

Pagina 3 • Veteran tapes • In het project 'The Veteran Tapes' van DANS en het Veteraneninstituut analyseren onderzoekers uit zeven verschillende disciplines dezelfde tapes waarop gesprekken zijn opgenomen met veteranen.

Pagina 4 • Psychologische data ont-sloten • Een grote hoeveelheid psychologische data is opgetast in het TestweekLab, dat uitkomsten van veertig jaar testen van eerstejaars psychologiestudenten bewaart. Nu komen ze vrij in een collaboratory.



JOHN THURING

Pagina 5 • Jacques van Maarseveen • Het 'historisch geweten van het CBS' wordt de statisticus wel genoemd die één van de drijvende krachten was achter de digitalisering van de Volkstellingen.

Pagina 6 • Brieven van Van Gogh • Het Van Gogh Museum en het Huygens Instituut bereiden een nieuwe wetenschappelijke editie voor van de brieven van Van Gogh. Die bieden een schat aan informatie over zijn artistieke ontwikkeling.

EN VERDER

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

KB vervangt TIFF door JPEG2000

De Koninklijke Bibliotheek (KB) heeft besloten de masters afkomstig uit haar digitaliseringsprojecten voortaan niet meer duurzaam te bewaren in het TIFF-formaat, maar in JPEG2000. Eigen onderzoek heeft uitgewezen dat JPEG2000 (lossless) circa de helft minder opslagcapaciteit vraagt. Omdat de KB de komende vier jaar zo'n veertig miljoen pagina's hoopt te digitaliseren, levert deze keuze een aanzienlijke besparing op terwijl de gebruiksmogelijkheden goed blijven. De opslag van de bits en de bytes is een belangrijke kostenpost bij duurzame opslag, zo overweegt de bibliotheek. Geheugencapaciteit wordt weliswaar steeds goedkoper, maar er is ook steeds meer capaciteit nodigen het langetermijn onderhoud daarvan is intensief, dus kostbaar. Het volledige onderzoeksrapport kan worden gedownload van de website van de KB. (IA)

www.kb.nl/hrd/digitalisering/documentatie/index.html

Onderzoek duurzaamheid Europese wetenschap

PARSE.insight is een nieuw project dat digitale duurzaamheid binnen de wetenschapssector in Europa gaat onderzoeken. Het is in maart dit jaar gestart en heeft een looptijd van twee jaar. Aan het project, dat wordt gefinancierd door de Europese Unie, doen negen partners mee uit verschillende Europese landen. De Koninklijke Bibliotheek heeft de leiding over de verkenningsfase.

Achtergrond van het project is de constatering dat het niet eenvoudig is om geschikte onderzoeksgegevens binnen verschillende wetenschappelijke disciplines en over landsgrenzen heen te vinden. Veel bronnen zijn al wel digitaal, maar van eenduidige beschikbaarstelling is nog geen sprake. Als het aan de Europese Unie ligt, komt hier verandering in door de ontwikkeling van een zogeheten *e-science infrastructure*, een Europees netwerk voor de wetenschap waarin informatie interdisciplinair met elkaar kan worden gedeeld. Op termijn zou dit moeten leiden tot een betere beschikbaarheid van bronnen en meer hergebruik van bestaande informatie. Interoperabiliteit, het mogelijk maken van onderlinge toegang en verkeer tussen verschillende bestaande bronnen, staat daarbij voorop. Maar wetenschappelijke bronnen voor de lange termijn beschikbaar houden betekent ook dat ze duurzaam bewaard dienen te worden. Een belangrijke vraag in dit project is in hoeverre onderzoeksinstellingen zich hiervan bewust zijn.

Momenteel ontbreekt een goed overzicht van de stand van zaken rond digitale duurzaamheid bij Europese onderzoeksinstellingen. Dat komt vooral door de grote verscheidenheid aan wetenschappelijke instellingen, onderzoeksgemeenschappen en samenwerkingsverbanden. PARSE.insight gaat dat landschap in kaart brengen. Daarnaast wordt onderzocht waar en hoe duurzaamheid van digitale onderzoeksdata en publicaties gestimuleerd moet worden om te voorkomen dat wetenschappelijke informatie verloren gaat. Op basis van de uitkomsten zal PARSE.insight advies uitbrengen aan de Europese Commissie en Europese instellingen als de European Science Foundation.

In het project werken negen Europese partners samen, waaronder de Koninklijke Bibliotheek (KB), de Science and Technology Facilities Council (STFC), de Deutsche Nationalbibliothek (DNB) en de onderzoeksinstellingen Max Planck Gesellschaft, CERN en ESA. De uitgeverwereld is vertegenwoordigd door de International Association of Scientific, Technical & Medical

(STM) Publishers. De KB heeft de leiding over de verkenningsfase die inzicht moet opleveren in de huidige en geplande activiteiten rondom digitale duurzaamheid. Via enquêtes en interviews wordt een representatieve selectie van instellingen in Europa bevraagd.

In de komende maanden wordt er gewerkt aan de ontwikkeling van een enquête. Deze zal gefaseerd worden

uitgezet, te beginnen in Nederland. Daarnaast leveren diepte-interviews met sleutelfiguren in de Europese onderzoekswereld de benodigde achtergrondkennis op om in de vervolgfase te kunnen bepalen wat er nodig is om wetenschappelijke informatie in digitale vorm te behouden. (Jeffrey van der Hoeven)

www.parse-insight.eu

Hollandse driemaster in de Sont



Een Nederlandse driemaster in de Sont, de zeearm tussen Denemarken en Zweden die de Noordzee met de Oostzee verbindt. Aan het begin van de zeventiende eeuw, toen dit schilderij werd gemaakt door de Haarlemse zee- en schepenschilder Hendrik Corneliszoon Vroom, domineerden reders uit ons land het vervoer over zee van bijvoorbeeld graan en eikenhout uit de Baltische landen. Dat blijkt uit de registers die werden bijgehouden door de ambtenaren die in het Deense Helsingør belast waren met het innen van de tolgeden. In 1420 verrees het

eerste fort bij de Sont om de Deense aanspraken op die gelden te ondersteunen. Honderdvijftig jaar later werd het door Vroom geschilderde kasteel Kronborg gebouwd, dat nu nog grotendeels overeind staat.

De Sonttolregisters, die lopen vanaf het eind van de vijftiende eeuw tot halverwege de negentiende, vormen een bron van kennis over de maritieme en economische geschiedenis van Nederland. Ze zijn te vinden in Kopenhagen, maar op microfilm ook

in het Nederlands Economisch Historisch Archief bij het Internationaal Instituut voor Sociale Geschiedenis in Amsterdam. De Rijksuniversiteit Groningen en het Fries historisch en letterkundig centrum Tresoor gaan de registers nu via internet bereikbaar maken. Er komt een elektronische databank met alle gegevens van de registers, die vanaf 2011 door iedereen kan worden geraadpleegd. DANS gaat een reservekopie onderhouden. NWO subsidieert het project. (MdG)

ECPA heft zichzelf op

De European Commission on Preservation and Access (ECPA), in 1994 opgericht om het behoud van documentair erfgoed te bevorderen, heeft zichzelf met ingang van afgelopen zomer opgeheven. Dat heeft de commissie in een brief aan relaties bekend gemaakt. Een reorganisatie binnen de Koninklijke Nederlandse Akademie van Wetenschappen, de thuisbasis van ECPA, was aanleiding voor de stap. 'Diepgaande veranderingen in het veld' hebben volgens de brief in de laatste vijftien jaar geleid

tot 'nieuwe omstandigheden, die vragen om nieuwe netwerken'.

Leden van ECPA waren de nationale bibliotheken van Portugal, Italië, Slovenië en Nedersaksen, de Koninklijke Bibliotheek van Denemarken, Koninklijke Academies van Noorwegen en Nederland, het Oostenrijkse Phonogrammarchiv, het Nationaal Archief van het Verenigd Koninkrijk en de informaticatak van het Spaanse grootwinkelbedrijf El Corte Ingles.

NCDD onderzoekt digitale duurzaamheid in Nederland

Omdat de zorg voor digitale data zowel internationale als nationale dimensies heeft, wordt parallel aan PARSE.insight onderzoek gedaan naar de status van digitale duurzaamheid in Nederland. De vorig jaar opgerichte Nationale Coalitie Digitale Duurzaamheid (NCDD) heeft zojuist een vooronderzoek afgerond voor een nationale verkenning die in het najaar zal starten. Die moet genoeg informatie opleveren om een nationale infrastructuur te bouwen voor digitale duurzaamheid voor de hele publieke sector – niet alleen de wetenschap, maar ook voor de overheid en het erfgoed. De NCDD streeft daarbij naar goede afstemming tussen de diverse sectoren en met internationale initiatieven.

www.ncdd.nl

AGENDA

12 september
Den Haag, Koninklijke Bibliotheek
Studiemiddag over massadigitalisering
Tijdens de studiemiddag van de Vereniging voor Geschiedenis en Informatica worden verschillende aspecten van massadigitalisering belicht. Van het voortraject en de uitvoering tot de daarop volgende ontsluiting, presentatie en archivering. Daarnaast wordt antwoord gegeven op vragen als: hoe verhoudt kwantiteit zich tot kwaliteit, en wat zijn de mogelijkheden en de beperkingen van massadigitalisering? Zie ook pagina 7 <http://www.den.nl/agenda/20080722115259/>

23 september
Leidschenveen, Centraal Bureau voor de Statistiek, Tinbergenzaal
Beperking en correctie van non-response in enquêtes
Op deze lezingenmiddag, georganiseerd door het Nederlandstalig Platform voor survey onderzoek (NPSO), staan presentaties centraal die ingaan op de ontwikkelingen in methodologie voor non-responsvermindering en -correctie, en de groeiende interactie tussen deze twee onderzoeksterreinen. <http://www.npso.net/agenda.html#p2>

29 en 30 september
London, British Library Conference Centre
iPRES 2008 – Fifth international conference on the preservation of digital objects
The fifth in the series of annual international conferences which bring together researchers and practitioners from around the world to explore the latest trends, innovations, thinking, and practice in digital preservation. <http://www.bl.uk/ipres2008/programme.html>

19-24 oktober
Ann Arbor, Michigan, Verenigde Staten
Workshop Digital preservation management, implementing short-term strategies for long-term problems
The *Digital Preservation Management* workshop series, organized by the Inter-university Consortium for Political and Social Research (ICPSR), is intended for those who are contemplating or implementing digital preservation programs in libraries, archives, and other cultural institutions. The goals of this initiative are to foster critical thinking in a technological realm and to provide the means for exercising practical and responsible stewardship of digital assets.

6 november
Amsterdam, World Trade Centre
Symposium Monitoring van Arbeid, vinger aan de pols van werkend Nederland
TNO organiseert dit symposium voor zowel wetenschappers als beleidsmakers bij overheid, uitvoeringsinstanties, sociale partners, brancheorganisaties en adviesorganen. Het levert u inzicht in de opbrengst en uitdagingen van monitoring van arbeid in Nederland.

13 november
Den Haag, Javagebouw NWO
Nederland in vergelijkend perspectief; Tweede Nederlandse workshop over het European Social Survey (ESS)
Doelgroep: wetenschappelijk onderzoekers, Aio's en research masterstudenten die met European Social Survey (ESS) data werken of willen gaan werken. De bijeenkomst, georganiseerd door NWO, VU en DANS, biedt onderzoekers de gelegenheid praktische informatie uit te wisselen over het gebruik van de ESS databestanden. In het afsluitend symposium wordt belicht hoe Nederland er volgens het ESS in vergelijkend perspectief voor staat. Aan deelname zijn geen kosten verbonden. http://www.dans.knaw.nl/nl/dans_symposia/2008_5/

20-22 november
University of Lisbon, Portugal
Private, do (not) enter; personal writings and textual scholarship. The fifth international conference of the European Society for Textual Scholarship
During three days, the European Society for Textual Scholarship Conference will discuss the technical, ethical, and legal issues raised at the time of editing or studying private writings. <http://ests.lisbon.conference.difusa.eu/index.html>

Digital humanities: bouwen of experimenteren

Het Finse stadje Oulu, vlakbij de poolcirkel, is een belangrijke vestigingsplaats van Nokia, en bovendien de zetel van het jaarlijkse wereldkampioenschap luchtgitaar. Eind juni dit jaar was de universiteit van Oulu gastheer van een wat substantieler gebeurtenis, het congres *Digital Humanities 2008*. Het is de belangrijkste jaarlijkse bijeenkomst op het vlak van Humanities Computing. Traditioneel zijn de belangrijkste onderwerpen voor deze conferentie het onderzoek van (literaire) teksten (auteursherkenning, stilometrie, text mining) en de presentatie van digitale edities. Bij dat laatste kan de nadruk liggen op de inhoud, op de gebruikte software, of bijvoorbeeld op de zoekmogelijkheden die een editie biedt.

Een interessante presentatie van Neil Audenaert gaf een abstract model voor 'Annotated Facsimile Editions'. Het model is gebaseerd op het werk aan verschillende belangrijke editieprojecten, en biedt een eerste aanzet voor een gemeenschappelijke software-architectuur voor facsimile-edities (edities die gebaseerd zijn op afbeeldingen, niet op tekst).

Een onderwerp dat enigszins aansluit bij de digitale editie is de presentatie van andere vormen van cultureel erfgoed. De opening keynote van Eero Hyvönen presenteerde CultureSampo, een rijkelijk gesubsidieerd project waarbij semantisch-webtechnieken worden gebruikt voor het ontsluiten van museumcollecties. Daarbij, zoals

de discussie naderhand ook toonde, bestaat altijd een spanningsveld tussen het opleveren van een werkend product en de nuances van het geesteswetenschappelijk onderzoek. Waar voor de geestwetenschapper alle onderscheidingen tijdgebonden en voorlopig zijn, heeft de bouw van systemen tijdloze en objectieve definities nodig.

Een verwant dilemma voor 'computing humanists' is het conflict tussen het bouwen van stabiele en bruikbare tools enerzijds en anderzijds de noodzaak om te experimenteren. De waarde van de ontwikkelde hulpmiddelen was het thema voor een aantal bijdragen. De reflectie op het vak kwam ook tot uiting in Edward Vanhoutte's aanzet tot een geschiedenis van Humanities Computing.

Een laatste onderwerp waarnaar veel aandacht uitging staat bekend als 'e-science'. Onder die verzamelnaam wordt gesproken over de mogelijkheden die ontstaan door grote collecties van data, gezamenlijk bewerkt en geanalyseerd, met

Wie gaat betalen voor de duurzaamheid van digitale data?

In Londen werd op 23 juni een conferentie gehouden over de kosten van digitale duurzaamheid. Binnen het LIFE2 project (Life Cycle Information for E-literature) wordt een model ontwikkeld om die kosten te kunnen berekenen. Toch blijft het lastig om concrete kosten te schatten. Zo zijn er veel variabelen in het spel, zoals de gestelde kwaliteitseisen. Bovendien is er nog steeds gebrek aan langeter-

min toetsingsgegevens. Keynote spreker Paul Courant van de University of Michigan ging in op de verantwoordelijkheid voor duurzame opslag. In economische zin vergeleek hij digitale data met de natuur: ook digitale data zijn (via internet) voor bijna iedereen toegankelijk en gebruik door de één doet niets af aan gebruiksmogelijkheden voor een ander. Bovendien zullen de vruchten van digitale duurzaamheid pas in de toekomst geplukt kunnen worden. Bedrijven zullen er daarom geen werk van maken; die hebben vooral een kortetermijnbelang. Lokale financiering is ook lastig, omdat er nauwelijks meer sprake is

van een directe relatie tussen degene die investeert (bijv. een bibliotheek) en degene die profiteert (bijv. internetgebruikers in de toekomst). Daarom moet de duurzame opslag van digitale data een kwestie worden van instellings- en sectoroverschrijdende overeenkomsten op hoog niveau. In Nederland wordt een dergelijke aanpak bepleit door de Nationale Coalitie Digitale Duurzaamheid en in Europa door de Alliance for Permanent Access to the Records of Science. De Koninklijke Bibliotheek, DANS (Data Archiving and Networked Services) en het Centraal Bureau voor de Statistiek participeren in deze coalities. (IA)

Op 21 mei vond op het Meertens Instituut in Amsterdam de eerste nationale FIDLR-Start workshop plaats. FIDLR staat voor 'Framework for the Interoperability of Dutch Language Resources'. FIDLR-Start beoogt de funding te leggen voor een later aan te vragen FIDLR project. Het doel is om een infrastructurele voorziening op te zetten voor het taalkunde-onderzoek in Nederland, waarbij de nadruk zal liggen op taalvariatie in tijd, ruimte en samenleving. FIDLR wil graag aansluiten bij andere lopende en komende initiatieven voor taalgebaseerde infrastructuur in Nederland en Europa.

Diverse wetenschappers presenteerden het eigen onderzoek en de wensen en eisen waaraan een infrastructuur zou moeten voldoen. Overeengekomen werd ten slotte om



KARINA VAN DALEN

Demonstratie van het nieuwste model van Nokia tijdens het afsluitende diner

behulp van gedeelde, massale computerfaciliteiten (het 'grid'). In dat kader werd ook CLARIN gepresenteerd, het project dat streeft naar een infrastructuur voor taaltechnologie. (zie e-data maart 2007) (PB)

www.ecl.oulu.fi/dh2008/index.html

www.life.ac.uk <http://paulcourant.net/about/>
www.ncdd.nl www.alliancepermanentaccess.eu

FIDLR: Taalvariatie in Nederland

de focus te richten op het centrale thema *Taalvariatie in Nederland*. Daarmee zou het project aan de ene kant een duidelijke focus krijgen en aan de andere kant dienstbaar kunnen zijn aan een zo breed mogelijk deel van de Nederlandse taalkundige onderzoeksgemeenschap. 'In de uitwerking komen zowel de benadering (geografisch, diachroon, sociaal) als de taal (talen in Nederland gesproken) aan bod' aldus directeur Hans Bennis van het Meertens Instituut.

Duidelijk werd ook dat FIDLR niet gaat over technische vernieuwing – daar is bijvoorbeeld CATCH voor – maar over vernieuwing van het onderzoek. Nog niet herkende taalkundige patronen zouden dankzij de FIDLR infrastructuur zichtbaar moeten worden. (Douwe Zeldenrust)

Zeer brede vraag naar Geodata

Er is in Nederland een zeer brede vraag naar geodata. Dat was een van de conclusies die getrokken kon worden na afloop van de studiemiddag 'Geodata op de kaart gezet', op 20 juni. De middag was georganiseerd door DANS in samenwerking met de faculteit Geowetenschappen van de Universiteit Utrecht en Geonovum, de organisatie die werkt aan de toegankelijkheid van publieke geo-informatie.

Uit een eerdere verkenning van DANS was naar voren gekomen dat het aanbod heel divers is, en de toegang tot de data voor wetenschappelijk onderzoekers vaak lastig. Het doel van de studiemiddag was dan ook om producenten en gebruikers van geodata bij elkaar te brengen, en te horen wat nu precies de wensen en belemmeringen zijn.

De middag bood onder meer presentaties van NWO-voorzitter Peter Nijkamp en Pieter Hooimeijer, hoogleraar Sociale Geografie aan de Universiteit Utrecht. Arda Riedijk, werkzaam bij Geonovum, verzorgde een presentatie over EduGIS, een project om het gebruik van geodata in

www.dans.knaw.nl/nl/dans_symposia/geo/

Discussie over Digital Author Identification

Veel reacties kwamen er deze zomer op een commentaar in The Lancet over een uniek identificatienummer voor onderzoekers. Het initiatief van wetenschappelijk informatiebedrijf Thomson voor een *Researcher ID* verdient steun, aldus de jonge Maastrichtse onderzoekers Jochen Cals en Daniel Kotz. De auteurs en hun universiteit waren blij met de belangstelling van het vermaarde wetenschappelijke tijdschrift. 'Je wordt niet wekelijks benaderd door The Lancet' citeerde een persbericht, dat werd opgepikt door Hester van Santen die er in NRC Handelsblad een berichtje over schreef: 'Geen namen, rugnummers!'.

'De mens wordt tot nummer gereduceerd' viel al snel te lezen op de discussielijst van informatieprofessionals NEDBIB. Spoedig bleek echter dat Cals en Kotz zich hadden vergist met hun vraag in de NRC, waarom iets dergelijks in de twintigste eeuw nog niet bestond. Het bestond in Nederland al wel en het heet DAI: Digital Author Identification. Directeur Wim Liebrand van de universitaire ict-koepel SURF stuurde een brief aan de krant. Sinds 2005 wordt er al aan zo'n nummer gewerkt door twee universiteiten plus de KNAW en het informatiebedrijf OCLC/PICA, meldde hij. Eind 2006 is begonnen met de

Nieuw project rond veteranenverhalen

Ook kwalitatief onderzoek kan heel goed worden gedaan op basis van bronnen die door anderen zijn verzameld. Dat wil het data-instituut DANS (Data Archiving and Networked Services) aantonen met het project *The Veteran tapes*, dat deze dagen van start gaat.



Zeven onderzoekers uit verschillende disciplines gaan op basis van dezelfde set interviews een publicatie verzorgen, ieder vanuit de invalshoek van het eigen specialisme. De bron waaruit ze allemaal putten is de interviewcollectie van het Veteraneninstituut in Doorn. Dat instituut is al sinds 2006 bezig met een grootschalig project dat na vier

uitrol over elf universiteiten plus de twee wetenschappelijke koepels. De instellingen hebben aan hun eigen onderzoekers een DAI toegekend en nu werken KNAW en SURF er hard aan om alle onderzoekpublicaties aan een DAI te koppelen, aldus Liebrand. In de wetenschapportal

jaar duizend interviews met Nederlandse veteranen moet hebben opgeleverd. Tot nu toe zijn tweehonderd interviews beschikbaar gekomen, die door het instituut worden gezien als een 'potentieel rijke bron voor onderzoekers'. DANS denkt er ook zo over. Het wil met *The Veteran Tapes* de noodzaak en meerwaarde onder de aandacht brengen van ontsluiting en

NARCIS zijn de nummers al te vinden. Ook DANS (Data Archiving and Networked Services) is bezig de DAI aan zijn datacollecties toe te voegen, blijkt bij navraag. (MdG)

@ www.thelancet.com
www.researcherid.com
www.narcis.info

hergebruik van kwalitatieve data.

De bedoeling is dat er aan het project onderzoekers deel zullen nemen vanuit de disciplines mondelinge geschiedenis, militaire geschiedenis, militaire psychologie/sociale psychologie, psychologie, antropologie, sociologie en taalkunde. Op basis van de voorkeuren van de deelnemende onderzoekers zullen 25 interviews worden geselecteerd en getranscribeerd. Van deze kerncollectie maken alle onderzoekers gebruik voor hun onderzoek. Het project moet leiden tot een publicatie en een symposium in het najaar van 2009. (Marion Wittenberg)

@ www.veteraneninstituut.nl
www.dans.knaw.nl

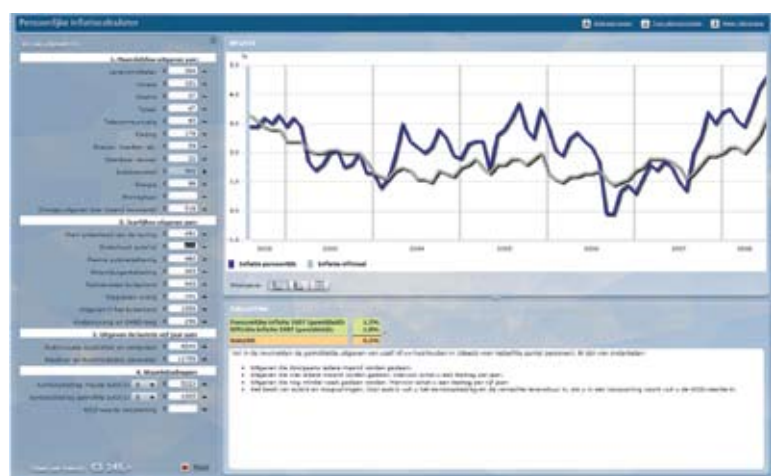
Je eigen inflatie berekenen

Sinds april van dit jaar staat de persoonlijke inflatiecalculator op de website van het Centraal Bureau voor de Statistiek (CBS). Hiermee kan iedereen in Nederland de gemiddelde jaarlijkse prijsontwikkeling berekenen van zijn eigen consumptiepatroon. Deze ontwikkeling wordt vervolgens vergeleken met de officiële inflatie.

Het CBS publiceert elke maand het inflatiecijfer over de afgelopen maand. Dit cijfer wordt berekend door de jaarlijkse prijsontwikkeling te meten van het consumptiepatroon van een *gemiddeld* Nederlands huishouden. Geen enkel huishouden heeft echter een gemiddeld consumptiepatroon. Daardoor is het mogelijk dat de gemiddelde prijsstijging die een huishouden of

individueel ervaart afwijkt van de officiële inflatie.

Nadat de eigen bestedingen op maand- of op jaarbasis zijn ingevuld, berekent de calculator de persoonlijke inflatie. Als in een huishouden veel auto wordt gereden en dus relatief veel wordt uitgegeven aan benzine, dan hebben de effecten van de prijsontwikkeling van autobrandstof meer invloed dan gemiddeld. Nu zijn de brandstofprijzen de afgelopen tijd zeer sterk gestegen, zodat de 'persoonlijke' inflatie van een huishouden waarin veel auto wordt gereden vanaf halverwege 2007 flink hoger zal uitvallen. De figuur toont het effect van sterk verhoogde uitgaven aan autobrandstof, terwijl alle overige uitgaven gelijk blijven. (Gert-Jan van Steeg)



@ www.cbs.nl/nl-NL/menu/themas/prijzen/cijfers/extra/pci.htm

CBS sluit eerste buitenlandse instelling aan met remote access



Claudia Cianfarani en Luigi Di Gennaro, medewerkers van het Italiaanse statistiekbureau (ISTAT), met op het bureau de vingerafdruklezers

In het Italiaanse Bergamo is onlangs de eerste buitenlandse *remote access* aansluiting op de gegevens van het Centraal Bureau voor de Statistiek (CBS) in gebruik genomen. De Universiteit van Bergamo had de primeur van de nieuwe voorziening.

'Zowel het Europese statistiekbureau Eurostat als de Europese Commissie wil dat over heel Europa *remote access*-verbindingen komen,' zegt projectleider Jan Mol van het Centrum voor Beleidsstatistiek van het CBS. 'Eigenlijk is het goed vervullen van de coördinatietaken die daarbij horen een grotere uitdaging

dan het zetten van de noodzakelijke technische stappen.' De techniek is inmiddels zo uniform dat het niet uitmaakt of een *remote access*-verbinding in Nederland of in het buitenland staat.

Om een en ander in een stroomversnelling te brengen heeft de Europese Commissie een tender uitgeschreven, waarin om een Europese *remote access*-voorziening wordt gevraagd. 'Het CBS heeft daarop ingeschreven,' aldus Mol. 'We willen bij dit soort projecten graag het voortouw nemen, want wij hebben hier al de nodige ervaring

DAREnet in NARCIS

Sinds juni van dit jaar is DAREnet, de toegangspoort naar alle open access publicaties van Nederlands universiteiten, onderdeel van de KNAW-portal NARCIS. DAREnet is sinds 2004 opgebouwd als netwerk van de universitaire repositories. NARCIS is de wetenschapportal van de KNAW. Naast de ruim 150 duizend full text beschikbare publicaties van DAREnet, waaronder ruim 20 duizend proefschriften van de voormalige Nationale Proefschriftensite, biedt NARCIS informatie over onderzoekers en lopend onderzoek. Ook naar de 2100 datasets uit het archiefsysteem EASY van DANS kan via NARCIS worden gezocht. De portal biedt nu ook een selectiemogelijkheid om te kunnen onderscheiden tussen datasets, publicaties en andere uitingen.

Assessment datakeurmerk mogelijk

Eind september wordt het assessment document voor het Datakeurmerk gepresenteerd op de iPRES conferentie in Londen. Daarmee komt het keurmerk, dat is ontwikkeld door DANS (Data Archiving and Networked Services) beschikbaar voor breder gebruik. 'De volgende stap is te komen tot een certificering procedure', aldus adjunct-directeur Henk Harmsen, die de presentatie in Londen voor zijn rekening neemt. DANS heeft sinds 2005 gewerkt aan het datakeurmerk, dat minimumeisen stelt voor (wetenschappelijke) data-archieven. In overleg met verschillende archieven en wetenschappelijke instellingen is eind 2007 een eerste versie in beperkte kring rondgestuurd. Begin dit jaar werd het *Data Seal of Approval* internationaal gepresenteerd. DANS werkt nu aan de internationalisering van het keurmerk door een assessment groep samen te stellen uit internationale data experts.

www.datakeurmerk.nl of
www.datasealofapproval.org

mee. Behalve in Denemarken is er in de meeste Europese landen nog een aantal barrières te nemen, soms op het gebied van wetgeving en soms meer gevoelsmatig. Het besef moet daar nog groeien dat *remote access* mogelijk is en dat het werkt'. Met het oog op de toekomst heeft het CBS plannen om de documentatie van bestanden ook in het Engels aan te bieden.

Sinds eind 2006 biedt het statistiekbureau aan onderzoekers de mogelijkheid om toegang tot microdata te krijgen zonder dat zij daarvoor naar het CBS hoeven te komen. Daartoe wordt op locatie een werkstation geïnstalleerd met vingerafdruklezers. De microbestanden blijven op een daarvoor ingerichte server bij het CBS. Via een internetverbinding op de werkplek van de onderzoeker kan hij op afstand analyses uitvoeren en de resultaten op zijn scherm bekijken. Inmiddels maken 38 Nederlandse instellingen gebruik van *remote access*. (LS)

Digitale Statenvertaling: vrijwilliger helpt wetenschap PETER BOOT

Een groep van meer dan honderd vrijwilligers onder leiding van Nicoline van der Sijs digitaliseerde de eerste druk van de Statenbijbel (1637). Oud-KNAW-president Frits van Oostrom presenteerde op 22 mei het resultaat. Het project is nog niet voorbij: met deelprojectleider Hans Beelen werkt Van der Sijs nog aan een aantal andere bijbelversies. *e-data&research* sprak met haar over de rol van de vrijwilligers en het belang van het project voor de wetenschap.

‘De motivatie van de vrijwilligers en de werkinzet is enorm, letterlijk onbetaalbaar’, benadrukt de projectleider. De inzet van de vrijwilligers kan volgens haar niet genoeg worden gewaardeerd. Daarbij is een goede organisatie van hun werk belangrijk. Van der Sijs toont zich vooral opgetogen over de snelheid waarmee is gewerkt: in minder dan een jaar zijn de 2.6 miljoen woorden overgetypt en nagekeken. ‘Dat kan doordat de vrijwilligers zich serieus genomen voelen’. Een duidelijke streefdatum en heldere afspraken motiveren mensen om aan de slag te gaan. Dankzij die goede afspraken blijkt het mogelijk om niet alleen de Statenvertaling van 1637 te digitaliseren, maar daarnaast nog een hele serie andere vertalingen. Van der Sijs’ organisatietalent werd tijdens de presentatie van het project geprezen door Herman Wiltink: ‘Ze heeft alles goed doordacht en op de lange termijn gepland, met aandacht voor het allerkleinste detail’.

De groep van vrijwilligers die zich voor het project aanmeldde was enkele malen groter dan verwacht, zodat Van der Sijs zich de luxe van een wachtlijst kon permitteren. De meeste vrijwilligers zijn niet speciaal godsdienstig gemotiveerd, aldus van der Sijs, maar eerder door belangstelling voor taal en cultureel erfgoed. Qua beroep was de groep vrijwilligers zeer gemêleerd. Alleen voor de correctoren heeft Van der Sijs vrijwilligers gekozen die beroepsmatig of op een andere manier affiniteit hadden met het correctiewerk.

Inmiddels is behalve de Statenvertaling van 1637 ook de Delftse bijbel uit 1477 digitaal beschikbaar, net als de eerste Friese Bijbelvertaling (uit 1943). In september volgt waarschijnlijk de lutherse bijbel (1648), in december de zgn. Deux-Aesbijbel (1562), volgend jaar volgt nog een aantal. Zo ontstaat een corpus aan bijbelvertalingen dat voor taalkundig onderzoek van groot belang kan zijn.

In haar artikel ‘Het boek der boeken nu ook digitaal’ (link onderaan dit artikel) somt Van der Sijs een aantal mogelijkheden op. De doorzoekbaarheid van de teksten is van belang voor het onderzoek naar de ontwikkeling van de Nederlandse woordenschat. Het wordt ook mogelijk om teksten te vergelijken op stijl en woordgebruik: zowel de verschillende bijbelvertalingen, maar ook tussen de bijbeltekst en literaire teksten. Daarmee zou iets duidelijk kunnen worden over de manier waarop de bijbelvertaling het literair Nederlands heeft beïnvloed. Er kunnen ook praktische toepassingen zijn: de beschikbaarheid van een groot corpus in zestiende- en zeventiende-eeuwse spelling maakt het mogelijk OCR (tekstherkenning) te verbeteren, omdat programma’s daarvoor gebruik maken van bekende spellingen.

Hoe kunnen onderzoekers bij de tekst? Op het moment zijn de gedigitaliseerde teksten beschikbaar



NFP PHOTOGRAPHY

Nicoline van der Sijs met de Statenbijbel

op de sites van het Instituut voor Nederlandse Lexicologie (INL), de DBNL en het Nederlands Bijbelgenootschap. De tekst is doorzoekbaar via www.biblia.net. Daar kunnen ook verschillende bijbelversies worden vergeleken. Deze mogelijkheden zijn echter beperkt tot de eigenlijke Bijbeltekst. De Kanttekeningen van de vertalers blijven hier buiten beschouwing. Het INL werkt aan andere mogelijkheden voor onderzoekers,

@ <http://www.inl.nl/images/stories/bijbels/statenvertaling1637/>

zoals een mogelijkheid om frequentielijsten af te leiden of concordanties te maken (lijsten van zoekwoorden in context). Onderzoekers die meer willen dan dat kunnen ook de hele tekst ter beschikking krijgen, maar moeten dan een contract tekenen dat ze de tekst alleen voor onderzoeksdoeleinden zullen gebruiken. De eerste belangstellenden hebben zich al gemeld.

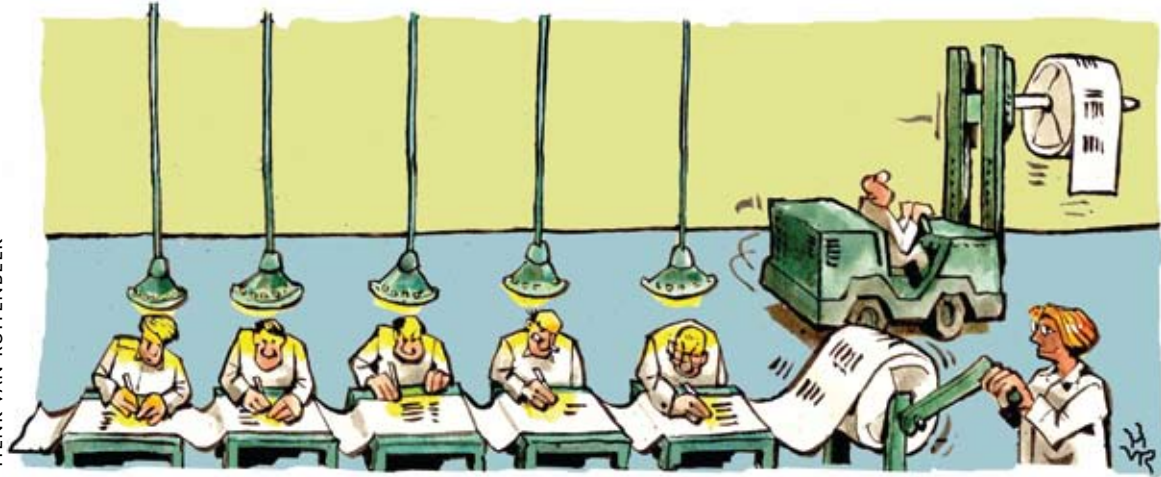
TestweekLab: een collaboratory met psychologische testdata

Al sinds veertig jaar organiseert de afdeling Psychologie van de Universiteit van Amsterdam (UvA) elk jaar de ‘testweek’. Vorig jaar ging een project van start om de grote hoeveelheid data die dat opleverde te ontsluiten.

De testweek is een propedeuseonderdeel van de opleiding Psychologie, waarin de psychologiestudenten worden getest. De naam dateert nog uit de tijd dat alle tests in één week werden gedaan; tegenwoordig vinden ze gedurende meerdere weken plaats.

Als methode van dataverzameling heeft hij twee unieke eigenschappen: er zijn zeer veel data beschikbaar en een aantal vragenlijsten wordt elk jaar afgenomen zodat van grote contuïteit kan worden gesproken.

In november 2007 is het project TestweekLab van start gegaan. Daarin is een begin gemaakt met het ontsluiten van data uit de testweek en het maken van een collaboratory daaromheen. Met de collaboratory worden twee doelen nagestreefd. Ten eerste wordt geprobeerd de heersende geslotenheid in het delen van elkaars datasets open te breken. Het online aanbieden van deze schat aan testdata moet in dat opzicht een voorbeeldfunctie vervullen. Het tweede doel is het ontsluiten van de grote hoeveelheid testdata, waarbij deze op een duurzame manier in het systeem worden opgeslagen.



Actieve community

Het project TestweekLab is onderdeel van het thema collaboratories binnen het programma SURFshare. De collaboratory die is gebouwd, ook genaamd TestweekLab, heeft als uitgangspunten toegang te bieden tot elkaars testdata en samen te werken op het gebied van psychologisch onderzoek. In eerste instantie zal TestweekLab speciaal voor data uit de testweek worden gebruikt. In de toekomst zullen de opgeslagen testdata ook voor onderzoekers buiten de UvA toegankelijk zijn. Omdat TestweekLab een grote dataverzameling zal ontsluiten, waarmee zowel verticale

(meerdere jaren testdata) als horizontale doorsnedes (meerdere tests in een jaar) gemaakt kunnen worden, wordt erop gerekend dat veel externe onderzoekers belangstelling zullen hebben voor het gebruik ervan.

Om het eerste doel te verwezenlijken – meer openheid rond data uit psychologisch onderzoek – is het belangrijk dat een actieve *community* ontstaat rondom de collaboratory. Een succesfactor voor een collaboratory is volgens Wikipedia de bereidheid van gebruikers om een gemeenschappelijk doel te verwezenlijken. Daarnaast is het cruciaal dat ze een *incentive* geboden krijgen

voor het leveren van een bijdrage. In het geval van TestweekLab moeten de testdata als trekpleister dienen voor actieve deelname. TestweekLab biedt online functionaliteiten die samenwerking rond de data mogelijk maken, zoals emaillijsten, een wiki en gedeelde documenten. De afdeling Psychologie van de UvA zal een eerste vulling van deze onderdelen verzorgen, waarmee de meerwaarde van het online aanbieden van datasets expliciet gemaakt wordt.

Uitbouwen met meer testdata

De collaboratory is gebouwd met behulp van *Sakai* en *Fedora*. Sakai is

een online *Collaboration and Learning Environment*. Het is een open source product dat een *plugin*-architectuur kent, wat wil zeggen dat externe partijen gemakkelijk kunnen meewerken aan uitbreiding van functionaliteit door modules te ontwikkelen (ook wel *Sakai tools* genaamd). Deze modules kunnen door beheerders van een Sakai-installatie optioneel geïnstalleerd worden. Voor TestweekLab wordt gebruik gemaakt van de functionaliteiten voor samenwerking die Sakai standaard aanbiedt. Daarnaast is speciaal voor TestweekLab een Sakai tool ontwikkeld die een interface biedt naar de Fedora repository, waarin de testdata worden opgeslagen. Deze worden in de vorm van SPSS-bestanden in de repository opgenomen. Met de Sakai tool kan de gebruiker zoeken naar testdata, browsen, en de metadata van de testen bekijken. Ook kan men testen downloaden en zelf testdata toevoegen.

In de nabije toekomst zal TestweekLab door de psychologen van de UvA verder uitgebreid worden met meer testdata. Voor de opslag van de testdata zal worden geprobeerd om aan het DANS Datakeurmerk te voldoen. Verder is de hoop erop gericht dat er onder psychologen voldoende belangstelling zal zijn om bij te dragen aan het online samenwerken naar aanleiding van psychologische testdata.

Historicus en CBS-veteraan Jacques van Maarseveen: 'Als je hier werkt ben je in de eerste plaats op de actualiteit gericht'

MARTIJN DE GROOT

'Het historisch geweten van het CBS'. Zo noemde socioloog Paul Klep hem bij de presentatie van het boek 'Twee eeuwen Nederland geteld' waarvan hij redacteur en medeauteur was. Zelf fronst Jacques van Maarseveen relativerend bij het horen van die kwalificatie. Maar een goed gesprek met hem geeft weinig aanleiding om er iets op af te dingen.

Het Centraal Bureau voor de Statistiek, het toneel van vrijwel zijn gehele loopbaan, verkeert in een staat van lichte ontreddeuring tijdens het interview. Dat is bijzonder voor het statige instituut met zijn inmiddels meer dan honderdjarige geschiedenis, maar een complete verhuizing is hier dan ook geen alledaags werk. Nog maar een paar weken scheiden de verzamelde statistici van die verhuizing. Dat betekent: overal dozen, en geen boek of cd is op de plaats waar je het hoopt te vinden. Van Maarseveen, die al sinds de jaren negentig niet meer in actieve dienst is bij het cijferconcern, heeft hier nog wel een bescheiden kamer en wie hem wil bellen heeft een goede kans op zijn vaste CBS-telefoonnummer. Het tekent zijn verknochtheid aan het instituut, maar vooral aan de projecten waaraan hij mocht werken sinds hij halverwege de jaren negentig met vervroegd pensioen ging.

'Budgetonderzoek was het moeilijkste dat er was'

Hoe het met zijn huisvesting verder zal gaan in het nieuwe onderkomen van het CBS in Leidschenveen weet Van Maarseveen nog niet. 'In het nieuwe gebouw zitten de mensen met vier op een kamer. Ze zijn aan het nagaan of ik nog ergens geplaatst kan worden, maar ik lever natuurlijk geen directe bijdrage meer aan het dagelijks statistisch werk. Ik leg momenteel als mederedacteur en -auteur de laatste hand aan een publicatie over de geschiedenis van de statistiek in Nederland en doe nog wat hand- en spandiensten voor het digitaliseringswerk waar Peter Doorn ook mee bezig is.' DANS-directeur Doorn en hijzelf waren hoofdrolspelers in het Volkstellingproject dat met medewerking van nog veel meer mensen sinds 1997 een aantal fases doorliep, en nog steeds niet helemaal is afgesloten. Het laatste hoogtepunt was de presentatie van het boek 'Twee eeuwen Nederland geteld', zo dik als een kindervuist, met artikelen van vooraanstaande wetenschappers op basis van onderzoek met de gedigitaliseerde Volkstellingencijfers. Van Maarseveen was redacteur en schreef er zelf drie bijdragen in.

'Peter Doorn kwam in 1997 voor het eerst langs bij het CBS met een

voorstel: kunnen we niet eens iets aan die volkstellingen doen? Er lagen natuurlijk tellingen vanaf 1795 met enorme mogelijkheden die niet konden worden benut als ze niet gedigitaliseerd zouden worden. En wat ook een rol speelde: ze werden heel frequent geraadpleegd, waardoor hun conditie zienderogen achteruit ging. Ze moesten dus hoognodig worden geconserveerd. Nu kwam op dat moment net het eeuwfeest van het CBS eraan, dat was in 1899 opgericht. Het zou dus wel mooi zijn om te beginnen met de telling van 1899 en zo is het ook gegaan. Alle gegevens van die telling zijn toen, cijfer voor cijfer, met de hand ingetikt. En daarnaast was er een zogenaamde mediumconversie, die inhield dat de publicaties van de tellingen vanaf het begin tot en met 1971 werden gescand en digitaal beschikbaar kwamen.' Het geheel werd op cd-rom uitgegeven ter gelegenheid van het jubileum. 'Hartstikke mooi', prijst van Maarseveen terwijl hij de schijf vergeefs zoekt op een van de laatste nog niet ingepakte boekenplanken.

In hetzelfde jaar verscheen een eerste publicatie met werk op basis van de vrijgekomen gegevens: 'Nederland een eeuw geleden geteld'. Wat hij toen nog niet wist, was dat de verschillende vervolgstadia, tot en met de complete digitalisering inclusief de virtuele tellingen van na 1971, hem nog bijna een decennium bezig zouden houden. 'Ik heb het met heel veel plezier gedaan' zegt hij nu, terugkijkend op de periode die buiten zijn actieve loopbaan viel maar niet weinig zal hebben bijgedragen aan de eretitel die hem door Klep werd verleend.

Binnen die loopbaan was er vaak niet veel tijd om zijn oorspronkelijke discipline te beoefenen. Afgestudeerd als economisch en sociaal historicus, kwam hij aan het eind van de jaren zestig eerst een paar jaar te werken op het toenmalige Economisch Instituut voor het Midden- en Kleinbedrijf. 'Ik kwam direct in het onderzoek terecht, consumentenonderzoek was dat, en daar ging ik mee door toen ik in 1971 naar het CBS kon om de afdeling Consumptie en Sparen op te zetten. Dat was een periode van groei voor het CBS en ik heb later nog verschillende andere onderzoeken op mogen zetten. Het onderzoek naar consumenten-



JOHN THURING

vertrouwen bijvoorbeeld, dat heette toen nog het consumenten conjunctuuronderzoek. Dat deden we op verzoek van Brussel. Ik weet nog dat de vraag toen was: moeten we het uitbesteden of gaan we het zelf doen? Het werd dat laatste, en het is nooit meer bij ons weggegaan,' peinst hij in de week waarin zijn voormalige werkgever juist alarmerende cijfers heeft vrijgegeven over het dalend consumentenvertrouwen in Nederland. 'Maar ook budgetonderzoek bijvoorbeeld, naar gezinsuitgaven. Dat was het moeilijkste onderzoek dat er was omdat degenen die voor de gegevens moeten zorgen echt een heel jaar lang volledig moesten meewerken. Niet alleen door hun dagelijkse uitgaven te melden, maar ook met hun verzekeringspremies, gas en water, huur, alles. Zulk onderzoek was in de jaren dertig al gehouden, maar later weer beëindigd. Toen ik bij het CBS kwam was *alle* kennis erover verdwenen. We hebben het toen helemaal vanaf de grond weer op moeten zetten'.

Ook het Woningbehoefteonderzoek, dat onlangs van het ministerie

van VROM weer terugkwam bij het CBS, viel in de jaren zeventig onder de afdeling waaraan Van Maarseveen leiding gaf. In de tussentijd ging het nog twee keer op en neer. 'In 1973, bij de splitsing van het CBS tussen Voorburg en Heerlen, ging het woningonderzoek naar het ministerie. In 1977 kwam het weer terug bij het CBS. Rond de eeuwwisseling ging het nog eens naar VROM en nu is het dus weer bij het CBS terecht gekomen'.

'Vaak zijn de uitkomsten wel gepubliceerd, maar zijn de onderliggende bestanden ook onderzoekbaar?'

Van afdelingschef werd hij leidinggevende van verschillende hoofdafdelingen, waaronder die voor Rechtsbescherming en Veiligheid, de Statistieken van Arbeid en Lonen, Financiële Statistieken, en ten slotte werd hij sectorhoofd Overheid tevens plaatsvervangend divisiedirecteur. Dat waren geen jaren om veel energie te steken in de beoefening van zijn oorspronkelijke vak, de sociale

en economische geschiedenis. 'Als je hier op die manier werkzaam bent dan ben je meestal heel erg op de actualiteit gericht. Maar ik ben toch de geschiedenis trouw gebleven, door aan een biografie te werken van Nicolaas Gerard Pierson, een van de grote economen die ons land heeft voortgebracht en die bovendien als minister nog aan de wieg van het CBS heeft gestaan.' Van Maarseveen promoveerde erop in 1981. Twee jaar later vroeg de Nederlandsche Bank om een bronnenuitgave over de bankier en politicus te maken. 'Ik heb daar een tijdlang een dag in de week aan gewerkt, maar dat werd me toch een klus!' Tien jaar en een overplaatsing naar een minder hectisch dienstonderdeel waren er voor nodig om de studie af te kunnen ronden. 'Drie zúlke dikke pillen kwamen eruit voort', zo beschrijft hij het eindresultaat dat in 1993 in druk verscheen. En nog een elegant klein boekwerkje, had hij eraan toe kunnen voegen want uit zijn studie van de negentiende-eeuwse econoom kwam nog een publicatie voort ter gelegenheid van het honderdjarig bestaan van de Centrale Commissie voor de Statistiek, het inhoudelijk sturings- en controleorgaan van het CBS. Het zou de opmaat blijken voor een reeks publicaties met een historische inslag waaraan de statisticus nadien meeschreef of anderszins een bijdrage leverde. Behalve de verschillende uitgaven rond de Volkstellingen hoort daarin bijvoorbeeld ook nog het eigenlijke jubileumboek van het CBS thuis, 'Welgeteld een eeuw', dat in het Nederlands en Engels verscheen.

Dat aan de gedreven CBS-er min of meer automatisch het etiket 'historisch' kwam te kleven is dus niet verwonderlijk. Tegelijk werpen zijn recente activiteiten om oude bestanden van het CBS aan de vergetelheid te ontrukken licht op het predicaat 'geweten' dat daaraan werd verbonden. 'Er zijn heel veel onderzoeken waarvan de uitkomsten wel zijn gepubliceerd, maar waarvan de onderliggende bestanden niet direct traceerbaar meer zijn of toegankelijk, zoals van de Arbeidskrachtentellingen. Ze zijn bewaard in een Bestand Opslag Systeem, maar dat wil niet zeggen dat ze ook nog beschikbaar zijn voor onderzoekers die ze nog willen inzien, of dat ze nog onderzoekbaar zijn. Datzelfde geldt voor de metadata: hoe zijn de gegevens van deze statistieken verzameld en opgeslagen, hoe zijn ze bewerkt, hoe zijn de variabelen benoemd? We zijn nu samen met DANS aan het bekijken: welke bestanden zijn écht belangrijk om te behouden en open te maken? Die willen we boven tafel krijgen.'

Van Gogh digitaal

MARIANNE PEEREBOOM

Eén van de topcollecties uit het Nederlandse papieren erfgoed is de correspondentie van Vincent Van Gogh, die behalve begenadigd kunstenaar ook een enthousiast briefschrijver was. Daarin is een schat aan informatie te vinden, onder andere over zijn artistieke ontwikkeling en de ontstaansgeschiedenis van zijn werken. Het Van Gogh Museum en het Huygens Instituut bereiden een nieuwe wetenschappelijke editie van de brieven voor, het resultaat van langlopend, intensief onderzoek.

In oktober 2009 zullen de uitkomsten van dit onderzoek worden gepresenteerd in een reeks van activiteiten en publicaties. Een rijk geïllustreerde gedrukte editie verschijnt in het Nederlands, Frans en Engels. Daarnaast komt er een wetenschappelijke digitale editie uit, eveneens een gezamenlijk project van het Van Gogh Museum en het Huygens Instituut. In deze web-editie worden naast de geannoteerde tekst ook de digitale facsimile's van alle brieven opgenomen en uitgebreidere annotaties dan in de gedrukte uitgave mogelijk is. Resultaten van het onderzoek zullen ook te zien zijn in een tentoonstelling in het Van Gogh Museum, waarin de kunstenaar zelf aan het woord komt. De bezoeker kan zien wat Vincent



Brief van Vincent van Gogh aan Theo van Gogh (met 1 schets), Van Gogh Museum, Amsterdam

van Gogh over zijn eigen werk dacht en schreef, en kan zijn schilderijen vergelijken met schetsjes ervan die hij in zijn brieven opnam of mees-tuurde.

Dit jaar wordt met subsidie van Metamorfoze de gehele collectie *Brieven en Documenten* van het Van Gogh Museum geconserveerd door middel van *preservation imaging*. Daarmee ontstaat een beeldarchief waaruit voor de gedrukte en digitale editie geput wordt. De collectie *Brieven en Documenten* bevat naast de correspondentie van Van Gogh (samen zo'n 800 brieven, 1200 afzonderlijke

briefbladen) ook circa vierhonderd familiefoto's en overige (familie-) correspondentie, notitieboekjes van Theo van Gogh, broer van de kunstenaar, en dagboeken van diens echtgenote Jo van Gogh-Bonger, lijsten met schilderijen voor tentoonstellingen, uitnodigingen, nota's en kwitanties, programmaboekjes, voordrachten, toespraken en krantenknipsels. De kern van deze collectie komt uit de nalatenschap van Theo van Gogh, Jo van Gogh-Bonger en hun zoon ir. Vincent Willem van Gogh.

Naast de gedrukte en digitale edities wordt uit de gedigitaliseerde *Brieven en Documenten* ook een collectie samengesteld voor het *Geheugen van Nederland* ([www.geheugenvan-](http://www.geheugenvan-nederland.nl)

Vlaanderen: onderzoeksproject multimediale data

'Bewaring & Ontsluiting van Multimedia in Vlaanderen' (BOM-VL) is een onderzoeksproject om strategieën te ontwikkelen voor het op lange termijn digitaal bewaren en ontsluiten van audiovisueel erfgoed in Vlaanderen. Het project loopt van januari 2008 tot juni 2009 en wordt ondersteund door de Vlaamse regering.

In Vlaanderen zijn er duizenden uren aan audio- en videomateriaal, of 'time-based media', opgeslagen op diverse analoge dragers. Deze collecties bevinden zich niet alleen bij de omroepen maar ook bijvoorbeeld bij culturele organisaties, erfgoedinstellingen, overheidsinstellingen. Volgens UNESCO behoort dit au-

diovisueel materiaal tot het meest kwetsbare cultureel erfgoed.

Het BOM-VL-project wil een langetermijnstrategie ontwikkelen voor het digitaal bewaren en ontsluiten van deze collecties in Vlaanderen. Aan culturele zijde neemt een brede waaier van organisaties deel: de Boekentoren (Universiteit Gent), Instituut voor Beeldende, Audiovisuele en Mediakunst (BAM), Vlaams Theater Instituut (Vti), Muziekcentrum Vlaanderen en FARO (Vlaams steunpunt voor cultureel erfgoed). Uit de mediasector participeren de Vlaamse openbare omroep VRT, en ook de commerciële en regionale omroepen. Het IBBT (Instituut voor Breedband Technologie) en het VRT Medialab zorgen voor de wetenschappelijke omkadering.

Binnen het BOM-VL project wordt een gemeenschappelijk innovatieplatform opgezet dat toelaat om stelselmatig bestaande kennis over archiveren en ontsluiten van multimediale data te toetsen aan de realiteit, samen met de onderzoekresultaten uit de verschillende werkpakketten van dit project. Het is verder de bedoeling van de partners om alle resultaten van het project beschikbaar te stellen aan alle spelers in het Vlaamse en Europese cultuur-, erfgoed- en mediaveld. (Jeroen Walterus)

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



[nederland.nl](http://www.geheugenvan-nederland.nl)). Daarmee wordt het materiaal voor een breed publiek ontsloten als onderdeel van het educatieve programma dat rond de brieven wordt ontwikkeld.

Via deze verschillende activiteiten en publicaties komt een unieke en kwetsbare collectie uit het Nederlandse erfgoed wereldwijd ter beschikking van verschillende geïnteresseerden: van kunsthistorici en andere wetenschappers tot iedereen die belangstelling heeft voor het leven en werk van Vincent van Gogh.

IMPACT: betere OCR voor de digitale bibliotheek

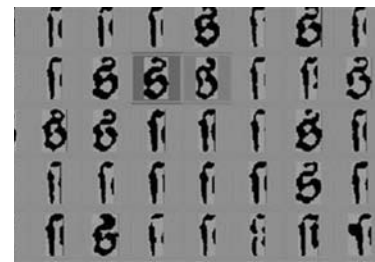
HILDELIES BALK

In januari 2008 ging het vierjarige Europese project Improving Access to Text (IMPACT) van start. In dit project werken vijftien partners uit bibliotheken, research en bedrijfsleven onder leiding van de Koninklijke Bibliotheek samen.

Doel van het project is om innovatieve oplossingen te vinden voor de verbetering van optische tekenherkenning (OCR), en methodes om het proces van massadigitalisering van gedrukte historische tekst in Europa te versnellen. De eerste resultaten zijn veelbelovend. In juni werd de eerste (interne) versie van het *overall design* van de IMPACT OCR software opgeleverd. Deze software is speciaal toegesneden op het digitaliseren van bibliotheekmateriaal. De kern is een krachtige OCR machine, ontwikkeld onder leiding van IBM Research in Haifa, die zich tijdens het OCR-proces kan aanpassen aan de tekst hij aantreft. Hij hoeft niet, zoals de bestaande OCR software, voorafgaand aan de digitalisering van ouder materiaal met moeilijk leesbare lettertypen en spellingsvarianten te worden 'getraind'. Daardoor verloopt het digitaliseringsproces sneller en goedkoper. De machine wordt geïntegreerd met een andere module van het IMPACT-project: de *Collaborative Correction*. Dit is een

uitgekiend gereedschap waarmee vrijwilligers zonder speciale kennis online kunnen meehelpen aan de correctie van OCR resultaten. Tekens waar de machine over twijfelt bij de OCR – bijvoorbeeld de oude schrijfwijze van de letter s, die er vaak uitziet als een letter f – worden automatisch naar de correctiemodule gestuurd voor beoordeling door een menselijke lezer. De feedback uit de correctie gaat direct terug naar de OCR machine, die zich aanpast aan de nieuwe informatie.

Binnen het uitgebreide verrijgingsprogramma dat door taalkundige partners in IMPACT wordt ontwikkeld, is ook voorzien in hulp van, nu wel specialistische, vrijwilligers. Het Instituut voor Nederlandse Lexicologie uit Leiden presenteerde dit voorjaar binnen het project het ontwerp van een databank met *Named Entities* (historische namen van plaatsen, personen enzovoort) die zal dienen voor de verbetering van de OCR resultaten. Hieraan gekoppeld is een online *Collaborative*



OCR herkenning van gotische s die op f lijkt

Environment, waar lijsten met eigenamen door specialisten kunnen worden toegevoegd.

De bibliotheken in het project werken hard aan de voorbereiding van de demonstratieprojecten, die begin 2009 moeten gaan draaien. Zo zal in de KB een proefopstelling komen van de IMPACT-tools met materiaal uit het KB-project Dagbladen Digitaal.

Op 6 en 7 april 2009 worden de eerste resultaten gepresenteerd op de internationale IMPACT conferentie in de Koninklijke Bibliotheek.

[@ impact@KB.nl](mailto:impact@KB.nl)
www.impact-project.eu

IBM ISRAEL SCIENCE AND TECHNOLOGY LTD



Brief van Vincent van Gogh aan Theo van Gogh (met 3 schetsen), Van Gogh Museum, Amsterdam

Taal- en Tekstdatabanken in kaart gebracht

DANS (Data Archiving and Networked Services) is dit voorjaar begonnen met het project 'Inventarisatie Taal- en Tekstdatabanken in Nederland'. Doel is om aanbevelingen op te stellen voor het opzetten, gebruiken en duurzaam bewaren van databanken op het gebied van taal- en tekst in Nederland. Daartoe worden de bestaande databanken op dit terrein geïnventariseerd. Ook worden de wensen van de Nederlandse taal- en letterkundigen ten aanzien van mogelijke verbeteringen in de data-infrastructuur van dit onderzoeksgebied in kaart gebracht. Zijn er leemtes en zo ja welke? Hoe is de situatie, gewenst en feitelijk, ten aanzien van

de langetermijn-bewaring van deze databanken?

Deze inventarisatie richt zich op alle in Nederland bestudeerde talen, maar beperkt zich wel tot bestanden die gevormd zijn in een *academische* onderzoeksomgeving. De belangrijkste werkzaamheden zijn deze inventarisatie, in samenwerking met het Europese project CLARIN (www.clarin.eu), en het afnemen van een beperkt aantal diepte-interviews met een representatieve groep onderzoekers 'in het veld'. Het project wordt in het najaar van 2008 afgesloten met een publicatie en een workshop met onderzoekers in het veld. Informatie: heiko.tjalsma@dans.knaw.nl

Workshop: data uit onderzoek vaker delen

Beleidsmakers in Nederland en in de wereld hebben al vaak uitgesproken dat het wenselijk is dat data uit wetenschappelijk onderzoek worden gedeeld, om redenen van wetenschappelijke aard maar ook om efficiënt met overheidsgeld om te springen. Onderzoekers blijken dat echter nog helemaal niet vanzelfsprekend te vinden. Een workshop van DANS (Data Archiving and Networked Services) en SURF Foundation op 1 juli moest aan het licht brengen wie de belanghebbenden precies zijn en hoe belemmeringen om data te deponeren kunnen worden weggenomen. Een kleine twintig beleidsmedewerkers, managers en onderzoekers boog zich in Utrecht over die vragen.

Wetenschappers zijn in principe bereid om hun onderzoeksdata met anderen te delen en in sommige disciplines is dat al sinds jaren gemeengoed, werd vastgesteld. Ook waar het nog geen gemeengoed is, komen er steeds meer initiatieven om gemeenschappelijk gebruik te stimuleren. Een gemeenschappelijke houding tegenover het delen van data is bij dat proces van belang. Maar waar liggen dan toch de drempels, en hoe kunnen die in de praktijk worden weggenomen?

Een belangrijke motivatie om data niet met anderen te delen is dat dat geen wetenschappelijke waardering oplevert. Het wordt veelal ervaren als een extra inspanning waar weinig



Opleiding moet aandacht besteden aan data delen

tegenover staat, zo constateerden verschillende deelnemers. Omgekeerd zou het goed zijn als duidelijker werd welke winst deponeren oplevert. Een co-auteurschap voor de dataproducent bij publicaties was een van de suggesties. Het beter checken van de wijze waarop data zijn gedeponerd bij de uitbetaling van de laatste subsidiegelden was een andere. Hier werd vooral een rol van onderzoeksfinanciers zoals NWO gezien.

Databewaarplaatsen missen veelal een duidelijke regeling of zelfs verplichting om onderzoekers ertoe aan te zetten om hun onderzoeksdata bij hen te deponeren. Ook dat bleek, naast actievare acquisitie, een aangrijpingspunt voor verbetering. Aan de andere kant werd betwijfeld of het opleggen van verplichtingen op het gebied van deponeren wel tot positief

resultaat zou kunnen leiden. Onderzoekers laten zich daar niet graag door beïnvloeden, zo werd naar voren gebracht. Aan de andere kant zou het de moeite waard zijn te investeren in een andere houding tegenover openheid rond de eigen wetenschappelijke bronnen. Een psycholoog bracht naar voren dat de angst dat er gebreken in het bestand zitten of dat anderen met de gegevens aan de haal gaan een hinderpaal vormen voor het delen van data. ‘Onderzoekers hebben blijkbaar geen vertrouwen in de eigen data’, reageerde een andere deelnemer daarop, die de bètawetenschappen als voorbeeld aanhaalde van het overwinnen van deze schroom. Meer aandacht voor data documenteren en delen in de opleiding zagen veel deelnemers als belangrijk om deze drempel te overwinnen. (LS)

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.

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 alle beveiligde microdata van het CBS kosteloos geleverd worden: <http://www.dans.knaw.nl/nl/data/cbs/overzicht/>

Data Centraal Bureau voor de Statistiek		Periode	Via
ABR	Algemeen bedrijfsregister	1994-1999	CBS
CIS	Community Innovation Study	1996, 1998 en 2000	CBS
AO	Arbeidsongeschiktheids registratie	2007	CBS
BUS	Bijstandsuitkeringen registratie	2005-2007	CBS
BUS-TRANS	Bijstandsuitkeringen transactie bestand	2005-2007	CBS
BFS-BDS	Bijstrandsfraude en bijstandsdebiteuren statistiek	2 ^e helft 2007	CBS
SRG	Statistiek re-integratie gemeenten	2007	CBS
WW	Werkloosheidsuitkeringen registratie	2007	CBS
RIO	Regionaal Inkomensonderzoek	2001 en 2005	CBS
MBO	Middelbaar beroepsonderwijs registratie	2004-2007	CBS
HO	Hoger onderwijs	1986-2007	CBS
POLS	Permanent Onderzoek van de Leefsituatie	2001, 2005 en 2006	CBS
DO	Doodsoorzaken	2007	CBS
Data van overige instellingen			
A. ter Wal – Project Verbreding A2: opgraving vindplaats 2	2005	DANS-EASY	
E. Westerman – Archeologische Projecten in het kader van de verbreding van de A2 (11 projecten)	2005	DANS-EASY	
Prof.dr J.L. Blussé van Oud-Alblas – South China Sea trade	1681-1792	DANS-EASY	
Inspectie van het Onderwijs – Kwaliteitskaart Voortgezet Onderwijs	2008	DANS-EASY	
Ministerie van Financiën – Financieel inzicht van Nederlanders	2007	DANS-EASY	
Sociaal en Cultureel Planbureau – Emancipatie Opinions	2006	DANS-EASY	
Ineke Nagel en Harry Ganzeboom – Jeugd en Cultuur, een multi-actor panelstudie 1998-2004	1998-2004	DANS EASY	
Beveiligde Microdata			
Centraal Bureau voor de statistiek – Consumentenconjunctuuronderzoek (CCO)	2007	DANS	
Centraal Bureau voor de Statistiek – Permanent Onderzoek van de Leefsituatie (POLS)	2006	DANS	

VGI Innovatieprijs 2008

De Vereniging voor Geschiedenis en Informatica looft sinds 2003 jaarlijks een prijs uit voor de meest vernieuwende ICT-toepassing op het terrein van de geschiedenis in Nederland of Vlaanderen. Creatieve oplossingen die het vakgebied promoten worden zo aangemoedigd. Tot de vorige prijswinnaars behoren: het Amsterdams Gemeentearchief (2003), het Ename Expertisecentrum (2004), Literair Gent (2005), Teleblik (2006) en Archeonet Vlaanderen (2007).

Bijzondere websites of standalone applicaties kunnen tot en met 25 september worden genomineerd voor de Innovatieprijs 2008. Ze moeten op innovatieve wijze een bestaande of juist nieuwe doelgroep aanspreken, bestaand ICT-gereedschap vervangen, vernieuwen en verbeteren of tools aanbieden voor historisch onderzoek of onderwijs. Je mag een eigen project aanmelden,

maar ook een applicatie van anderen die je een bekroning waard vindt. Stuur een e-mail naar juryvoorzitter Yola de Lusenet (y.de.lusenet@planet.nl), vermeld naam, adres en telefoonnummer van de inzender, de naam van de applicatie en de URL (indien aanwezig) en geef een toelichting op het innovatieve aspect van de applicatie.

Op 15 oktober worden de genomineerden bekend gemaakt. De deskundige jury nomineert drie toepassingen, die tijdens de VGI studiemiddag op 11 december 2008 een korte presentatie mogen geven. Daarna wijst de jury een winnaar aan, die naast een oorkonde het recht krijgt om het speciale ‘VGI Innovatieprijs 2008’-vignet te gebruiken, en een geldprijs van 500 Euro.

www.vgi-online.nl

VGI Studiemiddag: massadigitalisering

De Universiteitsbibliotheek van Gent, een van de grootste bibliotheken in Europa, is een samenwerkingsverband aangegaan met Google Book Search en zal honderdduizenden boeken digitaliseren. Zij zijn niet de enigen. Het programma Metamorfoze (voor behoud van papieren erfgoed) en de regeling Digitaliseren met Beleid van het Nederlandse ministerie van OCW subsidiëren grootschalige digitalisering van archieven en bibliotheken. ‘Beelden voor de Toekomst’ en de Databank Digitale Dagbladen richten zich alle op grote hoeveelheden materiaal. Kortom, er wordt massaal gedigitaliseerd in erfgoedland. Toch zijn er niet over alle ervaringen juichverhalen. Sommige ervaringen zijn te kostbaar om niet gedeeld te worden. Hoe verhoudt kwantiteit zich tot kwaliteit?

Op een studiemiddag van de Vereniging voor Geschiedenis en Informatica over Massadigitalisering worden verschillende aspecten belicht van het voortraject van grootschalige digitalisering, de uitvoering

en de daarop volgende ontsluiting, archivering etc.

Sprekers zijn (onder voorbehoud):

- Tanja de Boer, hoofd fysiek Beheer, Koninklijke Bibliotheek;
- Marc Holtman, Stadsarchief Amsterdam, auteur van ‘Digitalisering ontrafeld; Technische aspecten van digitale reproductie van archiefstukken’;
- Cees Klapwijk, Projectleider DBNL (Digitale Bibliotheek voor de Nederlandse Letteren);
- Ivo Zandhuis, zelfstandig adviseur van om. archiefinstellingen en net als Marc Holtman betrokken bij het opstellen van De BASIS van Digitaal Erfgoed Nederland;
- Astrid Verheusen, projectmanager Databank Digitale Dagbladen, Koninklijke Bibliotheek.

De bijeenkomst wordt gehouden in Den Haag op 12 september, van 14.00 tot 17.00 uur in de Koninklijke Bibliotheek. De toegang is gratis, na aanmelding bij Michelle van den Berk: berk@kitlv.nl.

COLOFON

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

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

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

Aan dit nummer werkten mee: Hildelies Balk, Hans van Brussel, Peter Doorn, Jeffrey van der Hoeven, Marianne Peereboom, Laurents Sesink, Gert-Jan van Steeg, Heiko Tjalsma, Caspar Treytel, Jeroen Walterus, Marion Wittenberg, Douwe Zeldenrust.

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

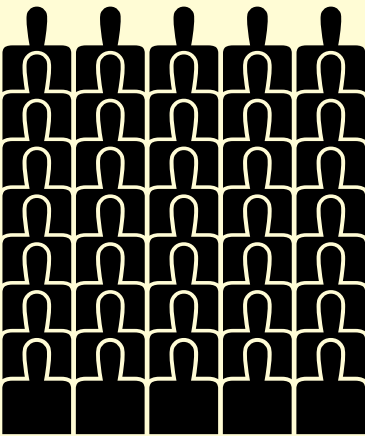
Vormgeving: Ellen Bouma

Productie: Uitgeverij Aksant, Amsterdam

Druk: Plantijn Casparie, Almere
ISSN: 1872-0374

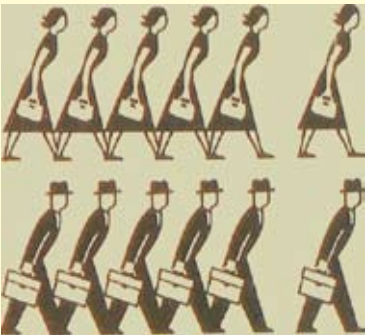
Een website voor de vader van de infographics

Steeds vaker maken de media gebruik van zogenaamde *infographics* om ingewikkelde zaken snel inzichtelijk te maken: één beeld zegt meer dan duizend woorden. Redacties van kranten, tijdschriften, radio, televisie en websites hebben er inmiddels specialisten voor in dienst als er al niet hele afdelingen voor zijn ingericht. Het past bij de beeldcultuur waarin we in toenemende mate leven. Maar is het allemaal wel zo nieuw?



Op 22 april vond in het Haags Gemeentemuseum een feestelijke bijeenkomst plaats ter gelegenheid van de lancering van www.gerdarntz.org. De website kan het beste worden omschreven als het depot van het werk van de 'vader' van de *infographics*: de Duitse kunstenaar Gerd Arntz. Samen met de Oostenrijkse wetenschapper Otto Neurath was hij in de jaren dertig en veertig van de vorige eeuw de grondlegger van de beeldstatistiek, in eerste instantie werkend vanuit Wenen en later, voor en tijdens de Tweede Wereldoorlog, vanuit Nederland en Engeland.

In Nederland raakt sindsdien de beeldstatistiek steeds meer ingeburgerd. Onderzoeksbureau NSS en het Centraal Bureau voor de Statistiek (CBS) gebruiken de afbeeldingen steeds vaker in steeds meer publicaties. In 1988 overlijdt Arntz. Na een tentoonstelling van zijn werk in de



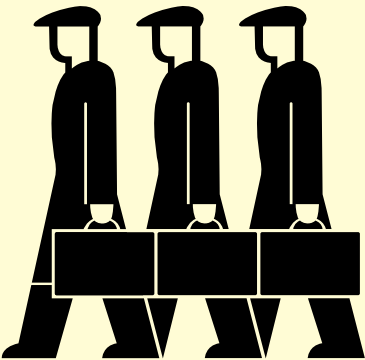
jaren zeventig is de bekendheid ervan sterk afgenomen hoewel hij in een kleine kring onverminderd populair blijft.

In het Internationaal Instituut voor Sociale Geschiedenis in Amsterdam wordt een archief bewaard van tal van documenten, waaronder correspondentie. Het beslaat de periodes 1934 – 1937 en 1946 – 1988.

Is het archief in Amsterdam zondermeer interessant voor de historici, intussen ligt er in de kelders van het Haagse Gemeentemuseum ook voor de kunstliefhebbers veel moois: in ladekasten liggen anno 2007 duizenden werken van Gerd Arntz te wachten tot er iets mee gebeurt. Ook in Den Haag loopt iemand rond met een plan het archief digitaal te ontsluiten;



geld is het probleem. De man met het idee heet Ed Annink van het Haagse ontwerpbureau Ontwerpwerk (www.ontwerpwerk.nl), hij wil uiteindelijk misschien wel alle vierduizend afbeeldingen digitaliseren en tentoonstellen op een speciale website. Hij heeft al afspraken gemaakt met allerlei belanghebbenden zoals NSS/Interview, de rechtsopvolger van de stichting waar Gerd Arntz ooit werkte, Peter Arntz (zoon van) en het Geheugen van Nederland dat wil helpen



met het scannen. Medio 2007 komt, voortvloeiend uit een contact tussen de gemeente Den Haag en het CBS, het verlossende woord: Den Haag betaalt mee en het CBS ook. Dat laatste instituut stelt wel als voorwaarde dat de gedigitaliseerde pictogrammen binnen het CBS weer in gebruik kunnen worden genomen, maar dat wordt alleen maar toegejuicht.

En zo kon het beuren dat op 22 april 2008 een bont gezelschap van liefhebbers het glas hief op de lancering van de website waarop het werk van Arntz nu nog lang en wereldwijd te zien zal zijn. Het valt ook zeker niet uit te sluiten dat bijvoorbeeld het CBS, nieuwe wegen zal willen vinden om de pictogrammen opnieuw te gaan gebruiken. Want wie naar de werken kijkt kan niet om de conclusie heen dat ze ook nu nog – of misschien wel juist nu – het begrip van statistieken kunnen vergemakkelijken. (Hans van Brussel)

Column

Peter Doorn

Het einde van de theorie?

Het kan u gemakkelijk zijn ontgaan, want NRC verkondigde het op de zwoele zomeravond van 24 juli: het einde van 'de wetenschapstheorie' is aangebroken. Modellen kunnen in de opruiming. Misschien minder erg dan het einde van de wereld, dat volgens de laatste inzichten in 2012 wordt verwacht, maar toch.



NRC ontleende het bericht aan een artikel van hoofdredacteur Chris Anderson uit Wired Magazine van 16 juli. Waarom de theorie wegkan? Omdat we nu leven in het petabyte-tijdperk van de datazondvloed. U gelooft het niet? Anderson: 'Out with every theory of human behavior, from linguistics to sociology. [...] Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves.' Er is nu iets beters dan de wetenschappelijke methode van Popper en A.D. de Groot: 'Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.' In drie weken tijd leverde het Anderson 69 sceptische reacties op van wetenschappelijke apologeten zoals ik, die zich wanhopig vastklampen aan een verouderd paradigma.

Ik heb drie bezwaren tegen Anderson's redenering. Ten eerste is er op veel gebieden helemaal geen data-overvloed. Dat geldt zeker voor de humaniora en sociale wetenschappen. Van het unieke historisch materiaal in de archieven is slechts een fractie digitaal. Er bestaan wel heel veel gegevens over de hedendaagse bevolking en bedrijven in allerlei registraties en digitale dossiers, maar veel daarvan is (nog) niet toegankelijk voor wetenschappelijk onderzoek.

Ten tweede is de behoefte aan theorievorming niet gebonden aan petabytes. Een voorbeeld uit een vakgebied en tijdperk zonder veel bytes. Karel van het Reve betoogde al in 1978 met sardonisch genoegen dat de hele literatuurwetenschap pretentieuze, deprimerend, oninteressant, perifeer en triviaal is. Ik vrees dat de digitale beschikbaarheid van literaire teksten de literatuurtheorie, bunk of niet, eerder heeft aangemoedigd dan verminderd.

Ten derde, wat wil je beginnen met een hele berg aan data zonder idee of methode om die te lijf te gaan? Natuurlijk verandert er heel wat in de wetenschap door de sterke toename in de beschikbaarheid van data. Ook in hele grote dataverzamelingen wordt doorgaans zeer doelgericht gezocht naar bepaalde verschijnselen of verbanden, ook in de deeltjesfysica. Uit een datamoeras stijgen de interessante patronen niet als giftige dampen vanzelf naar boven. Er zijn inderdaad datamining-technieken die onverwachte patronen in grote bergen data aan het licht brengen. Maar ook die zijn gebaseerd op theorie. De DANS-archieven bevatten duizenden bestanden. Zouden we meer te weten komen over samenleving, geschiedenis en cultuur door die te gaan minen? U bent van harte welkom om het te proberen.

www.wired.com/science/discoveries/magazine/16-07/pb_theory
www.nrc.nl/wetenschap/article1936868.ece/Het_einde_van_de_wetenschapstheorie

Peter Doorn is directeur van DANS (Data Archiving and Networked Services)

Gelezen

Ineke Stoop, Marion Wittenberg (eds): *Access panels and online research, panacea or pitfall? Proceedings of the DANS Symposium – Amsterdam October 12th, 2006; DANS Symposium Publications 4; Amsterdam, Aksant Academic Publishers, 2008; ISBN 978-90-5260-300-1; € 19,90*

In the previous century, survey data collection was often based on random samples from the population. Respondents were questioned in face-to-face or telephone interviews or by mail questionnaires. Nowadays, in many polls, people are invited to express their opinion through the internet. Is this technological progress, a methodological hornets' nest or both? In October 2006, DANS organized a symposium on this topic called 'Access Panels and online research, panacea or pitfall?'. The presentations and debates at the symposium reflected a mixture of enthusiasm, fierce criticism and scientific scepticism. This volume contains the papers of this day.



Patty Dempster: *Erfgoed in beeld, onderzoek naar beeldbanken van erfgoedinstellingen; Digitaal Erfgoed Nederland (DEN), 2008*

Analyse van erfgoedbeeldbanken op basis van desktop research en interviews met beheerders van erfgoedbeelddatabanken. Aan bod komen onder andere de wijze waarop een beeldbank kan worden opgebouwd,

aan welke voorwaarden moet worden voldaan, en hoe een beeldbank beter vindbaar kan worden gemaakt.

SURFnet: *Jaarverslag 2007; grensverleggend verbinden*

In het jaarverslag komen de belangrijkste ontwikkelingen van de diensten, projecten en organisatie van SURFnet in 2007 aan de orde; Utrecht, SURFnet, 2008.

Natasha Stroeker, Petra Schoen, Koos van Dijken: *Wegwijzer, ken uw virtuele kl@nt; Taskforce Digitale Toegankelijkheid Archieven, 2007; ISBN 978 90 78763 03 1*

Archiefinstellingen worden via internet steeds beter toegankelijk voor het publiek. Collecties worden gedigitaliseerd en via het internet aangeboden en er worden digitale producten en diensten ontwikkeld. Kennis over het publiek en zijn verwachtingen en ervaringen is onontbeerlijk om het digitale aanbod zo goed mogelijk op de behoeften te kunnen laten aansluiten. Maar hoe kun je deze kennis verzamelen? Welk soort onderzoek

hoort bij welk soort vraag van een archiefinstelling? Uit onderzoek van het Instituut voor Onderzoek van Overheidsuitgaven (IOO) blijkt dat in de archiefsector een brede behoefte bestaat aan een handzaam overzicht van soorten publieksonderzoeken, waarbij algemene achtergrondinformatie gegeven wordt en praktische handvatten worden aangeboden voor de toepasbaarheid. Om de medewerkers in de archiefsector te ondersteunen heeft de *Taskforce Archieven* besloten deze wegwijzer publieksonderzoek te ontwikkelen.

Mark, P.J. van der Loo: *An analysis of editing strategies for mixed-mode establishment surveys; Voorburg-Heerlen, Statistics Netherlands, 2008; CBS Discussion Paper*
For several of Statistics Netherlands' establishment surveys, respondents have the choice to respond on paper or electronically. This paper investigates the consequences for editing processes when two different data streams have to be handled. Special attention is paid to the question of how to quantitatively compare the quality of the data streams.