

**Pagina 3 • Boekentoren gescand**

Met de renovatie van het hoofgebouw in zicht sloot de bibliotheek van de Universiteit van Gent zich aan bij Google Library.

**Pagina 4 • Scholloopbanen**

Nu het nieuwe schooljaar is begonnen kunnen de eerste gegevens worden geoogst van het nieuwe cohortonderzoek COOL 5-18. 70.000 leerlingen worden in hun schoolloopbaan gevolgd met een 'basis dataverzameling'. Onderzoekers kunnen vragen toevoegen.

**Pagina 5 • CBS-methodoloog en bijzonder hoogleraar Ger Sniijkers**

De vorig jaar in Utrecht benoemde hoogleraar Ger Sniijkers zoekt naar collega's in de sociale wetenschappen om de bedrijvenstatistiek met een multidisciplinaire aanpak op een hoger plan te brengen.

**Pagina 6 • Virtuele erfgoedgemeenschappen in Vlaanderen**

Een digitale erfgoedroute op een handapparaat aanbieden voor leden van een erfgoedgemeenschap; dat wil het Vlaamse project Erfgoed 2.0.

**EN VERDER**

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

**Schuivende panelen in Europese landen**

## Data-archieven in beweging

Data-instituten in verschillende Europese landen zijn betrokken bij fusies en reorganisaties. Verbreding lijkt de trend in de sociale wetenschappen. De Britse Arts and Humanities Data Service wordt juist in zijn voortbestaan bedreigd.

In Zweden en Zwitserland zijn er plannen om het werkterrein van de data-archieven in de sociale wetenschappen uit te breiden binnen het kader van nieuwe organisaties, die aan een universiteit worden verbonden. De Zweedse Social Science Data Service (SSD) wordt omgevormd tot Swedish National Data Service (SND) en zal het onderdeel menswetenschappen, dat volgens directeur Iris Alfredsson altijd al tot het mandaat behoorde, met nieuw elan gaan aanpakken. Ook volksgezondheid komt in het pakket. Over de financiering van de plannen, inclusief een uitbreiding van zes naar rond vijftien medewerkers, wordt nog onderhandeld met de universiteit van Göteborg die als favoriet eindigde voor de 'hosting' van de SND. 'Er wordt nu eerst met hen gepraat, met als inzet een overeenkomst in oktober. Daarna komt misschien Stockholm in beeld', aldus Alfredsson. De universiteit wordt geacht een derde van het toekomstige budget bij te dragen.

Soortgelijke onderhandelingen zijn juist afgesloten in Zwitserland, waar het nationale sociaalweten-

schappelijke archief SIDOS wordt gekoppeld aan de universiteit van Lausanne en eveneens onder een nieuwe naam zijn horizon gaat verbreden. 'Het was wel even slikken voor Zürich', zegt interimdirecteur Reto Hadorn, 'maar we zullen er alles aan doen om ook hen financieel en anderszins betrokken te houden'. In Zwitserland worden organisaties op het gebied van surveyonderzoek geïntegreerd in het qua omvang verdubbelde instituut dat afdelingen krijgt voor diensten, dataverzameling en data-analyse. Hadorn ziet in de ontwikkelingen in Zwitserland en elders een bevestiging dat het belang van elektronisch databeheer steeds meer erkenning krijgt, maar waarschuwt ook dat die steun breekbaar

is: 'Gelden voor data concurreren altijd met gelden voor onderzoek; dat is een constante'.

Met die keerzijde wordt het Britse data-archief voor de mens- en geesteswetenschappen AHDS geconfronteerd, dat draait op financiële bijdragen van de onderzoeksraad AHRC en het Joint Information Systems Committee JISC. De onderzoeksraad kondigde eind maart aan na jaren de geldkraan dicht te draaien: 'Onderzoekers in de arts and humanities hebben de afgelopen tien jaar belangrijke kennis en deskundigheid op IT-gebied ontwikkeld, waardoor de oorspronkelijke context nu is veranderd', aldus een verklaring. Voor de archeologen blijft de raad een archief steunen,

maar dat zal onder een nieuwe vlag opereren. Medio juni meldde JISC dat het na het besluit van de AHRC niet als enige de lasten zou kunnen dragen. Ook JISC zou zijn bijdrage beëindigen.

Gerelateerde organisaties en bloggers bestrijden de motivatie van de onderzoeksraad: het belang van de AHDS ligt niet in ict-kennis maar in langetermijn archivering. Een internetpetitie op een speciale site van de Prime Minister was half augustus door 1028 Britten getekend. AHDS-directeur Sheila Anderson vindt het nog te vroeg om te zeggen of bij de AHDS echt per 1 april 2008 het licht uit gaat. 'Wij zoeken alternatieve bronnen en mogelijkheden om onze diensten op betaalde basis in de benen te houden'.

In Duitsland tenslotte is het Zentralarchiv voor sociaal onderzoek ZA dit jaar opgegaan in één organisatie met het informatiecentrum IZ en het centrum voor surveyonderzoek en methodologie ZUMA, onder de naam German Social Science Infrastructure Services GESIS. (MdG)

## International Polar Year 2007-2008



Het vierde International Polar Year sinds 1882 werd in maart officieel geopend, tegelijk in Parijs en een aantal arctische locaties zoals dit Ice Hotel in het Noord-Zweedse Jukkasjärvi. Vijftig duizend mensen uit 63 landen zijn betrokken bij 228 projecten om de beide polen en hun gedrag beter te begrijpen. De nadruk ligt op interdisciplinariteit, met een 'actieve inbreng van de sociale wetenschappen', volgens de mondiale wetenschapsraad ICSU die het onderzoekprogramma aanduidt als 'een van de meest ambitieuze internationaal gecoördineerde onderzoekspanningen ooit'. Om

de weelde aan data te beheren die uit het programma voortkomt is een internationale samenwerking van de grond gekomen met als Europese component IPYDLE (International Polar Year Data Legacy Europe), dat bij de Europese Commissie heeft aangeklopt voor financiering. Het Duitse Alfred Wegener Institut is pri-

mus inter pares binnen het IPYDLE-consortium. Nederlandse partners zijn het instituut voor zeeonderzoek NIOZ en DANS (Data Archiving and Networked Services), dat het beheer van de dataerfenis in de alfa- en gammasfeer coördineert. (MdG)

[www.ipy.org](http://www.ipy.org)

## Internationaal dataforum in oprichting

Op een conferentie in Peking in juni zijn stappen afgesproken die moeten leiden tot de officiële oprichting van een International Data Forum (IDF) in 2009.

Een internationaal forum zou nodig zijn om te helpen bij het ontwikkelen van databronnen op wereldschaal en van een internationale onderzoekagenda, zo meldt het conferentieverslag. Tegelijk zou het op moeten passen om niet het werk over te doen van organisaties die zich al inspannen voor coördinatie en uitwisselbaarheid van data. Het zou daarom verbonden moeten zijn met een al bestaande organisatie als de International Social Science Council. Het forum zou zich vooral nuttig kunnen maken door leemtes op te sporen tussen de bestaande databronnen en de behoeftes vanuit internationaal onderzoek, door databronnen op een rij te zetten, de beschikbaarheid van data in de wereld aan de orde te stellen en kennis te delen voor een wereldomspannende data-infrastructuur.

De bijeenkomst in Peking werd

bijgewoond door vertegenwoordigers van 22 landen, waaronder naast veel leden van de OECD ook rising economies als gastland China, India, Brazilië en Korea en ontwikkelingslanden als Bolivia, Zuid-Afrika en de Filipijnen. Financiers waren de nationale organisaties voor (sociale) wetenschap van de VS, het VK, Canada, Duitsland, Nederland en China.

Er werd in Peking een oprichtingscommissie benoemd met als Nederlands lid DANS-directeur Peter Doorn. De commissie kreeg de opdracht zichzelf uit te breiden met nieuwe leden die voor een betere spreiding moeten zorgen over landen en disciplines. In 2008 moet er een werkprogramma met financieringsplan liggen op basis waarvan in het voorjaar van 2009, tijdens het Global Social Science Forum, de officiële oprichting kan plaatsvinden. (MdG)

## e-data: gratis voor wie het wil

[e-data@research](mailto:e-data@research) is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Wilt u [e-data@research](mailto:e-data@research) als onderzoeker of student voortaan gratis ont-

vangen? Geef dan uw adresgegevens door aan het secretariaat: [edata@dans.knaw.nl](mailto:edata@dans.knaw.nl). U kunt het daar ook melden wanneer u geen prijs zou stellen op toezending.

## AGENDA

17 september

Amsterdam, Science Park, CWI-gebouw  
**BIG GRID Symposium** Het BIG GRID project richt een Nederlands Science Grid in voor alle wetenschappelijke disciplines – van hoge-energie fysica en life sciences tot alfa- en gamma-wetenschappen – die steeds meer data-opslag en rekenkracht nodig hebben. Tijdens deze officiële aftrap van het project worden de mogelijkheden besproken van toegang tot dat Grid voor geavanceerd wetenschappelijk onderzoek. [www.biggrid.nl](http://www.biggrid.nl)

25 en 26 september

SPINLAB, Vrije Universiteit Amsterdam  
**International workshop on Geo-ICT and the role of location within science.**  
 The objectives of the workshop are both to increase knowledge of the current use of Geo-ICT in science and to come up with (methodological) strategies to promote more effective use in the future.

2 oktober

Amsterdam – Persmuseum  
**Studiedag Krantendigitalisering**  
 Kranten digitaliseren vergt specialistische kennis en is kostbaar. Deze studiedag van de Nationale Werkgroep Krantendigitalisering heeft als doel het bevorderen van deskundigheid van erfgoedinstellingen die kranten digitaliseren. [www.den.nl/kennis/thema/kranten](http://www.den.nl/kennis/thema/kranten)

7-9 oktober

Ann Arbor, Michigan, USA  
**3rd International Conference on e-Social Science** ESRC National Centre for e-Social Science (NCESS) University of Manchester <http://ess.si.umich.edu/>

11 oktober

Antwerpen, Rubenianum  
**Studiedag Gebruikersonderzoek Digitaal Erfgoed**  
 Op initiatief van de Vereniging voor Geschiedenis en Informatica is er veel aandacht voor de vraag wie de bezoekers zijn van cultureel-erfgoed sites en de gebruikers van erfgoed informatie. Ook worden drie genomineerden voor de VGI innovatieprijs gepresenteerd en wordt de winnaar bekendgemaakt. Aanvang 13.00uur. Aanmelden: [berk@kitlv.nl](mailto:berk@kitlv.nl)

31 oktober

Den Haag, Koninklijke Bibliotheek  
**Studiedag: opzetten van beeldbanken**  
 Studiedag van de Taskforce Archieven, het Geheugen van Nederland, Erfgoed Nederland en DEN met inspirerende praktijkvoorbeelden en workshops over organisatorische, juridische en technische aspecten van het opzetten van een beeldbank. Bedoeld voor zowel 'beginners' als 'gevorderden' op beeldbank gebied binnen de erfgoedsector. [www.kb.nl](http://www.kb.nl)

1-2 november

Den Haag, Koninklijke Bibliotheek  
**Tools & Trends, Internationale conferentie over digitale duurzaamheid**  
 Zie pagina 7 in dit nummer

1-3 november

College Park, Maryland, USA  
**The TEI Members' Meeting TEI@20**  
 Twenty years of supporting the digital humanities [www.lib.umd.edu/dcr/events/teiconference/](http://www.lib.umd.edu/dcr/events/teiconference/)

22 november

Den Haag, NWO Java Gebouw  
**Nederland in Vergelijkend Perspectief – Symposium European Social Survey** DANS, VU, NWO  
 Symposium over de gegevens van de European Social Survey (ESS), met onder meer Nederlandse onderzoekers die betrokken zijn bij of gebruik maken van ESS. Vergelijkingen tussen Nederland en de andere Europese landen, maar ook tussen ESS en andere projecten. Doel is ook om sociaal-wetenschappelijk onderzoekers vertrouwd te maken met inhoud en gebruik van de rijke ESS gegevens. [www.dans.knaw.nl/nl/dans\\_symposia/en/http://ess.nsd.uib.no/](http://www.dans.knaw.nl/nl/dans_symposia/en/http://ess.nsd.uib.no/)

22-24 november

Vilnius, Vilnius University, Lithuania  
**Textual Scholarship and the Canon**  
 The 4th International Conference of the European Society for Textual Scholarship [www.textualscholarship.org/ests/conference07/index.html](http://www.textualscholarship.org/ests/conference07/index.html)

## De politiek van langetermijnarchivering

Een onderdeel van het succesvolle Duitse voorzitterschap van de Europese Unie was een conferentie over Europese samenwerking in digitale langetermijnarchivering. Strategische en praktische aspecten kwamen op 20 en 21 april aan de orde in Frankfurt. De Deutsche Nationalbibliothek was organisator; de Koninklijke Bibliotheek speelde een prominente rol. De conferentie had een politiek doel. Bijna alle bijdragen benadrukten hoe er invulling wordt gegeven aan het beleid van de EU voor digitale duurzaamheid. De bijdrage van Pat Manson, die als hoofd van de afdeling Cultural Heritage and Technology Enhanced Learning van de Europese Commissie de strategie van de EU voor digitale duurzaamheid uiteenzette, was daarom het meest interessant.

Manson noemde een aantal speerpunten: (1) de Commissie wil dat er meer wordt samengewerkt tussen en binnen lidstaten. Drempels rondom toegang en behoud kunnen niet in isolatie opgelost worden. Oplossingen zijn niet direct voorhanden dus moet er flink geïnvesteerd worden in Research en Development; (2) de Commissie ziet een kloof tussen onderzoekers in maatschappij- en geesteswetenschappen en informatici, die gedicht moet worden omdat beiden van elkaar kunnen profiteren;

## Van rollenspel naar gaming

Door de hausse in computergames ontstaat vooral in het onderwijs steeds meer belangstelling voor het gebruik ervan. Simulatie en gaming zijn op zich niets nieuws: de 'International Simulation and Gaming Association' hield in Nijmegen al haar 38e jaarlijkse conferentie (van 9-13 juli). Maar de rollenspel-achtige toepassingen die al decennia worden gebruikt om mensen in organisaties bewust te maken van bijvoorbeeld veranderingsprocessen worden steeds vaker vervangen door computerspellen.

Op de conferentie zochten mensen met zeer diverse achtergronden vanuit zowel de traditie ('first life') als de nieuwe computergestuurde toepassingen ('second life') naar de elementen die games en simulaties effectief maken. Uit diverse bijdragen bleek dat een belangrijk element van traditionele toepassingen is dat gewerkt wordt met één of meer facilitators die het proces begeleiden, sturen, feedback geven en reflectie uitlokken. Bij computertoepassingen ontbreken deze meestal en moet de computer deze taken overnemen. Dit is echter niet zo eenvoudig. Er blijken nog weinig richtlijnen te zijn voor elementen die in educatieve computergames het leren kunnen ondersteunen. Feedback is in veel games vaak te beperkt en alleen op



DIE DEUTSCHE BIBLIOTHEK

Het gebouw van de organiserende Deutsche Nationalbibliothek bij nacht

(3) de kosten van toegang tot data, opslag en hergebruik zijn momenteel nog niet te specificeren. Hoewel dit probleem niet makkelijk aan te pakken is, ziet de Commissie graag aanzetten daartoe; en (4) de Commissie constateert dat kennis wel bij enkele spelers aanwezig is en ziet deze graag wijder verspreid worden.

Behalve de strategie van de Europese Commissie kwamen twee hoofdonderwerpen aan de orde. Het eerste was 'Strategies and Cooperative Approaches of Long-term Pre-

servation in the European Union'. Centraal stond hier de creatie van een overkoepelende Europese organisatiestructuur voor digitale preservatie, waarin de KB een leidende rol speelt. Nationaal ligt de nadruk op het ontstaan van Coalities voor Digitale Duurzaamheid. Het tweede hoofdonderwerp was 'Cooperative

Long-term Digital Preservation Solutions in Europe'. Daarbij lag de nadruk op oplossingen voor digitale duurzaamheid, die momenteel met name worden ontwikkeld en gezocht in de criteria voor betrouwbaarheid en standaards en in workflows, procedures en tools. (LS, PB)

[www.langzeitarchivierung.de/eu2007/index.php](http://www.langzeitarchivierung.de/eu2007/index.php)

## 'Weet wat je meet en waarom'

Uitdagingen voor meetsceptici waren er 28 juni genoeg op het symposium 'Kwantitatieve benaderingen in het taal- en letterkundige onderzoek en elders in de geesteswetenschappen'. De bijeenkomst op het Meertens Instituut vormde de pendant van het in maart verschenen themanummer van het Tijdschrift voor Nederlandse Taal- en Letterkunde (TNTL) over hetzelfde onderwerp.

Voor wie nog weinig met kwantitatieve methoden heeft gewerkt was de middag zeer inspirerend. Frans Hinskens (Meertens instituut, VU) hield een verhelderende inleiding over kwantitatieve benaderingen in taal- en letterkundig onderzoek en ging daarbij ook in op de scepsis over zulke benaderingen binnen de alfawetenschappen. Daarna volgden eerst lezingen over het letterkundig onderzoek: Karina van Dalen-Oskam (Huygens instituut) over stijlonderzoek met Burrows' Delta en Max Louwerse (University of Memphis) over inhoudsanalyse met behulp van Latent Semantic Analysis (LSA). Louwerse daagde vooral de literatuurwetenschappers uit door in te gaan op de volgens hem nog steeds onder hen veel te prominente anti-empirische opvatting. Vanuit de taalwetenschap spraken vervolgens twee dialectologen en een psycholinguïst. Wietske Vonk (Max Planck Instituut) ging in op de laatstgenoemde discipline, waarbij

zij illustreerde hoe door het meten van oogbewegingen geconcludeerd kan worden dat al tijdens het lezen van een voornaamwoord eenduidige informatie over de verwijzing in het voornaamwoord van een zin onmiddellijk door de lezer gebruikt wordt.

De dialectometrie kwam aan bod in de lezing van Marco René Spruit (Meertens Instituut), die een onderzoek presenteerde naar de rol van geografie (de onderlinge afstand tussen twee dialecten) en inwonersaantallen in de verklaring van dialectvariatie. Roeland van Hout (Radboud Universiteit) ging tenslotte in op kwantitatieve analyses van taalvariatie. Op het symposium was ook ruimte voor het kwantitatieve echtheidsonderzoek van schilderijen (door J. van der Lubbe, TU Delft) en voor kwantitatieve stijl-analyse van muziek (door Peter van Kranenburg, Meertens Instituut).

Misschien wel de grootste pioniers in dit gezelschap waren de niet-taalkundigen, en dan met name de literatuurwetenschappers. Het is duidelijk dat binnen hun vakgebied de kwantitatieve benadering nog niet kan bogen op een lange traditie. 'Meten is weten', daar waren alle aanwezigen het wel over eens, maar vooral ook: weet wat je meet en waarom. (Katrien Depuydt)

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

[www.isaga2007.nl/](http://www.isaga2007.nl/)

# Gentse Boekentoren gescand door Google

De bibliotheek van de Universiteit van Gent was onlangs de eerste Nederlandstalige instelling die zich bij het Google Library programma aansloot. Bibliothecaris Sylvia van Peteghem is niet bezorgd over de kwaliteit van scans en metadata.

Zoekgigant Google is al lang niet meer alleen een zoekmachine voor het web. Naast populaire toepassingen als Google Earth en Google Maps werkt het bedrijf ook hard aan het toegankelijk maken van recente wetenschappelijke literatuur in Google Scholar. In het Google Print programma wordt samengewerkt met uitgeverij om hun uitgaven doorzoekbaar te maken. Voor geesteswetenschappers is daarnaast het Google Library programma van groot belang: daarin wordt het boekenbezit van grote bibliotheken gescand en met Optical Character Recognition (OCR) doorzoekbaar gemaakt. Oudere wetenschappelijke literatuur, primaire teksten, maar ook bijvoorbeeld veilingcatalogi, komen zo voor onderzoek ter beschikking.

Met de bibliotheek van de Gentse universiteit is één van de grootste bibliotheken van België nu toegetreden tot het Google Library project. Andere Europese partners zijn de universiteiten van Oxford, Lausanne en Madrid en de nationale bibliotheken van Catalonië (Barcelona) en Beieren (München). De universiteit van Keio (Japan) was onlangs de eerste partner

uit een niet-Europees taalgebied.

Alleen boeken waarop geen auteursrecht meer rust zullen worden gescand. Omdat de universiteit van Gent tot 1930 Franstalig was, is het Frans de belangrijkste taal (47% van de collectie van voor 1900), met Nederlands als goede tweede (22%). Voor de bibliotheek is digi-



JOHAN JACOBZ

De Boekentoren van de Gentse bibliotheek

tale beschikbaarheid van de collectie onder andere belangrijk omdat de hoofdvestiging, de door Henry van der Velde ontworpen Boekentoren, vanaf 2010 wordt gerenoveerd. Volgens hoofdbibliothecaris Sylvia van Peteghem zullen de digitale teksten rechtstreeks vanuit de catalogus toegankelijk worden gemaakt. Aanvankelijk worden alleen boeken uit de Boekentoren gescand; in een later stadium wellicht ook werken uit decentrale bibliotheken.

Het Google Library programma is niet onomstreden. De kwaliteit van de scans, de OCR en de metadata worden niet altijd positief beoordeeld. Van Peteghem heeft op dat punt geen zorgen, verzekert zij *e-data@research*: de Amerikaanse partnerinstellingen waarmee zij heeft gesproken zijn vol lof over Google en noemen zeer lage foutpercentages. De bibliotheek

houdt het recht de belangrijke stukken opnieuw te scannen wanneer een meer dan standaard kwaliteit gewenst is. De metadata zijn afhankelijk van wat de bibliotheek zelf aanbiedt. In eerste instantie streeft Gent naar wat zij noemt een 'no-nonsense catalogus'.

Grootscheepse digitalisatie van bibliotheekbezit wordt ook door andere partijen nagestreefd. De Open Content Alliance werkt samen met het Internet Archive, Microsoft, Yahoo en een aantal bibliotheken aan een gelijksoortig initiatief. De Franse nationale bibliotheek presenteerde onlangs een eerste versie van Europeana, een digitale mega-collectie in aanbouw. Google lijkt echter op het moment het meest aan de weg te timmeren. (PB)

<http://lib.ugent.be/>

## Sinds kort beschikbaar

Onderzoekers kunnen zowel bij DANS als bij het Centrum voor Beleidstatistiek (CvB) van het Centraal Bureau voor de Statistiek (CBS) terecht voor databestanden van het CBS. Bij DANS gaat het om beveiligde microdata die kosteloos geleverd kunnen worden voor secundaire analyses. Ze kunnen besteld worden via [info@dans.knaw.nl](mailto:info@dans.knaw.nl). Aan het gebruik van de onbeveiligde microdata via het CvB is

een aantal voorwaarden verbonden. Informatie over die voorwaarden is te vinden via <http://www.cbs.nl/microdata> -> *Werkwijze*. Het overzicht toont CBS databestanden die recent voor onderzoekers beschikbaar zijn gekomen. Een volledig overzicht van de bestanden met een documentatierapport van de meest recente versies vindt u op <http://www.cbs.nl/microdata> -> *Catalogi*.

	Naam	Periode	Beschikbaar via
SSB	Sociaal Statistisch Bestand Banen	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand Zelfstandigen	2005	CvB
SSB	Sociaal Statistisch Bestand overige arbeid en arbeid buitenland	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand arbeidsongeschiktheidsuitkeringen	2005	CvB
SSB	Sociaal Statistisch Bestand IOAW / IOAZ	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand overige uitkeringen en uitkeringen buitenland	2004 en 2005	CvB
POLS	Permanent Onderzoek Leefsituatie (basis)	1997 t/m 2005	CvB (1997-2004) DANS (2005)
POLS	Permanent Onderzoek Leefsituatie (modules)	1997 t/m 2005	CvB (1997-2004) DANS (2005)
MKB	Milieukosten bedrijven	1998 t/m 2004	CvB
SFGO	Statistiek Financiën Grote Ondernemingen	2005	CvB
LSO	Loonstructuur Onderzoek	1996 en 2002	CvB
EWL	Enquête Werkgelegenheid en Lonen	2005	CvB
EBB	Enquête Beroepsbevolking	2006	CvB
WW	Werkloosheidswet	2003 en 2006	CvB
WSW	Wet Sociale Werkvoorziening	2006	CvB
SRG	Statistiek Reïntegratie door Gemeenten	2006	CvB
BUS	Bijstandsuitkeringenstatistiek ook op transactiebasis	2006	CvB
BFS-BDS	Bijstandsfraude- en debiteurenstatistiek	2006	CvB
AO	arbeidsongeschiktheidsuitkeringen	2006	CvB
NEA	Nationale Enquête Arbeidsomstandigheden	2005	CvB
NVS	Nationale Verzuimstatistiek	2004 en 2005	CvB
EUSILC	European Community Statistics on Income and Living Conditions	2005 en 2006	CvB (2005) DANS (2006)
DO	Doodsoorzaken	2005	CvB
VMR	Veiligheidsmonitor Rijk	2005 en 2006	CvB
DHS	Household Survey	2005 en 2006	CvB
DNB	Huishoudens en personen	2005 en 2006	DANS
ICT	Huishoudens en personen	2005 en 2006	DANS
SEP	Sociaal Economisch Panelbestand (administratief)	2001 t/m 2003	DANS
CCO	Consumenten Conjunctuur Onderzoek	2006	DANS
B&U	Verleende bouwvergunningen burgerlijke en utiliteitsbouw	2006	DANS

## Liederbank gaat online

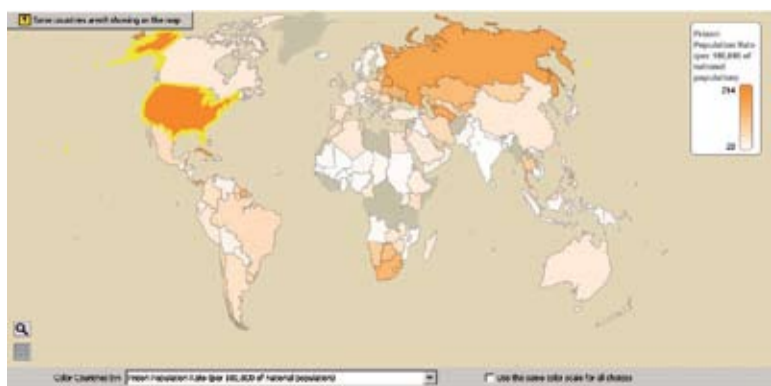
Op 22 juni is de internetversie van de Nederlandse Liederbank officieel geopend. Raadplegen en zoeken in de meer dan 125 duizend liederen van de middeleeuwen tot de twintigste eeuw is nu mogelijk op [www.liederbank.nl](http://www.liederbank.nl). Van elk lied worden naast de beginregel en de titel ook de wijsaanduiding vermeld (waarmee de melodie kan worden gevonden), het refrein, de strofenvorm en trefwoorden. Van enkele tienduizenden

liederen is de tekst beschikbaar; in veel gevallen ook originele audio-opnamen. De Liederbank wordt geproduceerd in het Documentatie- en Onderzoekscenrum van het Nederlandse lied op het Meertens Instituut. Zoeken op hoorbare melodieën is het oogmerk van het nu lopende project WITCHCRAFT, kondigde de geestelijk vader van de Liederbank prof. Louis Grijp aan in een interview in *e-data&research* nr 2-1.

## Data in beeld brengen 2.0

De website Many Eyes is een experiment van IBM op het gebied van visualisatie van data. Gebruikers kunnen kwantitatieve gegevens uploaden en kiezen voor een weergave als staafdiagram, bubble chart, tree map of, als dat toepasselijk is, in de vorm van een kaart met gekleurde gebieden of netwerkdiagram. Vaak bieden de visualisaties ook mogelijkheden om verder in te zoomen op

de data ('beperk de gegevens op het niveau van de VS tot enkele staten') of de weergave te manipuleren ('sorteer uitgaven op categorie van uitgave in plaats van op staat'). Het idee achter de site is dat deelnemers met elkaar discussiëren over inzichtgevende vormen van visualisatie. De figuur toont een van de vele visualisaties: het aantal gevangenen per 100.000 inwoners in de wereld. (PB)



### Leren digitaliseren

Opleidinginstituut GO organiseert vanaf november een zevendaagse cursus 'Digitalisering van culturele collecties'. De cursus gaat over het opstellen van projectplannen en over alle aspecten van de uitvoering. Er wordt voor een bepaalde 'case' ook een volledig projectplan geschreven. Onder meer wordt aandacht besteed aan terminologie, metadata, standaarden, auteursrecht, selectiebeleidbeheer en exploitatie, hard- en software en budgettering. GO opleidingen verkreeg voor de cursus medewerking van 'toonaangevende instellingen' als de Koninklijke Bibliotheek, DANS en het Nationaal archief. Informatie: [www.stgo.nl](http://www.stgo.nl).

### VGI innovatieprijs 2007

'Wie zijn de gebruikers van digitaal erfgoed?' Om die vraag draait een studiemiddag van de Vereniging voor Geschiedenis en Informatica (VGI) op 11 oktober in Antwerpen. Er wordt veel geïnvesteerd om de toegankelijkheid van bibliotheken, archieven, archeologisch onderzoek, museumcollecties en monumenten via internet te vergroten. Recente voorbeelden zijn de massale digitalisering van kranten en audiovisueel materiaal. Op deze studiemiddag is ruime aandacht voor de vraag voor wie dat gebeurt: wie zijn de bezoekers van culturele websites en de gebruikers van erfgoed informatie? Ook wordt de VGI innovatieprijs 2007 uitgereikt voor de meest vernieuwende ICT-toepassing op het terrein van de geschiedenis. Eerst worden drie genomineerde projecten gepresenteerd in het Rubenianum, Kolveniersstraat 20, aanvang 13.00 uur. Aanmelden: [berk@kitlv.nl](mailto:berk@kitlv.nl)

### Plasterk krijgt advies over archieven

Minister Plasterk van Onderwijs, Cultuur en Wetenschap krijgt dezer dagen advies van twee adviesraden over het beheer van overheidsarchieven. Hij vroeg hen in juni om zich daarover uit te spreken, als vervolg op de nota 'Informatie op Orde' van de bewindslieden Van der Laan en Pechtold, die door het kabinet een jaar eerder aan de Tweede Kamer werd aangeboden. De raad voor Cultuur en de Raad voor het Openbaar Bestuur moeten vooral adviseren over de verdeling verantwoordelijkheden, het toezicht, de kwaliteitsborging, archiefselectie en de uitvoerbaarheid van regels en procedures in het digitale tijdperk. Plasterk vroeg beide raden om voor 1 september te antwoorden.

### Onderwijsclassificaties in databank

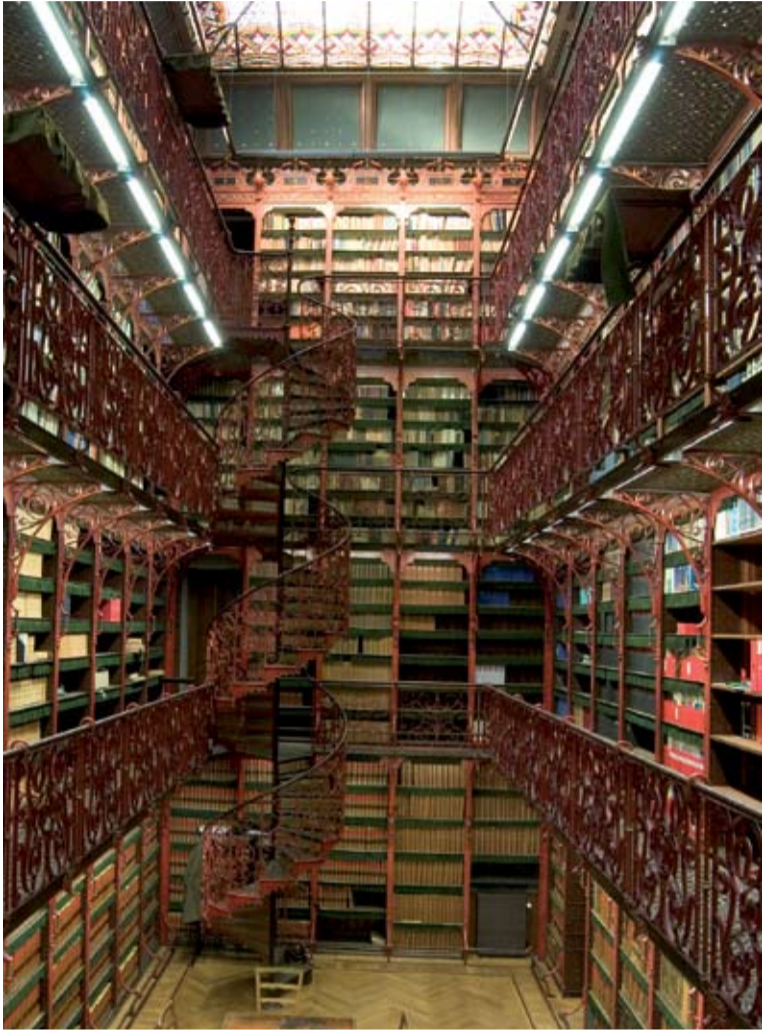
Het Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA) van de universiteit van Maastricht heeft een databank ontwikkeld waarin verschillende soorten onderwijsclassificaties zijn ondergebracht. In Nederland zijn verschillende systemen gangbaar voor zulke classificaties. Die worden met behulp van de databank eenvoudig vergelijkbaar. Het project, genaamd 'Koppeling tussen opleidingsclassificaties', werd financieel gesteund door data-instituut DANS. Het is mogelijk in de classificaties te zoeken, conversieschema's te genereren en labelsets te maken.

Ook registers van Handelingen ontsloten

HUIBERT CRIJNS EN TINEKE KOSTER

# Massadigitalisering is monnikenwerk

Na de lancering van de website [www.statengeneraaldigitaal.nl](http://www.statengeneraaldigitaal.nl) op 8 juni is het project voor de digitalisering van de Handelingen van de Staten-Generaal in volle gang. In juni waren nog alleen de jaargangen 1990-1995 in te zien en te doorzoeken. Maar in 2010 zullen de verslagen en vergaderstukken van de Eerste en de Tweede Kamer uit de hele periode van 1814 tot 1995 (2,3 miljoen pagina's) gedigitaliseerd en online zijn.



In het project Staten-Generaal Digitaal 1814-1995 werken de Tweede Kamer en de Koninklijke Bibliotheek samen. Doel is het conserveren en digitaal ontsluiten van de Handelingen. Behalve omvangrijk zijn de Handelingen ook complex en moeilijk toegankelijk, zelfs voor ervaren onderzoekers. Digitalisering maakt het mogelijk om via een website rechtstreeks in het materiaal te zoeken op onder andere vrije zoekterm, naam, jaartal, datum en documenttype.

Afgelopen zomer is een begin gemaakt met het ontsluiten van de registers van de Handelingen op de website. In de periode 1814-1979 zijn bij de meeste jaargangen registers verschenen. Vanaf 1979 hebben beide Kamers een geautomatiseerd zoekstelsel ontwikkeld. De registers bestaan uit alfabetische lijsten van begrippen (de zaakregisters) en personen (de persoonsregisters) met daarachter verwijzingen naar relevante pagina's in de Handelingen.

De registerpagina's worden eerst als afbeelding ingescand. Het tonen van een afbeelding alleen geeft echter geen meerwaarde boven het

origineel. Het wordt pas interessant als je via het register direct door kunt linken naar de passage in de Handelingen die je zoekt. Daartoe moet het origineel of de afbeelding omgezet worden naar computerleesbare tekst. De huidige OCR-software voor automatische tekenherkenning kan niet goed met historische lettertypes en taal omgaan en laat veel steken vallen, terwijl in de beknopte stijl van de registers elke letter, elk cijfer en elk leesteken telt. Volledig overtypen of handmatige correctie zal daarom voor de digitale ontsluiting noodzakelijk zijn. Bovendien is de opzet, lay-out en de methode van verwijzen niet consistent over de 165 jaar die de registers beslaan. Dat betekent dat deze informatie niet geautomatiseerd verwerkt kan worden.

Onder één lemma vallen vaak tientallen pagina's met verwijzingen, die vooral bestaan uit afkortingen gevolgd door nummers van documenten, pagina's en data. Het zijn eindeloze, bijna abstracte opsommingen van letters, Romeinse en Arabische cijfers en leestekens. Om links te kunnen creëren moeten

alle verwijzingen handmatig in een database ingevoerd worden. En dat is geen domme data-invoer: de datatypisten moet afdoende geïnstrueerd worden zodat zij begrijpen waar zij welke informatie moeten invoeren. De kleinste vergissing zorgt voor een niet werkende link.

Al dit werk gaat resulteren in een aparte zoekpagina op de website [www.statengeneraaldigitaal.nl](http://www.statengeneraaldigitaal.nl). De gebruiker kan kiezen uit een lijst van alle lemma's en deze in een zoekvraag combineren met een jaar of een periode. Als resultaat krijg je een lijst met alle vermeldingen van dat lemma in de registers van de betreffende jaren. Je kunt elke vermelding aanklikken, waarna je alle verwijzingen op je scherm krijgt naar passages in de Handelingen die op dit lemma betrekking hebben. Van elke verwijzing kun je rechtstreeks naar de desbetreffende pagina doorlinken. Door heen en weer te navigeren tussen de lijst met verwijzingen en de documenten is het mogelijk om in relatief zeer korte tijd met behulp van de registers de Handelingen grondig te doorzoeken.

Aan de digitale ontsluiting van historische bronnen gaat dus een enorme hoeveelheid monnikenwerk vooraf. Het spanningsveld tussen kwaliteit (nauwkeurigheid en volledigheid) en kwantiteit (2,3 miljoen pagina's in totaal, 17 duizend pagina's registers) is hierin een voortdurend aandachtspunt.

# Cohortonderzoek brengt nieuwe golf van kansen voor wetenschap

MARTIJN DE GROOT

Bijna vijfhonderd publicaties kwamen er voort uit PRIMA, een van de twee onderzoeken die vooraf gingen aan het deze zomer gelanceerde project COOL 5-18. PRIMA en het VOCL waren cohortonderzoeken voor respectievelijk het lager en voortgezet onderwijs. COOL omvat beide. De eerste gegevens worden in het nieuwe schooljaar binnengehaald. De eerste resultaten komen eind volgend jaar beschikbaar.

De indrukwekkende oogst van de voorgaande onderzoeken zal zeker een rol hebben gespeeld bij de beslissing van NWO om ook voor COOL 5-18 middelen uit te trekken, denkt dr. Lia Mulder, projectleider van het basisschoolgedeelte. 'Tal van onderzoek- en beleidsinstellingen hebben van de eerdere gegevens gebruik gemaakt en dat is ook nu weer de verwachting'.

COOL is een grootschalig cohortonderzoek: van veel mensen worden tegelijkertijd gegevens verzameld over een bepaalde tijdsspanne. In eerste instantie worden drie keer 70 duizend leerlingen van 5 tot 18 jaar in hun schoolloopbaan gevolgd door het primair-, voortgezet- en middelbaar beroepsonderwijs (mbo). In de tweede en derde ronde komen daar nog eens twintig duizend scholieren bij, waarbij speciaal wordt gelet op de cognitieve ontwikkeling (kennis en vaardigheden), de ontwikkeling van burgerschapscompetenties en de sociaal-emotionele ontwikkeling

(persoonlijkheid, motivatie en betrokkenheid).

Ook tal van andere aspecten en invalshoeken kunnen in het onderzoek worden betrokken. Volgens de onderzoekers gaat het om een 'basis dataverzameling', waaraan specifieke data kunnen worden toegevoegd op aanvraag van belangstellende onderzoekers met eigen onderzoeksvragen. Deze zogenaamde module-regeling wordt nog belangrijker dan in de eerdere onderzoeken, legt Mulder uit: 'Toen probeerden we zoveel mogelijk onderzoeksvragen alvast een plek te geven in de algemene dataverzameling. Maar vaak zagen we achteraf dat de gegevens die dat had opgeleverd niet beantwoordden aan de informatiebehoefte van de onderzoekers die zich meldden. Dan moest je toch weer met aanvullende vragenlijsten terug naar leerlingen of schoolleidingen, terwijl die in het eerdere onderzoek al behoorlijk belast werden. Om zeker te weten dat we alle vragen nu aan kunnen



NATIONALE BEELDBANK

zonder dat het teveel op de scholen drukt hebben we nu de basis dataverzameling versmald, met dus die moduleregeling als aanvulling'.

Een kleinere vragenlijst zou ook een kleiner budget kunnen betekenen maar daarvan blijkt geen sprake. Mulder: 'In theorie zou je daardoor minder kosten kunnen hebben, maar daar staat weer een andere verbetering ten opzichte van PRIMA en VOCL tegenover die behoorlijk arbeidsintensief is. We gaan nu ook de leerlingen volgen die uit

het cohort vallen om een of andere reden, zoals verhuizen, zittenblijven of verwijzing naar het speciaal onderwijs'. Voor het eerst kunnen nu namelijk de onderwijsnummers in het onderzoek worden betrokken, waardoor het mogelijk wordt de onderwijsloopbaan te volgen ook van leerlingen die niet meer op de scholen zitten die in het onderzoek zijn betrokken.

Bij de eigen gegevensverzameling wordt ernaar gestreefd in elke nieuweronde steeds dezelfde leerlin-

gen in beeld te houden. Maar na de basisschool verspreiden leerlingen zich over veel verschillende scholen en opleidingen. Om efficiënt te kunnen werken en om ook aan de betrokken scholen bruikbare resultaten te kunnen bieden zal een keus worden gemaakt uit de scholen voor voortgezet onderwijs waar de voormalige basisschoolleerlingen naar toe zijn gegaan. 'Dan gaan we door met die scholen waar een voldoende aantal leerlingen uit de vorige ronde, zeg bijvoorbeeld twintig, naar toe is gegaan. En rond die groepen nemen we dan de hele klas in de nieuwe ronde mee'.

Mulder zelf is projectleider vanuit het Instituut voor Toegepaste Sociologie (ITS) van de Radboud-universiteit, een van de deelnemers in het zogenaamde consortium dat COOL 5-18 tot een goed einde moet zien te brengen. Andere leden van dat consortium zijn het Kohnstamm Instituut dat samen met ITS de basisscholen voor zijn rekening neemt, toetsenproducent CITO en het Gronings Instituut voor Onderzoek van Onderwijs die samen het voortgezet onderwijs onder hun hoede hebben, en het CBS dat voor de onderwijsnummers zorgt. Data-instituut DANS zorgt voor de beschikbaarstelling van de resultaten.

CBS-methodoloog en hoogleraar Ger Snijkers:

MARTIJN DE GROOT

## ‘Multidisciplinair de kwaliteit van data verbeteren’

Een multidisciplinaire aanpak is een van de speerpunten in de visie van methodoloog Ger Snijkers, die eind maart zijn inaugurele rede uitsprak als bijzonder hoogleraar aan de Universiteit Utrecht. Maar hij heeft er meer. De verbinding tussen datapraktijk en wetenschap bijvoorbeeld. Afstand nemen van het beroep op autoriteit – een lange traditie in het vergaren van statistische gegevens. En allesoverkoepelend: ‘de notie dat er bij het verzamelen van data meetfouten kunnen optreden’.

Die notie is een terugkerend thema in het gesprek met Snijkers, die vier dagen in de week zijn werkplek heeft in de Heerlense vestiging van het Centraal Bureau voor de Statistiek (CBS). Welbeschouwd is zijn loopbaan bij het statistiekconcern, die begon met een stage als student politicologie aan de Universiteit van Amsterdam, een stapsgewijze toename tussen de sterke methodologische tradities van de sociale wetenschappen en de bedrijvenstatistiek die het lange tijd zonder die tradities moest stellen. In Voorburg begonnen als specialist in statistische beveiliging, stapte hij een aantal jaren later over naar de andere CBS-vestiging in Heerlen om het ‘vragenlab’ op te zetten voor de sociale statistieken. Hij bleef er ruim acht jaar. In 2000 ging hij naar de divisie Sociale en Ruimtelijke statistieken waar hij zijn proefschrift over onderzoek in het vragenlab afrondde. Bijna vier jaar geleden stapte hij over naar de divisie Bedrijfseconomische Statistieken, een vruchtbare bodem voor zijn inmiddels opgebouwde deskundigheid. En vorig jaar kwam de bijzondere leerstoel die Snijkers de kans geeft een steviger fundament te leggen onder de multidisciplinaire benadering waar hij zo hartstochtelijk voor pleit. ‘Met één been in de productiesector en met het andere in de wetenschap, dat is voor mij de ideale combinatie’.

*‘Nog maar twintig jaar geleden was het verzamelen van gegevens bij bedrijven vooral een administratieve bezigheid’*

‘Het bedrijvenonderzoek bestaat natuurlijk al heel lang bij het CBS en niet alleen daar, maar systematische aandacht voor de opzet van bedrijfsenquêtes en de kwaliteit van de metingen die je voor dat onderzoek doet, die is eigenlijk pas van de

laatste twintig jaar. In dat opzicht is mijn discipline nog heel jong,’ licht de methodoloog toe. ‘Nog maar zo’n twintig jaar geleden was het verzamelen van gegevens bij bedrijven vooral een administratieve bezigheid. Simpel gezegd: je stuurde de formulieren op, je kreeg ze ingevuld weer terug, je checkte een aantal antwoorden en dat was het. Er werd heel weinig gebruik gemaakt van de kennis over het opzetten van enquêtes die was ontwikkeld in de surveypraktijk van de sociale wetenschappen. En er werd sterk gedacht vanuit de opdrachtgever: ‘Dit wil ik weten en zo definieer ik die gegevens’. De vraag bijvoorbeeld of degene die de vragenlijst invult wel dezelfde definities hanteerde, en het besef dat je meetfouten kunt maken door de manier waarop je mensen met vragenlijsten benadert – dat komt uit de sociale wetenschappen.

*‘Er valt nog heel wat te winnen als er een betere coördinatie zou komen tussen de instanties die gegevens opvragen’*

Dat besef is heel belangrijk voor de kwaliteit van je conclusies’. Maar het is ook nog eens van belang voor een grooten genomen instituut als het CBS – weliswaar een ‘productieomgeving’ in de woorden van Snijkers, maar evengoed een omgeving waar de wens voortdurend bestaat om te werken aan de kwaliteit van de metingen.

En die wens wordt nog eens versterkt door factoren als de kosten, het imago van het bureau en de lastendruk. Het imago heeft last van de toch wel autoritaire aanpak waarvoor bij het bedrijvenonderzoek altijd is gekozen, aldus Snijkers. ‘Er wordt in de aanbestedingsbrieven altijd gewezen op de wettelijke verplichting om in te vullen. Daardoor gaan de respondenten het ook als een belas-



TWAN WIEMANS

**Snijkers: ‘Ik zoek collega’s die een bijdrage kunnen leveren. Dat zou tot nieuwe ideeën kunnen leiden’.**

ting voelen en daarmee raken we aan de derde factor: de administratieve lastendruk. De overheid heeft concrete doelen gesteld om die te verminderen en dat geldt natuurlijk ook voor het CBS’.

‘Nu zitten er aan die lastendruk twee kanten. Je hebt de werkelijke lasten en de lasten zoals ze worden gevoeld. De werkelijke administratieve lasten die het CBS veroorzaakt zijn niet hoog, zeggen wij vaak, maar ze kunnen nog verder naar beneden gevoeld. De werkelijke administratieve lasten die het CBS veroorzaakt zijn niet hoog, zeggen wij vaak, maar ze kunnen nog verder naar beneden omdat wij meer gebruik gaan maken van registerinformatie. Dat moeten we ook doen volgens de CBS-wet van 2003, die voorschrijft dat we allereerst gebruik maken van gegevens die door andere overheidsinstanties zijn verzameld, zoals de Belastingdienst. Op dat gebied valt nog heel veel te winnen als er een betere coördinatie zou komen tussen de verschillende instanties die gegevens opvragen bij bedrijven. Ik heb daar in mijn oratie afgelopen voorjaar ook

voor gepleit’. Vooruitgang aan dit front legt trouwens weer een extra accent op het belang van de kwaliteit van de door het CBS zelf verzamelde gegevens. ‘Die vullen de informatie aan die we aan de registers ontlene en je kan je voorstellen dat bedrijven bijvoorbeeld hun belastingaangiftes zo invullen dat ze er fiscaal gezien het beste vanaf komen’.

Zelfs al is tenslotte de objectieve lastendruk niet groot, dan wordt hij nog wel vaak zo gevoeld door de respondenten. ‘Dat beïnvloedt ook de mind-set waarmee ze de formulieren gaan invullen. Ze zullen zich er vaak niet schuldig over voelen om zich er zo gauw mogelijk vanaf te maken – begrijpelijk, maar wel een bron van meetfouten’, aldus Snijkers die daarom onderzoek doet naar het begrijpelijk en gemakkelijk maken van formulieren. ‘We hebben al proeven gedaan met internetenquêtes die zo gemakkelijk mogelijk in te vullen zijn. Dat is een nieuwe

aanpak, en we zijn nu aan het meten wat daarvan de resultaten zijn. De eerste indruk is dat de kwaliteit van de data er niet op achteruit is gegaan, en dat is al heel wat’. Dus ook niet verbeterd? ‘Dat is heel moeilijk te meten op deze korte termijn; ik denk nu dat we daarvoor een langere periode nodig zullen hebben. Maar wat wél duidelijk is, is dat de kosten aanzienlijk lager zijn, voor het CBS en de bedrijven’.

De door Snijkers en anderen opgebouwde kennis op dit gebied wordt nu gebruikt in het project ‘Begrijpelijke Formulieren’ van het ministerie van Binnenlandse Zaken dat bedoeld is voor de hele overheid. Maar behalve de kwaliteit van de formulieren zijn er ook andere factoren die de motivatie van de invullers beïnvloeden, zoals de manier waarop over de enquêtes wordt gecommuniceerd met de bedrijven. ‘Die vragen zich tegenwoordig af ‘Waarom moet ik dit invullen?’ en ‘Wat schiet ik ermee op?’. Daar kan je op ingaan door een incentive te gebruiken – een soort tegenprestatie te leveren voor de tijd die de invuller aan het formulier besteedt. Daardoor zou hij meer het gevoel kunnen hebben van wederkerigheid: ‘ik doe iets voor jou, jij doet iets voor mij’. Wij hebben dat niet zo lang geleden geprobeerd door aan een klein deel van de respondenten voor de kwartaalstatistiek terug te rapporteren: hun eigen cijfers, aangevuld met vergelijkbare cijfers voor de branche en algemene cijfers van het CBS. Ook daarvan weten we nog niet of het helpt om de kwaliteit van de data te verbeteren. Maar we weten al wel dat de reacties vanuit de bedrijven – het ging om horeca en uitzendbureaus – positief waren. Zo positief dat er is besloten om deze aanpak uit te breiden naar de hele steekproef van zo’n twintigduizend bedrijven’.

Maar ook andere aspecten van de communicatie rond de bedrijfsenquêtes, zoals de toon en de formulering van de brieven aan de aangeschreven bedrijven, kunnen meespelen. ‘De vraag is dus eigenlijk: hoe kunnen we onze enquêtes zo goed mogelijk laten aansluiten bij de bedrijven en, zeg maar, verkopen?’

De jonge hoogleraar is ervan overtuigd dat uit disciplines als de cognitieve psychologie, de sociale psychologie, en de communicatie- en informatiewetenschappen kennis gehaald kan worden die helpt om de kwaliteit van de metingen in het bedrijvenonderzoek te verhogen. ‘Ik probeer contacten te leggen met universitaire collega’s die misschien zeggen: ‘ik zou vanuit mijn discipline een bijdrage kunnen leveren’. Dat zou tot nieuwe ideeën voor een aanpak kunnen leiden, die we dan in de praktijk kunnen toetsen’.

Ger Snijkers (1963) werd in het voorjaar van 2006 benoemd tot bijzonder hoogleraar methodeleer van bedrijfsonderzoek aan de Universiteit van Utrecht. Hij hield zijn oratie eind maart van dit jaar (zie e nr 2-1). Snijkers blijft daarnaast als methodoloog werken bij het CBS, waar hij in 1987 in dienst kwam na een studie politicologie. In 2002 verscheen zijn proefschrift *Cognitive laboratory experiences; on pre-testing computerized questionnaires and data quality*.

## Focus

## Nederlands instituut voor onderzoek van de gezondheidszorg

LUUK SCHREVEN

Het Nederlands instituut voor onderzoek van de gezondheidszorg (NIVEL) doet toegepast onderzoek voor beleidsmakers in die sector. *edata&research* sprak met dr. Dinny de Bakker, hoofd van één van de onderzoeksafdelingen van het NIVEL.

Het NIVEL onderzoekt de effectiviteit en de kwaliteit van de gezondheidszorg in Nederland en met name van de daarop gerichte beleidsmaatregelen. 'Wij zijn bijvoorbeeld zeer geïnteresseerd in de gevolgen, bedoeld en onbedoeld, van de recente wijzigingen in het zorgstelsel in Nederland voor de verschillende partijen in de zorg', zegt De Bakker. De relaties tussen die partijen - zorgaanbieders, -gebruikers, -verzekeraars en de overheid - zijn daarbij een belangrijk aandachtspunt. De Bakker: 'Het NIVEL probeert door middel van diverse panels informatie te verzamelen over met name de gebruikers van zorg. Aan de aanbodkant maken we juist weer veel gebruik van databases'.

Nationaal is het NIVEL sterk verankerd in zowel de wetenschappelijke als de beleidsmatige wereld. 'Een aantal staffleden bekleedt een hoogleraarschap aan een universiteit', illustreert De Bakker, 'en daarnaast zijn er nauwe samenwerkingsverbanden met twee door de KNAW erkende onderzoeksscholen: CARE, waarin respectievelijk de universiteiten van Maastricht en Nijmegen en de Vrije Universiteit participeren, en Psychology and Health met de universiteiten van Utrecht, Leiden en Tilburg'. Op beleidsterrein is de financiering door en samenwerking met het Ministerie van Volksgezondheid, Welzijn en Sport belangrijk. Daarnaast is er een convenant met het RIVM en werkt het instituut onder andere samen met het Sociaal Cultureel Planbureau en het Centraal Bureau voor de Statistiek.

'Internationaal zijn we een partnerschap aangegaan met de Wereldgezondheidsorganisatie (WHO). Binnen de Europese regio is het NIVEL het enige instituut dat onderzoek doet op het gebied van de eerstelijnsgezondheidszorg. Daarnaast coördineren we de European Influenza Surveillance Scheme (EISS)', vertelt De Bakker. 'Dat is een internationaal netwerk van huisartsenregistraties en laboratoria dat de verspreiding van het griepvirus in kaart brengt en informeert over de samenstelling ervan, wat van belang is voor het maken van het vaccin'. Een ander internationaal netwerk waarin het NIVEL participeert is de European Association for Communication in Health Care (EACH), gericht op het verbeteren van de communicatie in de gezondheidszorg. 'Onze directeur Jozien Bensing is één van de pioniers op dit vlak. Niet voor niets heeft zij in 2006 de Spinozaprijs ontvangen,'



HERBERT WIGGERMAN

**Dinny de Bakker (NIVEL): 'In Europa het enige instituut met onderzoek in eerstelijnsgezondheidszorg'.**

aldus De Bakker.

Het NIVEL gebruikt veel eigen panels en creëert databases door op vrijwillige basis gegevens van zorgaanbieders af te tappen. Vaak kunnen die tegen kostprijs ook aan derden beschikbaar worden gesteld. 'Databestanden die met publiek geld tot stand komen, zullen ook zoveel mogelijk publiek goed moeten zijn,' vindt De Bakker. Op het gebied van de beschikbaarheid van gegevens zijn er

voor NIVEL wel verbeterpunten: 'De informatisering in de gezondheidszorg neemt steeds verder toe, maar daarbij wordt met de belangen van onderzoekers nog te weinig rekening gehouden'. Liefst zou NIVEL zoveel mogelijk gegevens van diverse bronnen in de gezondheidszorg direct aftappen. 'Daarbij is ook het koppelen en combineren van gegevensbronnen een aandachtspunt. Wellicht kan het CBS daarbij een rol spelen'.

## Virtuele gemeenschappen rond Erfgoed in Vlaanderen

Een digitale erfgoedroute op een handheld apparaat aanbieden aan bezoekers en belangstellenden - leden van een erfgoedgemeenschap. Dat is het oogmerk van het project Erfgoed 2.0 dat in april in Vlaanderen van start is gegaan. Verschillende partners hebben er een consortium voor opgericht onder leiding van het Interdisciplinair instituut voor Breedband Technologie (IBBT), een Vlaams onderzoeksinstituut gericht op ICT in het algemeen en de ontwikkeling van breedbandtoepassingen in het bijzonder.

Erfgoedinstellingen zoals archieven en musea, archeologische sites of sites over bouwkundig erfgoed maken deel uit van wat tegenwoordig een erfgoedgemeenschap wordt genoemd: een groep mensen die waarde hechten aan specifieke elementen en aspecten van het erfgoed en die dit erfgoed ook uitdragen. Een erfgoedgemeenschap bestaat uit gebruikers en bezoekers, professionals en vrijwilligers, organisaties en instellingen. Door de aanwezigheid van een website waarop de bezoekers van een erfgoedroute commentaar kunnen geven, foto's kunnen delen met anderen of nieuwe informatie kunnen geven over het erfgoed, kan er rond een erfgoedroute een virtuele gemeenschap ontstaan.

Het Erfgoed 2.0-project is bedoeld om onderzoek te verrichten naar een sociale interactieve location-based erfgoedbeleving via mo-

biele toestellen binnen een netwerk van erfgoed sites. De digitale content voor deze erfgoedroute moet komen uit onderling met behulp van een metadatamodel verbonden erfgoed-databanken. Het zullen echter nog steeds de makers van een route zijn die een route en verhaal zullen uitschrijven op basis van het bezoekersprofiel dat ze voor ogen hebben. Zij selecteren het digitaal erfgoed dat ze denken nodig te hebben en 'vertalen' dat zo dat het past in de user interface van de mobiele applicatie. Het project Erfgoed 2.0 ontwikkelt een theoretisch model en enkele instrumenten waarmee zo'n digitale erfgoedroute gemaakt kan worden.

Erfgoed 2.0 is een voor Vlaanderen unieke cross-over tussen onroerend-, roerend- en immaterieel erfgoed. Het metadatamodel zal aan erfgoedinstellingen uit de verschillende erfgoedsectoren de mogelijkheid bieden om gegevens met elkaar uit te wisselen. Ook de sociale interactie die Erfgoed 2.0 beoogt vertegenwoordigt een niet te onderschatten meerwaarde voor de erfgoedsector. Het is de bedoeling om in maart 2009 een prototype, waarin alle onderzoekspakketten zijn verwerkt, voor te stellen in een Vlaamse erfgoed site. (Bart De Nil en Gert Nulens)

[www.ibbt.be](http://www.ibbt.be)

## Zicht op Maastricht, een culturele biografie

Hedendaags Maastricht is winkelstad, culinaire stad, architectuurstad, congresstad, studentenstad... Kortom een veelzijdige stad in een fraai historisch decor. Elk gebouw, elk object en elke plek in de stad vertelt een verhaal.

Het geheel van al die verhalen en de samenhang ertussen vormt de culturele biografie van een stad.' Aldus de website [www.zichtopmaastricht.nl](http://www.zichtopmaastricht.nl), die sinds het najaar 2006 online is. Nu nog in het Nederlands; binnenkort volgen Engelse, Duitse en Franse vertalingen.

Jac van den Boogard is regisseur en eindredacteur van de Culturele Biografie van Maastricht die de website biedt. 'De Culturele Biografie toont de gelaagdheid in onze geschiedenis en geeft betekenis aan alle sporen uit ons verleden', zegt hij. 'Dat kunnen monumenten zijn, straten, pleinen, maar ook documenten, foto's of filmpjes. De Culturele Biografie geeft het verleden betekenis, maar niet zoals een digitaal museum want het 'decor' blijft in tact.' De Biografie toont de sporen in hun context, verbonden aan de historische identiteiten van de stad: Romeinse nederzetting, religieuze pelgrimsstad, vestingstad en vroege industriestad.



TWAN WIEMANS

**De Sint Servaesbrug over de Maas in Maastricht**

Om een groot publiek te kunnen bereiken is gekozen voor een digitale multimediale opzet, met een prikkelende en verrassende interactieve presentatie. De eerste laag van de site toont de multimediale presentatie

van historische items, in de tweede laag wordt het uitgebreide verhaal achter gepresenteerd, voor wie meer wil weten. In de derde laag vindt men de wetenschappelijke verantwoording: filmpjes, artikelen, boeken. De

vierde laag is de meest interactieve: Maastrichtenaren schrijven hun eigen geschiedenis: ze voegen zo hun eigen bijdragen toe aan de culturele biografie van hun stad: oral history, verhalen, filmpjes, foto's, alles vanuit de eigen geschiedbeleving.

Het doel van de website, die in opdracht van de Gemeente Maastricht tot stand is gekomen, is de culturele biografie van de stad toegankelijk te maken voor een groot publiek. 'De Culturele Biografie is gericht op Maastrichtenaren en alle andere geïnteresseerden', aldus Van den Boogard, maar dat is nog maar het begin: 'De Biografie nodigt geïnteresseerden uit om hun eigen geschiedenis te schrijven. Door het interactieve karakter kan men zelf de geschiedenis van de stad ontdekken, keuzes maken en dwarsverbanden leggen om de stad op zijn eigen manier, vanuit de eigen belangstelling te leren kennen.' (LS)

# Emulatie schenkt digitale documenten het eeuwige leven

JEFFREY VAN DER HOEVEN

Pong, Space Invaders en PacMan zijn de grootheden van de computerrevolutie. Zelfs nu nog worden afgeleiden hiervan op grote schaal gespeeld. Toch zijn ook de originele spellen nog springlevend. Dit komt door het gebruik van emulatie: een techniek die oude computers kan nabootsen op nieuwe systemen. De Koninklijke Bibliotheek en het Nationaal Archief hebben deze techniek omarmd en tot oplossing verheven voor duurzame toegang tot hun collecties.



'Retrocomputing' is hot, zo blijkt uit de honderden projecten op het gebied van (spel)computeremulatie die op internet te vinden zijn. Wie echter denkt dat de toepassing van emulatie alleen van nostalgische waarde is, heeft het mis. Naast spellen kunnen ook andere applicaties draaiend worden gehouden, zoals Wordperfect 5.1, Lotus 1-2-3 of Norton Commander 2.0. Digitale documenten die met dergelijke software ooit gemaakt werden en inmiddels verloren zijn gewaand kunnen nu weer in volle glorie bewonderd worden.

Sinds eind 2004 is er bij de Koninklijke Bibliotheek (KB) gewerkt aan een strategie voor digitale duurzaamheid op basis van emulatie. Al eerder, in 2000, toonde een kortlopend onderzoek aan dat emulatie van grote waarde kan zijn om het digitale erfgoed toegankelijk te houden. Toen begin 2005 een breder onderzoek werd afgerond bleek de boodschap helder: als we ons digitaal cultureel erfgoed willen laten overleven, dan is

emulatie de nog missende schakel. En zo werd halverwege 2005 een samenwerkingsproject voor emulatie tussen de KB en het Nationaal Archief gestart. Doel was het ontwikkelen van emulatiesoftware, een emulator, die zowel nu als in de toekomst in staat is oude documenten en software in originele vorm weer te geven.

Hoewel bestaande emulatoren dit ook kunnen, is de programmatuur sterk afhankelijk van hedendaagse computers. Aangezien de computer van vandaag net zo snel zal verouderen als die van twintig jaar geleden, is van duurzaamheid dus geen sprake. Om die reden hebben de KB en het archief gezamenlijk gewerkt aan een nieuwe, duurzame, maar ook flexibele emulator. Duurzaam omdat de software zo los mogelijk van het onderliggende computersysteem is ontwikkeld. Flexibel omdat

Ⓔ <http://dioscuri.sourceforge.net>  
[www.kb.nl/hrd/dd/dd\\_projecten/projecten\\_emulatie.html](http://www.kb.nl/hrd/dd/dd_projecten/projecten_emulatie.html)  
[www.planets-project.eu/](http://www.planets-project.eu/)

elk hardware component van de oude computer is nageemaakt als module in de emulator. Hierdoor wordt het mogelijk verschillende modules te koppelen, uit te breiden en aan te passen, net als de componenten in een echte computer.

Na ruim twee jaar is het emulatie project afgerond en begin juli 2007 is de eerste versie van de duurzame emulator opgeleverd. Het heeft de naam Dioscuri gekregen, een knipoog naar de Griekse mythe over de tweeling Castor en Pollux waarvan de een sterfelijk is en de ander onsterfelijk geniet. Met Dioscuri kunnen verloren gewaande documenten en programma's dus een nieuw bestaan krijgen. De emulator staat inmiddels online en is geheel gratis en open source.

De huidige versie van Dioscuri is in staat een Intel 8086 en 80286 computer nauwkeurig na te bootsen, geheel met ondersteuning voor toetsenbord, scherm en vele andere componenten. Inmiddels wordt er al gewerkt aan een uitgebreidere versie van Dioscuri die ook nieuwere computers als de 80386, 486 en Pentium moet kunnen imiteren. Deze doorontwikkeling vindt plaats als onderdeel van het Europese project Planets dat tot doel heeft een netwerk te bouwen van oplossingen voor langetermijn toegang tot digitale documenten.

# DANS: datalicensing voor bescherming rechten en privacy

Door de stormachtige ontwikkeling van het internet is het vaak onduidelijk hoe de rechten op informatie via dat net zijn geregeld. Data-instituut DANS heeft daarom voor zijn eigen archiveringsfaciliteit EASY datalicensing ontwikkeld.

De traditionele auteursrechten en privacybescherming waren niet berekend op de digitale mogelijkheden. De Auteurswet is nu aangepast, ook andere wetten zijn of worden dat en er is een aparte Databankenwet gekomen. Toch is het juridische antwoord op alle digitale uitdagingen zeker nog niet rond.

Toegankelijkheid via internet was een vereiste bij de ontwikkeling van het nieuwe DANS EASY systeem, waarin onderzoekers data kunnen plaatsen en raadplegen. Om copyright en eigendomsrechten te beschermen heeft DANS daarom nieuwe data-licenties ontwikkeld.

Uitgangspunt waren daarbij de principes van de Open Access beweging, die erop neerkomen dat onderzoeksdata én publicaties zoveel mogelijk zonder beperkingen via internet aan iedereen vrij beschikbaar worden gesteld. Toch kunnen er goede redenen zijn waarom data niet of niet onmiddellijk toegankelijk zijn. Privacygevoeligheid bijvoorbeeld, een tijdelijk embargo vanwege een komende promotie, of contractverplichtingen met financiers. De licenties kunnen daarmee rekening houden. Daarnaast gaan ze uit van de Nederlandse Gedragscode Wetenschapsbeoefening van de Ver-

eniging van Samenwerkende Nederlandse Universiteiten (VSNU), die vijf kernbegrippen hanteert: zorgvuldigheid, betrouwbaarheid, controleerbaarheid, onpartijdigheid en onafhankelijkheid.

De DANS-licenties zijn 'niet-exclusief': de eigenaar van de data blijft eigenaar en houdt alle vrijheid om ze ook elders beschikbaar te stellen. Data mogen niet commercieel hergebruikt worden, en een drietal toegangs categorieën geeft aan wie over de data mogen beschikken:

- Open access: onbeperkte toegang tot de data voor iedereen.
- Restricted access: toegang wordt voor maximaal twee jaar beperkt tot door de eigenaar aangewezen personen of instellingen.
- Other access: geen toegang via DANS EASY. Deze speciale categorie is bedoeld voor data die alleen bij DANS zijn gearchiveerd of die onder de Wet Bescherming Persoonsgegevens vallen.

Gegevensonderdiewet, waarmee vaak levende personen geïdentificeerd kunnen worden, mogen alleen onder strikte voorwaarden worden geraadpleegd en dan uitsluitend voor wetenschappelijk, statistisch of historisch onderzoek. DANS stelt hiervoor de Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek VSNU verplicht, die voor alle onderzoekers aan Nederlandse universiteiten en KNAW-instituten geldt en een klachten- en sanctieregeling bevat. (Heiko Tjalsma)



# Conferentie digitale duurzaamheid

Op 1 en 2 november organiseert de Koninklijke Bibliotheek (KB) Tools & Trends, een internationale conferentie over digitale duurzaamheid.

De conferentie wordt gehouden ter gelegenheid van het afscheid van Johan Steenbakkers, directeur e-Strategy en huisvestingsbeleid van de KB. Verscheidene internationale sprekers zullen de nieuwste ontwikkelingen op het gebied van digitale duurzaamheid en de laatste trends op het gebied van lange termijn archivering presenteren.

Op donderdag 1 november ligt de focus op de tools voor digitale duurzaamheid. Om de toekomstige toegang tot ons digitale erfgoed zeker te stellen, is lange termijn opslag slechts een eerste stap. Selectie, acquisitie, beschrijving,

karakterisering, datamanagement maar vooral ook de representatie in de toekomst zijn de andere onderdelen van het proces dat met de term digital resource management wordt aangeduid. Organisaties die verantwoordelijk zijn voor het duurzame beheer van digitale objecten, hebben daarvoor tools nodig. De internationale projecten waarin zulke tools worden ontwikkeld, waaronder PLANETS en CASPAR, zullen hun resultaten presenteren. In de verschillende sessies zal aandacht worden besteed aan karakterisering (identificatie, validatie) van digitale objecten, duurzaamheidsstrategieën

en preservation planning.

Vrijdag 2 november staat in het teken van trends op het gebied van digitale archivering. Archivering van verschillende soorten digitale objecten zal de revue passeren, waaronder publicaties, archiefbestanden, websites en wetenschappelijke data. Vervolgens zal aandacht worden besteed aan gemeenschappelijke problemen en oplossingen. Ten slotte wordt er vooruit gekeken: welke mogelijkheden zijn er voor samenwerking?

Meer informatie over de conferentie, het programma en de mogelijkheid tot online registratie is te vinden op [www.kb.nl](http://www.kb.nl). (MR)

## COLOFON

e-data@research is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen, verschijnend onder auspiciën van DANS, het Huygensinstituut, het Internationaal Instituut voor Sociale Geschiedenis 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: 7000. e-data@research is online te raadplegen op [www.edata.nl](http://www.edata.nl)

**Uitgever:** Edita-KNAW, Kloveniersburgwal 29, 1011 JV 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, Martijn de Groot (hoofd/eindredacteur), Marcel Ras, Jetske van der Schaaf, Luuk Schreven

**Aan dit nummer werkten mee:** Lex Borghans, Huibert Crijns, Katrien Depuydt, Jeffrey van der Hoeven, Tineke Koster, Hennie Leemkuil, Bart de Nil, Gert Nulens, Heiko Tjalsma, Evelien Walhout.

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

**Vormgeving en opmaak:** Ellen Bouma, Edita-KNAW  
**Productie en pre-press:** Edita-KNAW  
**Druk:** PlantijnCasparie, Almere

ISSN: 1872-0374

## Summer school in Ann Arbor

In het kader van mijn pas gestarte promotieonderzoek bij de leerstoel Cultuur in Brabant aan de Universiteit van Tilburg breng ik mijn zomer door in Ann Arbor, Michigan, USA. De stad Ann Arbor, zo'n veertig kilometer ten westen van Detroit nabij de grote meren, is volgens mijn reisgids a true regional research centre and a typical college town. En zo blijkt: hier is de grote Michigan State University gevestigd met ruim 46 duizend studenten, wat betekent dat ongeveer de helft van de bevolking van Ann Arbor student is.



**Evelien Walhout mailde haar ervaringen vanuit de ICPSR Summer School in Ann Arbor, plus met de mobiel genomen foto's, aan e-data**

Onderdeel van deze universiteit is het Inter-University Consortium for Political and Social Research (ICPSR) dat jaarlijks samen met onder meer het National Institute of Child Health and Human Development de summer school Longitudinal Analysis of Historical Demographic Data organiseert. De deelnemers aan deze cursus, voornamelijk PhD-studenten zoals ik, maar ook reeds gevestigde onderzoekers komen van over de hele wereld om hier te leren over de nieuwste onderzoeksmethoden op het gebied van de historische demografie. Deze zomer zijn er deelnemers uit onder meer India, Taiwan, Australië, Mexico, Noorwegen, Roemenië en de Verenigde Staten. Eén en ander levert tal van interessante onderzoeksprojecten op: van een vergelijking tussen structuren van huishoudens bij Laplanders aan de kust en in de binnenlanden van

negentiende-eeuws Noorwegen tot de invloed van veranderende politieke regimes op demografische processen in Oost-Europa, en de invloed van leefomstandigheden in de vroege



jeugd op het bereiken van zeer hoge leeftijden (85+ en 100+) in Canada. De colleges worden verzorgd door grote namen binnen de historische demografie waaronder George Alter (Indiana University), Myron Gutmann (University of Michigan) and Katherine Lynch (Carnegie Mellon University), wat de cursus op zichzelf al de moeite waard maakt.

Als ik dit schrijf is er één van de vier cursusweken voorbij. In deze week hebben we vooral kennisgemaakt met de grondbegrippen uit de historische demografie (gezinsreconstructie, sterftetafels) en de belangrijkste bronnen (bevolkingsregisters, volkstellingen) en hun gebreken (censoring). Daarnaast was er een introductie in het statistiekprogramma Stata met betrekking tot survival analyse en hebben we kennis mogen maken met de datasets die we komende weken zullen gebruiken voor analyse, te weten de Utah



Historical Database en de dataset met levenslopen uit het Belgische Sart.

In de komende tijd zullen onder meer de theorieën van Malthus en Hajnal de revue passeren en tal van modellen en analysetechnieken. Daarnaast zullen enkele gastsprekers

komen vertellen over hun nieuwste bevindingen en staan er ook nog enige sociale activiteiten op het programma waaronder een baseball wedstrijd in Toledo.

Het uiteindelijk doel van het ICPSR summer program is de posterpresentatie op de laatste cursusdag waarbij in groepjes van twee tot vier studenten antwoord wordt gegeven op onderzoeksvragen als waarom er in sommige gezinnen meerdere kinderen sterven en op de vraag of door sommige echtparen voorkeur wordt gegeven aan jongetjes ten nadele van meisjes. Dit soort vragen wordt beantwoord met behulp van bovengenoemde datasets en de aangereikte analysetechnieken. Enkele van deze posters zullen worden gepresenteerd tijdens een sessie van de Social Science History Association in november 2007 in Chicago.

Aan het eind van deze zomer zal iedereen hier zich weer op het eigen onderzoek gaan richten, gewapend met nieuwe kennis en – niet onbe-



langrijk – extra netwerk. Ikzelf zal me komende tijd bezig gaan houden met een onderzoek naar zuigelingen- en kindersterfte in de provincie Noord-Brabant in de periode 1880-1920. Aan de hand van data uit de Historische Steekproef Nederlandse bevolking (HSN) en aanvullend materiaal ga ik onder meer kijken naar verschillen tussen Brabant en de rest van Nederland en zoeken naar verklaringen voor de relatief ongunstige sterftecijfers voor zuigelingen eind negentiende eeuw in deze provincie. (Evelien Walhout) [www.icpsr.umich.edu/sumprog](http://www.icpsr.umich.edu/sumprog) [www.uvt.nl/faculteiten/fsw/onderzoek/bijzondereleerstoelen/brabant](http://www.uvt.nl/faculteiten/fsw/onderzoek/bijzondereleerstoelen/brabant) [www.iisg.nl/~hsn](http://www.iisg.nl/~hsn)

## Column

Lex Borghans

## Anders denken

*Het Sociaal Statistisch Bestand (SSB), de virtuele volkstelling van het CBS waarin informatie over de Nederlandse bevolking uit verschillende bronnen bijeen wordt gebracht, wordt langzaam realiteit voor onderzoekers. Deze data hebben een enorme potentie. Dit is echter niet vanzelfsprekend. Net zoals de eerste auto gebouwd werd als een koets, zo lijkt men nu vooral te kijken of administratieve data bestaande bronnen kunnen vervangen. Als je administratieve data echter vergelijkt met traditionele enquêtebestanden, dan heb je feitelijk veel minder informatie over veel meer personen. Vele honderden variabelen over misschien vijfduizend personen in bijvoorbeeld het Sociaal Economisch Panel (SEP) hebben plaatsgemaakt voor een handjevol variabelen over 17 miljoen mensen. Ieder statistiekboek laat echter zien dat je met informatie over vijfduizend personen vrij nauwkeurig uitspraken kunt doen over 17 miljoen mensen. Er lijkt dus sprake te zijn van een verlies.*



Dit betekent in de eerste plaats dat administratieve data de traditionele enquêtes niet kunnen vervangen. Maar met administratieve bestanden – zeker in de zeer doordachte opzet van het CBS – kunnen ook dingen die met de traditionele bestanden niet konden:

Allereerst kunnen we kijken naar zeer specifieke groepen zonder dat de steekproef te klein hoeft te worden. Als we de kappers uit het SEP nemen is er bijna geen steekproef meer over. In het SSB zitten er vele tienduizenden. Specifieke groepen kunnen dus zonder statistische problemen adequaat geanalyseerd worden.

Ten tweede kunnen verbanden tussen mensen veel beter worden onderzocht. Voor iedere observatie in het bestand is ook een observatie van zijn collega, buurman en broer of zus aanwezig. Onderlinge invloeden tussen mensen kunnen dus veel beter worden geanalyseerd en vergelijkingen veel beter worden gemaakt.

Ten derde is het probleem van uitval verdwenen. Bij een panel zijn we vaak na enkele jaren al een behoorlijk aantal mensen kwijt. Met administratieve data is er alleen sprake van uitval als mensen sterven of het land verlaten. Dit betekent dat we zeker op den duur werkelijke zeer-lange-termijn-effecten kunnen gaan bestuderen.

Door op een slimme manier gebruik te maken van deze mogelijkheden kunnen belangrijke maatschappelijke vragen onderzocht worden die vroeger onbeantwoordbaar leken. Om deze nieuwe mogelijkheden werkelijk te benutten moeten we op een andere manier gaan denken over de vraag welke informatie belangrijk is: minder clustering en meer detail in classificaties, gedetailleerde informatie over relaties tussen mensen en gegevens over omstandigheden die van invloed kunnen zijn op het latere functioneren van mensen. Omdat de potentiële waarde van informatie enorm is gestegen moeten we niet te zuinig zijn, eerder meer dan minder aanvullende informatie verzamelen en wat verzameld wordt zeker niet van het toeval af laten hangen. Vooral als we meer naar langetermijnverbanden willen kijken, bepalen de keuzes die nu gemaakt worden de onderzoeksinformatie die de komende decennia beschikbaar zal zijn.

Lex Borghans is hoogleraar arbeidseconomie en sociaal beleid aan de Universiteit Maastricht.

## Gelezen

Els Stronks, Peter Boot (eds.), assisted by Dagmar Stiebrl, *Learned love, Proceedings of the Emblem Project Utrecht Conference on Dutch Love Emblems and the Internet (November 2006)*. Edita-KNAW, Amsterdam, 2007; *DANS Symposium Publications 1*. ISBN 978 90 6984 510 4. € 20,00

Selected papers of a conference, held on November 6 and 7 2006, with focus on the Dutch love emblem and the process of digitising, complemented by a general introduction and additional papers by the editors. The funny, mysterious, moralizing, learned or pious combinations of word and image in the crossover emblematic genre form a delicate network of motifs and mottoes. This network served as a receptacle for the traditions of the European visual and literary arts. In its turn it influenced architecture, painting, love

poetry, children's books, preaching and interior decoration. The love emblem is one of the best known emblematic subgenres, first developed in the Low Countries and soon became popular all over Europe. In 2003 the Emblem Project Utrecht set out to digitise 25 representative Dutch love emblems books. The conference was held to celebrate the completion of that work.



UK Data Archive, *Across the decades, 40 years of data archives*, Colchester, University of Essex, 2007.

Een brochure over de ontstaansgeschiedenis van het Engelse sociaal-wetenschappelijk data-archief, de noodzaak van digitale duurzaamheid en de uitdaging om nieuwe tools te blijven ontwikkelen voor een continu veranderend datalandschap. Kosteloos downloaden: [www.data-archive.ac.uk/ukda40](http://www.data-archive.ac.uk/ukda40)

Maarten Evenblij, *Respect!, Onderzoek naar sociale cohesie in Nederland*, Aksant, Amsterdam, 2007. ISBN 978905260257 8. € 15,00

Uitkomsten van het grote NWO-onderzoeksprogramma 'Sociale Cohesie', dat in 2006 werd afgerond. Hoe kunnen mensen zich thuis voelen in de samenleving en wat bevordert of remt de binding van individuen of groepen met de maatschappij? Die groepen kunnen vrouwen, homoseks-

suelen en intellectuelen zijn, maar ook streng gelovigen, laagopgeleiden en lichamelijk gehandicapten. Sinds 1999 hebben tientallen onderzoekers zich over die vragen gebogen. Voor zowel de meer ingewijde lezer als de relatieve leek bevat *Respect!* een aantal eye-openers en conclusies om over na te denken. Wetenschappelijke conclusies die de afgelopen jaren ongehoord zijn gebleven en – op een rijtje gezet in *Respect!* – een nieuwe kans verdienen.

M.P.J.P. Vanderfeesten, G. van Westrienen. *A portal for doctoral E-theses in Europe. Lessons learned from a demonstrator project*. SURF-

*foundation, Utrecht, 2007.*

Uitkomsten van het gelijknamige pilotproject van SURFfoundation, het Britse Joint Information Systems Committee (JISC), en de Nationale Bibliotheek van Zweden. Praktische aanbevelingen voor het verbeteren van de interoperabiliteit tussen de service provider en degene die de gegevens beschikbaar stelt. Vooral nuttig voor instellingen die de resultaten van hun onderzoek beschikbaar willen maken voor de buitenwereld. Door hun materiaal op een gestandaardiseerde manier toegankelijk te maken kunnen ze landsgrenzen overschrijden. [www.surffoundation.nl](http://www.surffoundation.nl)