroep 'New approaches to European Women's Writing before 1900' in overleg over de mogelijkheden van een nieuwe literatuurgeschiedenis, juli 2007

## Volkstelling 1947

JAN JONKER

# Herkenning van handgeschreven tabellen

In een pilotproject in het kader van de digitalisering van de Volkstellingen zijn de mogelijkheden onderzocht van optische tekenherkenning bij het verwerken van de handgeschreven tabellen. Jan Jonker was bij dat project betrokken en doet verslag van de bevindingen.

Sinds 1997 hebben DANS en zijn voorgangers samengewerkt met het Centraal Bureau voor de Statistiek (CBS) om de uitkomsten van de Nederlandse volkstellingen digitaal toegankelijk te maken. Na de gedrukte publicaties van de volkstellingen vanaf 1795 tot en met de laatst gehouden telling van 1971, is het digitaliseren van de bij het CBS aanwezige transparanten en lichtdrukken aangepakt. Zo zijn onder meer in 2005 de circa dertig duizend beschikbare transparanten van de Volks- en beroepstelling 1947 gescand en als images toegankelijk gemaakt op [www.volkstellingen.nl](http://www.volkstellingen.nl).

Doel van het digitaliseringsproces is steeds om de cijfermatige gegevens van de tellingen in verwerkbaar vorm beschikbaar te stellen voor nieuw statistisch onderzoek. Dat vereist na het scannen een stap van 'herkenning' van de gegevens, uitlopend op representatie daarvan in tabellen, bijvoorbeeld in Excel. In de praktijk is dat meestal gedaan door een vorm van

data-entry van de gegevens, al dan niet rechtstreeks in Exceltabellen.

Voor Tabel 12 van de Volkstelling van 1947 zijn de actuele mogelijkheden onderzocht van optische tekenherkenning (*optical character recognition* of OCR). Dat is gedaan in een pilotonderzoek met gebruik van *eFlow*, een OCR-applicatie van Top Image Systems Benelux (TIS), uitgevoerd door Xerox Global Services. Uitdarend detail daarbij was dat de statistische gegevens in handschrift zijn ingevuld op de transparanten van deze Volkstelling.

In *TIS eFlow* worden automatisch de images van Tabel 12 geselecteerd uit de totale verzameling images van telling 1947. Images die niet voldoende worden herkend, worden alsnog handmatig geïdentificeerd. Per saldo zijn er meer dan duizend images van Tabel 12, namelijk voor alle gemeenten uit 1947 en voor de wijken van grotere gemeenten. Voor de images van Tabel 12 worden vervolgens de gegevens uit de tabelcellen bepaald

door weging van de resultaten van een aantal OCR-engines. In een daarop volgende *completion* fase worden de OCR resultaten visueel gecontroleerd, aangevuld en/of gecorrigeerd om *false positives* uit te sluiten. Hierbij wordt gebruik gemaakt van validatieregels voor regel- en kolomtotalen om de kwaliteit van de herkende en/of ingevoerde data te valideren.

De mogelijkheden van *TIS eFlow* voor dit doel zijn op zichzelf veelbelovend. In de eerste proeven van Xerox was zelfs sprake van 84% automatisch correct herkende tabelcellen. Wel bleken de resultaten bij een controle achteraf bij het data-instituut DANS nog relatief veel *false positives* te bevatten. Een van de oorzaken daarvan zijn kaderlijntjes om de tabelcellen in de transparanten. In samenwerking tussen DANS, TIS en Xerox is de *eFlow*-applicatie verfijnd, waarbij alle voor Tabel 12 mogelijke controletellingen over regels en kolommen van de tabel zijn gerealiseerd.

Daardoor zijn de resulterende Exceltabellen een zeer nauwkeurige weergave van de cijfers in de originele transparanten. De herkenningprocedure vergt op deze manier wel meer interactief werk in de *completion* fase voor validatie van de uitkomsten. Toch werd nog circa 70% van de tabelcellen geheel automatisch correct herkend.

Het pilotonderzoek met OCR van images van de Volks- en beroepstelling 1947 heeft laten zien dat een exacte beschrijving vooraf van de voor herkenning aangeboden tabellen onmisbaar is om de vereiste nauwkeurigheid te bereiken. Dat geldt in het bijzonder voor de per tabel mogelijke controletellingen. Op basis daarvan zou de *TIS eFlow* applicatie voor elke tabel in detail moeten worden gespecificeerd. Het laatste kost in de huidige situatie nog relatief veel gespecialiseerde arbeid. De OCR aanpak is daardoor nog niet kosteneffectief. Om die reden wordt voor de herkenning van de images uit dit volkstellingjaar alsnog gebruik gemaakt van 'traditionele' data-entry.

	Alleen-wonenden				Totale beroepsbevolking	
	V	M	V	M	V	M
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						

Data-entryscherm met validatieregels



# Jan Luiten van Zanden, historicus met datapassie

## ‘Kwantitatieve onderzoekers kunnen op elkaars schouders staan’

MARTIJN DE GROOT

‘Uiteindelijk is je ambitie om de hele wereld te verklaren. Dat is de heilige Graal. Die vind je nooit, maar je moet het blijven proberen’, zegt prof. dr. Jan Luiten van Zanden, hoogleraar Geschiedenis na de Middeleeuwen aan de Universiteit Utrecht. Hij heeft de lat hoog liggen, en zijn lijst van publicaties is ernaar.

Historicus. Econoom. Allebei tegelijk, en onverzadigbaar onderzoeker van alles wat zich in, tussen, en in de nabijheid van die beide gebieden bevindt. Als je het zo opschrijft klinkt het bijzonder, maar als je het aan Jan Luiten van Zanden zelf vraagt maakt hij er iets heel gewoons van. ‘Ik ben economisch historicus en wat ik doe draait eigenlijk altijd om dezelfde vraag: waarom zijn sommige landen beter ontwikkeld dan andere?’

Een ruime omschrijving maar die is ook wel nodig voor iemand die publiceert over de geschiedenis van de Koninklijke Shell, de landbouw in de negentiende eeuw, loonmatiging en het poldermodel, haringvisserij en walvisvaart in de Gouden Eeuw, de rol van de vrouw bij de opkomst van het kapitalisme, de geschiedenis van de ABN-AMRO bank, de boekenproductie per hoofd van de bevolking en nog veel meer.

In een nog niet gepubliceerd essay over de economische wereldgeschiedenis schrijft de Utrechtse hoogleraar dat hij probeert ‘wat orde aan te brengen in het onderzoek waaraan ik werk—in een poging om mezelf ervan te overtuigen dat achter de klaarblijkelijke chaos van projecten en papers structuur en betekenis schuilgaan’. Gepaste zelfspot natuurlijk, maar niet vrij van realiteitszin.

*Sommige onderzoekers zitten heel erg op hun eigen data of ze verwaarlozen het onderhoud*

‘Nou goed, ik ben wel nieuwsgierig’, geeft hij desgevraagd toe. ‘Dat betekent dat je veel dingen wilt weten, en dan verlies je de grote lijn wel eens uit het oog’. En daar ligt het eerste voorbeeld op tafel. Van Zanden haalt een velletje tevoorschijn uit een van de vele stapels papieren waarmee zijn kamer in het Internationaal Instituut voor Sociale Geschiedenis (IISG) is bezaaid. Er staat een staafdigram op



WILLIAM HOOGTEYLING

met leeftijden. ‘Op het ogenblik ben ik bijvoorbeeld met mijn Utrechtse collega Tine de Moor bezig met een studie naar *age heaping*: leeftijdstapelen’. Hij wijst naar langere staven in het diagram, die steeds worden afgewisseld met een paar kortere. ‘Mensen die minder gecijferd zijn hebben de neiging om hun leeftijd af te ronden. Als je ze vraagt hoe oud ze zijn zeggen ze bijvoorbeeld vijftienveertig als ze eigenlijk drieënveertig of zevenenveertig zijn. In de leeftijdsstatistiek zie je dat terug in hogere scores op de vijftallen. Nu blijkt dat de mate waarin die vijftallen uitschieten boven wat statistisch aannemelijk is, een hele goede indicator is voor de gecijferdheid van een samenleving. Zo kan je uit één gegeven dat vaak nog wel voorhanden is in een samenleving die misschien niet over uitgebreide statistieken beschikt, toch nog een heleboel informatie halen over die samenleving.’ En die informatie zegt dan weer wat over het zogenaamde *human capital*: zoiets als de scholingsgraad van een bevolking, die in de visie van Van Zanden een belangrijke rol speelt in de econo-

mische ontwikkeling. Van Zanden: ‘Soms ben ik bezig met dingen die maar een hele dunne band hebben met de grote lijn, maar die band is er dan wel’.

*Ik pleit ervoor dat data beschikbaar zijn voor referees bij tijdschriften*

En de grote lijn ook. Je zou kunnen zeggen dat Van Zanden zich in de loop van zijn wetenschappelijke carrière stap voor stap bewoog naar een hoger abstractieniveau. Zijn proefschrift, dat in de landbouwwetenschappen van de jaren tachtig als een belangrijk werk werd gezien, ging over de ontwikkeling van de agrarische sector in Nederland in de negentiende eeuw. Dat was geen keuze maar een kans: ‘Er liep een project bij de VU en er werd aan mij gevraagd of ik daar een deel van wilde doen. Wat ik in die studie deed was kwantificeren: met behulp van indicatoren meten van de groei in verschillende regio’s. Toen dat bleek te lukken lag de gedachte voor de hand: als het voor de landbouw kan, waarom dan niet voor de hele economie?’ De economisch historicus, inmiddels hoogleraar geworden, diende een aanvraag in bij NWO voor een groot project om de nationale rekeningen van Nederland over de negentiende eeuw te reconstrueren. Hij kreeg een pionierssubsidie en bracht het project in de jaren negentig met een reeks collega’s tot een goed eind, veel deelpublicaties en, aldus het laudatio bij de Spinozaprijs die hem in 2003

werd toegekend, tot een internationale doorbraak.

Intussen werd zijn blik echter alweer naar het Europese niveau getrokken en eigenlijk nog verder, naar de vergelijking tussen Europa en de rest van de wereld. ‘Dat onderzoek naar de nationale rekeningen liet zien dat Nederland welvarend en economisch voorlijk was geweest binnen Europa, althans tot aan de industriële revolutie. Maar Europa als geheel was natuurlijk op wereldschaal gezien ook welvarend en voorlijk. De vraag voor mij was steeds meer: wat maakt Europa bijzonder? Hoe komt het nu dat Europa het economisch zoveel beter doet dan andere delen van de wereld?’ Het was de opmaat voor de studie van de economische ontwikkeling van Indonesië, waarvoor Van Zanden in 2000 een onderzoekaanstelling kreeg bij het IISG.

Maar wisten we dan niet al dat Nederland zich zo goed kon ontwikkelen juist door de kolonisatie van Indonesië? ‘Dat is een gangbare verklaring,’ wimpelt de onderzoeker de veronderstelling snel af, ‘maar geen voldoende. Want daar gaat weer een nieuwe vraag achter schuil. Laat ik het zo zeggen: ‘Waarom konden wij wel Indonesië koloniseren en Indonesië niet ons? Als je dat gaat onderzoeken kom je aan die ene leidende vraag waar de verschillen in economische ontwikkeling vandaan komen’.

‘Uiteindelijk’, geeft hij tenslotte graag toe, ‘is je ambitie natuurlijk

om de hele wereld te verklaren. Dat is de heilige Graal. Die vind je nooit, maar je moet het blijven proberen’.

Een indicatie voor de richting waarin hij zoekt wil Van Zanden wel geven. ‘Het blijken steeds de zachte krachten die belangrijk zijn voor de ontwikkeling van een economie. Het poldermodel: overleg, tolerantie, burgerschap. Dat zijn ook de factoren die verantwoordelijk zijn voor de vooruitgeschoven ontwikkeling van Europa in de geschiedenis. Met een aantal collega’s in de wereld wil ik bestanden aanleggen om dat idee verdere onderbouwen. Voorbeelden documenteren van successen en van falen, en daarmee langzaam toe werken naar een idee over de werking van het hele wereldsysteem.’

Bij alles wat hij doet legt Van Zanden een passie voor cijfers aan de dag. ‘Ik ben altijd gefascineerd geweest door het getal’, zegt de man die bij het IISG als eerste een voor iedereen toegankelijke databank aanlegde met lonen en prijzen. Is dat eigenlijk niet een beetje de omgekeerde volgorde, eerst gegevens verzamelen en daarna pas zien wat je ermee kan bewijzen? ‘In de praktijk gaat het hand in hand. Tellen en theorie vormen, dat ketst voortdurend op en neer. Bij mij hebben theoretische benaderingen altijd een belangrijke rol gespeeld in de zoektocht naar data. Tellen vanuit een theoretisch en conceptueel kader, daar gaat het mij om. En het mooie is dat dat een cumulatief proces is. Een databank kan je steeds verder uitbouwen. Onderzoekers kunnen op elkaars verworven kennis voortbouwen, steeds verder toetsen, nieuwe dingen aantonen. Bij kwantitatief onderzoek kunnen onderzoekers op elkaars schouders staan. Daar komt bij dat digitalisering en de groei van het internet veel nieuwe deuren openen. Op den duur krijgen we toegang tot data die we in het verleden voor onmogelijk hielden’, vertrouwt Van Zanden, maar daar is wel een verandering van de spelregels voor nodig. ‘Sommige onderzoekers zitten heel erg op hun eigen data of ze verwaarlozen het onderhoud. Ik ken onderzoekers van naam die zonder blikken of blozen zeggen dat ze met hun data net zo omspringen als met hun belastinggegevens: na zes jaar gaat alles in de prullenmand. Het gaat de laatste jaren trouwens wel beter. De openheid en de bereidheid om te delen is veel groter.’

Er is echter nog veel meer mogelijk, zoals echte wetenschappelijke discussie op basis van de beschikbare data: ‘Ik pleit ervoor dat data beschikbaar zijn voor de referees bij wetenschappelijke tijdschriften. Daardoor zou wetenschappelijke discussie mogelijk worden op basis van de data onder een onderzoek. Helaas lopen die tijdschriften, vooral de Amerikaanse, daar niet in voorop. Daar is nog een lange weg te gaan op het gebied van openheid’.

Jan Luiten van Zanden (1955) studeerde in de tweede helft van de jaren zeventig economie en geschiedenis aan de Vrije Universiteit in Amsterdam. Hij promoveerde in 1985 op een proefschrift over de geschiedenis van de Nederlandse landbouw, begon direct daarna als docent aan de Erasmusuniversiteit en twee jaar later als hoogleraar aan de VU. In 1992 verhuisde hij naar de Universiteit van Utrecht. In 2001 begon hij naast zijn Utrechtse leerstoel als senior onderzoeker bij het Internationaal Instituut voor Sociale Geschiedenis. Van Zanden is lid van de Koninklijke Nederlandse Akademie van Wetenschappen. Hij kreeg in 2003 de Spinozaprijs van NWO.



## Focus

### SEOR

LUUK SCHREVEN

'SEOR is als zelfstandige bv onderdeel van de EUR Holding, waar de Erasmus Universiteit eigenaar van is. In de praktijk hebben wij vooral te maken met de Faculteit der Economische Wetenschappen die vertegenwoordigd is in de Raad van Commissarissen', zegt directeur Prof.dr. Jaap de Koning. Zijn bedrijf wil via onderzoek, advisering en onderwijs & training helpen sociaal-economische verschijnselen inzichtelijk te maken en wetenschappelijk verantwoorde oplossingen te vinden voor sociaal-economische problemen.

'SEOR moet zichzelf bedruipen, we werken dus op commerciële basis in opdracht van vooral de overheid, maar we hebben ook wetenschappelijke doelstellingen en we willen publiceren. Door onze organisatievorm kunnen we kruisbestuiving nastreven tussen ons toegepaste onderzoek en het wetenschappelijk onderzoek dat aan de faculteit wordt uitgevoerd. In samenwerking met de faculteit lukt het vaak om ook onze wetenschappelijke doelstellingen te realiseren.'

'Het gaat bij ons onderzoek bijvoorbeeld om beleidsevaluatie van maatregelen op het gebied van de arbeidsmarkt zoals het re-integratiebeleid. Denk daarbij aan vragen als: hoe is die re-integratie georganiseerd? Hoe wordt er uitvoering aan gegeven en wat is de effectiviteit van het beleid?', aldus De Koning. Maar het onderzoek van SEOR is niet alleen gericht op korte-termijnbeleid of acute vragen. SEOR voert ook meer exploratief onderzoek uit. 'Bijvoorbeeld om meer inzicht te krijgen in achtergronden van verschijnselen zoals werkloosheid. Hiervoor worden onder meer transities op de arbeidsmarkt in kaart gebracht en verklaard en worden mensen op lange termijn gevolgd.'

De opdrachtgevers van SEOR zijn in de eerste plaats op het niveau van de centrale overheid te vinden, maar er wordt ook steeds meer onderzoek verricht voor gemeenten. Daarnaast werkt SEOR voor sectororganisaties en bedrijven. 'Voor bedrijven verrichten we vooral analyse en onderzoek op het gebied van marktwerkingsvraagstukken', zegt De Koning. 'Op internationaal vlak werken we onder andere voor de Europese Commissie en de Wereldbank.' Regelmatig komt het voor dat SEOR onderzoek uitvoert in samenwerking met andere instellingen. 'We werken momenteel bijvoorbeeld samen met TNO, maar ook het Rotterdams Instituut voor Sociaal-wetenschappelijk Beleidsonderzoek, het RISBO is een van onze partners.'

Over de gebruikte gegevensbronnen is De Koning kort: 'Er zijn voor ons twee zeer belangrijke bronnen. Enerzijds zijn dat CBS data, zoals het Inkomens Panel Onderzoek (IPO) of de registraties van de UWV en CWI die via het Centrum voor Beleidsstatistiek toegankelijk zijn voor onderzoek. Anderzijds gaat het om de panels van de Organisatie



SEOR-directeur De Koning: 'Hopelijk komt er in de toekomst meer ruimte voor diepgravend, strategisch onderzoek'

voor Strategisch Arbeidsmarktonderzoek (OSA). Op het gebied van data heeft De Koning een belangrijke wens voor de toekomst: 'Het is essentieel voor ons type onderzoek dat de arbeidsmarktpanelen van OSA blijven bestaan en liefst uitgebreid en verbeterd worden.' Daarnaast is er de hoop en het vertrouwen dat het Centraal Bureau voor de Statistiek via zijn Centrum voor Beleidsstatistiek steeds meer data beschikbaar zal stellen voor onderzoek.

De eigen dataverzamelingen van SEOR zijn meestal niet zo omvangrijk. Voor specifieke onderzoeksprojecten worden wel eenmalige enquêtes uitgevoerd waarvan de gegevens soms ook buiten het betreffende onderzoek gebruikt kunnen worden. 'Een belangrijke verzameling die

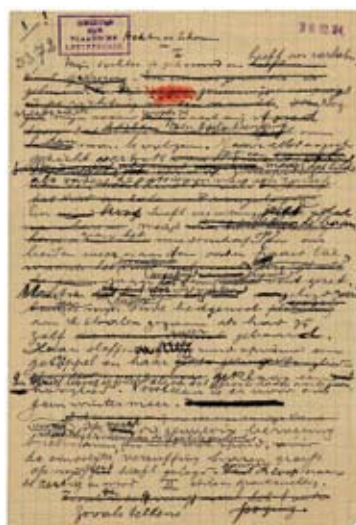
werd onderhouden door het Instituut voor Sociaal-Economisch Onderzoek (ISEO), inmiddels opgegaan in SEOR, komt uit het onderzoek naar de sociale positie en het voorzieningengebruik van allochtonen, het SPVA onderzoek,' aldus De Koning. 'Geanonimiseerde enquêtegegevens worden vaak via DANS beschikbaar gesteld voor verder onderzoek.'

De Koning en SEOR merken dat opdrachtgevers bekijven op budgetten voor onderzoeksopdrachten: 'Dit gaat vaak ten koste van de kwaliteit van het onderzoek en dat is spijtig. Hopelijk komt er in de toekomst weer meer ruimte voor diepgravend, strategisch onderzoek om op basis daarvan beleid te kunnen formuleren en niet slechts te toetsen.'

## Een kijkje in Elsschots keuken

Een wetenschappelijke uitgave van moderne literatuur biedt altijd een blik achter de schermen. De tekstbezorger maakt voor de lezer inzichtelijk hoe de tekst zich heeft ontwikkeld. In een genetische editie worden de verschillende stadia ook daadwerkelijk weergegeven. De lezer kan elke doorhaling en toevoeging zien, krijgt daardoor een beter begrip van de ontwikkeling van de tekst en, in het ideale geval, ook van de betekenis.

Extra spannend is het om dat procédé toegepast te zien op een tekst met de titel 'Achter de Schermen' – een tekst die zelf pretendeert het proces van schrijven inzichtelijk te maken. Willem Elsschot schreef zo'n tekst over de totstandkoming van de 'Opdracht' bij zijn roman *Tsjip*, en onlangs kwam een elektronische editie van die tekst uit, bezorgd door Peter de Bruijn, Vincent Neyt en Dirk van Hulle. Elsschot voelde zich miskend door de ontvangst van



*Tsjip*, en schreef in reactie 'Achter de Schermen', waarin hij laat zien hoe de 'Opdracht' onder zijn handen groeide. 'Achter de Schermen' moest het *l'art pour l'art* van *Tsjip* laten zien, en inzichtelijk maken hoe er aan de tekst gewerkt en geschaafd was.

De elektronische editie toont alle bestaande bronnen van de 'Opdracht' en van 'Achter de schermen'. Van alle bronnen is een afbeelding en

## Een nieuwe start voor het Text Encoding Initiative

Het Text Encoding Initiative (TEI) vierde onlangs zijn twintigste verjaardag met het uitbrengen van een geheel vernieuwde versie van de Guidelines, de richtlijnen voor tekstcodering die zijn voornaamste product vormen. Het TEI werd opener, constateert Peter Boot.

De eerste initiatieven voor een gezamenlijke standaard voor tekstcodering dateren van 1987. Om teksten door de computer te kunnen laten verwerken hadden taal- en letterkundigen de behoefte bepaalde aspecten van teksten gestructureerd te kunnen vastleggen. Het Text Encoding Initiative is opgericht om in die behoefte te voldoen. Met enig succes: de richtlijnen van het TEI zijn in gebruik bij veel digitale bibliotheken (als de DNB), in veel editie-projecten (zoals in de editie van Elsschot, zie elders op deze pagina's) en ook voor taalkundige corpora wordt TEI vaak gebruikt.

Zo langzamerhand had het TEI wel een beetje last van de wet van de remmende voorsprong: in 1987 bestond bijvoorbeeld het web nog niet. De codeertaal XML, die sinds 1999 niet meer weg te denken is uit de wereld van de IT, moest nog worden uitgevonden, net als Unicode, nu de algemeen geaccepteerde standaard voor het vastleggen van lettertekens in niet-romeins schrift.

Er bestond dus behoefte aan een nieuwe versie van de TEI Guidelines, waarbij de richtlijnen zouden worden aangepast aan de huidige stand van de techniek. Bovendien was het tijd de richtlijnen behoorlijk uit te breiden op die terreinen waar ze onvoldoende houvast voor de tekstcodering boden.

Die nieuwe versie, met als codenaam P5, is onlangs uitgekomen en op beide gebieden is belangrijk werk gedaan. Wat betreft de modernisering: er wordt gebruik gemaakt van veel faciliteiten van XML (voor de liefhebbers: bijvoorbeeld verschillende schemata en namespaces), en voor (intra- en inter-)documentverwijzingen, datums en talen worden internationale standaards gebruikt. Wat betreft de uitbreiding: er zijn onder andere onderdelen be-

schikbaar gekomen voor de codering van manuscriptbeschrijvingen, voor facsimiles en voor prosopografie (persoonsbeschrijvingen). Tegelijkertijd is belangrijk werk gedaan om het TEI opener te maken, in verschillende opzichten. De infrastructuur is zo aangepast dat de taal van de documentatie gewijzigd kan worden. De documentatie is inmiddels gedeeltelijk beschikbaar in verschillende Europese talen, maar ook in het Japans en Chinees. Een andere vorm van openheid, die ook mogelijk is geworden, is de integratie van andere codeertalen, zoals docbook (voor technische documentatie) of SVG (voor bepaalde grafieken).

De belangrijkste vorm van openheid is echter wel dat er een infrastructuur is gecreëerd waarmee gebruikers een eigen TEI deelverzameling kunnen definiëren, inclusief eigen documentatie, hernoemde codes, en waar nodig uitbreidingen. Met een online toepassing, *Roma*, kunnen zulke TEI extensies worden aangemaakt.

Ook institutioneel is er het nodige gewijzigd. Sinds enige jaren is het TEI een consortium met als leden de faculteiten, wetenschappelijke instituten en digitale bibliotheken die de richtlijnen van het TEI gebruiken. De gezamenlijke TEI-gebruikers kiezen het bestuur en de Technical Council, en bepalen zo de verdere ontwikkeling van de standaarden. Alle TEI software is onder een open source licentie beschikbaar.

Niet dat nu alle wensen zijn gerealiseerd. De TEI richtlijnen zijn, voor wie er mee aan de slag wil, nog steeds niet het meest toegankelijke leesvoer. Meer toegankelijk trainingsmateriaal en handleidingen zijn dringend gewenst. Dat is dan ook wat het TEI de komende periode wil gaan realiseren. (Peter Boot)

de correcties in het manuscript van 'Achter de Schermen' aannemelijk dat Elsschots voorstelling van zaken niet zonder meer betrouwbaar is. Het is een stukje theater dat Elsschot opvoert voor zijn lezers, zo suggereren zij. Dankzij deze editie hebben we nu alle bronnen bij de hand om te bestuderen hoe dat theaterstuk tot stand kwam.

'Achter de Schermen' is een uitgave van het Center for Manuscript Genetics (Universiteit van Antwerpen), het Huygens Instituut (KNAW) en het Centrum voor Teksteditie en Bronnenstudie (KANTL). (Peter Boot)

een tekst beschikbaar. Met behulp van TEI-codering (zie elders op deze pagina's) is het mogelijk gemaakt om de ontwikkeling van elke zin te volgen. Van het meest complexe manuscript is zelfs een 'topografische transcriptie' beschikbaar, een tekst die de plaatsing van de woorden op het papier volgt. Helaas bestaat er van de 'Opdracht' zelf geen manuscript meer.

We kunnen dus nu precies zien hoe Elsschot in 'Achter de schermen' het wilde voorstellen dat de 'Opdracht' gegroeid was – het echte bewijsmateriaal is voor altijd zoek. Toch maken de bezorgers aan de hand van

 [www.tei-c.org](http://www.tei-c.org)



## Data panelstudie Jeugd & Cultuur opvraagbaar

Onlangs zijn de data van het project Jeugd & Cultuur ondergebracht bij DANS. Een groot deel van deze gegevens is daarmee vrij beschikbaar gekomen voor geïnteresseerden.

Het project 'Jeugd & Cultuur' is een multi-actor panelstudie onder leiding van het voorgezet onderwijs. Onderzocht zijn onder meer het cultureel uitgaansgedrag (inclusief titels van bezochte voorstellingen en films, gelezen boeken, bezochte musea en tentoonstellingen), de houding tegenover kunst en cultuur en de kunstzinnige vorming in het onderwijs. Daarnaast hebben leerlingen gerapporteerd over sociale netwerken tussen klasgenoten en over hun vakkenpakket en schoolcijfers. Docenten, leerlingbegeleiders en random gekozen ouders hebben gegevens aangedragen over sociaal-culturele achtergronden en gedragingen.



NATIONALE BEELDBANK

Binnen het project zijn in 1998, 2000, 2001 en 2005 telkens nieuwe cohorten leerlingen uit de klassen drie, vier en vijf van het voortgezet onderwijs voor een eerste meting benaderd. Deze leerlingen kwamen van zeventig schoollocaties in veertien gemeenten in Nederland. Alle deelnemers waren tussen 14 en 17 jaar oud bij eerste ondervraging. Een deel van deze ondervraagde leerlingen is daarna via postale (soms telefonische) dataverzameling herondervraagd. De in 1998 voor het eerst ondervraagden zijn inmiddels vier keer herondervraagd, terwijl de in 2000 en 2001 ingevoegde cohorten tweemaal zijn herondervraagd.

Een deel van de gegevens is afkomstig van oud-leerlingen die tussen 1975 en 2000 examens hebben gedaan op de betrokken scholen en van hun leraren en ouders.

De dataverzameling als geheel is een unieke en omvangrijke bron van informatie over verschillende vormen van cultureel gedrag van jongeren in Nederland en daarvoor relevante sociale en onderwijskundige achtergronden. Te downloaden zijn gegevens uit de klassengewijze ondervraging in 1998, 2000 en 2001 (7500 leerlingen) en de herondervraging van een deel van deze leerlingen (2000) en hun ouders in 2000 en 2002. Ook de data over zevenhonderd oud-leerlingen (leeftijd 20-40 jaar) en hun ouders zijn beschikbaar. Gegevens over docenten en scholen zullen later gearchieveerd worden. Datzelfde geldt voor de gegevens van het panel uit 2004 en die van de recent toegevoegde cohort leerlingen uit 2005. (Harry Ganzeboom en Ineke Nagel)

Een deel van de gegevens is afkomstig van oud-leerlingen die tussen 1975 en 2000 examens hebben gedaan op de betrokken scholen en van hun leraren en ouders.

## Kickoff voor Europese infrastructuur

Op 14 en 15 januari kwamen zo'n vijftig vertegenwoordigers van data-archieven in Europa bijeen in het Engelse Glasgow om de start mee te maken van een zogenaamd *preparatory* project voor een nieuwe Europese data-infrastructuur. Het gaat om een *upgrade* van de data-infrastructuur van de Council of European Social Science Data Archives (Cessda). Die Council organiseert al dertig jaar een aantal data-archieven in de sociale wetenschappen in Europa. Op het ogenblik zijn meer dan twintig archieven lid. De data-infrastructuur die Cessda haar leden te bieden heeft is aan vernieuwing toe 'om te garanderen dat onderzoekers in de sociale wetenschappen kunnen beschikken over de hoogkwalitatieve databronnen die ze nodig hebben', aldus de website van de organisatie, 'onafhankelijk van de plaats waar onderzoekers dan wel data zich bevinden'.

## Workshop Social Software: wat levert het in de praktijk op?

De Vereniging voor Geschiedenis en Informatica (VGI) organiseert eind maart of begin april een workshop over Social Software. Tijdens deze studiedag worden mogelijkheden van Social Software getoond en worden vragen beantwoord als: wat is het, wat kun je er mee en wat hebben historici er aan? Is het alleen maar leuk voor privégebruik of hebben professionals er ook baat bij?

Afhankelijk van de wensen en het kennisniveau van de deelnemers wordt aandacht besteed aan sociaal bookmarken (del.icio.us), het uitlezen en bijhouden van RSS-feeds (via Google Reader of Bloglines), het maken van een sociaal netwerk (Ning) of het onderhouden van een wiki (via wetpaint). Eventueel kan er ook een blog (via Blogger), een fotopagina (via Flickr) of een Twitter-pagina

(via Twitter) gemaakt worden. Na de workshop kunnen de deelnemers zelf verder werken met de tools en software waarmee ze kennis hebben gemaakt.

De bijeenkomst is gratis voor VGI-leden; niet-leden betalen € 25. Het aantal plaatsen is beperkt. De precieze datum en de plaats van handeling waren bij het ter perse gaan van deze *e-data@research* nog niet bekend. Zodra dat wel het geval is worden ze vermeld op [www.vgi-online.nl](http://www.vgi-online.nl) waar ook meer informatie over de studiedag is te vinden. Aanmelden kan via [berk@kitlv.nl](mailto:berk@kitlv.nl).

In juni wil de VGI een studiedag wijden aan massadigitalisering. Voor september staat de uitreiking van de VGI-Innovatieprijs 2008 op het programma. (Michelle van den Berk)

## Kosten digitaal archief verkend

Digitaal archiveren wordt steeds belangrijker en gebeurt op steeds grotere schaal. Over de kosten die daarvoor gemaakt moeten worden is nog veel onduidelijk. Bij DANS werd een verkennend onderzoek uitgevoerd.

Voortbouwend op internationale projecten die deze kosten in kaart brengen, heeft DANS een pilotstudie uitgevoerd om te kunnen bepalen wat haar archiefdiensten kosten. In de studie is een kostenmodel gehanteerd, dat is gebaseerd op de functies die te onderscheiden zijn in het ISO-gecertificeerde Open Archival Information System (OAIS): inname, gegevensbeheer, archiefopslag, toegang, beheer en planning van de digitale bewaring. Die functies zijn gerelateerd aan de specifieke activiteiten van de belangrijkste *cost driver*: de tijdsbesteding van het personeel. Daardoor biedt het model de mogelijkheid om de verschillende activiteiten een weging mee te geven in de tijdsbesteding van de archivaris. Dit is relatief gemakkelijk kwantificeerbaar.

Een opwaardering van de Cessda infrastructuur was door het European Strategy Forum on Research Infrastructures (ESFRI) al opgenomen in een infrastructurele *Roadmap*, wat geldt als een aanbeveling aan de EU en de nationale regeringen om er financiële steun aan te geven. Met die steun in de rug heeft Cessda een aanvraag ingediend bij de Europese Commissie, die nu groen licht heeft gegeven voor de voorbereiding van een definitief voorstel. De bijeenkomst in Glasgow was het startsein voor die voorbereidende fase.

Het Nederlandse DANS (Data Archiving and Networked Services) was met drie mensen vertegenwoordigd op de bijeenkomst. Afgesproken werd dat DANS een bijdrage levert aan de werkpakketten over de mogelijkheden van GRID-technologie, de toegang tot en verspreiding van data-informatie en de harmonisatie en conversie van data. (MdG)

Het kostenmodel is getoetst door middel van een beperkte steekproef binnen de DANS-organisatie. De resultaten hiervan vormen een eerste indicatie van de benodigde tijd en de tijdsbesteding per dataset, en van de daarmee samenhangende kosten. Voor betrouwbare resultaten is verdere toetsing van het model nodig. Informatie: [Heiko.tjalsma@dans.knaw.nl](mailto:Heiko.tjalsma@dans.knaw.nl). (Kevin Heerema)

## Sinds kort beschikbaar

Het overzicht toont een aantal databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS en DANS. Een volledig overzicht van de CBS-bestanden is te vinden op [www.cbs.nl/microdata](http://www.cbs.nl/microdata).

De bij DANS beschikbare databestanden komen van diverse andere onderzoeksinstellingen. Deze kunnen kosteloos worden gedownload vanuit DANS EASY: <http://easy.dans.knaw.nl/dms>

Data Centraal Bureau voor de Statistiek	Periode	Via
Woningbehoefteonderzoek (WBO)	2002	CvB
Enquête Beroepsbevolking (EBB)	2000 t/m 2005	CvB
Inkomenspanelonderzoek (IPO)	1989 t/m 2004	CvB
Werkloosheidswet (WW)	eerste helft 2007	CvB
Bijstandsuitkeringenstatistiek (BUS)	eerste helft 2007	CvB
Bijstandsuitkeringen transactie (BUS-TRANS)	2005	CvB
Arbeidsongeschiktheidsuitkeringen (AO)	eerste helft 2007	CvB
Statistiek Re-integratie gemeenten (SRG)	eerste helft 2007	CvB
Nationale Enquête Arbeidsomstandigheden (NEA)	2006	CvB
Community Innovation Survey (CIS)	2004	CvB

Data van overige instellingen	Periode	Via
Sociaal en Cultureel Planbureau Ouderen in instellingen	2004	DANS (EASY)
Universiteit Leiden		
Politiek cynisme jongerenonderzoek NIBUD	2005	DANS (EASY)
Nationaal Scholierenonderzoek	2004 en 2006	DANS (EASY)
Fryske Akademie		
Onderwijs Fries op de basisschool	1988	DANS (EASY)
Taal-enquête Leeuwarder Courant	1991	DANS (EASY)
Taal in het grensgebied	1984	DANS (EASY)

### COLOFON

*e-data@research* is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van DANS, het Huygensinstituut, het Internationaal Instituut voor Sociale Geschiedenis, het Centraal Bureau voor de Statistiek, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen. Oplage: 7500. *e-data@research* is online te raadplegen op [www.edata.nl](http://www.edata.nl)

**Uitgever:** Uitgeverij Aksant, Postbus 2169, 1000 CD Amsterdam

**Redactieadres:** Postbus 93067, 2509 AB Den Haag; t (070)3494450 f (070)3494451 e [edata@dans.knaw.nl](mailto:edata@dans.knaw.nl)

**Redactie:** Peter Boot, Ivo Gorissen, Martijn de Groot (hoofd/eindredacteur), Marcel Ras, Jetske van der Schaaf, Luuk Schreven

**Aan dit nummer werkten mee:** Ronald Beelaard, Michelle van den Berk, Ingrid Dillo, Suzan van Dijk, Harry Ganzeboom, Kevin Heerema,

René van Horik, Jan Jonker, Ineke Nagel, Frank van Oort, Judith Rog, Mark Thissen, Caroline van Wijk, Huib Zuidervart, Joris van Zundert.

**Redactiesecretariaat:** Lucas Pasteuning, Jetske van der Schaaf

**Vormgeving en opmaak:** Ellen Bourma, Alkmaar

**Productie en pre-press:** Uitgeverij Aksant, Amsterdam

**Druk:** PlantijnCasparie, Almere

**ISSN:** 1872-0374



## Een Einzelgänger met gidsfunctie

'Een beetje eigenwijs'. Zo beschrijft projectleider Eddy van der Noord van het digitaal archief van de Leeuwarder Courant zijn eigen houding en die van zijn partners in het project. Toen ze in 2005 begonnen met hun idee om alle exemplaren van het oudste dagblad van Nederland digitaal beschikbaar te maken, hadden ze wel een paar voorbeelden maar die voldeden niet aan de eisen die hen zelf voor ogen stonden. Dus maar een eigen route uitgestippeld. 'We wisten dat de Koninklijke Bibliotheek met een project bezig was en dat Digitaal Erfgoed Nederland krantendigitalisering op de agenda had staan, maar wij wilden niet wachten', zegt Van der Noord die overigens de goede contacten met beide instellingen benadrukt. 'Er zit ook een vertegenwoordiger van onze stichting in de DEN-commissie over krantendigitalisering'.

Het begon met een A4-tje. 'Nee, het was een half A4-tje waarmee ik naar onze directie en hoofdredactie stapte toen ik in het voorjaar van 2005 het idee had gekregen. Die reageerden: 'Ziet er goed uit, Eddy, werk dat maar uit'. Ik was toen productontwikkelaar en het was mijn taak om nieuwe dingen te bedenken, die liefst iets verder gingen dan weer een nieuwe papieren bijlage.' En dat was het begin van het project dat vorig jaar oktober on-line een vliegende start beleefde met vijftig duizend bezoekers en 3,5 miljoen page views in de eerste twee en een halve week van zijn bestaan. Nog steeds is de belangstelling groot met een tamelijk constante score van twee miljoen page views per maand. Een 'eigenwijs' project met een mooi resultaat dat tot



**Van der Noord nog in zijn papieren archief**

op heden nog geen steekhoudende kritiek is tegengekomen.

Na het halve A4-tje volgde een document van enkele velletjes, herinnert Van der Noord zich. 'Ik was in die begintijd Lourens Oldersma tegen het lijf gelopen van het Fries Historisch Centrum Tresoar, die met een soortgelijk idee rondliep. Tresoar was toen al behoorlijk ver met digitalisering. Hij wilde af van die microfile readers in zijn studiezaal, waar de Leeuwarder Courant tot de meest geraadpleegde bronnen hoort. We kwamen snel op één lijn en dat is opmerkelijk in een samenwerking tussen een bedrijf en een stichting zonder winstoogmerk. We stelden een lijstje op met eisen, met referentieprojecten en met leveranciers. En we richtten samen de stichting op waarin het hele project nu nog steeds draait: de stichting Digitaal Archief Leeuwarder Courant.'

Twee belangrijke dingen gingen goed. In de eerste plaats konden financiers worden gevonden voor het project dat uiteindelijk eenmalig rond vier en een halve ton heeft gekost. Tresoar en de Leeuwarder Courant deden mee, het VSB-fonds, het Prins Bernhard cultuur Fonds, de provincie Friesland en het

*Je Maintiendrai* Fonds zagen brood in het archief. In figuurlijke zin dan, want het tweede dat goed ging was dat alle betrokkenen het erover eens waren dat het hier om een project moest gaan dat geen grote winst hoefde op te leveren. Ook de Leeuwarder Courant, waarvan de auteursrechten in het geding waren. 'De krant behoort toe aan de Friese samenleving, heb ik de directie en hoofdredactie voorgehouden, en daar waren ze het na enig nadenken mee eens'. Zo konden de auteursrechten nodig voor de digitalisering en digitale exploitatie in de nieuwe stichting worden ondergebracht en dat scheelde een aanmerkelijke financiële last. Het gebruik van het archief is gratis en zoals het er nu uitziet zal dat zo blijven, aldus Van der Noord.

De stichting ging in zee met het digitaliseringsbedrijf X-Cago in Roermond, dat zich qua mogelijkheden duidelijk onderscheidde. Inmiddels was het najaar 2005 en de rest is geschiedenis. Anders dan toen nog de norm was in de krantendigitalisering werd er *full colour* gescand. De inhoud van de 800 duizend pagina's die sinds 1752 verschenen wordt aangeboden op artikelniveau, met de mogelijkheid om de gehele pagina op te vragen voor wie een beeld van de context wil hebben.

De rol van Einzelgänger heeft nergens spijt opgeleverd, evalueert Van der Noord: 'Nee, we zijn tevreden dat we de vaart erin hebben kunnen houden en natuurlijk hebben we wel goed rekening gehouden met de technische eisen die in het DEN-register voorkomen. Dus we hebben de aansluiting met de andere projecten nog helemaal'. (MdG)

## Gelezen

**Milco Wansleeben en Marjolein van den Dries: *Wegwijzer Digitaal Deponeren Archeologie*; Amsterdam, Aksant, 2008; DANS Data Guide 3; ISBN 978 90 5260 295-0**



De Wegwijzer is speciaal geschreven voor de archeologische bedrijven die erop uit zijn de beschikbare data te gebruiken voor nieuwe gegevensverzamelingen en analyses. De handleiding bevat een praktisch gedeelte voor het deponeren van archeologische datasets in het archiveringssysteem EASY van Data Archiving and Networked Services (DANS). Ook wordt uitgebreid ingegaan op de perspectieven van digitale duurzaamheid, de organisatorische aspecten en toekomstperspectieven van het elektronische depot Nederlandse archeologie, auteursrechten en de geaccepteerde bestandformaten. De Wegwijzer is een flinke stap vooruit

in het proces naar het duurzaam toegankelijk maken van archeologische informatie. Door de snelle toegankelijkheid zullen de mogelijkheden sterk groeien voor wetenschappelijk onderzoekers en degenen die in de commerciële archeologie werken. De Wegwijzer is kosteloos te bestellen bij DANS: [info@dans.knaw.nl](mailto:info@dans.knaw.nl).

**Ray Siemens, Susan Schreibman (eds.): *A Companion to Digital Literary Studies*; Blackwell Publishing, 2007; ISBN 978 1 4051 4864 1**

This publication offers an extensive examination of how new technologies are changing the nature of literary studies. Through a series of specially commissioned articles by leading scholars, theorists, and writers creating born-digital literature, the text provides a thorough overview of the intersections between computing, literary studies, and new media. It takes a highly interdisciplinary perspective in its examination of scholarly editing and literary criticism, interactive fiction and gaming, multimedia and immersive environments, and born-digital literature. This Companion is the only comprehensive collection of seminal works available to meet the needs of this growing area of inquiry. It is a must-read for anyone wishing to understand, use, or create digital literature.

**Maurits van der Graaf, Kwame van Eijndhoven: *The European Repository Landscape, Inventory study into***

***present type and level of OAI compliant Digital Repository activities in the European Union*; Amsterdam, Amsterdam University Press, 2008; ISBN 978 90 5356 410 3**

Steeds meer universiteiten, onderzoeksinstituten en hbo-instellingen plaatsen hun kennis en resultaten in repositories zodat ze via internet vrij toegankelijk zijn. In deze studie wordt het Europese repositorylandschap in kaart gebracht. Het onderzoek laat zien dat in een zevental Europese landen al meer dan de helft van de universiteiten een repository heeft. Het DRIVER-project legt de basis voor een Europese repository-infrastructuur door een testbed te realiseren met ruim vijftig interoperabele repositories. Hiermee wordt het voor onderzoekers makkelijker om vrije toegang te krijgen tot wetenschappelijke resultaten. Het project wordt deels gefinancierd door de Europese Commissie.

**Jeurgens, K.J.P.F.M., A.C.V.M. Bongenaar, M.C. Windhorst (red.): *Gewaardeerd verleden, bouwstenen voor een nieuwe waarderingsmethodiek voor archieven, Rapport van de Commissie Waardering en Selectie*; Den Haag, Nationaal Archief, 2007**

In dit rapport biedt de Commissie Waardering en Selectie een nieuwe integrale doelstelling voor archiefselectie bij de overheid, met daarop afgestemde selectiecriteria plus bijbehorende inhoudelijke methodiek. Uitgangspunt is dat selectie en acqui-

## Column

Ronald Beelaard

# Wikipedia en wetenschap

*U bent vast wel eens op de site geweest van de on-line encyclopedie waar iedereen aan kan bijdragen. Zoek iets in Google en in negen van de tien gevallen staat er wel een Wikipedia-artikel op de eerste resultaatpagina. Mogelijk een artikel van de Nederlandstalige versie met haar 400.000 artikelen, zoniet dan wel van de Engelstalige versie met meer dan 2 miljoen artikelen.*

*Er iets aan bijdragen is zeer laagdrempelig. Dat is de kracht van Wikipedia en bepalend voor het succes, maar het is tevens de zwakte. Immers, de zojuist door u doorgevoerde aanpassing kan door anderen weer gewijzigd worden of in het ergste geval weer ongedaan worden gemaakt. Maar het kan evengoed de basis zijn voor verdere verbetering, want de volgende bezoeker van de pagina wordt misschien geïnspireerd door uw tekst en borduurt daarop dan weer verder. En als dat proces van voortdurend sleutelen en schaven maar lang genoeg doorgaat, wordt het artikel steeds beter. Het is dan niet meer de pen-nenvrucht van een enkele auteur, maar het gezamenlijk product van velen. En dat maakt Wikipedia uniek!*

*Die laagdrempeligheid is zeker van toepassing op jongeren. Opgegroeid met internet, vinden zij die bewerkknop heel snel en schromen ook niet om hem te gebruiken. Soms om iets te verbeteren of toe te voegen, soms om iets te vernielen. Herstellen is even eenvoudig als vernielen en in 99 van de 100 gevallen gebeurt dat ook binnen enkele minuten.*

*De laagdrempeligheid heeft er ook voor gezorgd dat Wikipedia informatie over de meest uiteenlopende onderwerpen bevat. Van de klassieke encyclopedische thema's tot artikelen over wat men in sommige hedendaagse subculturen belangrijk vindt. Sommigen ergeren zich daar wel eens aan: een serieus artikel over bijvoorbeeld de Tweede Wereldoorlog en een artikel over Idols in hetzelfde naslagwerk. Maar voor een on-line encyclopedie maakt dat niet uit. Opslagcapaciteit is schier oneindig en wordt ook steeds goedkoper.*

*Er zijn nog vele thema's uit de wetenschap onderbelicht of geheel ontbrekend. Voor een expert in het vakgebied waarschijnlijk een fluitje van een cent om daar wat aan te doen, maar er is kennelijk toch een drempel.*

*Waarom? Omdat u het beeld heeft dat Wikipedia een te 'volks' karakter heeft?*

*Onzin, gewoon doen is mijn advies. Grote kans dat u door de gemeenschap gewaardeerd zal gaan worden als expert op uw vakgebied. Wikipedia is nooit af en heeft uw inbreng nodig. De volgende keer dat u Wikipedia bezoekt, bent u dus uitgenodigd.*

Ronald Beelaard is gepensioneerd manager en zet zich onder meer als woordvoerder in voor Wikipedia.

sitie vanuit de optiek van het cultureel erfgoed twee kanten van één medaille vormen: de opbouw van een representatieve collectie staat centraal. Overheidsorganisaties kampen bij hun informatiehuishouding niet alleen met papieren achterstanden maar ook met de vluchtigheid van digitale informatie. De Nederlandse regering streeft daarom naar een andere aanpak van archiefselectie. Selectie is een van de delicate taken voor archivariissen en informatiebeheerders. Vernietiging is immers onomkeerbaar. Maar de keus voor blijvende bewaring brengt structurele inspanningen met zich mee voor toegankelijkheid en duurzaamheid, zeker in het digitale tijdperk.

**Matthijs Kalmijn, Marion Wittenberg (red.): *Multi-actor data in survey onderzoek, Bijdragen aan het DANS symposium van 17 januari 2007*; Amsterdam, AKSANT, 2008; DANS Symposium publicaties 3, ISBN 978-90-5260-292-9, € 19,90**

In de maatschappijwetenschappen is het verzamelen en analyseren van multi-actor data aan het opkomen. Ook multilevel modellen in de statistiek worden populairder. Tegelijk komen de prospectieve dynamische data (panels) op, waarbij ook steeds vaker multi-level modellen worden toegepast. Een belangrijke innovatie in het sociologisch onderzoek zou kunnen liggen in het combineren van deze ontwikkelingen in het survey-onderzoek. Daarmee wordt de deur geopend om vragen te beantwoorden over de manier waarop individuen beïnvloed worden door de resources, de normen en waarden, en de gedragingen van de personen in hun directe omgeving: klassieke sociologische vragen die tot nu toe onderbelicht zijn gebleven.

Begin 2007 organiseerden DANS en de Universiteit van Tilburg het symposium 'Multi-actor data in survey onderzoek: moeilijkheden en mogelijkheden' over dit onderwerp. In deze publicatie zijn de bijdragen aan dit symposium gebundeld.