

**Pagina 3 • Verrijkte publicaties •** SURFfoundation deed proefprojecten met verrijkte publicaties en laboratoria. De betrokken onderzoekers waren er goed over te spreken.

**Pagina 4 • Neurale netwerken •** Zoeken zonder de voor de hand liggende trefwoorden. Daarvoor gebruikt een Utrechtse zoekmachine netwerken van associaties, die zijn geïnspireerd op resultaten van herenonderzoek.



FREDDY RIKKEN

**Pagina 6 • DBNL •** Eind 2010 biedt de digitale bibliotheek net zo veel pagina's in de vorm van afbeeldingen en OCR als in de vorm van overgetypte teksten, voorziet hoofdredacteur René van Stipriaan.

**Pagina 6 • Aletta •** De rubriek Focus stelt scherp op Aletta, instituut voor vrouwengeschiedenis.



**Pagina 8 • Robbert Dijkgraaf •** Met de uitreiking van de DANS Dataprijs voor de deur reflecteert de KNAW-president over het belang van data delen.

## EN VERDER

Agenda .....	2
Nieuws .....	3
Achtergrond .....	4,6
Focus .....	6
Sinds kort beschikbaar .....	7
Column .....	8
Gelezen .....	8

## Aanloop voor Grote Sprong in humaniora

Verschillende instituten lopen zich warm voor de sprong voorwaarts die de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) in de geesteswetenschappen tot stand wil brengen. De voorstellen die worden voorbereid draaien om ict-toepassingen.

De KNAW gaat de komende vijf jaar 3,5 miljoen euro uittrekken om de geesteswetenschappen een impuls te geven in de richting van *computational humanities*. De toepassing van ict nam in die wetenschappen nog niet de hoge vlucht die in andere disciplines wel plaatsvond, constateerde de Akademie vorig jaar. Onderzoeksdirecteur Theo Mulder sprak van een 'grote sprong voorwaarts' die mogelijk zou worden bij een sterker beroep op 'de technologie van de natuur- en levenswetenschappen'. In februari van dit jaar werd bij het NIAS in Den Haag de laatste van twee workshops gehouden met vertegenwoordigers van de betrokken instituten, de Akademie en enkele externe deskundigen om de contouren uit te zetten van het gewenste programma. In elk geval moeten de eigen collecties van KNAW-instituten ten volle worden ingezet, werd daar geconcludeerd. Dat staat intussen ook in een recente

nota van de hand van Mulder, die het programma verder uitwerkt.

De computationele component zou uit zulke verschillende onderdelen kunnen bestaan als automatischedataverrijkingen met datatechnieken, visualiseringstechnieken, algoritmes voor het opzoeken van concepten en voor *machine-learning*, geografische informatiesystemen en meer. Ook zou de toepassing van collaboraties voor annotatie en van andere web 2.0-elementen moeten worden bevorderd. Voor de infrastructuur zou een beroep moeten worden gedaan op bestaande of in oprichting verkerende partners als CLARIN, Alfalab en DANS.

Kort na de workshop verspreidde de KNAW een *Call for Proposals*, die inmiddels al een veertiental voorlopige plannen en voorstellen heeft opgeleverd waarin de namen van Meertens Instituut, Huygens Instituut, Fryske Akademy, NIOD en KITLV (studies ex-koloniën) vaak voorkomen. Ook de Virtual Knowledge Studio tekent voor enkele voorstellen. Patronen en motieven in culturele uitingen zijn een gemeenschappelijk element. Dat is niet vreemd omdat de *Call for Proposals* daarom vraagt: juist die lenen zich goed om met behulp van ict boven water te worden gehaald. Onderwerpen variëren van de verspreiding van gebaren in de Europese schilderkunst tot de maatschappelijke rol van populaire muziek in Zuid-Oost Azië en van het herkennen van literaire kwaliteit tot visuele navigatie in internetbronnen als Wikipedia.

DANS (Data Archiving and Networked Services) bereidt een voorstel voor om archiefdata op het GRID toegankelijk te maken voor programma's, zodat het voor alle projecten mogelijk wordt om de te verwachten grote hoeveelheden data aan intensieve computerbewerkingen te onderwerpen. Bij GRID computing wordt de rekenkracht van een groot aantal computers op verschillende plaatsen gecombineerd in een snel netwerk. 'Op die manier wordt ook infrastructurele synergie uit het programma gehaald,' aldus indiener Dirk Roorda van DANS. 'De methoden die tot gepubliceerde resultaten hebben geleid, zijn dan ook in praktische zin herhaalbaar.'

Voorstellen kunnen nog tot 1 oktober worden ingediend. De KNAW wil in het najaar een definitieve keus

maken uit de ingediende voorstellen, zodat een vijftal daarvan in februari 2011 van start kan. (MdG)

## Dataprijs in wording



ILONKA ILIEVA

Terwijl de driekoppige jury zich nog beraadt over de DANS Dataprijs, legt kunstenaar Tjavar Iliev in zijn atelier de laatste hand aan de polyster mal voor de bronzen sculptuur die de uiteindelijke winnaar op 18 juni in ontvangst mag nemen (zie ook pagina 3).

Iliev (1950) volgde een opleiding als beeldend kunstenaar in zijn geboorteland Bulgarije. Hij vestigde zich in 1990 in Amersfoort, waar hij sindsdien naast vrij werk ook bronzen sculpturen maakt in opdracht. Tussen het moment van de foto en de 18<sup>e</sup> juni wordt van het model een mal gemaakt, een soort driedimensionaal negatief, die op zijn beurt wordt volgegoten met was. Daarna gaat het wasmodel naar de bronsgieterij waar het met keramische lagen wordt bedekt. Het keramiek wordt verhit, de was verwijderd, het brons in de keramische mal gegoten en na afkoeling en bevrijding uit de mal gepateneerd (op kleur gebracht) en op een sokkel gezet.



Het gebouw van het bureau van de KNAW in Amsterdam

## NWO-steun voor grote digitale projecten

De Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) heeft eind mei 1,4 miljoen euro toegewezen aan het project CLIO-INFRA, dat op wereldschaal kwantitatieve data bijeen moet brengen en beschikbaar maken over economische ontwikkelingen in de afgelopen vijfhonderd jaar. Die informatie moet helpen de vraag te beantwoorden waarom sommige landen rijk zijn en andere arm. In CLIO-INFRA wordt samengewerkt door het Internationaal Instituut voor Sociale Geschiedenis, DANS en de universiteiten

van Utrecht, Groningen, Tübingen en Debrecen.

Aan het Taalportaal stelde NWO 1,7 miljoen euro ter beschikking om een virtueel instituut te ontwikkelen dat informatie over de grammatica van het Nederlands en het Fries (klankleer, vormleer en zinsleer) bij elkaar brengt en beschikbaar maakt in beide talen plus het Engels. Het is een project van het Meertens Instituut, de Fryske Akademy, het Instituut voor Nederlandse Lexicologie en de universiteit van Leiden. (MdG)

## Politicooloog Aarts: 'Parlementair systeem heeft schokdempers nodig'



ROBERT HOETINK

Kees Aarts

'Het is nu relatief gemakkelijk om een partij op te richten en een zetel in de Tweede Kamer te krijgen. We hebben te weinig schokdempers in het systeem'. Dat zegt politicooloog prof.

Kees Aarts in een interview op pagina 5 van deze *edata@research*. Aarts presenteerde 21 mei in Den Haag samen met de co-auteurs de datagids *Dutch Parliamentary Election Studies* waarin de documentatie is samengebracht van een geïntegreerde dataset van het Nationaal Kiezersonderzoek tussen 1971 en 2006 (zie ook *Gelezen*, p.8). Staatssecretaris Bijleveld van Binnenlandse Zaken nam op een minisymposium over de verkiezingen in Perscentrum Nieuwspoor het eerste exemplaar in ontvangst.

## AGENDA

10 en 11 juni

Athens, Athena Research Centre  
*Scholarly information practices and requirements for arts and humanities digital infrastructures*

Expert Forum initiated by the Digital Curation Unit (DCU). Receiving feedback, nurturing critical discussion and reaching consensus on the empirical study and conceptual modelling work on scholarly information practices and requirements for scholarly e-infrastructure conducted by the DCU as part of DARIAH.

[www.dcu.gr](http://www.dcu.gr)

11 juni

Amsterdam, Historisch Museum  
*CATCH meeting – AGORA – Creating the historic fabric for and providing web-enabled access to objects in dynamic historical sequences*

Agora, a project of the Free University, the Institute for Sound and Vision and the Rijksmuseum, facilitates and investigates the access that both last-mentioned institutes offer to sources from the past. This should lead to a web-based interactive space for the interpretation of historical sources for both heritage professionals and interested citizens.

[www.nwo.nl/catch](http://www.nwo.nl/catch)

24 juni

Leiden, Museum Naturalis  
*Registers in sociaalwetenschappelijk onderzoek: mogelijkheden en valkuilen*  
Conferentie van CBS en VU Sociale Wetenschappen over het gebruik van registers. Economen over arbeidsproductiviteit bij ouder worden, sociologen over arbeidsmobiliteit, medici over het gebruik van registraties voor medisch onderzoek en criminologen over onderzoek naar criminaliteit. [www.fsw.vu.nl/](http://www.fsw.vu.nl/)

14 juni

Kerkrade, Rolduc Conferentiecentrum  
*Pre-conferentie Archief 2.0: Niet voor lurkers*  
Conferentie georganiseerd door Archief 2.0 over de behoefte aan een online platform voor archieven. [www.archief20.org/events](http://www.archief20.org/events)

7-10 juli

London, King's College  
*Digital humanities 2010, Cultural expression, old and new*  
This annual international conference for digital scholarship in the humanities will be hosted by the Centre for Computing in the Humanities and the Centre for e-Research, with the support of the School of Arts and Humanities, Information Services and Systems. <http://dh2010.cch.kcl.ac.uk/>

27 juni-3 juli

Bristol, University of Bristol  
*Summer School on Analysing the life course – Advanced methods for the analysis of complex event history data*  
Covering advanced topics of life courses, with a focus on methods for analysing recurrent events, transitions between states and interdependent events. Both descriptive techniques for exploring complex histories (sequence analysis) and statistical modelling approaches will be considered. [www.ccsr.ac.uk/](http://www.ccsr.ac.uk/)

6-9 juli

Madrid, Palacio de Congressos  
*5<sup>th</sup> International Conference on Open Repositories*  
Conference aiming to bring together individuals and organisations responsible for the conception, development, implementation and management of digital repositories, as well as stakeholders interacting with them, to address theoretical, practical and strategic issues.

13-20 augustus

Lausanne, University of Lausanne  
*Summer School on survey design and quality*  
Starting with an introduction to survey quality and the survey production process, the course will cover the challenges to quality of surveys designed to provide multi-national, multi-cultural or multi-lingual comparisons. Wherever possible practical exercises will be provided. [www.ccsr.ac.uk/qmss/summer/Switzerland](http://www.ccsr.ac.uk/qmss/summer/Switzerland)Digital Archiving in Genève:  
buigen voor Google

JULIA ROMIJN-WIXLEY

## Plenaire bijeenkomst in het International Conference Centre in Genève

Vier thema's kwamen aan de orde op de achtste European Conference on Digital Archiving in Genève: 'Professionele competenties in het digitale tijdperk', 'Wat te bewaren: hoe behouden we een staalkaart van de informatiemaatschappij?', 'E-archiving: reorganiseren van processen en bedrijfsmodellen', en 'Online toegankelijkheid: oplossingen en gevolgen'.

De openingspreker op deze conferentie, die werd gehouden van 28 tot en met 30 april in Genève, was onze eigen *éminence grise* Eric Keteelaar (Universiteit van Amsterdam) die sprak over *The inclusive archivist*: als een manusje-van-alles of 'universele mens' uit de renaissance is hij of zij overal bij betrokken, moet proactief handelen vanaf de creatie van documenten, moet bruggen bouwen tussen organisaties en gebrui-

kers, tussen de verschillende beroepen en alle belanghebbenden.

Daniel Pitti (University of Virginia) sprak over de recente XML-norm voor *Encoded Archival Context* (EAC) die informatie over organisaties, personen, en families die in 2010 is vastgesteld. Met de combinatie van EAD en EAC kunnen belangrijke organisaties, personen, of families (de archiefvormers) een eigen beschrijving krijgen, die kan worden gekoppeld aan alle voorkomens van die namen.De volgende dag begon met een plenaire lezing van Jason Baron (NARA) over e-discovery. Hij begon met een filmpje ([www.youtube.com/watch?v=bWBJWcsPpIM](http://www.youtube.com/watch?v=bWBJWcsPpIM)) dat de urgentie van het e-discovery probleem overduidelijk op de kaart zet: we worden nu al overspoeld met informatie die allemaal potentieel

bewijs kan zijn, en met traditionele methoden zijn we niet in staat om dit te verwerken.

De zeer interessante plenaire lezing van Steve Bailey (JISC) stelde de vraag in welke handen de toekomst van digitale archivering ligt, met als eenduidig antwoord: Google! Dit bedrijf biedt alle vormen van opslag aan en is zo groot dat de kans op een faillissement of overname zeer klein is. Als deze conclusie misschien onwenselijk maar niettemin onafwendbaar lijkt, zouden record managers hun beleid hierop moeten aanpassen. De dag werd afgesloten met een uitstekend galadiner met cultureel verantwoord geïndel – wel erg veel cultureel verantwoord geïndel. (Jaap Kamps)

[www.bar.admin.ch/eca2010/](http://www.bar.admin.ch/eca2010/)

## Kritiek rondom Wikipedia

'We hebben expres niet de Wikimedia Foundation Nederland uitgenodigd,' merkte Geert Lovink op in het openingswoord van de Critical Point of View meeting. Het was de tweede bijeenkomst van het Wikipedia Research Initiative, georganiseerd door het Institute of Network Cultures. De meeting vond plaats op 26 en 27 maart in de Openbare Bibliotheek van Amsterdam.

De toon was gezet voor een kritische benadering van de beroemde Web 2.0 encyclopedie Wikipedia, de zevende meest populaire website ter wereld. In de eerste van de zes sessies, gewijd aan Wikitheory, figureerden Foucault in de discursieve structuur van Wikipedia (Reichert), De Santos Sousa's blinde vlekken tussen regulering/order en emancipatie/ervaring (Hofmann) en wiki-lawyers (O'Neill). De tweede sessie, *Encyclopedia Histories*, was gewijd aan de vraag wat geleerd kan worden van historische precedenten van (collectieve) encyclopedieën zoals die van D'Alembert, Otlet e.a. (Reagle, van den Heuvel, Sullivan) en van historische kritiek zoals Flaubert's satire op woordenboeken en encyclopedieën. De derde

sessie, WikiArt, bevatte verbazingwekkend weinig beeldmateriaal en richtte zich meer op spanningen tussen artiesten en gevestigde orde van de Wikimedia Foundation. Meest relevant voor onderzoek naar data was de sessie Wikipedia Analytics. Ortega waarschuwde voor misbruik van metrisch onderzoek en besprak de evolutie van Wikipedia als dataset, Varghese Mathews belichtte de problemen van automatisering in content in Wikipedia pagina's en Stuart Geiger de macht van robots (BOTS politics): 'Natuurlijk heeft u de vrijheid om niet te tekenen, zoals de robot de vrijheid heeft namens u te signeren.' De sessie Designing Debate richtte zich op de vraag hoe controverse standpunten, bijvoorbeeld in het lemma Kosovo oorlog, zicht-



baar gemaakt kunnen worden binnen het 'neutral point of view'-beleid van Wikipedia. De zesde en laatste sessie Global Issues en Outlook hekelde de westerse oriënta-

Spaarzaam gebruik  
paradata goed  
voor veldwerk

Paradata zijn nuttig voor het monitoren en verbeteren van het veldwerkproces. Het verzamelen ervan kan echter belastend zijn voor de interviewers en zou alleen gedaan moeten worden wanneer er ook echt gebruik van wordt gemaakt. Dat bleek op een bijeenkomst van het Nederlandstalig Platform voor Survey Onderzoek (NPSO), op 11 maart in Brussel.

Paradata (ofwel procesdata) zijn data over het veldwerkproces, zoals het aantal contactpogingen bij mondelinge enquêtes of het invulgedrag bij webenquêtes. Ann Carton van de studiedienst van de Vlaamse regering leidde de bijeenkomst in met een overzicht van verschillende typen paradata en hun gebruiksmogelijkheden voor het verbeteren van survey-onderzoek. Ger Snijkers (CBS) gaf een voorbeeld van de analyse van toetsaanslagen bij het invullen van vragenlijsten (audit trails). Bedrijven blijken vragenlijsten in te vullen als het hen uitkomt, waarschijnlijk dus wanneer de gegevens beschikbaar zijn, en als ze gerappeld worden. Met audit trail data zou men verschillende invulprofielen kunnen onderscheiden en vragenlijsten kunnen verbeteren. Luc Schulpen van TNS Dimarso ging in op het gebruik van paradata in de praktijk, voor het monitoren van de voortgang en de efficiency van het dataverzamelingsproces (contactgegevens, financiële informatie). Als laatste vertelde Hideko Matsuo van de K.U. Leuven over het gebruik van non-respons surveys in de European Social Survey (ESS). Informatie van non-respondenten die een korte vragenlijst hebben beantwoord met kernvragen uit het onderzoek geeft inzicht in non-respons bias. In de afsluitende discussie bleek dat het laatste woord over het privacy-aspect bij het verzamelen van paradata in webenquêtes nog niet is gesproken. (Fannie Cobben)

tie van Wikipedia. Zoals elke andere encyclopedie laat Wikipedia weinig ruimte voor kritiek. Discussies wie het laatste woord heeft in de editing van lemmas: de Wikipediagemeenschap (Wikipedians) of de Wikimedia Foundation met de beschimpde supergod Jimmy Wales nemen niet zelden kinderachtige vormen aan. Echter juist omdat de inhoud en beleidsbepalingen van Wikipedia zo weinig ruimte bieden maakte de conferentie in Amsterdam onverwacht veel interessante kritische gezichtpunten niet in maar rondom Wikipedia zichtbaar. (Charles van den Heuvel)

# Vier nominaties voor Dataprijs

In de tweede week van mei heeft de jury van de DANS Dataprijs uit een achttiental inzendingen vier kandidaten geselecteerd. Aan een van deze vier wordt op 18 juni in Amsterdam de prijs uitgereikt door KNAW-president Robbert Dijkgraaf.

De genomineerde dataprojecten zijn van politologische, historische, sociologische en demografische komaf. Bij de 'Dutch Parliamentary Election Study' (DPES) gaat het om een cumulatieve dataset van verkiezingsonderzoeken over de jaren 1971-2006. De 'Historische Steefproef Nederland' (HSN) is een representatieve steekproef van bijna tachtig duizend personen die tussen 1812 en 1922 in Nederland zijn geboren. Het LISS-panel, eveneens genomineerd, omvat vijfduizend huishoudens met in totaal achtduizend personen, die maandelijks worden ondervraagd. De Netherlands Kinship Panel Study (NKPS) ten slotte biedt de gegevens van een grootschalig 'multi-actor' en 'multi-method' onderzoek naar familiebanden in Nederland.

De jury bestaat uit voorzitter Pieter Hooimeijer, hoogleraar Sociale Geografie en Demografie in Utrecht, Jan Karel Koppen, directeur beleidsontwikkeling en -ondersteuning bij NWO en Wiljan van den Akker, decaan van de Utrechtse faculteit Geesteswetenschappen. Jurysecretaris Heiko Tjalsma: 'Bij de keuze voor de nominaties heeft de jury alle cri-

teria meegewogen die voor de prijs gelden. Zo is er gekeken naar de innovatieve mogelijkheden en naar de vraag of er veel onderzoekers profijt van kunnen hebben. Maar ook de

kwaliteit van de documentatie is beoordeeld, en de mate waarin de wijze van archiveren en beschikbaar stellen innovatief zijn. Deze vier scoorden daarop het hoogst. Welke van de

vier op zijn beurt boven de andere uit stak maken we pas op de 18e juni bekend'.

Van de achttien inzendingen voor de DANS Dataprijs kwamen er vijf uit de historische hoek, vier uit de sociale wetenschappen en vier uit de archeologie. Ook waren er drie taalkundige inzendingen, een politologische en een demografische. (MdG)

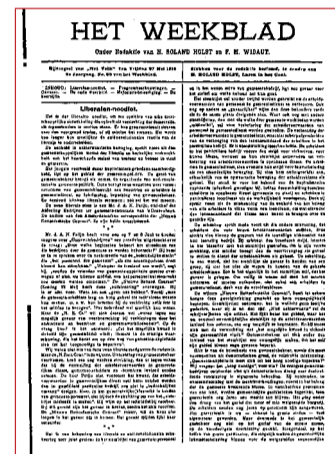
## Krantenwebsite vernieuwd

Op 27 mei heeft de Koninklijke Bibliotheek haar vernieuwde krantenwebsite [kranten.kb.nl](http://kranten.kb.nl) gelanceerd met een symposium over de waarde van kranten voor historisch onderzoek en online databanken.

De nieuwe website biedt om te beginnen een miljoen digitale krantenpagina's uit de periode 1618 - 1940, die beschikbaar zijn gekomen als resultaat van het project Databank Digitale Dagbladen. Tot eind volgend jaar, wanneer het project wordt afgerond, wordt het aanbod uitgebreid tot zo'n acht miljoen pagina's waaronder ook materiaal uit de na-oorlogse pe-

riode tot 1995. Zoeken in de databank is mogelijk op woorden in de tekst. De zoekresultaten worden gepresenteerd op artikelniveau. Er kan verfijnd worden gezocht op periode, plaats van uitgave, verspreidingsgebied (koloniaal, nationaal, lokaal/regionaal) en soort artikel (algemeen bericht, advertentie, familiebericht, afbeelding met onderschrift). Tot aan de vernieuwing bood de site zo'n 350 duizend pagina's van een viertal dagbladen in de periode 1910-1945: Het Centrum, NRC, het Vaderland en Het Volk. In de vernieuwde website kunnen naast deze dagbladen ook

nog zo'n zeventig andere titels worden doorzocht. (MdG)



## Tevredenheid over proef verrijkte publicaties

Onderzoekers die betrokken waren bij een SURFshare-project over verrijkte publicaties (meta-koppelingen naar onderliggende data, modellen

en formules) en collaboratories denken dat hun onderzoek beter wordt door het gebruik van zulke kanalen. Dat maakte SURFfoundation bekend bij de afronding van het project eind maart.

Onderzoek wordt toegankelijker en de resultaten kunnen beter worden hergebruikt en breder worden toegepast, aldus de betrokken onderzoekers volgens de universitaire ict-samenwerkingsorganisatie. Ook het werken met collaboratories, virtuele samenwerkingsomgevingen, viel bij de onderzoekers in goede aarde. 'Door online samen te werken kunnen onderzoekers bronnen delen, samen schrijven aan resultaten en deze zichtbaar maken voor de buitenwereld,' aldus SURFfoundation.

Basis voor deze conclusies was een achttal innovatieprojecten, waarin werd geëxperimenteerd met Verrijkte Publicaties en collaboratories. Daarbij werkten onderzoekers mee uit uiteenlopende disciplines zoals psychologen, juristen, natuurkundigen maar ook archeologen en sociale wetenschappers.

Er was wel enige weerstand te overwinnen, volgens SURFfoundation, maar toch waardeerden de deelnemers uiteindelijk de meerwaarde. Vooral de koppeling tussen een publicatie en onderliggende objecten werd hoog gewaardeerd. Het gaat hierbij niet om hyperlinks in een pdf, maar om koppelingen die ook

kunnen worden bewaard en voor computers begrijpelijk zijn.

Het koppelen van publicaties aan onderzoeksdata bleek wel tijdrovend omdat het vaak handmatig moet. Ook moeten er volgens het eindrapport heldere afspraken worden gemaakt over de relaties tussen de objecten zodat hergebruik ook werkelijk mogelijk is.

Ondanks het feit dat het nog om proefprojecten ging concludeert het rapport dat Nederland internationaal een vooruitgeschoven positie inneemt met verrijkte publicaties. Dat komt vooral doordat ons land beschikt over een goede infrastructuur van repositories, aldus SURFfoundation.

Op het gebied van verrijkte publicaties werden projecten uitgevoerd rond een e-Journal voor archeologen (JALC), een repository voor verrijkte publicaties over surveyonderzoek (DataPlus), betere communicatie door verrijkte publicaties (ESCAPE), verrijkte proefschriften (Proefschriften Plus) en kwalitatieve onderzoekbestanden (Veteran Tapes). (MdG)



## Standaardisatie persistent identifiers

Eind dit jaar zal er een centrale Europese website zijn voor persistent identifiers zoals ook DANS (Data Archiving and Networked Services) deze uit geeft ([urn:nbn](http://urn:nbn)). Gebruikers vinden hier informatie over de identifiers en kunnen ze vertalen in de actuele locatie van datasets of publicaties. De website is het resultaat van het project 'PersID' van acht Europese partners, waaronder uit Nederland SURFfoundation en DANS.

Persistent identifiers zorgen er voor dat bronnen in archieven duurzaam kunnen worden gerefereerd en gekoppeld. Omdat er meer systemen bestaan die dit op verschillende manieren aanbieden zijn de mogelijkheden en het gebruik nog beperkt. PersID gaat hier verandering in brengen door de gezamenlijke ontwikkeling van beleid, diensten en informatievoorziening.

## Oorlogsliefdekind

Op 16 mei begon bij de VPRO een reeks uitzendingen over het project 'Oorlogsliefdekind', waarin de verhalen zijn achterhaald van Nederlandse militairen en hun in 1949 achtergelaten kinderen en hun moeders in Indonesië. Het afgelopen jaar zijn de nooit eerder vertelde ervaringen van 'soldatenvaders' en 'soldatenkinderen' op de website [www.oorlogsliefdekind.nl](http://www.oorlogsliefdekind.nl) verzameld. Op 16 mei was er een live-uitzending van het programma *Onvoltooid Verleden Tijd* (OVT). Donderdag 1 juli wordt de tachtig minuten durende tv-documentaire *Tuan Papa* uitgezonden in het programma *Andere Tijden*. De interviews en allerlei verwante documenten van het project zijn gearchiveerd bij DANS.

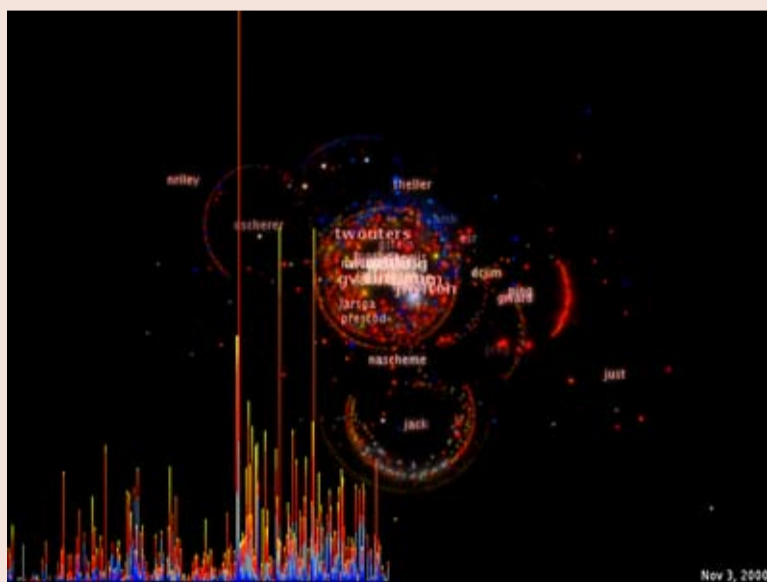
## Nieuwe site archeologen

De website van het e-depot Nederlandse archeologie (EDNA, [www.edna.nl](http://www.edna.nl)) heeft een grondige facelift ondergaan. De site biedt meer achtergrondinformatie over het e-depot, waaronder een sterk uitgebreide projectgeschiedenis en FAQ's: veel gestelde vragen. Het zoeken in het eigenlijke data-archief, dat onderdeel is van het doe-het-zelf-archief EASY van DANS (Data Archiving and Networked Services), is vanuit de website nog makkelijker gemaakt. De vormgeving van de site reflecteert nu ook de band met het data-instituut. [www.edna.nl/datasetsearch.html](http://www.edna.nl/datasetsearch.html)

## Connections connected

Bij DANS wordt gewerkt aan een database waarin een aantal verschillende databestanden over de Nederlandse handel in het Atlantisch gebied in de 17e en 18e eeuw wordt geïntegreerd. Deze bestanden, in de afgelopen twintig jaar door verschillende onderzoekers gevormd, zijn nu al afzonderlijk te raadplegen bij DANS in het kader van het project 'Dutch Atlantic connections: the circulation of people, goods and ideas in the Atlantic world': <http://easy.dans.knaw.nl> (zoek op 'Dutch Atlantic Connections'). De nieuwe gezamenlijke database, in de toekomst daar ook te vinden, zal de gebruiksvriendelijkheid voor onderzoekers verhogen.

## De groei van open source projecten



Michael Ogawa wilde visualiseren hoe open source software projecten zich ontwikkelen, en ontwikkelde daarvoor *code\_swarm*: een hulpmiddel om video's te maken die een beeld geven van de historische ontwikkeling van een project. De video's zijn gebaseerd op aantallen *commits*: het (terug)plaatsen van een programma-onderdeel in de gedeelde bibliotheek van programmacomponenten. Projecten houden bij wie welke programma's wanneer heeft gecommiteerd, en die gegevens kun je visualiseren. De afbeelding is

afkomstig uit een video die de ontwikkeling van de scripttaal Python laat zien. De video toont hoe de eerste jaren alleen Guido van Rossum, de (Nederlandse) bedenker van Python, actief is. Langzaam sluiten andere ontwikkelaars zich aan. De meesten dragen code bij (rood/geel), anderen documentatie (blauw). Sommigen hebben hun eigen deelproject, duidelijk onderscheiden van de kern van het project. Het getoonde moment is november 2000, en Python is dan inmiddels een populaire taal met een grote groep ontwikkelaars. (PB)

## Associatieve neurale netwerken voor de humaniora

Via associaties zoeken naar relevante tekstpassages waarin de meest voor de hand liggende trefwoorden niet voorkomen. Daaraan werkt de Utrechtse vakgroep Geschiedenis van de filosofie met een zoekmachine op basis van neurale netwerken.

Als wij complexe teksten lezen, doen we dat associatief: we willen bijvoorbeeld weten hoe Kant over vrijheid denkt, of op welke manier Descartes reageert op de door Hobbes geformuleerde kritiek op zijn werk. Hiervoor kan het belangrijk zijn om alle vindplaatsen te kunnen zien waar bij Kant het woord 'vrijheid' voorkomt, of waar de termen uit Hobbes' kritiek bij Descartes te vinden zijn. Maar niet alle vindplaatsen zullen relevant zijn, en het is denkbaar dat Kant inhoudelijk relevante dingen over vrijheid te zeggen heeft zonder het woord zelf, of daarvan direct afgeleide woorden, te gebruiken. Deze inhoudelijk relevante passages waarin het trefwoord zelf niet voorkomt zijn alleen *associatief*, op basis van semantische criteria te vinden.

Dit idee ligt ten grondslag aan een zoekmachine die gebaseerd is op *associatieve neurale netwerken voor de humaniora (ANNH)*, en die door de vakgroep Geschiedenis van de Filosofie aan de Universiteit Utrecht wordt toegepast op wijsgerig-historisch onderzoek. In die netwerken interageren neuronen met elkaar op

een manier die ontleend is aan het hersenonderzoek. In de hersenen zijn zenuwcellen - neuronen - met elkaar verbonden via synapsen. Op dezelfde manier zijn de onderdelen van een tekst met elkaar verbonden via de woorden in die teksten. Semantisch informatieve verbindingen zijn sterker dan andere, net zoals in de hersenen; op basis van de aan elk element van een tekst toegeschreven informatiewaarde, kunnen associaties tussen de tekstelementen worden berekend die voor inhoudelijk relevante resultaten zorgen.

Een belangrijk verschil met bestaande systemen om grote hoeveelheden tekst te analyseren is dat dit systeem geen informatie vooraf nodig heeft over de te doorzoeken teksten. Er ligt geen woordenlijst, thesaurus of expertsysteem aan ten grondslag. Het systeem kent zelf een relatieve informatiewaarde toe aan elk woord en maakt op basis daarvan associaties, afhankelijk van de teksten die doorzocht worden. Een zoekopdracht binnen één werk van een auteur levert dus heel andere associaties op dan diezelfde zoek-

opdracht binnen zijn gehele oeuvre. Het is ook denkbaar dat het systeem passages als resultaat levert waarin de oorspronkelijke zoekterm niet voorkomt, maar die toch op een semantisch niveau met de zoekterm verbonden zijn.

De zoekmachine is een belangrijke aanwinst voor de tekst- en edi-

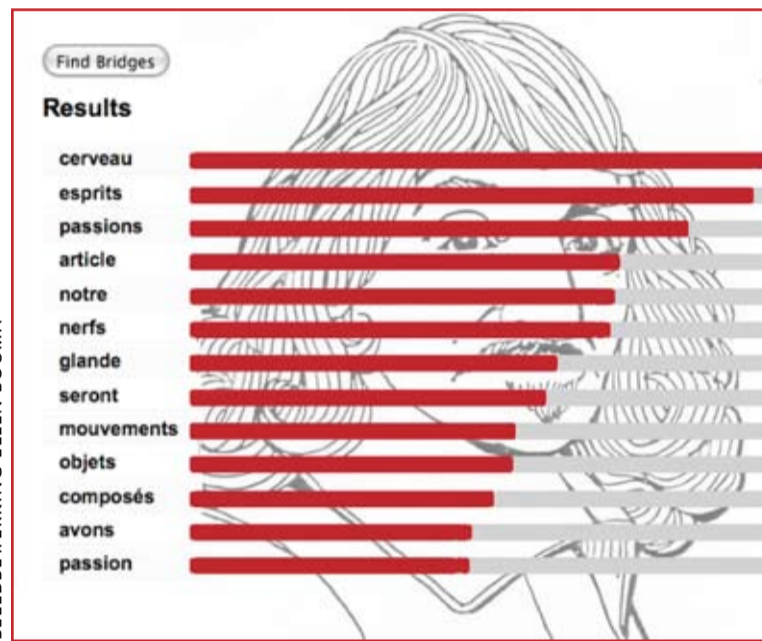
tiewetenschap. Traditionele vragen uit dit vakgebied, die tot op heden alleen na bijzonder arbeidsintensief onderzoek beantwoord konden worden, kunnen nu snel worden opgelost. Wie is de auteur van een tekst die zonder auteursnaam werd overgeleverd? Kunnen we iets zeggen over de datering van een onge-

dateerde tekst? Door zulke teksten associatief te relateren aan teksten waarvan auteur en datum wel bekend zijn, levert ANNH hele belangrijke informatie over auteurschap en datering.

Daarenboven biedt ANNH een nieuwe, innovatieve benadering om (al dan niet wijsgerige) conceptuele systemen te interpreteren. Welke concepten verbinden twee zoekwoorden op associatieve manier (zie illustratie)? Welke stukken in het werk van een denker hangen met welke andere stukken samen? Zijn er bijvoorbeeld binnen één werk heterogene elementen te ontdekken die in verband staan met verschillende andere teksten in het oeuvre van de auteur? Of, kleinschaliger geargumenteed: hoe groot is de mate van variatie in het taalgebruik van een filosoof? Leiden associatieve verbindingen steeds weer tot een relatief klein aantal van basistermen, of breidt zich het vocabulaire van een wijsgeer uit tot de meest exotische vertakkingen?

Op dit moment zijn de volledige werken van de filosofen Descartes, Schelling, Kant, Fichte, Schopenhauer, en andere belangrijke tekstbestanden uit de zeventiende eeuw aan het systeem toegevoegd. Ook bestaan er interfaces die het gemakkelijk mogelijk maken om nieuwe zoekfuncties te kunnen ontwikkelen, en nieuwe tekstbestanden aan het systeem toe te voegen.

ERIK-JAN BOS EN PAUL ZICHE



**Het resultaat van een zoekactie op geassocieerde woorden met het woordenpaar 'âme' (ziel) en 'matière' (materie) in het werk van Descartes.**

BEELDBEWERKING ELLEN BOUMA

## Naar een betere toegang tot tekst- en spraaktools

MARC KEMPS-SNIJDERS EN INEKE SCHURMAN

Het Nederlands-Vlaamse project 'Taal- en spraaktechnische Tools voor het Nederlands als Webservices in een Workflow (TTNWW)' maakt technische faciliteiten op het gebied van tekst en spraak toegankelijk voor een brede groep onderzoekers in de geestes- en sociale wetenschappen met geringe technische bagage. Daardoor moeten ze hun onderzoeksvragen beter en gemakkelijker aan kunnen pakken en vragen kunnen stellen die tot dusverre niet te beantwoorden waren.

TTNWW sluit aan bij CLARIN, een Europees project om een geïntegreerde en interoperabele infrastructuur te bouwen die onderzoekers eenvoudig toegang geeft tot taalkundige bronnen. Die bronnen kunnen geannoteerde mediabestanden zijn (vooreerst spraak en tekst), lexica en ontologieën, of het kan om technieken gaan zoals spraakherkenners, lemmatizers en parsers. Zulke bronnen, en de technieken om ze toegankelijk te maken, zijn voor het onderzoek van groot belang. Bij het beschikbaar maken ervan zijn vaak aanzienlijke kosten gemaakt voor het verzamelen, digitaliseren en

annoteren van bijvoorbeeld grote tekst- of spraakcorpora – werk dat grote vaardigheid en expertise vereist.

Die bronnen en technologie kunnen alleen volledig worden benut als ze op grote schaal eenvoudiger worden ontsloten met een robuuste en duurzame infrastructuur, zodat de onderzoekers niet zelf over (taal)technische knowhow hoeven te beschikken. Zo'n infrastructuur behelst de opslag en toegankelijkheid van data en services op grote schaal, het onderling aansluiten van dataformaten en het vermogen van de gebruikte tools om onderling samen te werken. Hierbij kan ge-

bruik worden gemaakt van middelen die beschikbaar worden gesteld door samenwerkende instituten uit heel Europa. Zo wordt voor onderzoekers de mogelijkheid geopend om een betere keuze te maken uit de beschikbare data en technieken om hun onderzoeksvragen te beantwoorden. Verschillende nationale initiatieven, waaronder CLARIN-NL en CLARIN-Vlaanderen, dragen inmiddels bij aan de realisatie van deze ideeën.

Naast CLARIN-NL wordt het project gefinancierd door het Vlaamse departement Economie, Wetenschap en Innovatie. In de afgelopen tien jaar was er al een nauwe sa-

menwerking tussen Nederland en Vlaanderen op het vlak van taal- en spraaktechnologie. Die kreeg vorm in het project Corpus Gesproken Nederlands en daarna in het bilaterale STEVIN programma voor taal- en spraaktechnologie. Deze samenwerking wordt nu voortgezet in het TTNWW-project. Naast het toegankelijk maken van faciliteiten voor onderzoekers is een belangrijk doel om de in Nederland en Vlaanderen gangbare de facto standaarden voor dataformaten en interfaces tussen tools en protocollen te promoten en

te toetsen aan Europese initiatieven zoals CLARIN.

Voor de tekstkant worden veel verschillende tools, waaronder voor aligering, automatische naamherkenning en coreferentie gecombineerd tot standaard workflows voor de beantwoording van onderzoeksvragen over historische teksten, forumdiscussies, vertaalde romans en archeologische data. Voor de spraakkant wordt gebruik gemaakt van spraakherkenners voor het ontsluiten van gesproken data op basis van toegankelijke en doorzoekbare tijdsynchrone transcripties.

Deelnemers aan TTNWW zijn de universiteiten van Tilburg, Antwerpen, Groningen, Utrecht en Twente, de Katholieke Universiteit Leuven, de Radboud Universiteit, Hogeschool Gent, het Max Planck Instituut voor Psycholinguïstiek, het Instituut voor Nederlandse Lexicologie, het Huygens Instituut, het Katholiek Documentatie Centrum, het Katholiek Documentatie- en Onderzoekscentrum voor Religie, Cultuur en Samenleving en Aletta.

Doorgewinterd verkiezingsonderzoeker Kees Aarts:

# ‘We hebben te weinig schokdempers in het systeem’

MARTIJN DE GROOT

Aan politieke dagkoersen heeft hij geen boodschap. Met des te meer belangstelling ziet hij die dagkoersen steeds meer invloed krijgen op de Nederlandse politiek. Politicoloog prof. Kees Aarts is al meer dan twintig jaar betrokken bij verkiezingsonderzoek. ‘Ik ga meer voor de lange termijn’.

Peilingsgewijs verloopt de aanloop naar de komende Tweede Kamerverkiezingen met horten en stoten. Partij van de Arbeid in glijvlucht en met Cohen weer in de lift, de VVD van Mark Rutte van het bijna niets tot grootste partij, de PVV en D66 weer op hun retour. Je zou het allemaal graag voorspeld willen hebben maar de praktijk is dat de opiniepeilers dat alleen met terugwerkende kracht kunnen doen. Het is een belangrijke reden dat Kees Aarts, hoogleraar Politicologie aan de Universiteit Twente, zich het liefst afzijdig houdt van dit type korte-termijnonderzoek. ‘Voorspellen doen we niet,’ zegt de man die al jarenlang bestuurslid en een van de woordvoerders is van het Nationaal Kiesonderzoek (NKO). ‘Maar wat we wel willen, is de ontwikkelingen van vandaag in een groter kader plaatsen. Wat betekenen de keuzes die worden gemaakt voor de democratie en omgekeerd? Over de steeds grotere onrust in het Nederlandse politieke landschap kan je interessante vragen stellen, maar die zijn niet te beantwoorden aan de hand van peilingen die per dag kunnen verschillen’.

*‘We voorspellen niet. We plaatsen ontwikkelingen in een groter kader’*

Het NKO is al sinds de jaren zeventig een bron van gegevens niet alleen over het stemgedrag van de Nederlandse bevolking, maar ook over tal van sociale en culturele factoren waarmee dat gedrag in verband kan staan. Met iedere verkiezing gaat een nieuwe ronde van het NKO gepaard en de kansen voor lange-termijnanalyses worden steeds mooier. Aarts zit dus precies op de goede plek. In

## Dutch Parliamentary Election Studies

Disseminatie van data is een van de grote ontwikkelingen van het digitale tijdperk volgens Kees Aarts, maar nog lang niet alle mogelijkheden worden benut. ‘We hebben een aantal kiezersonderzoeken liggen. Iedereen weet waar hij terecht kan (<http://easy.dans.knaw.nl>, red.) maar er wordt nog lang niet genoeg gebruik van gemaakt’. Ook om die reden is de datagids *Dutch Parliamentary Election Studies* uitgegeven, die op 21 mei in Den Haag werd gepresenteerd.



ROBERT HOETINK

Kees Aarts bij ‘Het torentje van Drienerlo’ op de campus van Universiteit Twente

Twente, waar hoogleraren als Jacques Thomassen vóór hem een klimaat hebben gevestigd van empirisch onderzoek en democratie-theorie. En bij het Nationaal Kiesonderzoek, dat de basis biedt om vragen te beantwoorden die Aarts’ onderzoeksbloed wél sneller doen stromen.

Net zoals dat grootschalige politieke survey is Kees Aarts’ wetenschappelijke loopbaan in een tamelijk rechte lijn verlopen. Constante factoren zijn zijn oriëntatie op empirisch onderzoek en de belangstelling voor verkiezingen. In de jaren zeventig en tachtig studeerde hij politicologie in Amsterdam. ‘Daar woedde in die tijd nog een methodenstrijd waarin het marxisme de boventoon voerde met kritiek op wat de burgerlijke wetenschap werd genoemd. Na een jaar voelde ik dat ik op een tweesprong stond: doorgaan met nog meer maatschappijkritiek of proberen een vak te leren.’ Het werd het laatste en Aarts richtte zich op de theoretische politicologie van hoogleraar Hans Daudt en op de methoden en technieken. De kleine maar sterk gemotiveerde vakgroep die zich daarmee bezig hield had goed stand kunnen houden in het weinig empirie-vriendelijke klimaat van de hoofdstedelijke politicologie. ‘Het toeval wilde dat die vakgroep onder Rob Mokken ook een centrum van verkiezingsonderzoek was geworden. Dat gebeurde op basis van surveys, die in het onderwijs ook een belangrijke plaats innamen’.

Daarmee was de basis gelegd voor een loopbaan waarin de rode draad door het rode potlood werd getekend.

Aarts greep de kans om in Enschede te promoveren, op een proefschrift over het functioneren van democratie en actiegroepen in stadswijken dat hij nu als een ‘uitstapje’ karakteriseert: ‘Verkiezingsonderzoek is wat me toch steeds is blijven boeien’. Hij kon direct daarna in Twente aan de slag als universitair docent en werd daar in het begin van de jaren negentig gevraagd om de voorbereiding ter hand te nemen van het NKO. Het begin was niet makkelijk, herinnert hij zich: ‘In die tijd zat het NKO in een periode van stagnatie, onder meer door de toentertijd moeizame samenwerking tussen het uitvoerende CBS en de betrokken universitaire onderzoekers. Maar in 1998 hebben mijn collega Henk van der Kolk en ik echt geprobeerd om er een heel nieuw modelonderzoek van te maken. We startten met serieus onderzoek naar de nonrespons, voegden elementen toe zoals het achterlaten van vragenlijsten bij de respondent om zelf in te vullen, en we verbeterden waar mogelijk de vragenlijst. We brachten de ontwikkelingen tijdens de campagne per week in kaart, we werkten met *gerandomiseerde*, dus van plaats wisselende vragenblokken, wat mogelijk was doordat Nederland als een van de eerste landen op grote schaal gebruik maakte van *computer-assisted personal interviewing*. En we konden

ook de datasets verrijken met externe registergegevens. Maar het werden hele saaie verkiezingen dus hoe goed we het ook hadden gedaan, er was weinig belangstelling voor onze resultaten.’

Dat was wel anders in 2002, het jaar waarin Nederland vlak voor de Tweede Kamerverkiezingen werd opgeschrikt door de eerste politieke moord. De veelgehoorde stelling dat niemand de opkomst van slachtoffer Pim Fortuyn en zijn lijst had kunnen voorspellen wil Aarts niet onweersproken laten. ‘Jacques Thomassen heeft in 2000 nog gepubliceerd over de groeiende kans dat we in Nederland met een soort Jörg Haider te maken zouden kunnen krijgen, en niet alleen hij’. En nu wordt het interessant, valt aan de Twentse hoogleraar te merken, want hier zie je nu zo’n lange-termijnontwikkeling die de achtergrond vormt van veel nieuwe kenmerken van het stemgedrag. ‘Je ziet dat de politieke cultuur steeds minder steunt op de binding tussen burgers en partijen. De groei van het islamitische bevolkingsdeel heeft daar ook mee te maken. Er zijn kansen voor allerlei ondernemende types zoals Fortuyn en Wilders die

electorale ruimte opeisen en die ook krijgen. Dat is het grote verhaal.’

Het politieke landschap wordt dus steeds instabieler. Wat je als politicoloog onder die omstandigheden wél zou kunnen doen, aldus Aarts, is ‘aan de politici voorhouden welke ontwikkelingen er zijn, zowel onder kiezers als onder partijen, laten zien welke kant het op gaat als er niets gebeurt en hoe ze zich aan de haren uit het moeras kunnen trekken. Dan gaat het over de inrichting en de laagdrempeligheid van de politiek. Het is nu relatief gemakkelijk om in Nederland een partij op te richten en een zetel in de Tweede Kamer te krijgen. Daar zou misschien wat meer demping in moeten komen. We hebben te weinig schokdempers in het systeem’.

*‘Je ziet dat de politieke cultuur steeds minder steunt op de binding tussen burgers en partijen’*

De data van de achtereenvolgende kiezersonderzoeken zijn beschikbaar gemaakt via DANS. Voor Kees Aarts is dat beschikbaar maken geen extraatje, maar een wezenlijk onderdeel van het project. Zo veel mogelijk onderzoekers moeten kunnen profiteren van de gegevens die het onderzoek oplevert. ‘Wat dat betreft is er in de politicologie veel veranderd. Toen ik begon met onderzoeken was een survey een afgerond iets. Veldwerk, dataset, klaar. Natuurlijk kon je zo’n survey wel eens koppelen aan ander onderzoek, maar dat was dan toch een heel project. Eerst is met het kiezersonderzoek de stap gemaakt van afzonderlijke surveys naar langlopende programma’s, en vervolgens heeft de digitale revolutie het mogelijk gemaakt om steeds vaker en meer gegevens uit verschillende bronnen aan elkaar te koppelen.’ Zelf heeft Aarts op dit gebied genoeg ambitie. ‘Er zijn zoveel data waar je de gegevens van het kiezersonderzoek aan kan koppelen. Denk maar aan inkomensanalyses, aan uitkomsten van specifiek campagne-onderzoek, analyses van het kabinetsbeleid, inhoudsanalyses van partijprogramma’s, en gegevens van andere landen. Op dat laatste gebied gebeurt trouwens al heel veel. ‘Met het NKO werkten we van meet af aan samen met collega’s uit landen als Oostenrijk, Noorwegen, Zweden maar ook Australië en Canada. De politicologie heeft zich ontwikkeld van landenstudies naar internationaal onderzoek. Het verkiezingsonderzoek is echt de meest internationale tak van de politicologie.’

Prof. dr. Kees Aarts (1959) studeerde politicologie in Amsterdam en promoveerde in 1990 in Enschede op een proefschrift over ‘Bodemverontreiniging en collectieve actie’. Hij begon in dat jaar als docent aan de Universiteit Twente (UT), waar hij in 2005 hoogleraar Politicologie werd. Hij werd in 1993 bestuurslid van de Stichting Kiesonderzoek Nederland en bleef daaraan in verschillende functies verbonden, onder meer als een van de *principal investigators* in 1994, 1998 en 2006. Sinds 1 januari van dit jaar is hij directeur van het Institute for Innovation and Governance Studies van de UT.

## Focus

RONALD VAN DER BIE

## Aletta, instituut voor vrouwengeschiedenis

Aletta is het studieceterium en data-archief voor vrouwengeschiedenis, met bijzondere aandacht voor vrouwen uit migrantengroepen.



DIEDERIK VAN DER LAAN

Dr. Saskia Wieringa, directeur van Aletta: 'We wilden het beeld vestigen dat we een dynamische club zijn.'

Aletta is in 1935 opgericht als Internationaal Archief voor de Vrouwenbeweging (IAV) en als zodanig het zusje van grote broer Internationaal Instituut voor Sociale Geschiedenis (IISG) met wie het lange tijd gehuisvest was aan de Herengracht. De oprichtsters, de feministen Rosa Manus (1881-1943), Johanna Naber (1859-1941) en Willemijn Posthumus-van der Goot (1897-1989), echtgenote van de IISG-oprichter Nicolaas W. Posthumus (1880-1960), wilden het erfgoed van de eerste feministische golf bewaren toen de vrouwenbeweging in de jaren twintig wat in de luwte was geraakt.

De naamsverandering in 2009 was onderdeel van een beleid om het imago op te poetsen. Dr. Saskia Wieringa, directeur van Aletta: 'Als archief en bibliotheek hadden we de roep stoffig en belegen te zijn. We vonden de lancering van de website en de vernieuwing van het dienstpakket het juiste moment om jongere generaties bij het werk te betrekken. We wilden het beeld vestigen dat we een dynamische club zijn.' Aletta bekt beter dan archief. Aletta is een herkenbare naam sinds Alletta Jacobs, de eerste vrouwelijke arts en strijdster voor vrouwenkiesrecht, in de nationale canon is opgenomen en alle jongeren in het voorgezet onderwijs haar kennen.

## Oral history testimonies

Archief en bibliotheek is Aletta nog steeds. Bij het instituut werken ongeveer dertig vaste medewerkers. Er zijn vier hoogleraren aan het instituut verbonden, die onderzoek doen op het gebied van etniciteit, vrouwengeschiedenis, seksualiteit en interpersoonlijk geweld.

Collectievorming is het nieuwe beleidsspeerpunt. Aletta wil zoveel

mogelijk verhalen van vrouwen verzamelen en opslaan in een digitaal videoarchief. Het vastleggen van deze *oral history testimonies* is hard nodig. Wieringa: 'De vrouwen van de Tweede Feministische Golf zijn minder formeel, minder georganiseerd dan de oudere generatie feministen. Van hen is minder vastgelegd. Willen wij hun verhalen leren kennen, dan moeten we ze zelf aan het woord laten.' Medewerkers van Aletta leggen de verhalen vast in lange interviews en op film. De levensverhalen van de vrouwen kleuren de geboekstaafde, officiële geschiedenis op een persoonlijke manier.

## Dolle Mina en CPN-vrouwen

Vier oral-history-projecten zijn afgerond: over Dolle Mina, over vrouwenhulpverlening, NSB-vrouwen en hun dochters, en over CPN-vrouwen en hun dochters. Met een subsidie van het ministerie van VWS heeft Aletta 30 april negentwintig dagboeken en brievencollecties van vrouwen met hun oorlogsherinneringen elektronisch

ontsloten. De vele tientallen uren interviews en videofilms zijn integraal te doorzoeken met de modernste spraakherkenningstechnologie, die in samenwerking met de Universiteit Twente is ontwikkeld. Wieringa: 'Je tikt, zeg eens wat: abortus in, en de computer zoekt in het filmbestand al die fragmenten op waar vrouwen daar iets over zeggen.' Wieringa weet het niet zeker, maar denkt dat zij met Aletta de eerste is die haar collectie op deze manier doorzoekbaar heeft gemaakt.

## Online archief

Anno 2010 beschikt Aletta, al jaren gevestigd in de fraai gerestaureerde Gerardus Majellakerk in Amsterdam-Oost, over ruim een kilometer documenten, bijna honderdduizend boeken en een grote collectie foto's, muziekwerken, enkele honderden egodocumenten (dagboeken en brieven) en digitale bronnen. Het heeft een omvangrijk online archief met video-interviews en een *full-text on line* beschikbare collectie historische boeken en tijdschriften, websites, databases, digitale dossiers en digitale lespakketten voor het onderwijs.

Met het aanleggen en ontsluiten van de digitale bronbestanden heeft Aletta zich geprofileerd als actieve kennis- en bronnenproducent. Wieringa: 'We zijn inmiddels zo geavanceerd dat we niet alleen kennisproducent, maar ook kennismakelaar zijn. Wij weten hoe en waar je moet zoeken in die enorme rijstebrijberg van digitale bestanden en kunnen bemiddelen tussen de onderzoeker die kennis zoekt en de archiefwereld.'

 [www.aletta.nu](http://www.aletta.nu)

## Boekgeschiedenis in Utrecht en in de hele wereld

De jaarlijkse bijeenkomst van de Nederlandse Boekhistorische Vereniging vond dit jaar met een overvol programma plaats in het Academiegebouw in Utrecht. Alle universiteiten waar boekwetenschap wordt onderwezen, hadden een vertegenwoordiging afgevaardigd met een inleider (meestal de lokale hoogleraar) en een jongere (in een enkel geval een oudere) onderzoeker. Dat leverde een divers beeld op, en daarnaast waren ook nog eens wetenschappers uit Antwerpen en Leuven uitgenodigd.

De focus liep uiteen van materiële bibliografie via de distributie naar

de receptie. Dat ligt voor de hand: boeken werden gemaakt, verkocht en gelezen, verschillende activiteiten waarvan de wetenschapper zich het hoe en waarom afvraagt. De bezigheden concentreerden zich op de lokale boekproductie (Leiden) maar waaierden ook uit over de hele wereld (Mexico en Peru, de activiteiten van Jezuiten).

Het was een boeiende dag die duidelijk maakte dat de boekwetenschap ook in het digitale tijdperk springlevend is. Daarbij viel op hoe belangrijk de computer en digitalisering zijn voor de moderne boek-

## Nieuwe ontwikkelingen bij de DBNL

PETER BOOT

Tien jaar na de lancering neemt de Digitale Bibliotheek voor de Nederlandse letteren (DBNL) een aantal grote stappen. *e-data@research* sprak met hoofdredacteur René van Stipriaan.

Jarenlang was de DBNL een vertrouwde leverancier van digitale tekst, waar weliswaar een steeds indrukwekkender hoeveelheid tekst ter beschikking kwam, maar qua functionaliteit niet zoveel leek te veranderen. 2010 lijkt echter het jaar te worden waarin de digitale bibliotheek zichzelf opnieuw uitvindt.

Het jaar begon met de aankondiging van een vernieuwde en beter doorzoekbare website. Volgens Van Stipriaan wordt de website sindsdien 'zeker de helft meer' gebruikt. Enigszins verstoep in de aankondiging van de vernieuwde website zat de mededeling dat de DBNL al lang niet meer alleen maar de met de hand overgetypte en betrouwbare teksten bevat waarom de bibliotheek bekend staat. Eén miljoen pagina's staan ter beschikking in de vorm van afbeeldingen en ongecorrigeerde OCR (optical character recognition, door de computer gelezen en omgezet in bewerkbare tekst). Van Stipriaan: 'We zijn overgestapt op een beleid met twee snelheden. Grote hoeveelheden tekst worden

vooruit gescand en via OCR ter beschikking gesteld. Eind dit jaar zijn er twee miljoen pagina's ongecorrigeerde tekst, ongeveer net zo veel als de hoeveelheid pagina's die op de klassieke manier beschikbaar is. Eind 2012 zullen er tien miljoen pagina's zijn gescand. De betrouwbare tekst en de webweergave op basis van XML volgt later'. De DBNL komt hiermee tegemoet aan de vaak gehoorde wens het tempo te verhogen (zie bijvoorbeeld Ewoud Sanders in *e-data* van december 2007).

Een volgend nieuwtje volgde in april: de beschikbaarstelling, in samenwerking met het Fries historisch en letterkundig museum Tresoar, van een collectie Friese literatuur. De boeken worden gepubliceerd op Sirkwy, het virtuele huis van de Friese literatuur, en op de site van de DBNL. Wellicht het meest opmerkelijke onderdeel van het bericht was de aankondiging dat de teksten ook via *printing-on-demand* ter beschikking komen. Van Stipriaan: 'En ze werden vanaf dag één besteld. Voor de rechtenvrije delen van de rest van de collectie willen we dit binnenkort ook mogelijk maken'. Bovendien komen binnenkort diverse titels ook ter beschikking als e-book, in vijf formaten, waaronder EPUB en Mobipocket.

Voor het komend najaar staat er nog meer op stapel. Van Stipriaan: 'Bijvoorbeeld de lancering van een specifieke site gewijd aan Menno ter Braak, met het volledige werk en een toelichtende encyclopedie met vijfduizend lemma's, speciaal voor deze site ontwikkeld. Daarnaast komen we nog met een tijdbalk als ontsluitingsmechanisme voor gedateerd materiaal en een thesaurus op Bijbelverzen'.

Wellicht de belangrijkste ontwikkeling, aangekondigd voor eind 2010, is de creatie van een afgesloten subsite waarop ongeveer tweeduizend romans uit de periode 1945-2000 digitaal ter beschikking komen. De DBNL werkt daaraan samen met een aantal grote uitgeverij. Bibliotheken en individuele personen zullen tegen betaling toegang tot de site kunnen krijgen. Van Stipriaan: 'De boeken zullen als *e-book* op leasebasis ter beschikking worden gesteld en kunnen dankzij *printing-on-demand* onder andere via het Centraal Boekhuis worden besteld'.

De stormachtige ontwikkelingen op het gebied van digitale bibliotheken zijn, zoveel is duidelijk, nog niet ten einde. In een volgende *e-data* gaan we nader op die ontwikkelingen in.

 [www.dbnl.org](http://www.dbnl.org)  
[www.sirkwy.nl](http://www.sirkwy.nl)



FREDDY RIKKEN

René van Stipriaan, hoofdredacteur van DBNL

# Humanities Lab op ICTDelta



MARION WITTENBERG

Joris van Zundert (Huygens) presenteert de online editiewerkomgeving e-Laborate

Op 18 maart vond in Rotterdam de ICTDelta plaats, een jaarlijks evenement op het gebied van ict-onderzoek en -innovatie. DANS (Data Archiving and Networked Services) organiseerde daar de sessie Humanities Lab die was bedoeld om te laten zien wat er mogelijk is met *humanities computing*: taal- en letterkunde, geschiedwetenschap, archeologie, musicologie

en kunstgeschiedenis gebruiken allemaal ict om onderzoekgegevens vast te leggen, te analyseren, te visualiseren en te presenteren.

Waarover gingen de toespraken van koningin Wilhelmina op Radio Oranje? Arjan van Hessen van de Universiteit Twente liet zien hoe au-

diovisueel materiaal veel beter wordt ontsloten als met spraakherkenning wordt gewerkt en verschillende digitale bronnen worden geclusterd. Is dit schilderij een echte Van Gogh? Erik Postma van de Universiteit van Tilburg demonstreerde hoe een digitale analyse van canvas en schilderijen kunsthistorici kan helpen in hun oordeelsvorming. Andere projecten die aan de orde kwamen gingen over tekstanalyse, het koppelen van data aan Geografische Informatie Systemen (GIS), *datamining* en verrijkte publicaties.

's Middags demonstreerde een aantal sprekers hun software in de kleurrijke DANS-stand op de demomarkt. Tussen stands over robotica, geavanceerde medische apparatuur en hightech gaming liet DANS zien welke nieuwe mogelijkheden er ontstaan wanneer reeds beschikbare onderzoekgegevens voor hergebruik worden opengesteld. (Marion Wittenberg)

@ [www.ictdelta2010.nl/ICTD/Het-nieuwe-lab-Humanities-lab.html](http://www.ictdelta2010.nl/ICTD/Het-nieuwe-lab-Humanities-lab.html)

## NCDD publiceert strategische agenda

De Nationale Coalitie Digitale Duurzaamheid (NCDD) publiceert in juni haar strategische agenda. Die gaat over het verzekeren van duurzame toegang tot digitale gegevens in de publieke sector in Nederland, zowel in de wetenschap als voor overheid/archieven, media en het cultureel erfgoed.

De strategie is gebaseerd op twee assen: een aanpak per sector en een sectoroverschrijdende aanpak. De eerste wordt aangevoerd

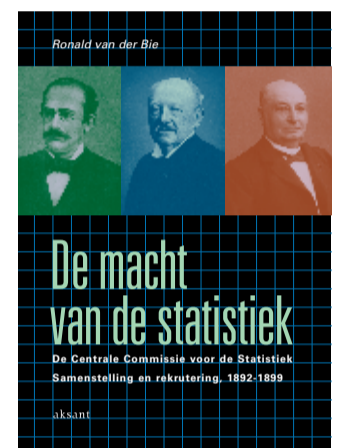
door voortrekkerorganisaties. Voor de wetenschap zijn dat DANS en de Koninklijke Bibliotheek. De sectoroverschrijdende aanpak wordt gecoördineerd door de NCDD zelf. Uiteindelijk zal een landelijk dekkend netwerk ontstaan waarin alle informatie duurzaam toegankelijk gehouden kan worden die belangrijk genoeg wordt geacht om te bewaren. (IA)

@ [www.ncdd.nl](http://www.ncdd.nl)

## Hoe de statistiek invloed kreeg op het overheidsbeleid

Tot aan het einde van de negentiende eeuw waren het vooral beschavingswerkers die de 'redelijke opvoeding van het volk' ter hand namen. Maar zij bleken niet bij machte om oplossingen te vinden voor de problemen die het sociale oproer van de jaren tachtig veroorzaakte. De arbeidsenquête van 1886 toonde aan hoe treurig het gesteld was met de arbeidsomstandigheden in de werkplaatsen en de rechteloosheid van de arbeiders, en leidde tot nieuwe wetgeving. De statistiek werd het noodzakelijke instrument van een staat die de ambitie had om sturing te geven aan de samenleving.

In 1892 leidde dit tot de oprichting van de Centrale Commissie voor de Statistiek (CCS). Ronald van der Bie, zelf werkzaam bij het CBS, schreef onlangs de geschiedenis van de Commissie. Van der Bie noemt de oprichting van de CCS het culminatiepunt van een reeks buitenparlementaire 'opiniërende' interventies in de sociale kwestie. Tegelijkertijd betekende de oprichting van de Commissie bij Konink-



lijk Besluit van 6 oktober 1892 een overwinning op de jarenlange tegenwerking van de subsidiegevers, die de statistiek maar 'overbodige omhaal' vonden. Het maken van statistieken zou in 1899 worden opgedragen aan het Centraal Bureau voor de Statistiek. De professionalisering van de dataverzameling, de statistiek, is dan een feit geworden. (Tineke in 't Veld)

Ronald van der Bie, *De macht van de statistiek. De Centrale Commissie voor de Statistiek, samenstelling en rekrutering (1892-1899)* Amsterdam: Aksant, 2009

## Sinds kort beschikbaar

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

van diverse andere onderzoeksinstellingen. Deze kunnen kosteloos worden gedownload vanuit DANS EASY: <http://easy.dans.knaw.nl>. Via DANS kunnen ook de beveiligde microdata van het CBS kosteloos geleverd worden: [www.dans.knaw.nl/nl/data/cbs/overzicht/](http://www.dans.knaw.nl/nl/data/cbs/overzicht/)

Centraal Bureau voor de Statistiek	Periode
Integraal-Inkomensbestand	2006-2008
ict-gebruik huishoudens en personen	2005-2007
AOW-uitkeringen	2009
Community Innovation Survey	2004, 2006
Nationale enquête arbeidsongeschiktheid	2008
EU-SILC	2007, 2008
Perinatale Registratie Nederland	2004, 2005
<b>Productiestatistieken:</b>	
Fysiotherapeuten	2006-2007
Verloskundigen	2006-2007
Milieudienstverlening	1993-2001
Commerciële diensten	1993-1999
<b>Via DANS EASY</b>	
<b>Archeologie</b>	
Westervoort De Schans Geldersoorth Opgraving (Archeologisch Diensten Centrum)	2009
De Oertijd van de Dru (Rijksdienst voor het Cultureel Erfgoed)	2007
Archeologische onderzoeken Valkenburg ZH (Rijksdienst voor het Oudheidkundig Bodemonderzoek)	1985-1988, 1994, 1998, 2000
Heiloo-Hoog en Laag (Amsterdams Archeologisch Centrum/projectenbureau)	2006
<b>Geschiedenis</b>	
Akkers van Margraten. Erfgoed van de Oorlog; interviews afgenomen in het kader van het project Getuigen Verhalen (Ruim Kader Filmproducties)	1940-1945
Duitse bezetting op Texel, de 'Russenoorlog'; Erfgoed van de Oorlog; interviews afgenomen in het kader van het project Getuigen Verhalen (Stichting Traktor)	1940-1945
Herinneringen aan mijn Amsterdamse zusje/broertje; Erfgoed van de Oorlog; interviews afgenomen in het kader van het project Getuigen Verhalen (Limburgs Geschiedkundig en Oudheidkundig Genootschap)	1940-1945
VOC Opvarenden – Amsterdamse Kamer-update (Drs. A.J.M. van Velzen, Prof. dr. F.S. Gaastra)	18e eeuw
<b>Sociale wetenschappen</b>	
WOoN – Woononderzoek Nederland (Ministerie van VROM)	2009
International Crime Victims Survey (J.N. van Kesteren, Tilburg University)	2000 en 2005
European Social Survey – round 4 (via Norwegian Social Science Data Service – NSD)	2008
Aanvullend Voorzieningen Gebruiksonderzoek – AVO (CBS beveiligd microbestand)	2007
POLS Basis (CBS beveiligd microbestand)	2008
POLS Gezondheid (CBS beveiligd microbestand)	2008

## Nieuwe mogelijkheden in BNTL

Aan de Bibliografie van de Nederlandse Taal- en Literatuurwetenschap (BNTL) is sinds kort de mogelijkheid van RSS-feeds toegevoegd. Daardoor kunnen belangstellenden zich abonneren op informatie over recent aan de BNTL toegevoegde publicaties of op bepaalde berichten (bijvoorbeeld alle publicaties over Couperus). RSS-feeds kunnen worden gelezen in newsreaders zoals Google Reader of FeedDemon in mailprogramma's, maar ook opgenomen worden in andere websites.

De toevoeging van RSS-feeds is een van de vernieuwingen in een update van de e-bibliografie die het Huygens Instituut op 19 april lanceerde. Een andere vernieuwing is dat bij het zoeken naar publicaties voortaan ook treffers in de tekst worden getoond als een bestand met de volledige tekst aanwezig is. Binnenkort volgt een actie om van zo veel mogelijk publicaties de volledige tekst beschikbaar te stellen. Daarvoor zal de medewerking van auteurs en redacties worden gevraagd.

Ook het invoeren van gegevens over publicaties is wat eenvoudiger gemaakt. De redactie hoopt dat gebruikers van de BNTL zich zullen aanmelden om (eigen) ontbrekende publicaties in te voeren. Mede dankzij de nu beschikbaar gekomen RSS-functionaliteit kan dat de zichtbaarheid van de publicaties alleen maar ten goede komen. (PB)

@ [www.bntl.nl/](http://www.bntl.nl/)  
[www.google.com/reader/](http://www.google.com/reader/)

### COLOFON

*e-data&research* is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van DANS, het Huygensinstituut, het Centraal Bureau voor de Statistiek, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen.

Oplage: 9100.

*e-data&research* is online te raadplegen op [www.edata.nl](http://www.edata.nl)

**Uitgever:** Stichting Uitgeverij e-data&research, Postbus 93067, 2509 AB Den Haag

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

**Redactie:** Inge Angevaere, Ronald van der Bie, Peter Boot, Martijn de Groot (hoofd/eindredacteur), Thijs Hermsen, Jetske van der Schaaf

### Aan dit nummer werkten mee:

Erik-Jan Bos, Fannie Cobben, Suzan van Dijk, Robbert Dijkgraaf, Paul Dijstelberge, Charles van den Heuvel, Jaap Kamps, Marc Kemps-Snijders, Ineke Schuurman, Tineke in 't Veld, Marion Wittenberg, Paul Ziche.

### Redactiesecretariaat:

Lucas Pasteuning, Jetske van der Schaaf

### Vormgeving en opmaak:

Ellen Bouma

**Productie:** Uitgeverij Aksant, Amsterdam

**Druk:** Ten Brink, Meppel

**ISSN:** 1872-0374

## MEEGEMAAKT

## Internationale samenwerking en het paardentram-paradigma

Suzan van Dijk

Aangespoord door artikelen in *e-data&research*, en aangelokt door de digitaliserings- en internationaliseringsgelden, willen ook geesteswetenschappers wel eens grootschalig samenwerken. Voor bepaalde onderwerpen is het trouwens noodzakelijk. Toch doen zich op dat front problemen voor, waaraan men aanvankelijk niet gedacht had. Naar aanleiding van een brede internationale samenwerking op het gebied van vrouwenliteratuur ga ik hier op enkele aspecten in – rekening houdend met wat Hugo Brandt Corstius tien jaar geleden schreef: ‘Het is met vrouwenliteratuur als met de paardentram’. Deels zijn de te signaleren problemen misschien terug te brengen tot dit gegeven, want: ‘U denkt misschien bij “paardentram” aan een tram die door paarden wordt getrokken. Maar een paardentram is ook een tram waarin paarden zich van de wei naar de rennen laten rijden. Of waar een olijk gemeentebestuur paardenlijven op heeft laten schilderen...’. De materie is, kortom, ingewikkeld en kan van verschillende kanten worden benaderd.

Een aanvraag voor zo'n meestal Europese samenwerking wordt veelal voorbereid door een groep van vijf à tien collega's. In onderling overleg schrijven zij een onderzoeksprogramma dat in het ideale geval inderdaad voor financiering in aanmerking komt. In dat overleg houdt men rekening met mogelijke, en soms bekende behoeften van andere, later uit te nodigen collega's. Gepoogd wordt om recht te doen aan de te verwachten diversiteit van perspectieven. Zo was het althans in ons geval.

Inmiddels is onze samenwerking volop op gang gekomen, en is men in twintig landen druk doende om de eerste werkbijeenkomst voor te bereiden. De gedachte was dat eenieder daarbij enthousiast gebruik zou maken van enkele reeds bestaande, volgens ons heel redelijk functionerende

online tools. In de praktijk blijkt het echter anders te werken. Ten eerste doet zich het bekende probleem voor van de tegenover ict nog steeds zeer aarzelende geesteswetenschapper, en ten tweede hadden wij zelf misschien het 'paardentram-paradigma' nadrukkelijker moeten expliciteren.

‘Waarom stappen onderzoekers in een digitaliseringsoperatie?’, vroegen wij ons regelmatig af toen wij mailtjes begonnen te ontvangen van collega's uit diverse landen, waarin zij berichtten dat (1) bepaalde gegevens niet konden worden ingevoerd, (2) *idem* voor gegevens uit een bepaald land, (3) *idem* voor gegevens die afhankelijk lijken van bepaalde andere, nog niet ingevoerde, gegevens, en vooral (4) dat zij hun wachtwoord waren vergeten. ‘Waarom bekijken onderzoekers niet eerst de reeds aanwezige content?’, was een andere vraag, toen bleek dat heel wat al ingevoerde informatie ten tweeden male aan de database werd toegevoegd. Op onze vraag wáár precies de Instructies tekort schieten kwam geen antwoord.

De problemen zijn niet alleen van technische aard. Sommige collega's, die alleszins in staat lijken om de data in te voeren, doen dat niet omdat ze vrezen dat hun bronnenmateriaal door anderen voortijdig (vóór de publicatie van hun conclusies) ontdekt wordt. Ze gaan er kennelijk van uit dat dit materiaal slechts voor één inter-

pretatie vatbaar is. *Quod non* ongetwijfeld, ook al betreft het vrouwenliteratuur.

Wat onze eigen paardentram-benadering van vrouwenliteratuur betreft, concludeerden wij dat wij wellicht wat optimistisch zijn geweest. Waarschijnlijk ligt er een voorbereidende taak voor geesteswetenschappers die rond een enigszins breed onderwerp een samenwerking op touw willen zetten. Van het paardentram-effect is niet iedereen zich bewust, en daardoor ook niet van de noodzaak om de eigen oogkleppen eens kritisch te bekijken.

Omdat wij nog aan het begin van onze samenwerking staan houden wij de moed erin: gezien de verklaringen van 'nieuwe' collega's die willen toetreden, en ook toelichten waarom zij dat willen, is daar alle reden toe. Het belang dat wordt gehecht aan vrouwenliteratuur en aan de bestudering ervan is per land zeer verschillend: het is niet alleen literair-historisch maar ook, zeker in Oostelijk Europa, politiek. Dit spoort op zichzelf weer aan tot het realiseren van de samenwerking. De 'technische' problemen zullen, in ons geval, dan ook wel worden overwonnen.

Suzan van Dijk is verbonden aan de faculteit Geesteswetenschappen van de universiteit Utrecht en gastonderzoeker bij het Huygens Instituut



## Gelezen

Bojan Todosijevic, Kees Aarts, Harry van der Kaap: *Dutch Parliamentary Election Studies, Data Source Book 1971-2006*; Den Haag, DANS, 2010; *DANS Data Guide 7*, ISBN 978 94 90531 04 1

General elections for the Second Chamber of Parliament constitute one of the most important elements of democracy in the Netherlands. Insight into the backgrounds, motives, goals and behaviour of the Dutch voters is therefore essential for understanding and appraising the functioning of democracy. Since 1971, the Dutch political science community has conducted large scale, nationwide election surveys around every parliamentary election. All resulting datasets have been extensively documented and archived, and placed at other researchers' disposal through DANS. This Data Guide provides the documentation of an integrated dataset based on the Dutch Parliamentary Election Studies of 1971-



2006. It thus provides directly accessible information on change and stability in electoral behaviour and political orientations in the Netherlands. The data

are concentrated on the more frequently asked questions of these Dutch Parliamentary Election Studies.

Michiel Hildebrand: *End-user support for access to heterogeneous linked data SIKS; Dutch Research School for Innovation And Knowledge Systems, 2010; SIKS Dissertation Series no. 2010-22*

The openness of the Web allows data from different sources to be combined, enabling information access in ways unforeseen by the original providers. The rise of linked data adds a new dimension to this information revolution. Using Semantic Web representation languages, structured information becomes more easily acceptable by machines, enabling emergent technologies to reuse, integrate and process information. The foundations for the Web of data are now in place and institutions are starting to share linked data. In this thesis, Michiel

## Column

Robbert Dijkgraaf

## Dataprijs

Een oude regel in de wetenschap zegt dat je je onderzoeksgegevens altijd beschikbaar moet houden voor wie je conclusies zou willen checken. Eigenlijk is die ene regel al genoeg om elke discussie over het delen van data snel te beslechten. Je zou daar nog aan toe kunnen voegen dat het delen van data niet alleen goed is om elkaar te controleren, maar ook om elkaar te stimuleren. Dan heb je al twee sterke argumenten.



Maar de tijden zijn veranderd, en die verandering brengt in een stijgend tempo nieuwe overwegingen mee. Overwegingen die allemaal in dezelfde richting wijzen. Om te beginnen de digitalisering, die het behalve wenselijk ook veel beter mogelijk heeft gemaakt om gegevens uit te wisselen. De oprichting van het Steinmetzarchief in de jaren zestig van de vorige eeuw laat zien dat de sociale wetenschappers dat voordeel al wilden verzilveren in het vroeg elektronische tijdperk. Maar toen het elektronische tijdperk eenmaal was aangebroken, kwam er in veel disciplines een wereld aan nieuwe mogelijkheden bij: om data te delen en bij elkaar te voegen, en om veel grotere hoeveelheden data te verzamelen en verwerken dan tot dan toe denkbaar was geweest. Om dat te illustreren hoef ik de Geneefse deeltjesversneller maar te noemen, die niet alleen de samenwerking van duizenden onderzoekers vereist, maar ook zelf voor een groot deel de technologie heeft ontwikkeld die deze samenwerking mogelijk maakt.

Het elektronische tijdperk verandert de wetenschap niet alleen in kwantitatief maar ook in kwalitatief opzicht. Zo zie je in de humaniora, waar de KNAW juist dit jaar een impuls geeft aan de toepassing van ict in het onderzoek, dat die toepassingen nu al leiden tot hele nieuwe vormen van analyse en samenwerking.

Intussen wordt de ons omringende maatschappij hectischer, ingewikkelder, veeleisender en kritischer. Dat betekent dat er meer druk op de wetenschap komt te staan, dat er grotere problemen op ons bordje worden gelegd en dat we kritischer worden beoordeeld - op onze conclusies door iedereen, getuige ook de discussies over de al of niet beschikbare data van het klimaatpanel IPCC - en door de overheden op de kosteneffectiviteit van ons onderzoek. Ook dat alles wijst in de richting van meer en, waar mogelijk, interdisciplinair samenwerken, meer krachten bundelen en meer openheid. Dus data bundelen en data beschikbaar maken.

Het elektronische tijdperk heeft ons niet alleen maar goeds gebracht. Door de snelle ontwikkeling van hard- en software blijken digitale gegevens heel wat vergankelijker te zijn (in letterlijke zin) dan de data die we vroeger op papier bewaarden. Er is dus veel meer aandacht voor duurzaamheid nodig om te zorgen dat de vruchten van digitaal datadelen zich ook werkelijk laten plukken.

Moet ik meer zeggen om uit te leggen waarom ik op 18 juni graag zelf de DANS Dataprijs in Amsterdam wil overhandigen aan de alfa- of gamma-onderzoeksgroep waarvan de beste, meest toegankelijke en duurzaamste dataset is genomineerd?

Prof. dr. Robbert Dijkgraaf is president van de Koninklijke Nederlandse Akademie van Wetenschappen

Hildebrand investigates access to heterogeneous linked data from an end-user perspective. Based on case studies in the cultural heritage domain he explores how domain experts can benefit from semantically-rich linked data. With the development and evaluation of prototype systems he demonstrates how a number of specific tasks can be supported. Hildebrand shows that generic solutions can be successfully applied to linked data, by using the semantic relations in the data combined with careful configuration of the search algorithms and interface components.

Tony Hey, Stewart Tansley, Kristin Tolle (Eds.): *The fourth paradigm, data intensive scientific discovery*; Redmond

Washington, Microsoft research, 2009; ISBN 978 0 9825442 0 4

Increasingly, scientific breakthroughs will be powered by advanced computing capabilities that help researchers manipulate and explore massive datasets. The speed at which any given scientific discipline advances will depend on how well its researchers collaborate with one another, and with technologists, in areas of eScience such as databases, workflow management, visualization, and cloud computing technologies. In this publication, the collection of essays expands on the vision of pioneering computer scientist Jim Gray for a new, fourth paradigm of discovery based on data-intensive science, and offers insights into how it can be fully realized.