

Genegeerd materiaal brengt licht in middeleeuwse handschriften

# Paleografen zetten hun horloges weer gelijk

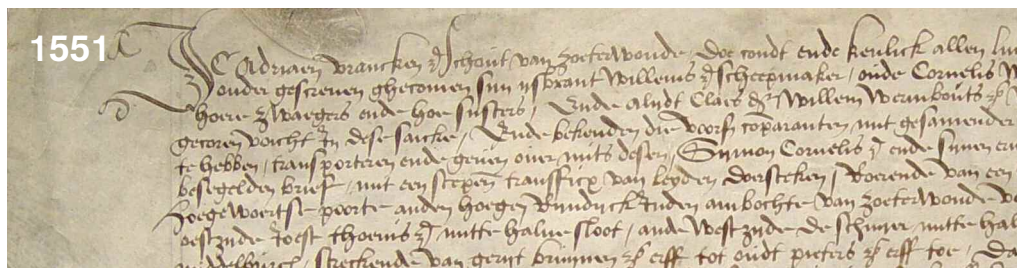
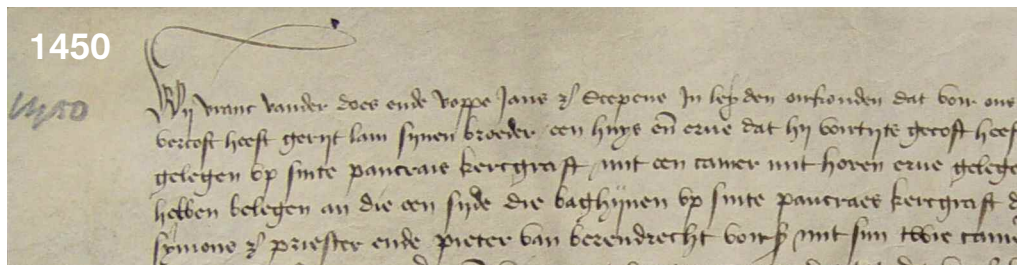
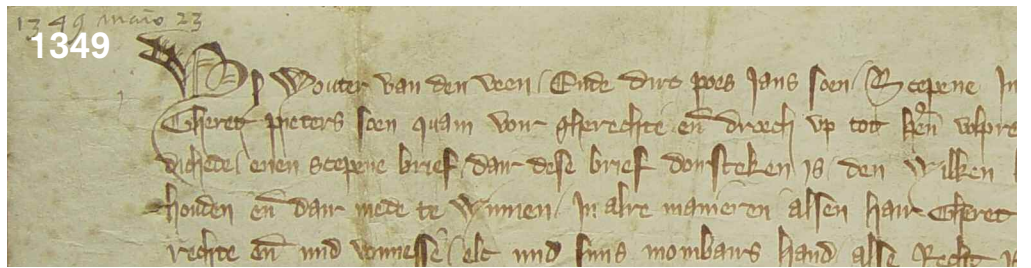
Waar en wanneer werd dat middeleeuwse handschrift geschreven? Een vraag voor de computer. *Jan Burgers en Lambert Schomaker*

*Lambert Schomaker*

Onderzoekers die zich voor de middeleeuwen interesseren, hebben vaak een probleem. De primaire bronnen die ze bestuderen, handschriften, dragen geen aanduiding van plaats en tijd. Daarmee wordt hun onderzoek bemoeilijkt, want de vraag naar het ontstaan is meestal de eerste die aan een bron wordt gesteld. Gewoonlijk wordt die vraag beantwoord op basis van schriftkenmerken, en daartoe is slechts een handvol gespecialiseerde paleografen in staat. Maar ook die komen lang niet altijd tot eensluidende conclusies, omdat hun expertise vaak meer is gebaseerd op intuïtie dan op objectieve criteria, juist vanwege het gebrek aan voldoende gedateerde manuscripten.

## Paleografische meetlat

Door NWO is onlangs subsidie toegekend aan een project om te komen tot een meer objectief, geautomatiseerd procedé van datering en lokalisering van laat-middeleeuws schrift. In het kader van dit project, getiteld *MPS, the Medieval Palaeographic Scale for dating and localizing handwritten manuscripts using digital palaeography*, aangevraagd door de Universiteit van Amsterdam, de Rijksuniversiteit Groningen en het Huygens Instituut voor Nederlandse geschiedenis, zullen een historicus en een informaticus een paleografische 'meetlat' construeren met daarin de dateer- en lokaliseerbare elementen in laatmiddeleeuws schrift (1300-1550). Dit zal gebeuren op basis van materiaal dat door paleografen tot nu toe vrijwel volledig is genegeerd, te weten oorkonden en andere archivalische documenten, die meestal precies gedateerd en goed te lokaliseren zijn. Bovendien zijn zij vaak ge-



Typische handschriften uit 1349, 1450 en 1551 bron: Leids Archief, Stadsarchief I

schreven door dezelfde (professionele) scribenten die ook de ongedateerde handschriften vervaardigden, veelal in vergelijkbare schrifttypen.

## Doorbraak voor mediëvistiek

Aan de hand van dit corpus, samengesteld in enkele stadsarchieven, zullen de onderzoekers de meetlat vervaardigen. Daarbij maken zij gebruik van de traditionele paleografische expertise, en werken daarnaast aan software die het schrift op een dieper niveau analyseert. Dit nieuwe terrein van de 'digitale paleografie' heeft al veelbelovende resultaten opgeleverd: door de Groningse groep van Schomaker is bijvoorbeeld programmatuur ontwikkeld voor de automatische schrijveridentificatie in veertiende-eeuwse oorkonden. Het dateren is een vergelijkbaar probleem, gericht

op de ontdekking van algemene, in de tijd ontwikkelende stijlgroepen. Wanneer de 'meetlat' werkt, zou dat een doorbraak betekenen voor het brede veld van de internationale mediëvistiek. Ook niet-specialisten zullen in staat zijn tijd en plaats van vervaardiging van hun bronnen vast te stellen, in ieder geval bij benadering, en de inzet van de computer zal dit tijdrovende aspect van het onderzoek versnellen. Nu bovendien steeds meer afbeeldingen van middeleeuwse handschriften op internet verschijnen, zal het mogelijk zijn de *palaeographic scale* breed in te zetten en al doende te verfijnen. En deze belangrijke stap in de digitale paleografie brengt ons weer dichterbij de heilige graal op dat terrein, namelijk programma's die de computer in staat stellen handschrift te lezen.

## Een mooi tweede leven voor oude data

Afgestoft en opgepoetst: van bureaula naar Datacuratie-service. *Erica Renckens*

Elke wetenschapper heeft wel eens een harde schijf (of misschien zelfs diskette) met oude data. Zelf verzamelde tekstbestanden die ooit, jaren terug, essentieel waren voor eigen onderzoek. Maar de focus van dat onderzoek is in de loop der tijd verschoven en de data komen niet meer van pas. Dat betekent niet dat de data waardeloos zijn. Andere wetenschappers kunnen de data misschien nog goed gebruiken. Maar dan moeten ze wel weten dat de data bestaan en moeten ze er toegang toe krijgen. Dat is het doel van de Datacuratieservice (DCS), gevestigd bij de Radboud Universiteit in Nijmegen, sinds afgelopen januari

officieel van start. "Het bestuur van CLARIN-NL moest regelmatig aanvragen voor standaardisatie afwijzen, terwijl de data in kwestie wel een aanvulling op de CLARIN-infrastructuur zouden zijn", vertelt Nelleke Oostdijk, projectleider van de Datacuratieservice. "Dankzij de DCS kunnen deze data nu toch worden afgestoft en opgepoetst voor een tweede leven."

## Eén standaard

Momenteel neemt Oostdijk, samen met coördinator Henk van den Heuvel, de Dutch Bilingual Database onder handen. "Het grootste deel van het werk bestaat uit de inventarisatie voorafgaand aan de conversie", legt Oostdijk uit. Niet alle data liggen op dezelfde locatie: een deel ligt bij het Max Planck Instituut,

een deel bij het Meertens Instituut. "Vervolgens converteren we de data naar één CLARIN-standaard", aldus Oostdijk. "Want dat werkt het fijnste voor toekomstige gebruikers."

## Licentieconstructie

Mogen alle data zomaar worden hergebruikt? "Nee, en dat is zeker een probleem. Tegenwoordig is het bij de meeste onderzoeksaanvragen een vereiste dat alle IPR (*Intellectual Property Rights*, red.) geregeld zijn, maar vroeger werd daar niet naar gekeken. De wetenschapper verzamelde de data voor eigen gebruik; het regelen van de rechten, zodat andere wetenschappers er ook mee zouden kunnen werken, kostte alleen maar moeite en leverde de onderzoeker zelf geen voordeel op. We werken aan een oplossing. Je

kunt bijvoorbeeld denken aan een licentieconstructie, waarbij de gebruiker alleen toegang tot de data krijgt als hij een licentie ondertekent", aldus Oostdijk.

De projectleider heeft goede hoop voor de toekomst van DCS. Onderzoekers met oude data kunnen DCS benaderen, maar DCS gaat ook zelf op zoek naar interessante data. Het bestuur van CLARIN-NL beslist vervolgens welke datasets onder handen worden genomen. "Ik ga er vanuit dat, als we hebben laten zien verantwoord om te gaan met IPR-kwesties, de meerwaarde van de Datacuratieservice binnen de geesteswetenschappen breed gedragen zal worden."

[www.ru.nl/letteren/datacuratie-service/](http://www.ru.nl/letteren/datacuratie-service/)

## E-DATA & RESEARCH

Jaargang 6 | nummer 4

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

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: DANS, CentERdata, CBS, CLARIN-NL, Huygens ING, Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica.

## INHOUD

3 Waar sta jij in het netwerk van onderzoekers?

3 Multicultureel panel kleurt kennis over onze cultuur

4 Laten zien hoe goed je database of dataset al is

5 Schuyt over integriteit onderzoeksdata



6 Schatten van de Bodleian Library onder de loep

6 E-science binnenkort in hogere versnelling

7 16,5 miljoen Hollanders in acht variabelen

8 Biografisch portaal wordt semantisch netwerk

8 Ewoud Sanders kan sneller scannen



Scan deze QR code met een smartphone om de website van E-data te bezoeken.  
[www.edata.nl](http://www.edata.nl)



## Variatie in derde ronde CLARIN-NL

Erica Renckens

Dinsdag 6 maart presenteerde CLARIN-NL tijdens haar kick-off bijeenkomst in het Instituut voor Beeld en Geluid de projecten uit de derde ronde. De middag begon met positief nieuws: een week eerder had CLARIN officieel de ERIC-status gekregen, een Europese rechtspersoon voor onderzoekinfrastructuur. Bovendien heeft CLARIAH, een samenwerkingsproject met DARIAH, een plekje gekregen op de Nationale Roadmap Grootschalige Onderzoeksfaciliteiten. Het CLARIAH-voorstel is niet direct gehonoreerd, maar de aanvragers hebben wel één miljoen euro gekregen om hun plannen te verbeteren. Ten slotte maakte Hans Westerhof, adjunct-directeur van Beeld en Geluid, bekend dat zijn instituut zijn audiovisuele bronnen beschikbaar stelt aan de CLARIN-infrastructuur, zodat geesteswetenschappelijk onderzoek met deze data mogelijk wordt.

Vervolgens kregen alle dertien projecten uit de derde ronde kort de tijd om hun plannen uit de doeken te doen. Voor deze ronde had het bestuur van CLARIN-NL voor de eerste keer expliciet uitnodigingen verstuurd naar onderzoekers uit vakgebieden die tot nu toe onderbelicht waren gebleven. Het resultaat van deze aanpak was goed zichtbaar in de onderwerpen van de projecten, die sterker gevarieerd en toegespast waren dan in eerdere rondes. Zo zal bijvoorbeeld in het PILNAR-project een corpus worden samengesteld met moderne pelgrimverlagen van reizen naar Santiago de Compostela. De onderzoekers van het Migmap-project gaan werken aan een module waarmee het mo-

gelijk wordt om de migratie van de bevolking interactief in kaart te brengen. Letterkundigen gaan in het Namespace-project tools ontwikkelen waarmee de rol van eigennamen in moderne romans onderzocht kan worden. Alle projecten hebben een looptijd van maximaal één jaar.



Hans Bennis, directeur van het Meertens Instituut, opent de bijeenkomst foto Erwin Verbruggen

## Workshop over ict en bijbelwetenschap

Dirk Roorda

Hebreeuwse en Griekse bijbelmanuscripten hebben een eeuwenlange geschiedenis van overdracht en vertaling. Maar waaraan moeten tekstdatabases voldoen om oude teksten systematisch te kunnen bestuderen? De *Lorentz Workshop Biblical Scholarship and Humanities Computing* van 6 tot en met 10 februari bracht bijbelwetenschappers en ict-ers bij elkaar, met name ook om te zien hoe digitale methoden nog meer kunnen betekenen voor dit complexe veld.

De complexiteit komt onder andere door de verschillen in handschrift. De Hebreeuwse tekst kent minder handschriften. Hierdoor is het relatief eenvoudig om tot een standaard teksteditie voor analytisch onderzoek te komen. Aan Griekse zijde

zijn echter meer dan 5.000 handschriften, die een complex verhaal vertellen over de oorsprong van het Nieuwe Testament. Hier is de wens om analyse al op individuele manuscripten te doen en dan de resultaten bij elkaar te brengen en te visualiseren. Het is immers zonde om pas analyse te kunnen doen op een gestandaardiseerde tekst. Naast het constateren van de verschillen openbaarde de workshop ook overeenkomstige problematieken. Zo is de digitale toegang tot bronmateriaal niet voldoende om nieuwe onderzoeksvragen aan te pakken. De bronnen zijn vaak verpakt in bijbelsoftware waarvan de ontwikkeling niet gedreven wordt door de wensen van de collegezaal, laat staan het onderzoek. Een andere constatering is dat onderzoekers zelf *open source tools* willen ontwikkelen waarmee bronnen geanalyseerd en in beeld gebracht kunnen worden. De resultaten hiervan moeten direct beschikbaar zijn om nieuw werk te inspireren, een voorwaarde voor netwerkeffecten. De conclusie van



Poster Lorentz workshop

de workshop is dat werd gesproken over digitale wetenschapsbeoefening, een stap verder dan gedigitaliseerd werken. Het is nu de vraag of uitgeverij van edities deze vrije toegang tot bronnen voldoende kunnen faciliteren. Een aanzet om data toegankelijk te krijgen, is de deponeering van de WIVU dataset in het online archiveringssysteem EASY van DANS door prof. Eep Talstra. Deze tekst database van het Oude Testament kan voor niet-commercieel gebruik gedownload worden. Een demo applicatie bij DANS geeft een indruk van de inhoud ervan.

[www.lorentzcenter.nl/lc/web/2012/480/info.php3?wsid=480&venue=Oort](http://www.lorentzcenter.nl/lc/web/2012/480/info.php3?wsid=480&venue=Oort)  
<https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:48490>  
<http://demo.datanetworkservice.nl/qaq>

## Birnbaum's buttons: 'Just shut up and code'

Peter Boot

Aldus de provocerende tekst op de buttons die slavist en programmeur David Birnbaum uitdeelde op het slotsymposium van de Europese COST actie Interedition. Het symposium werd gehouden op 19 en 20 maart en georganiseerd door Huygens ING. Centraal in Interedition stond de ontwikkeling van gedeelde hulpmiddelen voor de bouw van digitale edities van historische en letterkundige bronnen.

Een belangrijk onderwerp vormde het spanningsveld tussen enerzijds kleine groepjes programmeur-onderzoekers die in korte 'sprints' werken aan concrete hulpmiddelen, en anderzijds grote infrastructuurprojecten die vooral een voorwaar-

denscheppende rol hebben, zoals DARIAH en Project Bamboo. 'Niet praten maar maken', was de geest van de bijeenkomst, en ook de achterliggende gedachte van Birnbaums buttons.

Veel van het werk in Interedition had betrekking op het thema collatie: het analyseren van de verschillen en overeenkomsten tussen versies van teksten. Eén van de gepresenteerde hulpmiddelen stelde de onderzoeker in staat om, na een eerste automatische analyse van de verschillen, de verschillen naar verschillende inzichten te classificeren en eventueel te negeren. Een aantal andere projecten werkten aan heldere presentatie van tekstverschillen op het web.

Een centraal onderwerp in veel discussies over teksteditie is de potentiële rol van vrijwilligers in het editiewerk. Eén presentatie besprak het gebruik van anonieme hulpkrachten die werden ingeschakeld via Amazons 'Mechanical Turk'-programma. (Nog) niet te automatiseren taken ('Human Intelligence Tasks') worden tegen microbetalingen uitgevoerd door mensen die zich daarvoor bij Amazon aanmelden. Men had met enig succes korte verhalen laten coderen op de aanwezigheid van bepaalde verhalende elementen, zoals plot en personages. Toch liepen de kosten wel op. Joris van Zundert, projectleider van Interedition, benadrukte dat de periode van *funding* wel is afgesloten, maar dat wordt gezocht naar manieren om het samenwerkingsverband te kunnen voortzetten. Een doel is daarbij om ook mensen met minder ontwikkelde programmeervaardigheden bij Interedition te betrekken.

[www.interedition.eu/?page\\_id=212](http://www.interedition.eu/?page_id=212)  
[www.mturk.com/mturk/welcome](http://www.mturk.com/mturk/welcome)

## AGENDA

### 5 juni • Amsterdam

#### Symposium HISGIS Amsterdam (1832-1852), Eerste resultaten en perspectieven

Dit symposium behandelt de voortgang van de GIS-matige digitalisering van het oudste kadaster van (groot) Amsterdam, de koppeling ervan aan vroege huisnummerregistraties én de ruimtelijke ontsluiting van de bevolkingsgegevens van ca. 1850. Daarbij zullen de onderzoeks- en verrijkingperspectieven voor Amsterdam zelf aan bod komen en wordt nader ingegaan op de uitbreiding ervan tot andere steden en gemeenten in de provincies Noord-Holland en Zuid-Holland.

[www.hisgis.nl](http://www.hisgis.nl)

### 6 juni • Amsterdam

#### Lezing 'Mining, Mapping, and Accelerating Science and Technology'

Tijdens deze lezing zal Katy Börner, huidig visiting professor bij de KNAW, meer vertellen over nieuwe benaderingen en handvatten die toegang tot informatie, netwerken van onderzoekers en onderzoeksmangement helpen

verbeteren. Ook een overzicht van dataservices en handvatten op het gebied van het visualiseren van onderzoeksdata worden besproken.

[www.dans.knaw.nl/content/symposia](http://www.dans.knaw.nl/content/symposia)

### 12 - 15 juni • Kopenhagen (Denemarken) IS&T Archiving Conference

The Imaging Science & Technology Archiving Conference offers an opportunity for imaging scientists and those working in the cultural heritage community (curators, archivists, librarians, etc.) as well as in government, industry, and academia, to come together to discuss the most pressing issues related to the digital preservation and stewardship of hardcopy, audio and video.

[www.imaging.org/ist/conferences/archiving/index.cfm](http://www.imaging.org/ist/conferences/archiving/index.cfm)

### 27 - 30 juni • Tartu (Estland)

#### 41st LIBER Annual Conference

LIBER is the main research libraries network in Europe. The theme of its annual conference is: "Mobilising the knowledge economy for Europe". On June 27 the APARSEN project will host a workshop about "European e-infrastruc-

tures: optimizing the business relationship between libraries, archives and data centres".

[www.utlib.ee/liber2012/index.php](http://www.utlib.ee/liber2012/index.php)

### 2 - 6 juli • Oxford (Groot-Brittannië) Digital Humanities@Oxford Summer School (DHOXSS) 2012

DHOXSS delegates will be introduced to a range of topics suitable for researchers, project managers, research assistants, and students who are interested in the creation, management or publication of digital data in the humanities.

<http://digital.humanities.ox.ac.uk/dhoxss/>

### 9 - 13 juli • Edinburgh (Schotland)

#### 7th intern. conference on open repositories

The theme and title of this 2012 conference – Open Services for Open Content: Local In for Global Out – reflects the current move towards open content, 'augmented content', distributed systems and data delivery infrastructures.

<http://or2012.ed.ac.uk/>

### 16 - 22 juli • Hamburg (Duitsland)

#### Digital Humanities Conference

Digital Humanities is the annual international

conference of the Alliance of Digital Humanities Organizations (ADHO). ADHO is an umbrella organization whose goals are to promote and support digital research and teaching across arts and humanities disciplines, drawing together humanists engaged in digital and computer-assisted research, teaching, creation, dissemination, and beyond, in all areas reflected by its diverse membership.

[www.dh2012.uni-hamburg.de](http://www.dh2012.uni-hamburg.de)

### 6 - 8 september • Sheffield (Groot-Brittannië) Digital Humanities Congress

The purpose of the Digital Humanities Congress is to promote the sharing of knowledge, ideas and techniques within the digital humanities. Digital humanities is understood to mean the use of technology within arts, heritage and humanities research as both a method of inquiry and a means of dissemination.

[www.shef.ac.uk/hri/dhc2012](http://www.shef.ac.uk/hri/dhc2012)

### 1 - 5 oktober • Toronto (Canada)

#### iPRES 2012

9th International Conference on Preservation of Digital Objects.

<http://ipres.ischool.utoronto.ca/>



**Sommige dingen zijn alleen in beeld begrijpelijk. De relaties tussen 8.000 hoogleraren en universitair hoofddocenten aan 14 universitaire instellingen bijvoorbeeld.** *Andrea Scharnhorst*

Visiting professor Katy Börner licht toe

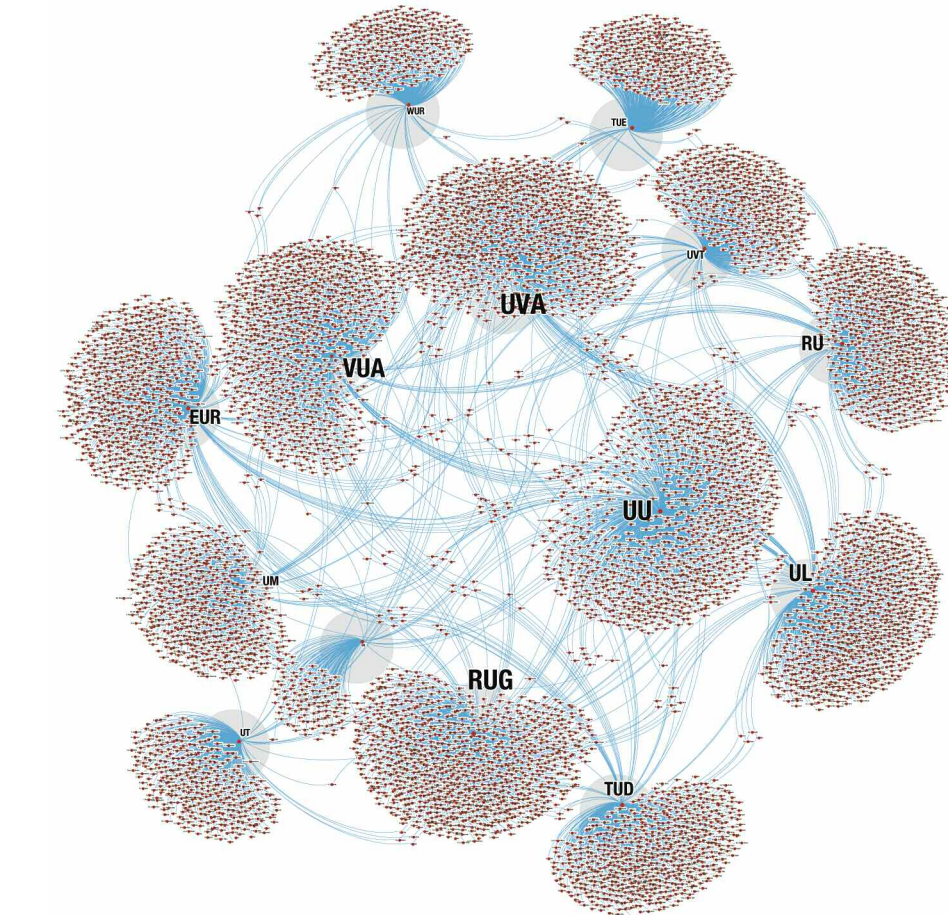
# Facebook voor onderzoekers

Die gecompliceerde relaties worden met VIVO een stuk overzichtelijker. VIVO is een open semantische webapplicatie en ontologie-editor ontwikkeld op de Cornell University, VS. Met deze applicatie wordt informatie uit gegevensopslagsystemen van wetenschappelijke instellingen geïntegreerd en omgezet naar webpagina's en interactieve visualisaties. Visiting professor bij de KNAW Katy Börner licht de applicatie met een voorbeeld toe.

“De afbeelding laat een netwerk zien van 14 instellingen met hun ruim 8.000 hoogleraren en universitair hoofddocenten. Het netwerk kent twee soorten knopen: de universiteiten en de onderzoekers. De grootte van de institutionele knopen is geschaald op basis van het aantal dienstbetrekkingen van personen bij deze instellingen. Maar ook de grootte van de knopen van onderzoekers is geschaald: hoe meer betrekkingen, hoe groter de knoop en het lettertype.”

## Meerdere dienstbetrekkingen

“Het grootste knooppunt is de Universiteit van Amsterdam (UvA). Veel van de 930 hoogleraren en universitair hoofddocenten aan de UvA hebben twee dienstbetrekkingen, één hoogleraar heeft zelfs banden met drie universiteiten. Deze personen worden als het ware ‘gedeeld’ en vormen de lijnen in de afbeelding tussen de universiteiten. Wel zijn de meeste hoogleraren aan één universiteit verbonden. Het netwerk dat zo ontstaat, is een uiting van ‘Mode 2’: daar waar men voorheen loyaal was aan één universiteit, vindt nu een verschuiving plaats naar transdisciplines en een toenemende mobiliteit in academische loopbanen. Uiteindelijk leiden de ‘gedeelde docenten’ tot



*Dienstverbanden tussen de 14 universiteiten illustratie Katy Börner en Michael J. Stamper*

een nationaal netwerk van Nederlandse onderzoekssamenwerking. En dat is interessant. Bijvoorbeeld om internationaal te laten zien dat Nederlandse universiteiten op persoonsniveau met elkaar verbonden zijn, samenwerken.”

## Dezelfde taal

Ook over de toepassing voor individuele onderzoekers heeft Börner nagedacht. “NARCIS biedt 24/7 een online overzicht van onderzoekers in Nederland. Deze applicatie zorgt

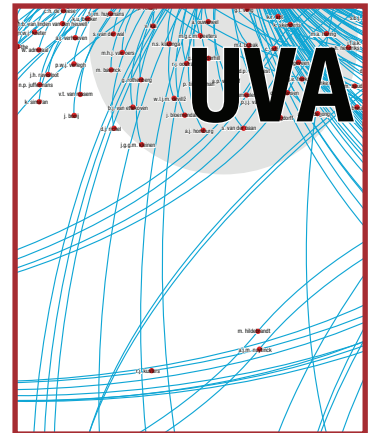
ervoor dat dit overzicht in één oogopslag zichtbaar is.” Börner onderzoekt momenteel met het Centre for Science and Technology Studies en DANS welke deskundigheidsgebieden de hoogleraren uit NARCIS voor hun rekening nemen en in hoeverre deze gebieden met elkaar zijn verbonden. “De getoonde visualisatie is een experiment, gebaseerd op een momentopname. Het laat zien dat je de content van NARCIS en andere onderzoeksinformatiesystemen wereldwijd kunt uitwisselen, be-

ter kunt ontsluiten, visualiseren, doorzoekbaar kunt maken. Om zo het potentieel van Nederlandse onderzoekers in het internationale wetenschapslandschap visueel tastbaar te maken. Bovendien spreekt iedereen dezelfde taal bij een afbeelding. Het laat zien waar je staat in de wetenschappelijke wereld.” Iedereen heeft toegang tot VIVO om eigen VIVO-applicaties te ontwikkelen en analyses te maken.

[vivo.sourceforge.net](http://vivo.sourceforge.net) <http://vivoweb.org>

## Maak kennis met Börner

Op 6 juni houdt Katy Börner de lezing ‘Mining, Mapping, and Accelerating Science and Technology’. Onderdeel van deze lezing zijn voorbeelden van visualisatie. Aanmelden voor deze gratis lezing is verplicht, en kan via [www.dans.knaw.nl](http://www.dans.knaw.nl).



Legenda:

**WUR** Wageningen Universiteit en Researchcentrum; **TUE** Technische Universiteit Eindhoven; **UvA** Universiteit van Amsterdam; **UvT** Tilburg University; **VUA** Vrije Universiteit Amsterdam; **RU** Radboud Universiteit Nijmegen; **EUR** Erasmus Universiteit Rotterdam; **UU** Universiteit Utrecht; **UuM** Universiteit Maastricht; **UL** Universiteit Leiden; **OU** Open Universiteit; **UT** Universiteit Twente; **RUG** Rijksuniversiteit Groningen; **TUD** Technische Universiteit Delft

Onderwijs is een brug naar andere culturen

# Online panel onderzoekt acculturatie

**Bij acculturatie neemt een groep de cultuur over van een andere groep. Naast opvattingen die dit proces beïnvloeden, meet het LISS migrantenpanel verschillende opvattingen over de tijd. Met interessante resultaten.** *Eric Balster*

Acculturatie is het proces waarbij de cultuur van een groep wordt overgenomen door een andere groep. Psychologen zijn zeer geïnteresseerd in dit proces en in de diversiteit in attitudes die dat proces beïnvloeden. Niet alleen binnen, maar juist ook tussen de verschillende groepen. Binnen het LISS migrantenpanel kunnen de opvattingen achter attitudes worden gemeten, onder zowel autochtone als allochtone Nederlanders, en dat geeft interessante resultaten.

De eerste resultaten van onderzoek in het migrantenpanel laten zien dat de persoonlijkheidsverschillen tussen autochtonen en allochtonen kleiner zijn dan we al dachten. “En dat geldt ook voor de etnische groepen onderling”, stelt Van de Vijver. De verschillen in opvattingen zitten vooral bij de tevredenheid met de relatie/het huwelijk, religie en verdeling van huishoudelijk werk. De afstand die de autochtone bevolking ervaart tot etnische groe-

pen heeft grote invloed. Hoe groter de afstand, des te minder deze allochtonen zich thuis zullen voelen in een land. Daarbij is het zeer interessant om te kijken hoe mensen zich identificeren: voelen ze zich Nederlander, Turk, Turkse Nederlander, Nederlandse Turk? Door de panelopzet kan dit nu voor de verschillende generaties over de tijd gemeten worden. Van de Vijver: “Onderzoek laat zien dat onderwijs een belangrijke brug is naar de an-

dere, in dit geval Nederlandse cultuur.”

## Het panel

Het migrantenpanel bestaat sinds eind 2010 binnen het MESS project naast het reguliere LISS panel. De doelstellingen van dit panel zijn om acculturatieprocessen en multiculturalisme te bestuderen, en om inzicht te verkrijgen in de methodologie van het opbouwen en onderhouden van een internetpanel van migrantengroepen. Het panel is opgezet op basis van een kanssteekproef, getrokken door het CBS uit het bevolkingsregister, gestratificeerd naar etnische afkomst. De kracht van het panel voor onderzoekers is de brede, representatieve samenstelling (veel etnische groepen zijn vertegenwoordigd). De vakgroep Crossculturele Psychologie van de Universiteit van Tilburg

is naast CentERdata en het Centraal Bureau voor de Statistiek (CBS) de andere partner in dit project.

## De data

Panelleden vullen sinds oktober 2010 online vragenlijsten in. De verzamelde data zijn na registratie beschikbaar via de LISS data website. Bovendien kunnen onderzoekers een voorstel indienen om kosteloos eigen data te verzamelen in het migrantenpanel. Is er dan niets meer te wensen? Ja, toch wel: alle vragen zijn in het Nederlands en dat leidt waarschijnlijk toch tot uitval van bepaalde respondenten binnen de etnische groepen (differentiële uitval). Meertaligheid zou daar een oplossing voor kunnen zijn. “Er is dus nog ruimte voor verbetering”, concludeert Van de Vijver.

[www.lissdata.nl/dataarchive](http://www.lissdata.nl/dataarchive)

## Fons van de Vijver

Van de Vijver is hoogleraar crossculturele psychologie aan de Universiteit van Tilburg en bijzonder hoogleraar aan North-West University (Potchefstroom, Zuid-Afrika) en University of Queensland (Brisbane, Australië). Hij is auteur van meer dan 350 publicaties op allerlei deelterreinen van de crossculturele psychologie, zoals acculturatie, multiculturalisme en intelligentie. [www.fonsvandevijver.or](http://www.fonsvandevijver.or)



*foto Mary Lou Emmert*



Kees Mandemakers is nog steeds blij met zijn onderscheiding

# ‘Ben jij dat, van die Dataprijs?’

In oktober dit jaar wordt de Dataprijs weer uitgereikt. Nomineren kan tot 1 juli aanstaande. Maar wat brengt de prijs, wat levert het op? We vragen het Kees Mandemakers, winnaar van de Dataprijs 2010.

Heidi Berkhout

Mandemakers kreeg de prijs voor zijn werk aan de Historische Steekproef Nederland (HSN). “De HSN was door een collega uit de HSN stuurgroep ingediend, ik wist in eerste instantie van niets. Toen we hoorden wie de andere genomineerden waren (Pearl Dykstra met de Netherlands Kinship Panel Study, Marcel Das met het LISS panel en Kees Aarts met het Nationaal Kiezeronderzoek, red.), voelden we ons vereerd!”

“Ik denk dat de kracht van de HSN er vooral in ligt dat het een initiatief is voor en door onderzoekers uit verschillende disciplines binnen de sociale en historische wetenschappen. Samen hebben we een representatieve database van negentiende- en twintigste-eeuwse levensgeschiedenissen opgezet. Een steekproef uit geboorteakten uit de periode 1812-1922 dienen als basis, dat zijn ongeveer 80.000 geboorteakten. Door aan deze akten gegevens uit andere archieven toe te voegen, wordt on-



Applaus voor Dataprijs winnaar Kees Mandemakers (midden) foto Johannes Abeling

derzoek mogelijk naar de levensloop van diverse groepen Nederlanders. Mijlpalen zijn onze data-releases: zo telt de release uit 2010 levensloopgegevens van ongeveer 37.000 onderzoekspersonen. Inclusief de ouders, kinderen en andere personen waarmee onderzoekspersonen hebben samengeleefd en gewoond, zijn dit ruim 400.000 personen.”

## Interessant om 4 redenen

Voor onderzoekers is de HSN om vier redenen interessant. De HSN zorgt voor een representatieve dataset voor onderzoek naar maatschappelijke processen in de negentiende- en twintigste-eeuw. Denk hierbij aan onderzoek naar sociale mobiliteit, naar de oorzaken van sterfte. Ook functioneert de HSN-basisset als controlegroep(en) waar onderzoekers andere onderzoekspopulaties mee kunnen vergelijken. Bovendien bouwt de HSN expertise op over data-invoer en -beheer, handig voor onderzoekers die daar niet aan toekomen. Als laatste is de HSN interessant omdat onderzoekers de eigen onderzoeksprojecten op de HSN-

basisset kunnen voortbouwen. Een goed voorbeeld is het onderzoek van Hilde Bras en Ineke Maas waarbij de levenslopen van broers en zussen van bestaande onderzoekspersonen worden gereconstrueerd.

## Onverwachte complimenten

Mede door de Dataprijs weten meer mensen van de HSN, vertelt Mandemakers. “En nog steeds krijg ik uit onverwachte hoek complimenten, mede door de publiciteit rondom de Dataprijs. ‘Ben jij dat, van die Dataprijs?’, wordt dan gevraagd. Dat is leuk!”

Inmiddels is de HSN aanvoerder van het European Historical Populations Samples Network (EHPS). Een netwerk van de European Science Foundation (ESF) waarbinnen 20 landen samenwerken om tot een gemeenschappelijke data infrastructuur te komen met informatie over personen, gezinnen en huishoudens. “Ieder land heeft wel een database met historische persoonsgegevens, al varieert de kwaliteit en kwantiteit enorm. Onze eerste halfjaarlijkse bijeenkomst in januari was

## IISG en HSN

Het Internationaal Instituut voor Sociale Geschiedenis (IISG), onderdeel van KNAW, is één van 's werelds grootste documentatie- en onderzoekscentra op het terrein van de sociale en economische geschiedenis. Het instituut bezit ruim 3.000 archieven, 1 miljoen boeken en tijdschriften en een grote collectie beeld- en geluidsmateriaal. Het IISG onderhoudt verschillende grote databestanden waaronder die van de Historische Steekproef Nederlandse bevolking (HSN).

Kees Mandemakers is senior onderzoeker op het IISG en bijzonder hoogleraar 'Large Historical Databases' bij de Erasmus School of History, Culture and Communication (ESHCC).

belangrijk; we wilden kennismaken, kennis uitwisselen en taken binnen dit nieuwe netwerk verdelen. Bestaande gegevensbanken moeten bijvoorbeeld hun data omzetten naar een standaard die iedereen gaat gebruiken. Fysieke aanwezigheid was belangrijk, je maakt makkelijker afspraken als iedereen er is. De frequentie van de bijeenkomsten zorgt dat de gang erin blijft. Het prijzengeld van de Dataprijs gaf ons net dat beetje extra, we konden het event breder opzetten dan in eerste instantie mogelijk was. Inmiddels zijn negen werkgroepen succesvol van start gegaan.”

## Niet te bescheiden

“De Dataprijs is echt de moeite van het winnen waard. Naast de prijzen geeft het erkenning. Onderzoekers zijn vaak bescheiden. Denken vaak ‘te klein’, dat het werk (nog) niet de moeite van het vertellen waard is. Maar binnen elke dataset gebeurt wel iets – en vaak is dat iets unieks. Laat maar zien hoe goed je database of dataset al is! Wij zien de prijs als een kroon op ons werk, zijn er trots op!”

[www.iisg.nl/hsn](http://www.iisg.nl/hsn)

## In Mandemakers' voetsporen?

Ook in 2012 wordt de Dataprijs uitgereikt. Een prijs voor onderzoekers die onderzoeksgegevens bij elkaar brengen, documenteren en toegankelijk maken voor de wetenschap. De prijs bestaat uit een bedrag van € 7.500, een sculptuur en een geheel verzorgd diner voor de onderzoeksgroep. Indienen kan tot 1 juli via [www.dans.knaw.nl](http://www.dans.knaw.nl)

**Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij CBS, CentERdata, Huygens ING (Huygens Instituut en Instituut voor Nederlandse Geschiedenis) en DANS.**

### Centraal Bureau voor de Statistiek

Een volledig overzicht van de CBS-bestanden staat op [www.cbs.nl/microdata](http://www.cbs.nl/microdata)

- Productiestatistiek: autohandel, bouw-nijverheid, commerciële diensten, delfstoffenwinning, detailhandel, energie, groothandel, industrie en transport, 2009
- Algemene Nabestaandenwet, 2001-2006
- Bijstandsuitkeringenstatistiek, registraties, 2001-2006
- Gemeentelijke Basisadministratie Personen, 1995-2011
- Mobiliteitsonderzoek Nederland, 2009
- Integrale veiligheidsmonitor, 2010
- ICT huishoudens en personen, 2008
- Onderzoek verplaatsingsgedrag in Nederland, 2010

### CentERdata LISS Data Archive

De LISS-bestanden zijn kosteloos beschikbaar via [www.lissdata.nl/dataarchive](http://www.lissdata.nl/dataarchive)

### Studies LISS panel

- Family and Household - Wave 3 (CentERdata), maart 2010

## SINDS KORT BESCHIKBAAR

- Left- and right-handedness (Buser, T.; Oosterbeek, H.; Plug, E.), mei 2010
- Using Anchoring Vignettes to Examine the Effect of Socioeconomic Status and Child Health (Delaney, L.; Hudson, E.), mei 2010
- Decreasing Trust in Democratic Governments - Wave 4 (Wisse, E.), maart 2010
- Work and Schooling - Wave 4 (CentERdata), april 2011 / mei 2011
- Economic Situation: Housing - Wave 4 (CentERdata), juni 2011 / juli 2011
- Tilburg Consumer Outlook Monitor - 2011 March (Yabar, J.; Pieters, R.; Leenheer, J.), maart 2011
- Economic Situation: Income - Wave 4 (CentERdata), juni 2011 / juli 2011
- Are Investors Driven by Sentiments? - 2010 November, December, 2011 February, March, June and July (Kaplanski, G.; Levy, H.; Veld, C.; Veld-Merkoulova, Y.), 2010 november, december, 2011, februari, maart, juni en juli
- Initial Questionnaire - 2010, 2011 (CentERdata), januari t/m december 2010, 2011
- Politics and Values - Wave 5 (CentERdata), december 2011 / januari 2012
- Help Resources (Houwen, K. van der), september 2011

- The Guardianship Survey (Reynald, D.), augustus 2011
- Values Questionnaire (Vijver, F. van de), april 2011
- Social contacts (Houwen, K. van der), augustus 2011
- Social Integration and Leisure - Wave 1 (CentERdata), februari 2011
- Retrospective Childhood (Vijver, F. van de), september 2010

### Huygens ING

Het Huygens ING heeft de volgende publicaties gedigitaliseerd en op [www.historici.nl](http://www.historici.nl) gezet:

- Handel met Frankrijk 753-1585, 4 delen. Bronnen tot de geschiedenis der Oostindische Compagnie in Perzie, 1611-1638, 1 deel
- Nederlandse oorlogen op zee 1652-1676, 2 delen
- Oorkonden van de abdij Kloosterrade 1108-1381, 1 deel
- Oorkonden van Noord-Brabant 690-1312, 4 delen
- De correspondentie van Willem van Oranje, bundel ten gevolge van de openstelling van de databank op 12 april 2005, 1 deel
- De kolonie en dekolonisatie, congresbundel

### DANS EASY

De bestanden zijn kosteloos beschikbaar via <http://easy.dans.knaw.nl>

### Sociale Wetenschappen

- Kostas Gemenis and Roula Nezi, University of Twente, Department of Public Administration, The 2011 political parties expert survey in Greece
- M. de Voogd, Sociaal en Cultureel Planbureau, Arbeidsvraagpanel 1989-2010
- CBS, Leefsituatie Index Onderzoek 2010 (beveiligd microbestand)

### Historische Wetenschappen

- A.C.M. Kappelhof, Roggeprijzen Breda 1535-1870
- J.H.A. van der Velden; International Institute for Social History IISG - Amsterdam, Strikes in the Netherlands, 1810-1995; Stakingen in Nederland, 1810-1998

### Archeologie

- ADC, Boarnsterhim Terp Groot Palma: een archeologische begeleiding, 2010
- Universiteit Leiden en NWO, Preliminary report on the 2010 excavations at Tell Satu Qala, Erbil Province, Kurdish Region, Iraq, 2010

### Geodata

- Kadaster, Top10NL, februari 2012



Kees Schuyt, voorzitter van de KNAW-commissie Onderzoeksgegevens:

# ‘Archiveer en deel de data om ze te kunnen controleren’

**Diederik Stapel publiceerde jarenlang artikelen op basis van gefingeerde onderzoeksgegevens. Na het rapport van de commissie Levelt over de fraude stelde de KNAW een Commissie Onderzoeksgegevens in die na de zomer zal rapporteren. Voorzitter Kees Schuyt licht alvast een tipje van de sluier op. Inge Angevaere**

**Was de kwestie-Stapel de enige aanleiding voor de KNAW om een commissie Onderzoeksgegevens in te stellen?**

“Nee”, zegt Schuyt met nadruk, “juist niet. Er waren al voornemens bij de KNAW om een commissie in te stellen die zou onderzoeken hoe in diverse gebieden data beheerd en gearciveerd worden. De kwestie Stapel heeft de opdracht versneld en veranderd. De verandering zit in een verruiming van de opdracht: ook bezien hoe de integriteit van onderzoekers verbeterd kan worden. Dat zijn twee zaken die goed uit elkaar moeten worden gehouden.”

“De fraudezaak is natuurlijk het meest spectaculair. Maar in de dagelijkse praktijk is verbetering van datamanagementpraktijken veel belangrijker. We hebben verscherpte onderlinge controle nodig, niet per se om fraude tegen te gaan, maar om ervoor te zorgen dat er beter wetenschappelijk onderzoek wordt gedaan en dat daar beter over gerapporteerd wordt. Onderzoeksgegevens moeten goed gearchiveerd én gedeeld worden opdat collega-wetenschappers de mogelijkheid hebben om de resultaten te controleren, zowel tijdens het onderzoek als na publicatie.”

**Hebben digitalisering en informatisering invloed op de omgang met onderzoeksgegevens?**

“Ja, zonder meer. Het grote verschil is dat onderzoekers vroeger veel vaker zelf hun gegevens verzamelden.” Schuyt benadrukt dat hij vooral de praktijk kent uit zijn eigen discipline, de sociologie. “We gingen de samenleving in, om te kijken en te luisteren vooral. En we werkten in teamverband. Bij de lunch en tijdens de wekelijkse formele bijeenkomst besprak ik met mijn collega’s en medewerkers onze methodes en resultaten. We controleerden elkaar en hielden elkaar scherp. We wisten precies wat onze cijfers waard waren en rapporteerden daarover. Nu vraagt een onderzoek-



Prof.dr.mr. C.J.M. (Kees) Schuyt foto Inge Angevaere

ker bij het CBS een dataset aan, hij krijgt een enorme hoeveelheid gegevens en gaat daarmee aan de slag, achter zijn computer, met behulp van statistische bewerkingen die voorgebakken zijn.”

“Een ander voorbeeld: biologen staan bekend om hun zeer nauwkeurige waarnemingen. Voor zijn proefschrift heeft Maarten ’t Hart vele, vele uren doorgebracht met het observeren van stekelbaarsjes. Nu maakt men veel gebruik van computersimulaties. De interessante vraag is nu of die simulaties dezelfde bewijswaarde hebben.”

Schuyt is er de man niet naar om te oordelen over de ene of de andere methode. “Maar je moet wel weten wat je data waard zijn, wat je

**“Integriteit is het goede doen zonder dat iemand staat te kijken”**  
(C.S. Lewis)

## De taken van de KNAW-commissie onderzoeksgegevens

- de in verschillende wetenschapsgebieden bestaande praktijken op het gebied van gegevensverzameling en -verspreiding in kaart brengen
  - na te gaan op welke manier onderzoekers en hun werkgevers ervoor zorgen dat normen op het gebied van wetenschappelijke integriteit worden nageleefd
  - adviseren hoe met name jonge onderzoekers ertoe gebracht kunnen worden om wetenschappelijk integer om te gaan met onderzoeksgegevens.
- Het rapport van de commissie zal na de zomer verschijnen.

## INTERVIEW

“De KNAW wilde weten of er in het onderwijs wel genoeg aandacht wordt besteed aan dit onderwerp. Het antwoord daarop is: nee, nauwelijks. Op de middelbare scholen knippen en plakken ze alles bij elkaar. Dat mag ook wel, maar dan moet je de bronnen uitgebreid citeren. Een cursus datamanagement met ethische elementen, die combinatie zou erg sterk zijn. Je moet de notie van wetenschappelijke integriteit overdragen aan jonge generaties. En de beste definitie van integriteit vond ik bij C.S. Lewis: ‘Integriteit is het goede doen zonder dat iemand staat te kijken’. Dat is een intrinsieke eigenschap. Die kun je niet bevorderen door met al te veel sancties te gaan dreigen. Enigszins paradoxaal is dat je wel beter op elkaar moet letten.”

**Maar er is zo weinig tijd, zeggen wetenschappers, de Colleges van Bestuur eisen dat zij zo veel mogelijk publiceren.**

“Dat vind ik een beetje onzin. We hebben hier te maken met hoog-intelligente mensen. Die moeten met die druk om kunnen gaan. Is een langdurig en gedegen onderzoek niet veel meer waard dan een berg artikeltjes? Trou-

wens, de beoordelaars van Stapels artikelen voor de wetenschappelijke tijdschriften (peer reviewers) hebben zich er ook te gemakkelijk van afgemaakt. Als zij hun werk grondig hadden gedaan, hadden ze ontdekt dat er duidelijke fouten in Stapels werk zaten, dat bleek uit het Leveltrapport.”

“Waar het mij om gaat, is dat we zorgvuldiger moeten zijn. Als onderzoekers moeten we beter op elkaar letten. Uit Engeland zijn wetenschappers bekend die in hoog aanzien staan ondanks het feit dat ze zelf weinig (maar wel heel goede) publicaties op hun naam hebben staan. Zij bleken wel in staat om zeer goed gefundeerde kritiek te leveren op het werk van collega’s voordat het wordt gepubliceerd. Dat is goud waard, dat zou ook in Nederland gewaardeerd moeten worden.”

**Wat kunnen instellingen doen om een goed klimaat voor integere wetenschap te bevorderen?**

“Men zou in functioneringsgesprekken niet alleen aandacht moeten hebben voor publicaties, maar ook voor de kwaliteit van onderzoeksgegevens. En bij visitaties zou men – wat nu helemaal niet gebeurt – kunnen toetsen of een instelling voldoende aandacht heeft voor de controle op onderzoeksgegevens, bijvoorbeeld door collega’s. En onderzoeksgroepen zouden jaarlijks hun praktijken met onderzoeksgegevens kunnen bespreken. De VSNU en KNAW kunnen hierin een belangrijke rol spelen. En je moet wat doen aan het onderwijs. Vooral in de researchperiode.”

**Moeten er algemene regels voor datamanagement komen?**

“Nee, daarvoor zijn de verschillen tussen de disciplines te groot. De Commissie bekijkt wel of good practices uit het ene wetenschapsgebied ook kunnen werken in andere gebieden.”

**In de onderzoeksopdracht aan de commissie spreekt de KNAW vooral van de “jonge” wetenschapper. Is die fraudegevoeliger dan oudere onderzoekers?**



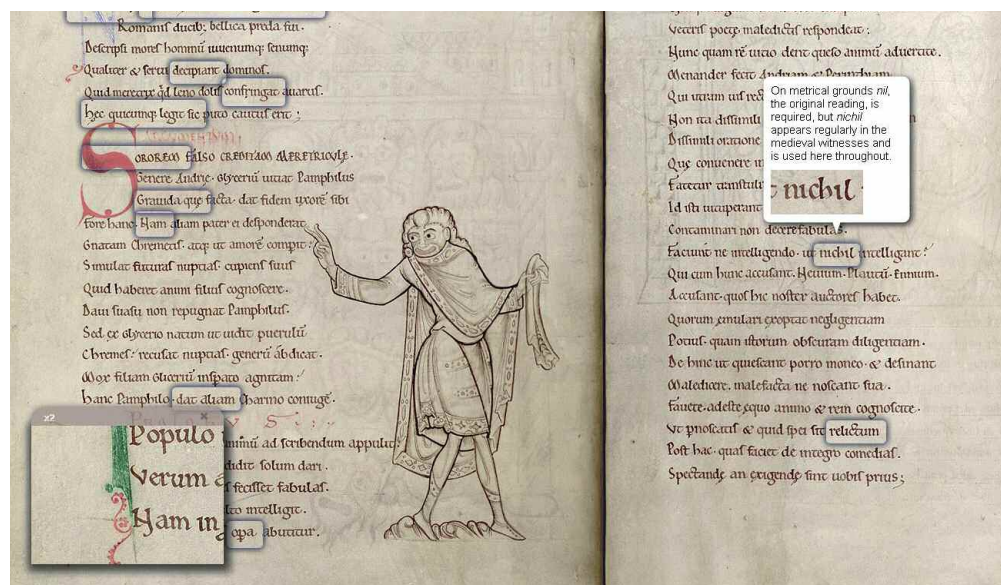
# Fascinerend prachthandschrift

**Dat niet alle dataparels online verkrijgbaar zijn, bewijst een dvd met een twaalfde-eeuws handschrift van zes Latijnse komedies van Terentius. Met vergrootglas en 'hot' knop! Mariken Teeuwen**

De Bodleian Library (Oxford) heeft in september 2011 een tweede deel uitgebracht in de serie Bodleian Digital Texts Series: een dvd met daarop een facsimile editie van Oxford, Bodleian Library, MS Auct. F. 2. 13, een twaalfde-eeuws handschrift van zes Latijnse komedies van Terentius. Net als het vorige deel in de serie (The Vernon Manuscript) bevat de dvd de reproductie van één van de schatten van de Bodleian Library: een rijk geïllustreerd prachthandschrift.

## Textual scholarship

Deze digitale editie dient twee broodheren: aan de ene kant is het een item voor de museumwinkel van de Bodleian Library, een commercieel product dat mikt op een breed publiek, in een reeks 'grote' handschriften uit hun collectie. Daarom ook is gekozen voor een uitgave op dvd, en niet online. Aan de andere kant is het een staaltje van *textual scholarship*, want de dvd biedt alles wat een wetenschappelijk geïnteresseerde zich wenst: kleurenfoto's van hoge kwaliteit, *hyperlinked* transcriptie van de hele tekst, annotaties bij de tekst met tekstkritisch commentaar, vertalingen met hyperlinks naar de transcripties en foto's, een historische introductie en kunsthistorisch commentaar. Aan de set illustraties van dit handschrift is een set uit verwante handschriften toegevoegd ter vergelijking en analyse. Bovendien heeft de tekst een zoekfunctie, die een snelle lijst hits oplevert waarnaar kan worden doorgelinkt. Wellicht was de doelgroep van deze tweede heer



Een inkijkje: links het vergrootglas en rechts de 'hot' knop

beter gediend met een online publicatie, om op termijn andere digitale facsimiles eraan toe te voegen, maar voor dit moment was een vrije online publicatie niet mogelijk.

De komedies van Terentius dateren uit de tweede eeuw voor Christus, en bleven populair gedurende de hele middeleeuwen. Ze werden gebruikt als schoolteksten en zijn overgeleverd in meer dan 600 handschriften. Een dozijn ervan is geïllustreerd. Het rijk geïllustreerd prachthandschrift is vermoedelijk vervaardigd in een Benedictijns klooster in St. Albans.

## Pop-up informatie

De dvd biedt een compleet inzicht in de middeleeuwse setting van de tekst: we krijgen een goed inzicht in de tekstoverlevering, door *pop-up* informatie over afwijkende vormen en varianten, ook de middeleeuwse annotaties die aan de tekst zijn toegevoegd en die een complexe, eigen tekstoverlevering hebben, zijn getranscribeerd en

voorzien van editoriaal commentaar. Daarnaast is de middeleeuwse beeldtraditie van deze antieke teksten tot in de puntjes navolgbaar.

De dvd laat zich gemakkelijk en intuïtief gebruiken. Het is mogelijk om te bladeren, tegelijkertijd de transcriptie erbij te openen, door te klikken naar de vertaling, etcetera. Zonder uitgebreide instructie klik je gemakkelijk door van het een naar het ander en terug. Mooi gevonden zijn de *magnifyer*knop, die je een zwevend vergrootglas geeft dat je over de geopende pagina kunt laten glijden, en de *hot*knop, die de gebruiker per pagina wijzen op bijzonderheden door de passages te markeren met kadertjes. Als de gebruiker één van de kadertjes met de muis aanwijst, verschijnt er een pop-up schermje met commentaar. De dvd is verkrijgbaar bij de Bodleian Library Bookshop voor GBP 66,00. Een zeer geslaagd voorbeeld van digitale tekstontsluiting!

[www.bodleianbookshop.co.uk](http://www.bodleianbookshop.co.uk)

## DASISH brengt ESFRI initiatieven bij elkaar

17, 18 en 19 januari vond bij 'Swedish National Data Service' in Göteborg de kickoff meeting plaats van het Europese project DASISH.

DASISH staat voor 'Data Service Infrastructure for the Social Sciences and Humanities' en wordt gefinancierd door de Europese Commissie vanuit het FP7 programma. Het project loopt drie jaar en beoogt de ESFRI ('European Strategy Forum on Research Infrastructures') initiatieven binnen de Social Sciences and Humanities voor onderzoeksinfrastructuren bij elkaar te brengen. Het gaat om de volgende onderzoeksinfrastructuren: CLARIN (Common Language Resources and Technology Infrastructure), DARIAH (Digital Research Infrastructure for the Arts and Humanities), CESSDA (Council of European Social Science Data Archives), ESS (European Social Survey) en SHARE (Survey on Health, Aging and Retirement in Europe). Het doel is om kruisbestuiving en synergie te behalen bij de ontwikkeling van de infrastructuren die worden opgezet binnen deze initiatieven, zoals gezamenlijke activiteiten op het gebied van dataontsluiting, het delen van data, datakwaliteit en het archiveren van data. Ook kan er synergie worden bereikt ten aanzien van gerelateerde juridische en ethische aspecten. (EB)

<http://dasish.eu/>

## Het Netherlands eScience Center (NLeSC) moet het e-niveau van de wetenschap omhoog krikken. E-data sprak met Jacob de Vlieg, directeur van NLeSC. Peter Boot

NLeSC is opgericht door NWO en SURF met de opdracht e-('enhanced') science in Nederland van de grond te tillen. Enhanced science is data-intensief, multidisciplinair, en ondenkbaar zonder krachtige en gedeelde computer georiënteerde methoden. "Wetenschappers uit uiteenlopende disciplines kunnen van elkaar leren", zo steekt De Vlieg van wal. "Geesteswetenschappers doen aan *text mining*, biologen doen hetzelfde als ze de letters van de genetische code willen kraken."

## Kennis over big data

"Netwerkorganisatie NLeSC wil bij elkaar brengen wat op ICT gebied is ontwikkeld, om vervolgens deze methoden en concepten in een nieuwe context te kunnen hergebruiken. Innovatie ontstaat vaak op de grenzen van disciplines. Wij helpen samenwerkingsverbanden tussen disciplines van de grond te tillen. De eerste serie projecten betreft onder andere klimaatonderzoek, hersenonderzoek, scheikunde



Directeur NLeSC Jacob de Vlieg foto Leo van Velzen

## De wetenschap kan data-intensiever

en sterrenkunde. Er is één geesteswetenschappelijk project, BiographyNed." (zie artikel op pagina 8) De Vlieg vervolgt: "Een voorbereidende commissie heeft een aantal data-intensieve wetenschapsgebieden geselecteerd die rijp leken voor eScience, en daaronder bevindt zich dus ook uitdrukkelijk het terrein van de eHumanities. In die gebieden is een aantal projecten geformuleerd.

In de toekomst zullen we *calls* voor projecten laten uitgaan. De eerste is zojuist aangekondigd en heeft als thema *converging sciences*. De centrale vraag is hoe kennis over *big data* uit het ene domein kan worden toegepast in het andere. En big betekent hier niet alleen een hoog volume, maar ook geproduceerd in hoog tempo of in grote variëteit. Ik verwacht hier ook belangstelling vanuit de sociale wetenschappen,

bijvoorbeeld van onderzoekers van social media. Met de KNAW zijn we in gesprek over mogelijkheden voor een thematische call op geesteswetenschappelijk gebied."

## Duurzame infrastructuur Zijn deze calls het voornaamste beleidsinstrument van NLeSC?

"De calls zijn belangrijk, maar losse projecten alleen leveren geen duurzame infrastructuur. We hebben daarom daarnaast een team van eScience engineers, mensen gespecialiseerd in zowel een wetenschappelijk domein als ICT. Sommige engineers werken bij het NLeSC zelf, anderen zijn gedetacheerd bij onderzoeksgroepen. Dit levert kruisbestuiving tussen disciplines en projecten, en bovendien biedt het een carrièreperspectief en opleiding aan deze nieuwe wetenschappers. In de projecten die NLeSC financiert, zal altijd ook een eScience engineer werkzaam zijn."

## En wat gebeurt er met software die in de projecten tot stand komt?

"Voor langdurig onderhoud aan software hebben wij niet de juiste schaalgrootte. De software zelf komt vrij ter beschikking, beheer kan mis-

schien worden ondergebracht bij instellingen als SARA (de ict-infrastructuurdienstverlener voor universiteiten en hogescholen). Maar we zien ook mogelijkheden voor commerciële *spin-offs*, die services leveren."

## Wat is de rol van de integrators?

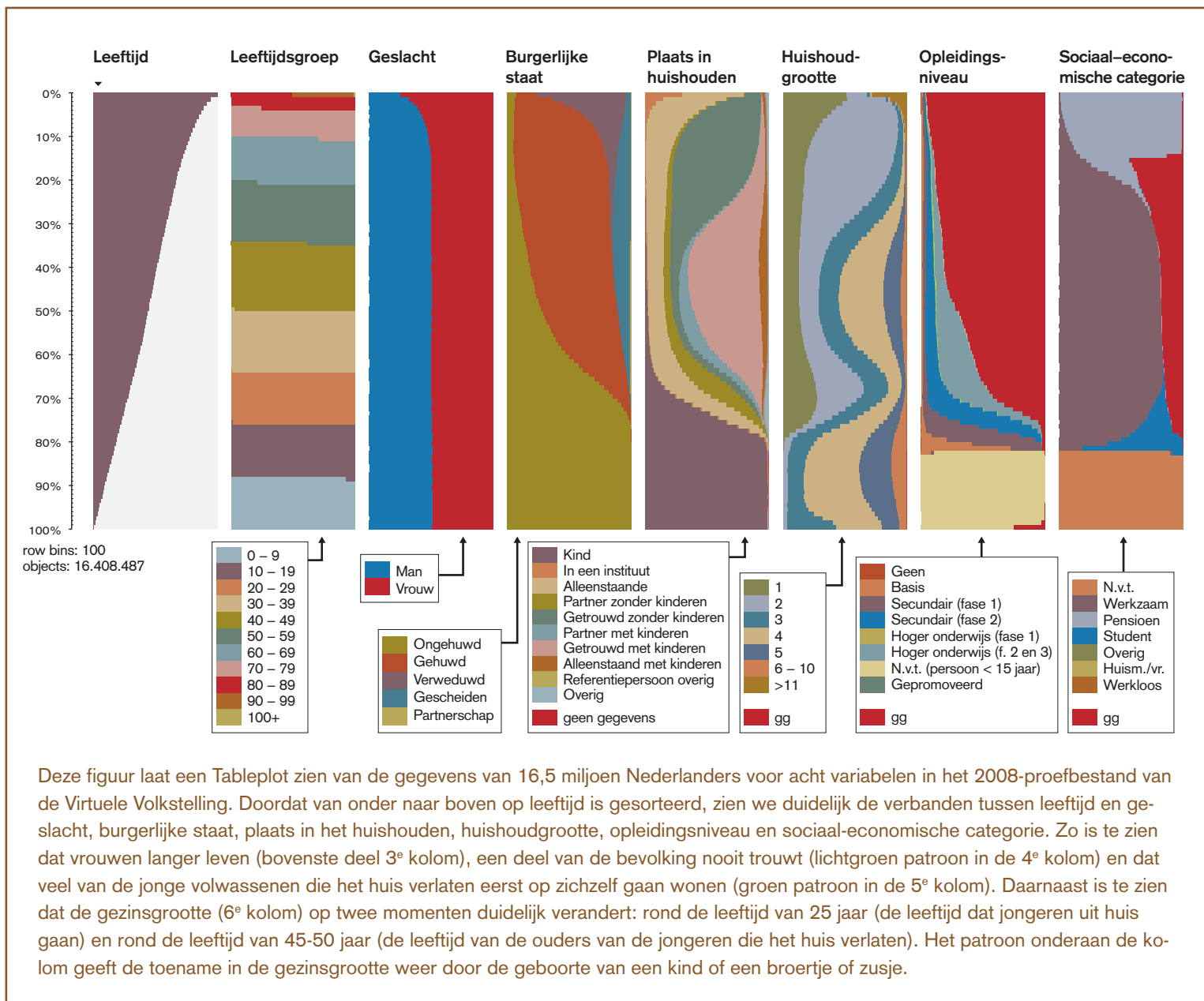
"Integrators zijn wetenschappers, uit verschillende vakgebieden, die een reputatie hebben op het gebied van data-intensief onderzoek. Zij kunnen een verbindende rol vervullen tussen disciplines. De integrators zullen in hun netwerk ook fungeren als evangelist voor het concept van eScience."

Concluderend benadrukt De Vlieg nogmaals het belang van samenwerking. "Euro-commissaris Kroes opende onlangs de bijeenkomst Open Infrastructures for Open Science, en sprak daarbij over het belang van delen, openheid, en samenwerking. In het tijdperk van massieve data hebben we daarvoor geavanceerde ICT nodig. Ik ben ook erg blij dat Jos Engelen, directeur van NWO, bij ons in het bestuur zit om die missie van samenwerking te helpen uitdragen."

[www.esciencecenter.com](http://www.esciencecenter.com)

FOCUS





# Complete dataset kleurrijk weergegeven in één figuur

**Snel inzicht in de kwaliteit van gegevens en mogelijke dataproblemen met behulp van een Tableplot. Het CBS heeft er inmiddels ervaring mee. Piet Daas en Eric Schulte Nordholt**

Het Centraal Bureau voor de Statistiek (CBS) gebruikt voor de statistiekproductie steeds vaker administratieve bestanden die grote hoeveelheden gegevens bevatten. Denk hierbij aan bronnen zoals de Gemeentelijke Basisadministratie, Inschrijfgegevens werkzoekenden (van het UWV WERKbedrijf) en de Polisadministratie (van het UWV).

Eén van de grootgebruikers van administratieve bestanden is Eric Schulte Nordholt en zijn Virtuele-Volkstellingsteam. De volkstelling wordt namelijk tegenwoordig virtueel uitgevoerd. Nadat in 1971 de laatste traditionele Volkstelling in Nederland werd gehouden, waarbij iedere inwoner een vragenlijst moest invullen, worden nu grote hoeveelheden gegevens uit allerlei administratieve bronnen en reeds uitgevoerde enquêtes gecombineerd (zie E-data&Research 3, jaargang 3). Controle van de correctheid van de gelegde koppelingen en de resulterende set van gegevens is belangrijk maar lastig. De Volkstelling brengt namelijk gegevens in allerlei bestanden bij elkaar en is een belangrijk ijkpunt in de vergelijking van gegevens tussen landen, met name binnen Europa. Het feit dat de gekoppelde bestanden een aanzienlijke omvang hebben, maakt controleren ex-

tra lastig. Voor de Virtuele Volkstelling van 2011 moet gedacht worden aan een dataset van ongeveer 50 variabelen voor zo'n 17 miljoen personen! Dat zijn enorm veel gegevens.

## Alles in één

Om de kwaliteit van dergelijke grote bronnen goed en snel inzichtelijk te maken, is recentelijk door de CBS-onderzoekers Martijn Tennekes, Piet Daas en Edwin de Jonge een visualisatiemethode ontwikkeld. De methode, gebaseerd op een aanpak die oorspronkelijk door Duitse onderzoekers is bedacht, maakt het mogelijk de inhoud van een geselecteerd aantal variabelen in een groot databestand in één figuur volledig weer te geven. Een dergelijk figuur wordt in de literatuur een *Tableplot* genoemd. Belangrijkste verbeteringen die de CBS'ers aan het originele ontwerp hebben toegevoegd, zijn: het weergeven van de gebieden waar gegevens ontbreken, de mogelijkheid om de getoonde schaalverdeling en het aggregatieniveau van de gegevens aan te passen en het kunnen gebruiken van kleurpaletten. Daarnaast is de ontwikkelde software als *open source* beschikbaar (als 'R-package') waardoor het door iedereen kan worden gebruikt en (waar nodig) kan worden uitgebreid en aangepast.

Een Tableplot maakt patronen in de verdeling van variabelen en de samenhang tussen variabelen zichtbaar. Maar ook het ontbreken van gegevens is duidelijk. Zo laat de laatste kolom, sociaal-economische categorie, zien dat gegevens ontbreken (rode kleur) van personen die niet studeren, niet in loondienst werken en niet gepen-

sioneerd zijn (zoals zelfstandigen). Deze informatie is in geen enkele publieke administratie te vinden en wordt met steekproeven verzameld, wat de hoeveelheid beschikbare informatie beperkt. Opvallend in kolom 8 is verder dat een aanzienlijk deel van de ouderen boven de 65 jaar nog blijkt te werken (de groene uitloper in bovenste deel van 8<sup>e</sup> kolom).

## Sorteervariabele vereist

Alle genoemde bevindingen zijn, na enige oefening, direct uit het figuur te halen. Met meer traditionele methoden, zoals staafdiagrammen en tabellen, lukt dat niet zo snel en gemakkelijk. Zeker niet bij grote bestanden. Met behulp van een Tableplot kan dat wel. Andere toepassingen van Tableplots door de CBS-onderzoekers laten zien dat dergelijke figuren ook als hulpmiddel voor de controle van de interne verwerking van gegevens te gebruiken zijn. De betrokken CBS'ers verwachten dan ook dat er op het CBS in de nabije toekomst veel vaker van Tableplots gebruik zal worden gemaakt. Momenteel worden BTW-gegevens van de Belastingdienst en de Polisadministratie (loon- en uitkeringsgegevens) van het UWV op deze manier bekeken. Een dergelijke aanpak is ook voor andere onderzoekers van grote databestanden nuttig. Belangrijkste eis voor een succesvolle toepassing is wel dat in de dataset een geschikte sorteervariabele, zoals leeftijd (bij persoonsgegevens) en omzet (bij bedrijfsgegevens), aanwezig is.

<http://cran.r-project.org/web/packages/tabplot>  
<http://cran.r-project.org/web/packages/tabplot GTK>

## KORT

### Wetenschapper mag zelf e-boeken kiezen

Leden van de Koninklijke Bibliotheek kunnen sinds kort 200.000 wetenschappelijke e-boeken lenen. De KB sloot daartoe een contract met het Australische Ebook Library. Publicaties worden *patron driven* aangeboden: de bibliotheek betaalt pas wanneer het boek ook daadwerkelijk wordt geraadpleegd. Zo wordt de collectie wetenschappelijke monografieën van de KB niet langer *just-in-case* maar *just-in-time* en op basis van concrete behoeftes van onderzoekers opgebouwd. Die moeten wel inloggen om bij het aanbod te komen. (SC)

### CATCHPlus: van demo tot software

Eind juni loopt CATCHPlus af. In CATCHPlus zijn enkele CATCH-demo's doorontwikkeld tot software die breed bruikbaar is in de erfgoedwereld. Musea en archieven krijgen bijvoorbeeld toegang tot een Art Recommender en software die automatisch trefwoorden suggereert aan documentalisten. De ontwikkelde tools en diensten zullen eind november gepresenteerd worden tijdens een congres, dat in samenwerking met DEN georganiseerd zal worden. Meer informatie over dit congres is binnenkort te vinden op [www.catchplus.nl](http://www.catchplus.nl). (ER)

### Data Portal maakt zoeken eenvoudiger

Met de DANS Data Portal is het mogelijk om datacollecties van zo'n 75 (inter)nationale organisaties te doorzoeken. Het gaat om omvangrijke collecties van organisaties zoals: het Dataverse Network, Google Public Data Explorer, World Bank, EUROSTAT, Medieval Song Network, European Union Statistics on Income and Living, Geodan, Educatief Gisportaal, Archiefwebsites Nederlandse Politieke Partijen, Werkgelegenheidssysteem LISA, en Europese en Amerikaanse data-archieven. De data zijn afkomstig uit de volgende wetenschappelijke disciplines: geesteswetenschappen, archeologie, ruimtelijke wetenschappen en gedrags- en maatschappijwetenschappen. De portal kan worden uitgebreid met sites door een mail te sturen naar [info@dans.knaw.nl](mailto:info@dans.knaw.nl) met de link naar de site. (HB)

# BiographyNED scheept orde in chaos

**Kan een analytische tool voor het Biografisch Portaal ons verleden ordenen door mensen aan gebeurtenissen te koppelen? Els Kloek en Susan Legêne**

In feite is ons verleden chaos: een openhoop van ontelbare gebeurtenissen, waar geschiedschrijvers iets zinnigs van proberen te maken. Het duiden van het verleden is zowel het uitdagende als het aantrekkelijke van de geschiedwetenschap, vinden historici. De vraag is: kunnen moderne technieken (*tools*) worden ontwikkeld bij pogingen om orde in die chaos te scheppen, bijvoorbeeld door verbindingen te tonen tussen mensen en gebeurtenissen? Of kunnen we technieken ontwikkelen om in de gegevens uit het verleden netwerken, patronen of relaties te ontdekken die het eenvoudige menselijke oog nooit kon ontwaren?

## Dankbaar gebruik

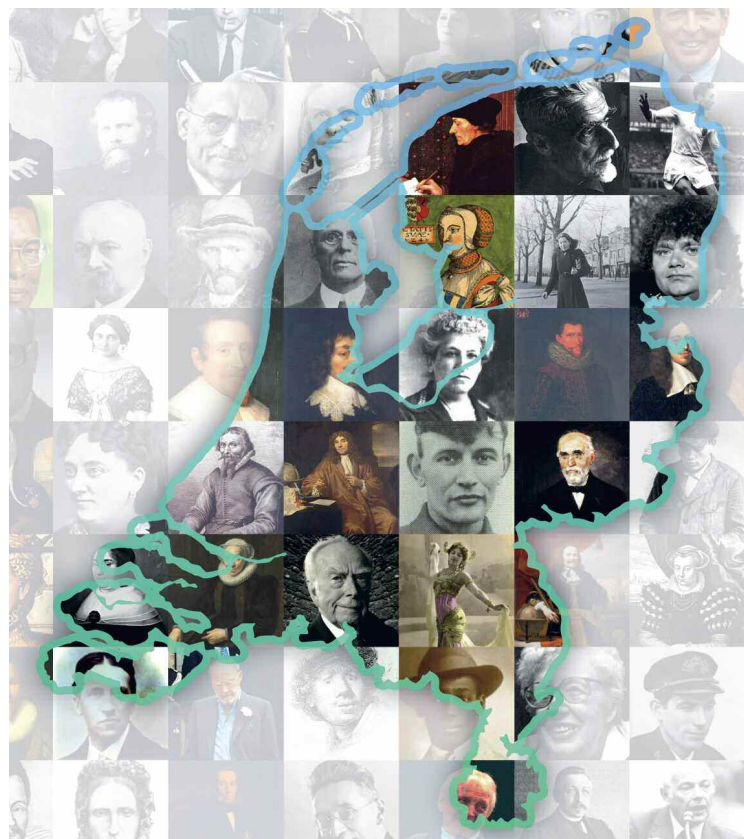
Dit zijn vragen waar het onderzoeksprogramma BiographyNED, onlangs gehonoreerd door het nieuwe Netherlands eScience Center (NleSC, zie rubriek Focus), zich de komende drie jaar mee wil gaan bezighouden. Het project maakt daarbij dankbaar gebruik van de tienduizenden biografieën die eerdere geschiedschrijvers bijeenbrachten in gespecialiseerde naslagwerken. Het e-Science Centre wil interdisciplinair onderzoek bevorderen dat

effectief gebruik maakt van bestaande data en technieken en een bredere inzet van ICT technologieën mogelijk maakt. Die doelstelling sluit goed aan bij BiographyNed, waarin computerwetenschappers en historici van de VU, het Huygens ING en e-Science Centre zullen samenwerken.

## Meer mogelijkheden

Het onderzoeksprogramma BiographyNed neemt het Biografisch Portaal van Nederland als uitgangspunt. Dit portaal, dat in 2010 werd gelanceerd, verbindt een groot aantal online en/of gedigitaliseerde biografische naslagwerken en andere datasets met elkaar. De aldus verzamelde biografische informatie werd in verschillende tijden verzameld en geordend. Sommige biografische collecties gaan over specifieke groepen, zoals kunstenaars of predikanten, andere hebben een meer algemene noemer als 'bekende Nederlanders' of 'vrouwen'. Het Biografisch Portaal verbindt momenteel deze gegeven via een beperkt aantal metadata. Doel van BiographyNed is de vergroting van de mogelijkheden van het portaal voor historisch onderzoek. Welke geschiedenis duikt op uit die meer dan 100.000 persoonbeschrijvingen en portretten?

Meer specifiek wil BiographyNed gaan onderzoeken of uit het Biografisch Portaal gegevens over mensen en gebeurtenissen, hun reizen en verblijfplaatsen en hun netwerken aan elkaar kunnen worden ge-



63.000 Nederlanders op biografischportaal.nl illustratie Jelle Hellinga

## Historici en informatici werken aan BiographyNED

**Prof.dr. Guus Schreiber**, Informatica, Faculteit der Exacte wetenschappen, VU; **Prof.dr. Piek Vossen**, Taal en Communicatie, Letteren Faculteit VU; **Prof.dr. Susan Legêne**, Geschiedenis, Letteren Faculteit VU; **Dr. Els Kloek**, Biografisch Portaal, Huygens ING; **Dr. Jelle Gerbrandy**, Biografisch Portaal, Gerbrandy Software

koppeld. Als het portaal dit met ICT inzichtelijk kan maken, vinden we misschien antwoord op een breed scala aan vragen over elitevorming in Nederland. Wat waren de knooppunten waar mensen met macht en invloed elkaar tegenkwamen, in welke netwerken functioneerden zij? Met dergelijke vragen als sturing ontwikkelt BiographyNed een 'project demonstrator' die een semantische laag over de zeer uiteenlopende data van het huidige Biografisch Portaal heenlegt. Dat

begint met een pilot gebaseerd op een kwalitatieve selectie van personen, waarvoor ondermeer de virtuele portretgalerij die het Rijksmuseum ontwikkelt, het uitgangspunt vormt. Wie geschilderd werd zit in de regel immers ook opgeslagen in het Biografisch portaal. Het project zal in het najaar van 2012 beginnen; vacatures voor twee post-docs zijn door de VU in mei gepubliceerd.

[www.biografischportaal.nl](http://www.biografischportaal.nl)

## GELEZEN

*Users, narcissism and control tracking the impact of scholarly publications in the 21st century*

door Paul Wouters en Rodrigo Costas

In dit rapport analyseren de auteurs zestien tools en methodologieën om de impact van wetenschappelijke publicaties te meten. Voorbeelden zijn Google scholar en Microsoft Academic research. Met hulp van deze tracking tools is te volgen hoe onderzoeksuitkomsten gelezen, geciteerd, gebruikt en ingezet worden. Dit inzicht biedt nieuwe kansen en uitdagingen voor de beoordeling en kwaliteit van de wetenschap. Het rapport bevat een aanzet tot een samenhangend onderzoeksprogramma waarin meetinstrumenten worden gerelateerd aan reeds bestaande indicatoren voor wetenschappelijke impact en levert hiermee een bijdrage aan de ontwikkeling van beter bruikbaar gereedschap voor de wetenschappelijke wereld. (RvH)

[www.surf.nl/nl/publicaties/Pages/Users\\_narcissism\\_control.aspx](http://www.surf.nl/nl/publicaties/Pages/Users_narcissism_control.aspx)

## COLUMN

### Gezocht: eenvoudige richtlijnen voor laagdrempelige digitalisering

Onlangs stond er een paginagroot artikel in NRC Handelsblad over e-boeken. Kortste samenvatting: het is geen marginaal verschijnsel meer. Kun je ook zelf e-boeken maken, vroeg NRC Handelsblad zich terzijde af. Ja, "met een gewone scanner, speciale software en veel geduld." Om de lezers van dienst te zijn, legde de redactie in een kadertje in vier stappen uit hoe je zelf een boek kunt digitaliseren. Eerste punt: "Klem een boek met de rug naar boven in een werkbank. Zaag de rug af met een cirkelzaag." De andere punten zal ik u besparen, die waren net zo onzinnig. Want heb je werkelijk een werkbank en cirkelzaag nodig om de bladzijden van een boek los te snijden? Welnee, in principe kom je een heel eind met een liniaal en stanleymes. En als je het professioneler wil aanpakken kun je het

best een zogenoemde stapelknijder kopen, een apparaat om een stapel papier mee te guillotineren. Voor de tekstherkenning werd in de NRC de verkeerde software aangeraden. De redactie bleek zich voornamelijk te hebben gebaseerd op een verouderd filmpje op internet van een goedbedoelende amateur die zo'n acht uur nodig had om één boek te digitaliseren. Hoewel ik vaste medewerker ben van NRC Handelsblad, kon ik het niet laten om in een ingezonden brief te melden dat dit toch allemaal wel een stuk sneller en beter

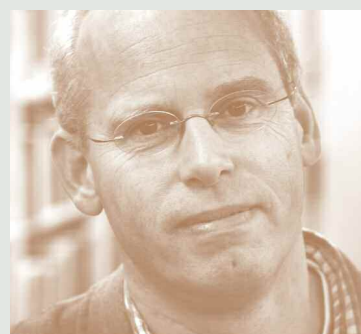


foto Leo van Velzen

kan. Ik scan zelf al jaren en als je het een beetje efficiënt aanpakt, kun je een boek van 220 pagina's binnen 5 minuten volledig digitaal doorzoekbaar maken.

Die brief leverde een paar mailtjes op waar ik van schrok. Zo meldde zich een universiteitsmedewerker die de hele dag proefschriften scant. Zoals u wellicht weet zijn veel universiteiten en wetenschappelijke instellingen relevante publicaties aan het digitaliseren om die op internet toegankelijk te maken.

Ook deze medewerker sneed eerst de bladzijden los, want dat is nu eenmaal de goedkoopste en vaak ook beste aanpak, maar alleen voor het scannen was hij per proefschrift toch zeker zo'n twintig minuten bezig. Kon dat werkelijk sneller?

Ja, met de juiste apparatuur. Iemand anders, eveneens van een

wetenschappelijke instelling, meldde geschrokken dat ze daar kennelijk al heel lang de verkeerde software gebruikten voor tekstherkenning.

Ik zie dit nu al jaren. Ik ken instituten waar mensen al jarenlang publicaties aan het scannen zijn zonder echt te weten hoe dit moet, op de verkeerde apparatuur, met de verkeerde tekstherkenningssoftware.

Is dan niet makkelijk te vinden hoe het wel moet? Metamorfoze beschikt over goede richtlijnen voor digitale conservering en ook bij het Stichting Digitaal Erfgoed Nederland (DEN) is van alles te vinden.

Maar bij mijn weten ontbreekt het aan een actueel, simpel gidsje waarin kort en helder wordt uitgelegd hoe je een efficiënte workflow opzet, welke software en

hardware je het best kunt gebruiken, hoeveel metadata je minimaal nodig hebt, wat je met eenvoudige middelen kunt doen aan digitale duurzaamheid, welke mogelijkheden er zijn als je boeken of tijdschriften niet wil of kunt versnijden, enzovoorts.

Geen digitalisering op het allerhoogste niveau, maar praktisch en laagdrempelig, want vaak is dat al voldoende.

Dat zo'n gidsje er nog niet is, vind ik zorgelijk, want het betekent dat er nu veel tijd en geld wordt verspild. Om nog maar te zwijgen over de soms belabberde kwaliteit van de gedigitaliseerde publicaties.

**Ewoud Sanders**

Taalhistoricus en journalist. Sanders is vaste medewerker van onder meer NRC Handelsblad en Onze Taal