

Pagina 3 • Oude publicaties CBS

Het CBS is bezig met een project om oude publicaties te digitaliseren en via internet toegankelijk te maken. Een half miljoen pagina's staan op het programma.

Pagina 4 • DANS bouwt resolver

Een resolver verbindt een blijvende identificatiecode met de plek op het internet waar de inhoud achter die code zich bevindt. DANS heeft zo'n resolver gebouwd.



Pagina 5 • Franciska de Jong

De Twentse hoogleraar taaltechnologie Franciska de Jong combineert taalkunde en statistiek bij het werken aan multimediazoektechnologie, bijvoorbeeld om in audio-interviews te kunnen zoeken.

Pagina 6 • Vergrijzingsonderzoek

Netspar richt zich op de sociale en economische vraagstukken rond een vergrijzende bevolking. Door middel van onderzoek, onderwijs en kennisuitwisseling.

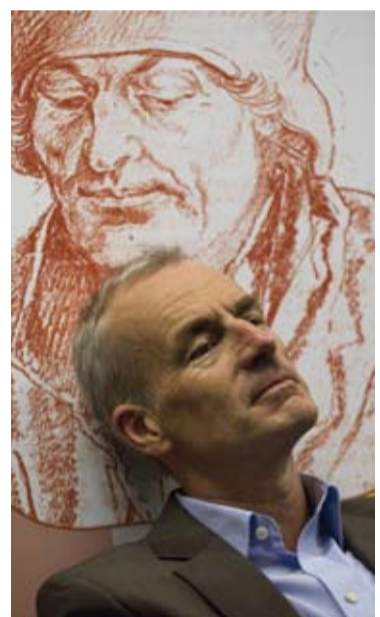
EN VERDER

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

Spinoff van nieuwe BNTL: samengebruik van bibliografieën

Op 24 april is in Den Haag de nieuwe Bibliografie van de Nederlandse Taal- en Literatuurwetenschap (BNTL) gelanceerd. De 'officiële handeling' werd verricht door Karina van Dalen, verantwoordelijk bij de Nederlandse host van de bibliografie, het Huygens Instituut, en directeur Marcus de Schepper van de Belgische partner, het Bureau van de Bibliografie van de Neerlandistiek.

De nieuwe BNTL is de opvolger van de versie die sinds 1970 werd opgebouwd vanuit het Huygens Instituut en de Vlaamse Bibliografie voor de Neerlandistiek. Hij bevat Nederlandse, Vlaamse en buiten-



Een tevreden Huygens-directeur Henk Wals

landse titels van publicaties over Nederlandse taal- en letterkunde uit de periode 1940 tot heden (dus geen primaire literatuur), waar mogelijk met een link naar de digitale volledige teksten.

Die BNTL is sinds de jaren tachtig van de vorige eeuw digitaal maar daarmee nog niet gemakkelijk toegankelijk, legt Van Dalen uit. De bibliografie werd toen ondergebracht in het bibliotheekstelsel Pica van OCLC, 'maar daar konden alleen abonneementhouders bij. De versie die sinds vorig najaar intern draait en die nu officieel is gelanceerd, is voor iedereen toegankelijk en dat is een groot verschil. Eén van de aspecten waar wij ons op verheugen is dat vanaf het komend najaar ook gebruikers gegevens kunnen invoeren'. Daarmee wordt tegelijk een grotere inhoudelijke dekking bereikt en iets gecompenseerd van het verminderde budget voor de BNTL.



De Schepper en Van Dalen verrichten de opening

Van Dalen: 'Het systeem is in deze nieuwe versie vereenvoudigd. Er zijn minder velden om in te vullen. Daardoor kunnen we nu veel sneller titels invoeren.' Voor de mogelijkheid om extern in te voeren wordt een flinke belangstelling verwacht. Die mogelijkheid heeft ook de vorm van een product aangenomen: de Collaborative Online Bibliography (COB), die al operationeel is voor de Friese Bibliografie en waarvan een Engelse versie wordt onderzocht voor de naamkundebibliografie van de International Council of Onomastic sciences (ICOS).

De nieuwe BNTL is in ontwikkeling, benadrukken de makers. De bedoeling is om regelmatig mogelijkheden aan het systeem toe te voegen. Op dit moment kan gelijktijdig worden gezocht op informatie uit de titels (auteur, jaar, titel) en toegevoegde informatie (trefwoorden, samenvattingen en andere informatie). Ook kan er geavanceerd worden gezocht op specifieke onderdelen. De digitale teksten waarnaar wordt verwezen zullen later dit jaar ook doorzoekbaar zijn. Dagelijks worden titels en teksten door de interne redactie toegevoegd. (MdG)

Data uit beleidsonderzoek voor rijk: verplicht archiveren

Gegevens uit beleidsgericht onderzoek in opdracht van de rijksoverheid, 'die zich lenen voor meervoudig gebruik', moeten voortaan worden gearchiveerd bij DANS (Data Archiving and Networked Services). Dat staat in een modelovereenkomst die is verbonden aan de nieuwe Algemene Rijksvoorwaarden voor het verstrekken van Opdrachten tot het verrichten van Diensten (ARVODI).

De nieuwe voorwaarden zijn op 16 maart in de Staatscourant gepubliceerd. Ze leggen vast hoe er met de rijksoverheid zaken gedaan kunnen worden op alle gebieden waar die overheid gebruik maakt van diensten door derden. Een van die gebieden is het uitvoeren van onderzoek voor het beleid. Daarvoor is een aparte modelovereenkomst gemaakt. Die vraagt van de opdrachtnemers naast deponeren van de data bij DANS dat ze tevoren hun onderzoek aangeven bij de Nederlandse Onderzoeksdatabank NOD, die wordt onderhouden door de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW). Ook wordt het deponeren van het onderzoeksrapport bij het e-Depot van de Koninklijke Bibliotheek verplicht gesteld. Voor het deponeren van data maakt de modelovereenkomst

wel het voorbehoud dat die moeten zijn 'geschoond van vertrouwelijke gegevens, gerubriceerde gegevens en gegevens waarvan de publicatie op grond van de wet- en regelgeving achterwege moet blijven'. Data die door die zeef komen moeten 'binnen drie maanden na publicatie van het eindrapport door de opdrachtnemer (...) aan Data Archiving and Networked Services (DANS) worden aangeboden', aldus de modelovereenkomst. De Algemene Rijksvoorwaarden zijn opgesteld onder verantwoordelijkheid van het ministerie van Algemene Zaken. De uitwerking van de voorwaarden voor de levering van beleidsgericht onderzoek is gebeurd door het departement van Onderwijs, Cultuur en Wetenschap (OCW). Dat verzekert desgevraagd dat zowel de voorwaarden als de daaraan verbonden

modelovereenkomsten bedoeld zijn voor gebruik door alle departementen. DANS is nog in gesprek met het ministerie van OCW over de praktische uitwerking van het verplicht deponeren. (MdG)



Minister Plasterk, verantwoordelijk voor de voorwaarden voor beleidsgericht onderzoek

Microsoft add-in beste conversietool PDF/a

De Microsoft add-in geeft de beste resultaten bij het converteren van MS Office formaten naar het veel duurzamere PDF/a formaat. Dat blijkt uit onderzoek van de Koninklijke Bibliotheek. PdfMaker van Adobe is een goede tweede. Voor wie minder hecht aan de exacte weergave van het originele uiterlijk maar meer aan structuur, bleek conversie via OpenOffice een goed alternatief. Tools die een printerdriver gebruiken om een PDF te converteren, zijn geschikt voor een natuurgetrouwe weergave van het uiterlijk en platte tekst, maar niet voor gebruikers die de structuur van het document belangrijk vinden. Overigens was geen van de geteste tools in staat om een testbestand te

converteren tot een valide PDF/a, wat betekent dat geen van de conversieresultaten voldeed aan de officiële PDF/a-grammatica.

De Koninklijke Bibliotheek besloot om een aantal conversietools te testen omdat het bestandsformaat waarin wordt opgeslagen grote invloed heeft op de houdbaarheid van digitale informatie. Eerder onderzoek had al aangetoond dat het PDF-formaat voor archiveren, PDF/a, in die zin veel betere toekomstperspectieven biedt dan de MS Office formaten. Bij deze test werden de resultaatbestanden beoordeeld op de drie eigenschappen *inhoud*, *uiterlijk* en *structuur*.

Informatie: wouter.kool@kb.nl. (IA)

AGENDA

3 en 4 juni

Utrechts Archief, Utrecht
Symposium Koninklijke Vereniging van Archivarissen in Nederland
Thema: 'In perspectief...'. Welke kant gaat het archiefwezen op? Wat zijn belangrijke ijkpunten in de nabije toekomst? Aan de hand van drie presentatielijnen 'Zichtbaarheid', 'Good Governance', en 'Bestel en Professionaliteit' werkt een keur aan sprekers dit thema uit. www.kvan.nl/

18-20 juni

Manchester Conference Centre, Engeland
4th International Conference on e-Social Science
The conference aims to bring together leading international representatives of the social science, e-infrastructure/cyberinfrastructure and e-Research communities, in order to improve mutual awareness, harmonize understanding and instigate coordinated activities to accelerate research, development and deployment of powerful, new research methods and tools for the social sciences and beyond.
www.ncess.ac.uk/events/conference/

20 juni

Universiteit Utrecht
Geodata op de kaart gezet – toegang tot geodata voor onderzoekers
DANS studiemiddag, georganiseerd in samenwerking met Geonovum en de Faculteit Geowetenschappen van de Universiteit Utrecht. Er zijn in Nederland tal van producenten, aanbieders en gebruikers van geografische databestanden. Het gaat hierbij om overheidsdiensten, bedrijven en wetenschappelijk onderzoekers. Maar het aanbod is onoverzichtelijk en de toegang tot geodata is voor wetenschappelijk onderzoekers vaak lastig. Tijdens de studiemiddag wordt aandacht besteed aan het aanbod van en de vraag naar geodata, vooral voor wetenschappelijk onderzoek.
www.dans.knaw.nl/nl/dans_symposia/

25-29 juni

University of Oulu, Finland
Digital Humanities
This is a joint conference between the Association for Literary and Linguistic Computing, the Association for Computers and the Humanities, the Society for Digital Humanities/ Société pour l'étude des medias interactifs. It is the largest conference in the field of Humanities Computing.
www.ekl.oulu.fi/dh2008/

23 juni – 31 augustus

University of Michigan, Ann Arbor, VS
ICPSR Summer School in Quantitative Methods of Social Research
Founded in 1963, the ICPSR Summer Program is recognized throughout the world as the preeminent forum for basic and advanced training in the methodologies and technologies of social science research. It serves a diverse multidisciplinary and international constituency. The general instructional philosophy emphasizes the integration of methodological strategies with the theoretical and practical concerns that arise in research on substantive social issues. www.icpsr.umich.edu/sumprog/

24-27 juni

Universiteit van Bern, Zwitserland
Archiving 2008
The Archiving Conference brings together an international gathering of experts from industry, academia, governmental institutions, non-profit organizations, archives, libraries, museums, and research labs who share a commitment to addressing the challenges of archiving modern and historic materials in an increasingly digital world. <http://imaging.org/conferences/archiving2008/>

21-22 Augustus

Universiteit Essex, Engeland
GIS in Historical Research
This will be the first major European conference concerned with the use of Geographical Information Systems (GIS) in historical research. The main aim of the conference is to demonstrate how GIS can make a contribution to our understanding of the geographies of the past. www.hgis.org.uk/HGIS_conference/index.htm

22-23 augustus

Universiteit Tilburg
MESS Workshop – Measurement and Experimentation with Internet Panels – Innovative features of Internet interviewing
Internet interviewing is coming of age and it is a good time to focus on new possibilities with this interviewing mode. A workshop on the state of the art of Internet interviewing, in particular focusing on various new and innovative aspects.
www.centerdata.nl/en/LISSpanel

Levenslopen in de 19^e en 20^e eeuw

Op vrijdag 11 april 2008 werd bij het Internationaal Instituut voor Sociale Geschiedenis (IISG) in Amsterdam het programma *Life Courses in Context* officieel afgesloten met een klein congres. Met behulp van dit programma is de afgelopen vijf jaar onder andere gewerkt aan de reconstructie van de levensloop van veertigduizend Nederlanders die werden geboren tussen 1863 en 1922.

In Amsterdam presenteerden diverse wetenschappers hun onderzoeksbevindingen op basis van een eerste release van twintigduizend levenslopen. Historici, demografen en sociologen gebruikten de nieuwe data om interessante vragen te beantwoorden over levenslopen van de Nederlandse bevolking in de 19e en 20e eeuw. Interessant was daarbij om te zien hoe met verschillende methoden en invalshoeken naar inzicht werd gezocht in factoren die van invloed zijn op de kans om in sociale klasse te stijgen. De Nijmegen historicus Onno Boonstra liet in zijn presentatie zien dat de



Koffie tijdens het levenslopesymposium

invloed van onderwijs, gemeten aan de hand van alfabetisme, op intra- en intergenerationele beroepsmobiliteit slechts beperkt was in de negentiende eeuw. Verheffing door onderwijs vond wel plaats in culturele, maar niet in economische zin. In de presentatie van Ulbe Bosma, onderzoeker bij het

IISG, werd aangetoond dat de keuze om naar Indië geen sociale stijging met zich meebracht. Op basis van zijn onderzoek zou kunnen worden geconcludeerd dat met die keus in ieder geval de bestaande sociale status behouden bleef. De presentaties op de dag illustreerden een grote

verscheidenheid aan gebruik van de data. In de afsluitende discussie wees de Amsterdamse socioloog Harry Ganzeboom op de methodologische aspecten van het gebruik van de data. Hij pleitte daarbij voor een genuanceerde toepassing van bestaande analysetechnieken. (LS)

Interviews uit de kast

'Interviews als bron van wetenschappelijk onderzoek', dat was het onderwerp van het symposium op 14 maart georganiseerd door DANS in samenwerking met het Veteranen Instituut, dat zelf met een groot interviewproject bezig is, en de Universiteit Twente waar prof. Franciska de Jong multimedia-onderzoek doet met interviews als materiaal.

De onderzoekers die interviewdata verzamelen of gebruiken vormen een heterogene groep met uiteenlopende onderzoekstradities wat betreft bijvoorbeeld de interpretatie van de inhoud en de beschikbaarstelling/ontsluiting van deze data voor hergebruik. De grote gemene deler van deze groep onderzoekers is dat zij de neiging hebben, na hun noeste arbeid van verzamelen en interpre-

teren, de data voor onbepaalde tijd in de 'kast' op te bergen. Het doel van deze studiedag was het inventariseren van de mogelijkheden om interviews 'uit de kast' te halen.

Keynotespreker Louise Corti van het Engelse data-archief liet zien dat in Engeland de afgelopen tien jaar veel succes is geboekt met het duurzaam archiveren en beschikbaar stellen van kwalitatieve onderzoeksdata. Tony Hak, dagvoorzitter van het symposium, meent dat er een open discussie aangegaan moet worden met de onderzoekers die zich vandaag de dag nog steeds sceptisch opstellen tegenover duurzaam archiveren.

Naast discussie bleek er ruimte voor suggesties hoe zelfs de data van de sceptici uit de kast gehaald

kunnen worden. Zo zou een data-archief het goede voorbeeld kunnen geven door onderzoek te doen met data uit eigen archief. Dit om te demonstreren hoe data nieuw leven ingeblazen kan worden door hergebruikt te worden. De onderzoekers plukken hier de vruchten van door citations en co-auteurschappen. Tijdens de studiedag was er verder aandacht voor ICT mogelijkheden om interviewdata te analyseren, voor

beheer en ontsluiting van interviewcollecties en voor methodologische aspecten van onderzoek met kwalitatieve data. Dankzij de sprekers en de aanwezigen werd het symposium 'Interviews uit de kast' een informatieve en inspirerende bijeenkomst. Samen met een aantal onderzoekers gaat organisator DANS bekijken hoe het archiveren van dit type onderzoeksmateriaal verder gestimuleerd kan worden. (Marion Wittenberg)

www.dans.knaw.nl/interviewsuitdekast

Open Repositories

De *Open Repositories* conferentie wordt elk jaar op een andere locatie gehouden. Dit jaar was Southampton University de setting voor de conferentie met 450 deelnemers: een moderne Engelse universiteit met veel groen, fluitende vogels en een lekker zonnetje. Het drukke conferentieschema liet toch nog ruimschoots de ruimte om te netwerken.

Presentaties over zo uiteenlopende onderwerpen als repository-architectuur, web 2.0 en juridische aspecten werden gehouden in de zalen van verschillende faculteiten. Naast bekende universiteiten, bibliotheken en instellingen waren ook grote bedrijven zoals Sun en Microsoft dit keer van de partij. Sun presenteerde bijvoorbeeld het *Honeycomb* systeem, een opslagmedium dat zeer schaalbaar, betrouwbaar en snel zou zijn. Daarnaast was er natuurlijk ook speciale aandacht voor de Fedora-, ePrints- en Dspace-communities, waarbij de hoofdspekers aan het woord kwamen. Alle drie kondigden nieuwe versies van hun software aan.

Een interessant en leuk experiment dat dit jaar voor het eerst werd gehouden was de zogenaamde 'repository developer challenge'. Er kon vijftigduizend dollar gewonnen worden door het team dat de meest innovatieve repository software werkend kon krijgen binnen twee dagen. Meer dan twintig teams deden enthousiast mee. Dat de uiteindelijke winnaar voor het gebruik van het nieuwe OAI-ORE protocol gekozen had om een Fedora en een ePrints repository te verbinden was niet verbaasd. Aan OAI-ORE werd namelijk speciaal de gehele laatste dag gewijd. Het nieuwe Object Reuse and Exchange protocol zal er voor zorgen dat de inhoud en semantiek van repositories en het web in het algemeen in de toekomst geheel transparant zullen zijn.

De OR2008 conferentie was goed georganiseerd, leuk en vol met interessante informatie en mensen. Volgend jaar zal de conferentie in Atlanta, USA, worden gehouden. (Lodewijk Bogaards)

<http://oro8.ecs.soton.ac.uk/>
<http://www.openarchives.org/ore/>

Social software: let op je privé-gegevens

Op vrijdagmiddag 18 april werd in het instructielokaal van de Universiteitsbibliotheek in Leiden een *hands-on* studiemiddag gehouden over Social Software. Web 2.0 deskundige Martijn Hulst (<http://www.martijn-hulst.nl>) leidde de aanwezigen door de wijde wereld van de sociale media, van networking sites als Hyves en LinkedIn, microblogging services als Twitter tot kennisuitwisselingsdiensten als Wikipedia en del.icio.us. De voor de hand liggende vraag bij de opkomst van dergelijke services is of sociale software alleen 'voor de leuk' is of ook voor professioneel gebruik nuttig kan zijn. Hoewel sociale media vooral voor privé-gebruik

wordt ingezet, wisten enkele deelnemers uit eigen ervaring sprekende voorbeelden te noemen van hoe zij via *online communities* moeilijk te krijgen vakliteratuur hadden weten op te sporen. Het uitwisselen van bookmarks bleek voor sommigen een effectieve methode om interessante sites op het spoor te komen. Maar sociale software heeft ook een schaduwzijde; steeds meer informatie over gebruikers belandt soms – met of zonder toestemming – in het publieke domein. De slotconclusie van de middag: blog, twitter of flickr er op los maar weet wat je doet en ga bewust om met je privé-gegevens. (Edwin Klijn)

CBS bibliotheek digitaliseert collectie

De bibliotheek van het Centraal Bureau voor de Statistiek (CBS) voert een digitaliseringsproject uit om grote delen van collectie van het bureau via het web toegankelijk te maken. Ook veel kwetsbare en moeilijk te raadplegen statistische publicaties uit de 19^e eeuw worden opgenomen. De geplande omvang van het project bedraagt een half miljoen pagina's.

Eerder werden in samenwerking met het opgeheven NIWI de Nederlandse Volkstellingen tussen 1795 en 1971 gedigitaliseerd (www.volks-telling.nl). Daarop komt nu een vervolg door de digitalisering, in samenwerking met DANS, van een aantal bedrijfstellingen (1930-1978) en gedetailleerde gegevens op wijk-niveau uit de volkstellingen, woningtellingen en beroepstellingen (1930-1960). Daarnaast worden in het ambitieuze project andere veel gebruikte CBS publicaties uit de collectie gedigitaliseerd.

Tot de al gescande publicaties horen de *Jaarcijfers van het Koninkrijk der Nederlanden* (1881-1967) en het *Statistisch Zakboek* (1899-1988), en cijfers die gebruikt werden voor het opstellen van de Nationale Rekeningen (sinds de jaren vijftig). De eerste twee geven een goed statistisch overzicht van vele onderwerpsgebieden. Die cijfers zullen worden aangevuld door een reeks van historische publicaties uit de 19^e eeuw, waaronder *Lobatto's jaarboekje* (1826-1849) en zijn opvolgers onder titels als *Statistisch jaarboek voor het Koninkrijk der Nederlanden*. De bedrijfstellingen geven naast de

PROVINCIE	GROEBELASTING Rijk, Prov. en Gem.		PERSONELE BELASTING Rijk, Provincie en Gem.		PATENT-RECHT	Hoofdelijke onslag, belasting op landerijen, boorschroeven en trekkieren (N.-Brabant)	Gehel. bedrag aan directe belastingen.
	Gehoord.	Opgehoord.	Rijk	Provincie en Gem.			
Noordholland	1,400,200	700,000	3,004,997	1,276,600	2,150,000	9,300,000	10,254,740
Zuidholland	1,600,000	1,270,000	4,150,000	1,110,000	2,000,000	10,254,740	10,254,740
Utrecht	300,000	400,000	870,000	200,000	200,000	2,100,000	2,100,000
Zeeuws-Vlaand.	250,000	250,000	600,000	100,000	100,000	1,000,000	1,000,000
Noord-Brabant	470,000	1,000,000	1,000,000	1,000,000	100,000	3,000,000	3,000,000
Limburg	100,000	100,000	200,000	100,000	100,000	1,000,000	1,000,000
Gelderland	500,000	500,000	1,000,000	1,000,000	100,000	2,000,000	2,000,000
Overijssel	300,000	300,000	600,000	100,000	100,000	1,000,000	1,000,000
Drenthe	100,000	100,000	200,000	100,000	100,000	1,000,000	1,000,000
Groningen	100,000	100,000	200,000	100,000	100,000	1,000,000	1,000,000
Friesland	100,000	100,000	200,000	100,000	100,000	1,000,000	1,000,000
Totaal 1878	6,040,000	4,170,000	14,540,000	4,100,000	8,870,000	41,700,000	41,700,000
Rijk	3,000,000	6,000,000	2,100,000	4,100,000	—	16,000,000	16,000,000
Provincie	300,000	1,000,000	800,000	—	100,000	2,000,000	2,000,000
Gemeente	1,000,000	500,000	11,500,000	—	8,700,000	20,000,000	20,000,000

Tabel uit 'Jaarcijfers omtrent bevolking, landbouw, handel, belastingen, onderwijs enz.' van de Vereniging voor Statistiek in Nederland, 1882

volkstellingen een goed beeld van de economische activiteit in Nederland. Op deze manier ontstaat een compleet beeld van de 19^e en 20^e eeuw in cijfers.

Net als bij het eerste Volkstellingenproject doet zich bij dit materiaal de vraag voor hoe deze data gedigitaliseerd en toegankelijk gemaakt kunnen worden. Cijfers moeten betrouwbaar zijn, maar bij tekstherkenning treden altijd

onnauwkeurigheden op. De bedrijfs- en volkstellingen die nu gedigitaliseerd worden, bestaan voor een groot deel uit handgeschreven cijfers die in een voorgedrukte tabel zijn ingevuld. OCR van dit materiaal is mogelijk, maar controle op de resultaten brengt de kosten sterk omhoog. Daarom is uiteindelijk besloten tot complete data-entry van deze publicaties. Dit materiaal zal in Excel formaat en als afbeelding beschikbaar gesteld worden. Gelet op de ervaringen van gebruikers met volkstellingen.nl zal de raadpleging en gebruik van de data gebruik-

Friese bestanden naar DANS

De Fryske Akademy is sinds begin dit jaar bezig een groot aantal bestanden uit taalsociologisch onderzoek bij data-instituut DANS onder te brengen. Een zevental is daar al beschikbaar. Nog eens vijftien volgen binnenkort, 'als de belangrijkste onderdelen die voor het begrip nodig zijn in het Nederlands zijn vertaald' aldus documentalist-historicus Peter van der Meer van het KNAW-instituut. Het gaat om resultaten uit eigen onderzoek dat de Akademy sinds de jaren tachtig uitvoerde naar verschillen in het gebruik van de Friese taal en de oorzaken daarvan: 'Wie spreekt wat tegen wie en wanneer?'. 'Wij kwamen tot de conclusie dat het voor de toegankelijkheid van de gegevens beter is om ze via een gespecialiseerd data-archief aan te bieden', licht Van der Meer toe. 'Er zullen dan meer mensen gebruik van maken dan wanneer wij materiaal het alleen op onze eigen website aanbieden'. (MdG)

kersvriendelijker worden. De tekst van de Excelbestanden wordt volledig doorzoekbaar via één interface en de gemeentenamen worden direct gekoppeld aan een historische standaard gemeentecode, de zogenaamde 'Amsterdamse code'. Ook manipulatie van de images (zoals roteren en inzoomen) zal mogelijk zijn. Begin volgend jaar zullen de eerste data te raadplegen zijn via www.cbs.nl. (Henk Laloli)

derzoek en erfgoedinstellingen door een verbreding naar de geesteswetenschappen'. Daar zijn onderzoekers vaak afhankelijk van erfgoedcollecties terwijl in erfgoedinstellingen veel collectiebeheerders werken met een achtergrond in de geesteswetenschappen. Voor het toegankelijk maken van de collecties blijven intussen informatici nodig en dat maakt de opzet van deze nieuwe ronde 'tamelijk uniek', volgens Schlarmann: 'Het gaat er echt om de verschillende disciplines beter met elkaar te laten samenwerken'. Multidisciplinaire teams met erfgoedbeheerders, informatici en beoefenaars van de geesteswetenschappen worden aangemoedigd zich te melden met voorstellen waarin een of meer van drie hoofdlijnen herkenbaar zijn: naast de interactie tussen databronnen of de 'interoperabiliteit van grootschalige en gedistribueerde bronnen' zijn dat 'visualisatie en simulatie' en 'historische dimensies en moderne interpretaties'. Er is drie miljoen euro beschikbaar voor deze ronde, die wordt ge-

3TU Datacenterum

De technische wetenschappen krijgen een eigen datacenterum voor digitale datasets. Het zogenaamde 3TU.Datacenterum wordt opgezet door de bibliotheken van de universiteiten van Delft, Eindhoven en Twente. Voor de alfa- en gammadata wordt al in duurzame opslag voorzien door DANS (Data Archiving and Networked Services). Het initiatief van de drie technische universiteiten zorgt ervoor dat die functie voortaan ook voor technische en bètawetenschappen beschikbaar is. Hun datacenterum bouwt voort op de ervaringen die de drie al opdeden met het DARELUX-project, waarin werd geëxperimenteerd met de duurzame opslag van hydrologische meetgegevens in een Luxemburgs stroomgebied.

Basisvoorziening geo-informatie

Binnen vier jaar moet Nederland een goed werkende, duurzame en intensief gebruikte basisvoorziening geo-informatie hebben. Dat staat in de nota GIDEON van de stichting Geonovum en het programmabureau Ruimte voor Geo-informatie. De nota werd geschreven in opdracht van het GI-beraad, een interdepartementaal overleg van ambtenaren, om een visie te geven op de verdere ontwikkeling van de infrastructuur voor een geo-basisvoorziening. De bedoeling is dat de minister van VROM de nota aanbiedt aan de Ministerraad en vervolgens aan de Tweede Kamer. Daarbij zal worden voorgesteld om jaarlijks aan de Tweede Kamer verslag uit te brengen over de voortgang.

Geografisch zoeken naar data-sets archeologie

Sinds maart van dit jaar kunnen archeologische datasets in het archiveringssysteem EASY van DANS worden gezocht met behulp van Google maps en Google Earth. Projecten waarvan de locatie bekend is kunnen nu op een kaart of satellietfoto worden gezocht en geraadpleegd.

Archeologische gegevens zoeken via een geografisch informatiesysteem (GIS) bestaat al langer, bijvoorbeeld bij zie besloten Archis en het openbare KICH (Kennisinfrastuctuur Cultuurhistorie). Geografisch zoeken was echter nog geen onderdeel van EASY, dat onderdak biedt aan het e-Depot van de Nederlandse Archeologie (EDNA). Het gaat hier om een experiment. Op termijn wordt besloten of en hoe die wordt voortgezet.

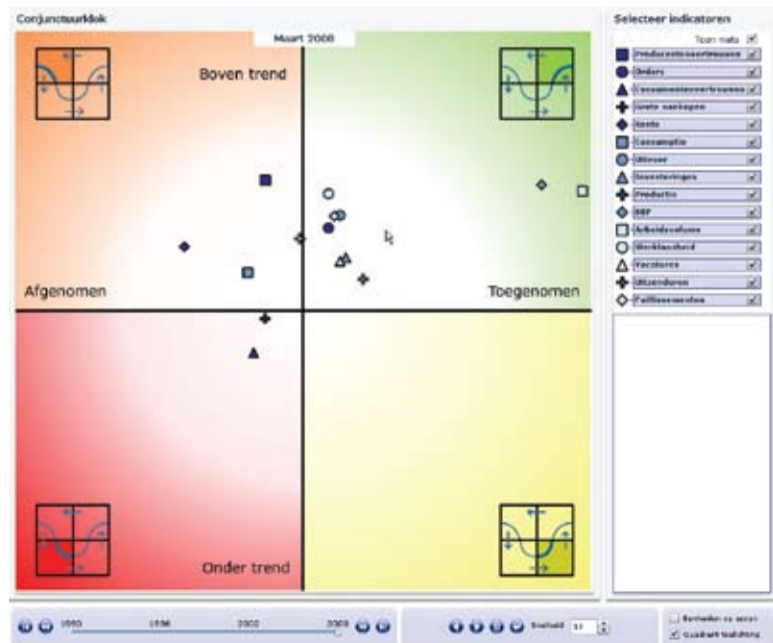
financierd door de NWO-gebieden Exacte wetenschappen en Geesteswetenschappen samen met het ministerie van OCW. Het programma gaat er vanuit dat met elke toekenning een bedrag van 500 tot 640 duizend euro is gemoeid en houdt rekening met een viertal aanvragen dat kan worden gehonoreerd. Dat is ruim berekend, geeft Schlarmann desgevraagd toe, maar 'er is ook wat speelruimte nodig als smeerolie, juist omdat het hier een nieuw soort samenwerking tussen disciplines betreft'. Vooraanmeldingen voor de subsidieronde moeten uiterlijk 28 juni binnen zijn bij NWO. (MdG)

Conjunctuurklok tikt

De Conjunctuurklok is eind 2005 door het Centraal Bureau voor de Statistiek (CBS) geïntroduceerd en heeft zich in korte tijd een centrale plaats verworven in de conjunctuurberichten van Nederland. Van vijftien indicatoren wordt elke maand in één plaatje en in onderlinge samenhang getoond of het 'goed' of 'slecht' gaat (onder of boven de trendmatige ontwikkeling over een lange periode) en of het 'beter' of 'slechter' gaat (stijging of daling ten opzichte van de voorgaande periode). Door de foto's van alle maanden achter elkaar te laten zien wordt een film getoond van het verloop van de conjunctuur. Op

de CBS-website kan dat interactief worden gedaan voor de periode vanaf 1990 tot heden.

Een belangrijk kenmerk van de Conjunctuurklok is dat het gaat om real-time gegevens. De klok laat zien hoe men op een bepaald moment met de toen beschikbare gegevens over de conjunctuur dacht. Als door bijstellingen achteraf een bepaalde indicator een andere waarde heeft gekregen dan is dat in de klok zichtbaar in de maand dat deze nieuwe informatie beschikbaar kwam en dus niet in de periode waarop de bijstelling betrekking heeft. (Ben van Cleef)



CATCH 2008: Aandacht voor databronnen

Een betere interactie tussen databronnen is een van de drie hoofdlijnen in de nieuwe ronde van het subsidieprogramma CATCH (Continuous Access To Cultural Heritage). Binnen die hoofdlijn geeft het programma als belangrijke onderzoekrichtingen onder meer hergebruik en ontdebelling van data, kwaliteit van data, evolutie en diversiteit van data. Bij de betere interactie of 'interoperabiliteit' moet worden gedacht aan het bouwen van systemen waardoor grote databronnen 'gemakkelijker met elkaar kunnen praten zonder zelf te hoeven veranderen', aldus woordvoerder Marjolein Schlarmann van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO), die het programma beheert.

Vanuit CATCH, dat in 2005 werd gelanceerd, wordt tot nu toe een tiental projecten gefinancierd. Daarin ligt het accent vooral op samenwerking tussen erfgoedinstellingen en onderzoekers in de informatica. Die formule is volgens de website van NWO heel geslaagd maar 'nu wordt het tijd om een volgende impuls te geven aan de interactie tussen on-

derzoek en erfgoedinstellingen door een verbreding naar de geesteswetenschappen'. Daar zijn onderzoekers vaak afhankelijk van erfgoedcollecties terwijl in erfgoedinstellingen veel collectiebeheerders werken met een achtergrond in de geesteswetenschappen. Voor het toegankelijk maken van de collecties blijven intussen informatici nodig en dat maakt de opzet van deze nieuwe ronde 'tamelijk uniek', volgens Schlarmann: 'Het gaat er echt om de verschillende disciplines beter met elkaar te laten samenwerken'. Multidisciplinaire teams met erfgoedbeheerders, informatici en beoefenaars van de geesteswetenschappen worden aangemoedigd zich te melden met voorstellen waarin een of meer van drie hoofdlijnen herkenbaar zijn: naast de interactie tussen databronnen of de 'interoperabiliteit van grootschalige en gedistribueerde bronnen' zijn dat 'visualisatie en simulatie' en 'historische dimensies en moderne interpretaties'. Er is drie miljoen euro beschikbaar voor deze ronde, die wordt ge-

derzoek en erfgoedinstellingen door een verbreding naar de geesteswetenschappen'. Daar zijn onderzoekers vaak afhankelijk van erfgoedcollecties terwijl in erfgoedinstellingen veel collectiebeheerders werken met een achtergrond in de geesteswetenschappen. Voor het toegankelijk maken van de collecties blijven intussen informatici nodig en dat maakt de opzet van deze nieuwe ronde 'tamelijk uniek', volgens Schlarmann: 'Het gaat er echt om de verschillende disciplines beter met elkaar te laten samenwerken'. Multidisciplinaire teams met erfgoedbeheerders, informatici en beoefenaars van de geesteswetenschappen worden aangemoedigd zich te melden met voorstellen waarin een of meer van drie hoofdlijnen herkenbaar zijn: naast de interactie tussen databronnen of de 'interoperabiliteit van grootschalige en gedistribueerde bronnen' zijn dat 'visualisatie en simulatie' en 'historische dimensies en moderne interpretaties'. Er is drie miljoen euro beschikbaar voor deze ronde, die wordt ge-

Er is drie miljoen euro beschikbaar voor deze ronde, die wordt ge-

Taaltechnologe Franciska de Jong over multimediale spraaktechnologie

‘Dit onderzoek is een investering in de toekomst’

MARTIJN DE GROOT

‘Technologie veroudert, maar de kennis die er achter zit blijft relevant’. Dat zegt prof. dr. Franciska de Jong, leider van een aantal onderzoekprojecten waarin spraakherkenning een grote rol speelt. De techniek om gesproken woord doorzoekbaar te maken komt steeds verder, blijkt uit een gesprek met de Twents-Rotterdamse taaltechnologe.

‘Voor onderzoekers is het belangrijk om over data te kunnen beschikken, en dat die data ergens worden ondergebracht waar anderen er ook iets mee kunnen doen’. Franciska de Jong kan zo’n uitspraak met enig gezag doen want haar eigen onderzoek consumeert enorme hoeveelheden data - zij spreekt glimlachend van tekstcorpussen die honderden miljoenen woorden tellen. Bovendien is juist zij bezig met projecten die belangrijke historische en eigentijdse bronnen in gesproken vorm voor een groot publiek beschikbaar en tekstueel doorzoekbaar maken. De onlangs gelanceerde website *buchenwald.nl*, met 38 volledig doorzoekbare audio-interviews met overlevenden, is er een voorbeeld van.

Toch heeft haar uitspraak niet daar betrekking op, maar op de rol van de Virtual Knowledge Studio (VKS), waarvan zij sinds de zomer van 2007 de Rotterdamse vestiging bestiert. ‘Een goede organisatie van het proces waarbij data worden opgeslagen, bewerkt en gedeeld’, gaat ze verder, ‘is een gemeenschappelijk belang van onderzoekers, beheerders en gebruikers van dat materiaal, want de beschikbaarheid en uitwisselbaarheid van zulke grote hoeveelheden data is een relatief nieuw fenomeen. De ontwikkeling van nieuwe samenwerkingsverbanden die dat vraagt en de methodologische vernieuwing die erbij komt kijken, dat ligt op het onderzoeksterrein van de VKS’.

Het beschikbaar komen van grote elektronische datasets heeft het vak enorm veranderd

De Erasmus Studio, zo wordt deze vestiging genoemd, is ondergebracht in een paar kamers op de zevende verdieping van een betonnen studeerkolos van de gelijknamige Universiteit. De lift is traag, de hallen groot, de gangen zijn kaal en grijs. Als je niet wist dat in de jaren zeventig van de vorige eeuw zulke bouwwerken her en der waren verrezen, zou je de stijl voor Rotterdams aan kunnen zien: hier rest de bewoners weinig anders dan studeren en onderzoeken. ‘De Amsterdamse VKS is sterk op de methodologie georiënteerd’ antwoordt



WILLIAM HOOGTEYLING

ze op de vraag naar het verschil. ‘Dus dan is het logisch dat je niet steeds met hele praktische resultaten kunt zwaaien. Hier in Rotterdam ligt de nadruk misschien meer op de doenerige aspecten van het onderzoeksterrein, en er zijn hier veel studenten. Wij werken samen met een aantal onderzoeksgroepen, zoals de econometristen en bijvoorbeeld met de medische faculteit in onderzoek naar het ontsluiten van medische informatie met behulp van thesauri.’ En dan vloeit het gesprek automatisch door in de richting waar ook de passie van de hooggeleerde taaltechnologe ligt. ‘Als je zoekt naar bijvoorbeeld de samenhang tussen een gen en een bepaalde ziekte, dan is het meestal zo dat daar bergen literatuur over te vinden zijn. Nu is er aan de EUR technologie ontwikkeld waardoor je in die tekstbestanden kan zoeken welke concepten met elkaar in verband worden gebracht die voorheen niet eerder in samenhang zijn besproken. Dan ben je nieuwe kennis of nieuwe inzichten op het spoor. Nu die technologie werkt, gaan we natuurlijk proberen of ze ook voor andere wetenschapsgebieden te gebruiken is.’

Voor dit soort zoektechnieken is naast veel taalkundige data ook veel statistiek nodig en dat is precies de combinatie die de Jong het meeste boeit, vooral als er dan ook nog een bruikbaar resultaat uit voortkomt. Aan de volkswijsheid dat een mens over alfa- of beta-talenten mag be-

schikken heeft ze geen boodschap. ‘Je ontwikkelt je op de gebieden waar je aandacht aan geeft. Ik had affiniteit met talen en wiskunde en kwam als student terecht in de taalkunde – een tamelijk exact vak binnen de letteren, met logica, mathematische taalkunde en computationele linguïstiek in het pakket. Na mijn studie kon ik bij het Natlab van Philips aan de slag in een onderzoeksgroep die zich richtte op machinaal vertalen. Een fantastische tijd. We bouwden met een flink aantal mensen aan een systeem, maar dat werkte vooral op basis van grammaticaregels die in een modulaire opzet over talen heen aan elkaar werden gekoppeld. Daar kwam dus geen statistiek aan te pas. Er kwam een werkend systeem uit, maar een product is het nooit geworden’.

Eigenlijk diende pas in de jaren negentig van de vorige eeuw de kans zich aan op veel betere en krachtiger taalmachines, aldus de Jong, en dat was een direct resultaat van het beschikbaar komen van grote elektronische dataverzamelingen. Dat maakte de opkomst van statistische modellen mogelijk ‘en daardoor is het vak enorm veranderd’. De Jong verwierf in 1992 een leerstoel in de taaltechnologie, die ze nog steeds bekleedt. Ondergebracht tussen Twentse informatici bleek die een mooie basis te bieden om juist op het statistische spoor verder te gaan. In de spraakherkenning, het gebied waar De Jong zich sindsdien steeds

verder in ging verdiepen, speelt de statistiek een doorslaggevende rol, licht ze toe. ‘Als je met spraakherkenning bezig bent dan moet je een akoestisch model ontwikkelen waarmee je meer dan honderdduizend woorden van elkaar kunt onderscheiden en in een transcriptie omzetten. En het gaat dan om spontane spraak dus dan moet zo’n model de verschillende uitspraakvarianten van één woord ook kunnen herkennen. Maar woorden lijken soms heel erg op elkaar. Mensen weten dan toch het onderscheid omdat ze de context van die woorden meenemen. Dus dat moet zo’n spraakherkenner ook doen. Daarvoor ontwikkel je een taalmodel, dat is een statistische representatie van al die woorden en hun onderlinge verbanden. Zo’n model levert dan de informatie op dat er bijvoorbeeld is gesproken over een ‘Italiaans restaurant’ en niet over een ‘Italiaans restant’.

De bedoeling is dat we van elkaar kopiëren, want daar worden alle deelnemers beter van

De Jong is wel bereid om een knap stukje werk te demonstreren dat uit

de activiteit van haar Twentse onderzoeksgroep Human Media Interaction (HMI) is voortgekomen: een showcase op de HMI-website biedt de mogelijkheid om in de acht-uur-journaals van de laatste twee weken te zoeken. Dat gebeurt in feite in de tekst die na afloop van elke uitzending automatisch wordt voortgebracht door de spraakherkenner die haar onderzoeksgroep afleverde. Een zoekactie met de woorden ‘kelder’ en ‘Oostenrijk’ levert binnen een paar seconden een reeks verwijzingen op, met tekstweergave van de relevante passages en de mogelijkheid om het betreffende fragment te bekijken. Een overtuigend resultaat, ook met de foutjes die er nog in de weergegeven tekst blijken te zitten – vooral waar de spraakherkenner zich niet ontziet om ook Duits gesproken fragmenten in geschreven Nederlands om te zetten.

Kan zo’n mooi product niet op de markt gebracht worden, om zo wat terug te verdienen van de investeringen die er in de loop van de jaren in zijn gedaan? ‘Dat zou misschien wel kunnen’, denkt de Jong, ‘maar niet kostendekkend. Dan zou het veel te duur zijn want daarvoor zijn de kosten tot nu toe te hoog geweest. De technologie is dan ook in ‘open source’ beschikbaar gesteld, zodat anderen nieuwe toepassingen kunnen ontwikkelen. Voor ons zelf is het eigenlijk alleen interessant om daaraan te werken als er een nieuw wetenschappelijk perspectief aan verbonden is. En dat is natuurlijk wel het mooie van dit soort projecten. De kennis die je hebt ontwikkeld gaat niet verloren. Daar wordt steeds op doorgebouwd. Wij doen met onze spraakherkennings- en multimediazoektechnologie mee met internationale competities waar je kan vergelijken met anderen, hoe die het hebben gedaan. Wat is hun foutenpercentage? En de bedoeling is dat we van elkaar kopiëren, want daar worden alle deelnemers beter van.’

‘Een investering in de toekomst’ is het vooral, en dat geldt ook voor het *access to oral history*-project CHoral waarvoor de Jong financiering kreeg binnen het cultureel erfgoed-programma CATCH van de Nederlandse Organisatie voor Wetenschappelijk onderzoek. En trouwens voor CATCH als geheel: ‘Doordat binnen CATCH de samenwerking wordt gestimuleerd, krijg je een hele generatie onderzoekers en softwareontwikkelaars met de kennis om de infrastructuur van de toekomst vorm te geven’.

Franciska de Jong werkte na haar studie taalwetenschap als onderzoeker bij het Philips NatLab in Eindhoven. Sinds 1992 is ze hoogleraar taaltechnologie aan de Universiteit Twente. Rond die tijd werd ze ook onderzoeker aan de Universiteit Utrecht en later bij TNO in Delft. In juni 2007 trad ze aan als managing director van de Virtual Knowledge Studio in Rotterdam. Sinds 2004 is zij lid van het NWO-gebiedsbestuur Geesteswetenschappen.

FOCUS

Netspar

Eind vorig jaar werd bekend dat onderzoeksnetwerk en maatschappelijk topinstituut Netspar een vervolgsubsidie van 14 miljoen euro zou krijgen van de overheid en de pensioen- en verzekeringssector. *e-data&research* liet zich voor deze aflevering in de reeks Focus informeren door wetenschappelijk directeur Arthur van Soest van Netspar, dat zich richt op de sociale en economische uitdagingen van de vergrijzende bevolking. Dit doet het door middel van onderzoek, onderwijs en kennisuitwisseling.



doen twaalf landen mee. Er worden rond dertigduizend personen geënquêteerd (in Nederland circa 2.500). De gegevens zijn via CentERdata in Tilburg gratis beschikbaar voor de onderzoeksweld. Daarnaast gebruikt Netspar voornamelijk de grote dataverzamelingen van het Centraal Bureau voor de Statistiek (CBS) zoals het Sociaal Statistisch Bestand en het Inkomenspanelonderzoek. Samen met het Centraal Planbureau en het CBS wordt onderzoek uitgevoerd naar de gevoeligheid voor financiële prikkels bij VUT-regelingen. Van Soest: 'We gebruiken de mogelijkheid om via Remote Access te werken, met aansluitingen onder meer in Tilburg, Rotterdam en Utrecht. Via het CBS kunnen we tegenwoordig ook gebruik maken van het Gezondheid Statistisch Bestand (GSB) waarmee weer hele nieuwe onderzoeksvragen te beantwoorden zijn.'

Voor de toekomst bestaat er volgens Van Soest 'een duidelijke behoefte aan longitudinale, internationaal vergelijkbare, paneldata die gekoppeld kunnen worden aan administratieve gegevens. Daarmee kunnen we mensen voor een langere periode volgen en beslissingen analyseren over de ouderdagsvoorziening, de participatie op de arbeidsmarkt en de rol van gezondheid in de sociaal-economische positie van mensen.'

Wetenschappelijk directeur Arthur van Soest van Netspar: 'Behoeft aan internationaal vergelijkbare paneldata.'

'Netspar is in april 2005 gestart', zegt Van Soest, 'nadat onze mededirecteur Lans Bovenberg in 2003 de Spinozaprijs won. De anderhalf miljoen euro die daarmee was gemoeid, is als startsubsidie gebruikt om het instituut van de grond te krijgen. Daarna konden we subsidie krijgen voor de periode 2005 - 2008. Uit de subsidie die nu beschikbaar komt vanuit overheid en bedrijfsleven, blijkt dat het instituut breed gedragen wordt, zowel binnen de wetenschap als de maatschappij. Hiermee kan Netspar in ieder geval tot april 2011 voortgezet worden.'

Thema's voor Netspar zijn bijvoorbeeld de economie van de vergrijzing, waarvoor een masterprogramma opgezet is, en de relatie tussen pensionering en gezondheid en het daaruit voortvloeiende arbeidsmarktgedrag. Netspar zet wetenschappelijk onderzoek uit en werkt nauw samen met verschillende universiteiten. Doelen zijn daarbij om sociale innovatie te stimuleren, een brug te slaan tussen de sociale wetenschappen en de financiële en publieke sectoren, een uitgebalanceerd onderwijsprogramma op te zetten voor verschillende doelgroepen en multidisciplinair onderzoek op het gebied van pensioenen en pensionering te stimuleren. 'In onze eerste drie jaar was de focus vooral gericht op de nationale positionering,' aldus Van Soest. 'We werken nauw samen met de zeven Nederlandse universiteiten die in Netspar participeren, maar ook met het Centraal Planbureau en het Centraal Bureau voor de Statistiek. Maar in de komende periode gaan we ons ook meer internationaal oriënteren. We deden dat al op kleine

schaal door samen te werken met onderzoekers over de hele wereld, onze research fellows, maar in de periode 2008-2011 willen we dat verder uitbouwen. Netspar is door de European Science Foundation gevraagd om het voortouw te nemen in het opstellen van een onderzoeksagenda op het terrein van vergrijzing en pensionering. Daarvoor werken we onder andere samen met het Center for Economic Policy Research (CEPR) in Londen.'

Op het gebied van dataverzameling is Netspar betrokken bij de Survey of Health, Aging and Retirement in Europe (SHARE). Aan dit onderzoek

LUUK SCHREVEN

Toekomst voor Historische Steekproef Nederland

Op vrijdag 11 april 2008 werd in Amsterdam het programma *Life Courses in Context* afgesloten met een wetenschappelijk congres (zie ook pagina 2). Na afloop van dit eerste project in de humaniora dat een NWO-Groot subsidie kreeg blijkt er aan perspectieven voor de toekomst geen gebrek.

'Het *Life Courses in Context* project brengt enerzijds de levenslopen in beeld van veertig duizend Nederlanders geboren tussen 1863 en 1922,' zegt Kees Mandemakers, drijvende kracht achter de Historische Steekproef Nederland (HSN) bij het Internationaal Instituut voor Sociale Geschiedenis (IISG). 'Anderzijds bieden de gedigitaliseerde volks- en beroepstellingen uit de periode tussen 1859 en 1947 een maatschappelijke achtergrond waartegen de levenslopen geprojecteerd kunnen worden.' In *Life Courses in Context* worden namelijk twee dataverzamelingen gecombineerd: de HSN, waarin de levenslopen op basis van de burgerlijke stand en de bevolkingsregisters zijn verzameld, en de gedigitaliseerde volkstellingen. 'Voor de HSN zijn we ongeveer twintig jaar geleden met een plan begonnen, zegt Mandemakers, 'en vanaf begin jaren negentig zijn de eerste pilots in de gegevensverzameling uitgevoerd. Het doel is om een longitudinale database in te richten waarmee maatschappelijke vragen over een langere termijn beantwoord kunnen worden. In totaal hebben we nu circa 75 duizend levenslopen ingevoerd.'

Op het congres in Amsterdam werden analyses gepresenteerd op

basis van een eerste release van twintigduizend levenslopen uit de HSN databank. 'We plannen dit jaar nog twee data releases, één in juli en één in december. Daarnaast willen we in oktober een versimpelde dataset via het internet beschikbaar stellen.

Voor de toekomst ziet Mandemakers verschillende mogelijkheden om de HSN uit te breiden: 'In de eerste plaats willen we de levenslopen tussen 1812 en 1863 toevoegen zodat de HSN dan de periode van 1812 tot en met 1922 bestrijkt. Daarnaast willen we de samenwerking met de burgerlijke stand database Genlias uitbouwen om over tien jaar mogelijk uit te komen op een complete gezinsreconstructie van Nederland. Daarvoor is subsidie bij de KNAW aangevraagd. Ook is er NWO-subsidie beschikbaar gekomen om samen met vergelijkbare programma's in het buitenland tot een gezamenlijk internationaal datamodel te komen. Tenslotte hebben we een nieuwe NWO-Groot aanvraag ingediend, SPAN geheten (Spanning Past and Present). Het gaat daarbij om een samenwerkingsverband met het Nederlands Interdisciplinair Demografisch Instituut om op basis van de HSN en de *Netherlands Kinship Panel Study* een grote geïntegreerde database te maken met voor de HSN ook meer informatie over de moeders en directe familieleden van de onderzoekspersonen.' Als het SPAN-project de gewenste subsidie krijgt zal een waardevolle longitudinale database ontstaan met informatie over levenslopen tussen 1812 en 1990. Eind juni volgt de uitslag. (LS)

Geodata voor de wetenschap: studiedag en survey

Twee initiatieven op het gebied van geo-informatie zijn de studiemiddag 'Geodata op de kaart gezet' op 20 juni en een survey-onderzoek naar de stand van zaken op het gebied van de behoefte en het gebruik van geodata.



Geodata of geo-informatie is een verkorte naam voor geografische databestanden of informatie met een ruimtelijke component, een verwijzing naar een plek op de aarde. Het gaat bijvoorbeeld om informatie over

een gebied, een perceel, een kanaal of een stad. Geodata wordt uiteraard meestal in de vorm van een kaart gepresenteerd. In een geografisch informatie systeem (GIS) kunnen digitale kaartlagen worden gecom-

bineerd met elkaar en met tabellen die informatie over de ruimtelijke objecten bevatten, zoals bodemgebruik, bevolkingsaantallen, eigendom of vervoersgegevens. Geodata zijn van groot belang voor onderzoek en beleid.

Er zijn in Nederland tal van producenten, aanbieders en gebruikers van geografische databestanden, bijvoorbeeld overheidsdiensten, bedrijven en wetenschappelijk onderzoekers. Maar het aanbod is onoverzichtelijk en de toegang tot geodata is voor onderzoekers vaak lastig. Data Archiving and Networked Services (DANS) tracht de toegankelijkheid en het gebruik van geodata voor wetenschappelijk onderzoek te stimuleren.

Op de studiemiddag, die DANS in Utrecht organiseert samen met de faculteit Geowetenschappen in die stad en GeoNovum, kijken vooraanstaande sprekers uit binnen- en buitenland naar de manieren waarop geodata ter beschikking wordt gesteld en gebruikt. In een afsluitende forumdiscussie komt de vraag aan de orde hoe de toegang tot geodata voor de wetenschap verder kan worden verbeterd. Dit gebeurt aan de hand van een inventarisatie en survey van geodata. De survey is uitgevoerd in de laatste twee weken van mei 2008. Beide initiatieven zijn onderdeel van het DANS-project 'AlfaGeo' dat gedeeltelijk gefinancierd wordt door het project Ruimte voor Geo-Informatie. (René van Horik)

DANS archiveert data voor de Universiteitsbibliotheek Amsterdam

Data-instituut DANS en de Bibliotheek van de Universiteit van Amsterdam (UBA) hebben onlangs een overeenkomst afgesloten om zorg te dragen voor de duurzame archivering van een aantal digitale imagecollecties van de bibliotheek.

Het Digitaal Productiecentrum van de UBA ontwikkelt elektronische publicaties en onderzoeksdatabanken op basis van onder andere de bijzondere collecties van de bibliotheek. Zo is een aantal beeldatbanken gemaakt waarin digitale afbeeldingen van collecties met hun beschrijvingen online beschikbaar worden gesteld, waaronder kerkelijke collecties, collecties van het universiteitsmuseum en een archeologische collectie.

Bij de UBA bestond een groeiende zorg over het behoud van de grote hoeveelheden *moeder-images*, die op verschillende plaatsen en op verschillende dragers werden bewaard. Er was behoefte aan een centrale en veilige opslag van dit materiaal, bestaande uit hoge resolutiescans die vooral voor reproductiedoeleinden worden gebruikt. Vanwege de daar aanwezige expertise op het gebied van digitale archivering is besloten de



Negentiende-eeuwse prent 'ter herinnering aan het 350-jarig feest der Hervorming' van A.B.H. Braakensiek uit het UBA-archief.

images bij DANS onder te brengen.

DANS ziet het archiveren van de *image*-collecties als het verkennen van een mogelijk nieuwe vorm van dienstverlening: het archiveren van grote hoeveelheden digitaal materiaal in opdracht van andere wetenschappelijke instellingen. Het

project wordt gebruikt om een goed inzicht te krijgen in de (technische) behoeften (opslag, backup), de nodige metadata, de werkzaamheden, het beheer en de kosten. Na drie jaar zullen beide partijen kijken of de samenwerking wordt voortgezet. (Marjan Balkestein)

@ <http://dpc.uba.uva.nl/beelddatabanken>

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/dms>

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
Algemeen bedrijfsregister	1967-2006	CvB
Productie industrie, import en export	1993-2006	CvB
Community Innovation survey	2002	CvB
Statistiek financiën van kleine ondernemingen	1993-2006	CvB
Voortgezet onderwijs	2003-2006	CvB
Middelbaar beroepsonderwijs (MBO) diploma en examengegevens	2004-2005	CvB
Wet Werk en Inkomen naar Arbeidsvermogen (WIA)	2006-2007	CvB
Communautaire Statistiek voor inkomens en Levensomstandigheden (EUSILC)	2006	CvB

Data van overige instellingen	Periode	Via
B. Goudswaard		
Archeologie van de Betuweroute	2002	DANS EASY
F.B. Koek		
CLIWOC – Climatological Database for the World's Oceans 1750-1850 (release 2.1)		DANS EASY
Sociaal en Cultureel Planbureau SCP		
AVO-Sportonderzoek	2003	DANS EASY
SEOR – Erasmus Universiteit		
Sociale Positie en Voorzieningengebruik		
Allochtonen (nieuwe migrantengroepen)	2003	DANS EASY
OSA – Universiteit Tilburg		
OSA Arbeidsvraagpanel	1991-2006	DANS EASY

Beveiligde microdata	Periode	Via
Centraal Bureau voor de Statistiek		
Leefsituatie Allochtone Stedelingen	2004/2005	DANS
Centraal Bureau voor de Statistiek		
Bouwvergunningen burgerlijke- en utiliteitsbouw	2007	DANS

Een Nederlandstalig platform voor surveyonderzoek

Surveyonderzoek is al vele jaren een belangrijke methode van dataverzameling. De basisprincipes ervan zijn al tientallen jaren hetzelfde; de praktijk verandert echter. De nieuwe ontwikkelingen vragen voortdurend om herijking van de praktijk van surveyonderzoek en surveymethodologie. Dit was reden voor de oprichting van het Nederlandstalig Platform voor Survey-Onderzoek (NPSO).

Op 9 april verzorgde het platform in samenwerking met Data Archiving and Networked Services (DANS) haar eerste symposium: 'Wat is goed surveyonderzoek?' Zo'n honderdvijftig deelnemers uit wetenschap, overheidsstatistiek en commercieel marktonderzoek luisterden naar de openingslezing van directeur Paul Schnabel van het Sociaal-Cultureel Planbureau (SCP) en de afsluitende gesproken column van wetenschapsjournalist Hans van Maanen. Daarnaast waren er veel interessante bijdragen van sprekers uit wetenschap, bedrijfsleven en overheid. Zo sprak Jelke Bethlehem van het Centraal Bureau voor de Statistiek over 'Surveys voor officiële statistieken; betrouwbaarheid tot elke prijs?' en Frans Louwen van GFK



ELLEN BOUMA

Panel Services over 'Goed surveyonderzoek; illusie of haalbare kaart?'. Ook was er nog ruimte voor enkele discussies onder leiding van Ineke Stoop van het SCP. De hoge opkomst en vele positieve reacties op deze dag onderstreepten de behoefte aan een Nederlandstalig platform. Internationaal zijn er weliswaar vele platforms waar surveyonderzoekers samenkomen. Surveys blijken echter vaak sterk nationale en cultuurgebonden invloeden te kennen. Een overzicht van standaarden, *best practices* en aandachtspunten bij de keuze van een surveymethode, en de mogelijkheid van collegiaal advies en wederzijdse ondersteuning werken kwaliteitsverhogend, juist bij een continue verandering van de praktijk van het surveyonderzoek.

Het Nederlandstalig Platform voor Survey-Onderzoek is een initiatief van Jelke Bethlehem, Barry Schouten en Ineke Stoop. Het platform wil graag iedereen samenbrengen die zich bezig houdt met de methodologie van surveyonderzoek om zo een actieve uitwisseling van kennis en ervaring tot stand te brengen tussen academia, marktonderzoekbureaus en de 'official statistics'. De NPSO wil in ieder geval een vaste jaarlijkse lezingendag verzorgen en ad hoc lezingenmiddagen over specifieke thema's. Daarnaast gaat ze haar website verder uitbreiden. In de weken na het symposium houdt de NPSO een enquête om de meningen te peilen en suggesties te vragen van alle aanwezigen en geïnteresseerden. De resultaten zijn te volgen op www.npso.net. (Barry Schouten)

Een infrastructuur voor harmonisatie en conversie van sociaalwetenschappelijke data

Steeds meer data worden verzameld in internationale projecten. Daarom is de vergelijkbaarheid en harmonisatie van data een belangrijk onderwerp geworden in het sociaalwetenschappelijk onderzoek. Een probleem bij de verschillende harmonisatieprojecten is dat de kennis die bij deze projecten wordt opgedaan vaak na afloop van zo'n project verloren gaat. Begin april was er in Parijs een workshop waar verschillende vooraanstaande onderzoekers op het gebied van dataharmonisatie hun ideeën presenteerden. Dat gebeurde in het kader van het Preparatory Phase Project van de Council of European Social Science Data Archives (CESSDA PPP). Verschillende problemen rond het harmoniseren van sociaal-structurele variabelen zoals beroep en opleiding, geografische variabelen en attitudevariabelen passeerden

de revue. Duidelijk werd dat er geen consensus bestaat over de manier waarop harmonisatie moet worden aangepakt, maar dat er verschillende ideeën en stromingen bestaan met ieder hun eigen strategie. De projectmedewerkers van CESSDA PPP gaan deze ideeën verder uitwerken met het doel om een Europese infrastructuur te ontwikkelen waar kennis, tools en producten van harmonisatieprojecten worden verzameld, verrijkt en toegankelijk gemaakt. Het Nederlandse data-instituut DANS (Data Archiving and Networked Services) heeft binnen het project als taak om te kijken op welke manier hierbij een (internationale) vragendatabank ontwikkeld moet worden. (Marion Wittenberg)

@ <http://www.nsd.uib.no/cessda/ppp/about/>

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: Uitgeverij Aksant, Postbus 2169, 1000 CD Amsterdam

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

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

Aan dit nummer werkten mee: Marjan Balkestein, Lodewijk Bogaards, Ben van Cleef, Maarten

Hoogerwerf, René van Horik, Edwin Klijn, Henk Laloli, Michiel Leenaars, Barry Schouten, Marion Wittenberg

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Vormgeving en opmaak: Ellen Bouma

Productie: Uitgeverij Aksant, Amsterdam

Druk: PlantijnCasparie, Almere

ISSN: 1872-0374

Historische informatie over elke plek in Nederland

'Historische informatie over elke plek in Nederland'. Dat biedt de website WatWasWaar, product van samenwerking tussen een groot aantal erfgoedinstellingen. 'Het WatWasWaar-avontuur begon in 2005 met het ontwikkelen van een toegankelijke manier om de oudste kadasterkaart van Nederland te ontsluiten' aldus Job Gerlings van WatWasWaar. Deze kadasterkaart uit 1832 was kort daarvoor gedigitaliseerd. Het bleek een lastige klus om deze toegankelijk te presenteren. En dus werd er een projectgroep aan het werk gezet om deze en andere kaarten gebruikersvriendelijk op het internet te zetten.



Militaire kaart Bijerland 1850, bron Nationaal Archief

believen in- en uitzoomen. Ook kun je bekijken hoe een perceel in de loop van de tijd van eigenaar is gewisseld door de gedigitaliseerde kadasterboeken te raadplegen. Voor je het weet dwaal je zo zomaar een aangenaam uurtje door historisch Nederland. Het geheel werkt dankzij een zelfontwikkelde *open source* GIS-applicatie die vele informatie-soorten aan kan en is gebouwd 'in een tijd waarin Google Maps en Google Earth nog niet bestonden' zo vertelt Job.

In WatWasWaar neemt een groot aantal erfgoedinstellingen deel, waaronder het Nationaal archief, het Kadaster en gemeentearchieven. Ze werken ook mee aan het ontwikkelen van standaarden door Digitaal Erfgoed Nederland. Hierbij is WatWasWaar – hoe kan het ook anders – vooral gefocust op de locatie-aspecten. Alle informatie die betrekking heeft op een geografisch punt moet locatiekenmerken krijgen, zoals een postcode of een x-y-coördinaat. De heilige graal van WatWasWaar is immers om alle historische gegevens over Nederland die aan een locatie verbonden zijn via haar website toegankelijk

te maken. Het richt zich op een eerste kennismaking met het materiaal via watwaswaar.nl en een gerichte doorverwijzing naar gedetailleerde informatie van andere erfgoedinstellingen zoals musea en bibliotheken. WatWasWaar zoekt daarom de samenwerking op met organisaties als het Geheugen van Nederland en het instituut Beeld en Geluid. Hoog op de verlanglijst staat ook om de Atlas van Willem Blaeu uit de zeventiende eeuw van het Regionaal Archief Leiden op de website te zetten. Mede geïnspireerd door de kortgeleden bezochte conferentie *Museums on the Web* in Canada barst de projectgroep van de nieuwe ideeën. 'Wat bijvoorbeeld straks mogelijk zou kunnen zijn', vertelt Job enthousiast, 'is dat je de WatWasWaar-informatie koppelt met je TomTom en je via je navigatiesysteem informatie krijgt over interessante historische gebouwen die je passeert. Maar dat is nog toekomstmuziek. Op kortere termijn zijn we op zoek naar een koppeling met Google Earth en met het ontwikkelen van een API (Application Programming Interface - IG) die ervoor zorgt dat je de kaartinformatie op je eigen website gebruikt waarna je er zelf toepassingen mee kunt maken.' Aan ideeën geen gebrek dus. WatWasWaar



Amsterdam grachtengordel, Kadasterkaart 1832

zoekt daarom volop naar nieuwe deelnemers, want ze zijn nog lang niet klaar met hun avontuur. (Ivo Gorissen)



Openings scherm WatWasWaar

Deze projectgroep die de toepasselijke naam WatWasWaar draagt, bestaat uit vier personen die opereren vanuit een kamer in het Amsterdamse stadsarchief aan de statige Herengracht. Ze laten zich ondersteunen door een vormgever en een softwarebureau. In september 2007 lanceerden ze de website watwaswaar.nl, met daarop onder meer de eerder genoemde kadasterkaart uit 1832 en een aantal militaire kaarten.

In de opzet om het materiaal toegankelijk te presenteren zijn ze bij WatWasWaar geslaagd. Wil je een bepaald gebied bekijken dan kom je daar door te klikken op de kaart van Nederland of door een zoekterm op te geven. Vervolgens kun je voor de kaart van je interesse kiezen en naar

Gelezen



Sesink, L., R. van Horik, H. Harmsen (red): *Datakeurmerk, kwaliteitsrichtlijnen voor digitale onderzoeksdata in Nederland*; Den Haag, DANS, 2008 (pp.18)

Om te zorgen dat gearchiveerde data in de toekomst nog te vinden, te herkennen en te gebruiken zijn, kreeg DANS bij de oprichting van KNAW en NWO de opdracht om een datakeurmerk te ontwikkelen: richtlijnen voor het duurzaam opslaan en toegankelijk houden van data. De in dit document geformuleerde kwaliteitsrichtlijnen zijn van belang voor onderzoekers en instellingen die digitale onderzoeksbestanden aanleggen, voor organisaties die onderzoeksbestanden archiveren en voor gebruikers van onderzoeksdata. De uitgave kan kosteloos worden besteld via info@dans.knaw.nl. Downloaden kan ook, vanaf de Datakeurmerk website: <http://www.datakeurmerk.nl/>

Gideon, basisvoorziening geo-informatie Nederland, Visie en implementatiestrategie 2008-2011; Den Haag, Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer, 2008.

In opdracht van het ambtelijke Geo Informatie-beraad hebben Geonovum en Ruimte voor Geo-informatie (RGI) zich gebogen over de verdere ontwikkeling van een nationale geo-informatie infrastructuur (NGII). De nota bevat een visie op de verdere ontwikkeling van de NGII. Daarnaast gaat GIDEON ook in op de uitvoering en de implementatie. De verdere bestuurlijke verankering van GIDEON wordt ter hand genomen door de ministeries van VROM en Binnenlandse Zaken en Koninkrijksrelaties. De nota wordt respectievelijk aangeboden aan de Ministerraad en de Tweede Kamer.

Wichers Hoeth, L., onder redactie van Ch. Gielen e.a.: Kort begrip van het intellectuele eigendomsrecht; Deventer, Kluwer, 2007 - ISBN 9789013043341

Dit is een up-to-date gebrachte versie van een al langer bestaand handboek op het terrein van de intellectuele eigendomsrechten. Het is in de eerste plaats een naslagwerk en bevat een uitgebreide verzameling verdragen, Europese richtlijnen en Nederlandse wetgeving samen met de hierop gebaseerde jurisprudentie. Het is daardoor een rijke bron van informatie,

niet alleen voor juristen maar voor iedereen die met intellectuele eigendomsrechten te maken heeft. Een terrein dat bij uitstek volop in beweging is wordt gevormd door de digitale eigendomsrechten. Auteursrechten op digitale publicaties en software krijgen dan ook veel aandacht en aan het databankrecht is een apart hoofdstuk gewijd.

Leeuw, Edith D. de, Joop J. Hox, Don A. Dillman: International handbook of survey methodology; Psychology Press, 2008 - ISBN 9780805857528

Taking into account both traditional and emerging modes, this comprehensive new handbook covers all major methodological and statistical issues in designing and analyzing surveys. With contributions from the world's leading survey methodologists and statisticians, this invaluable new resource provides guidance on collecting survey data and creating meaningful results. Intended for advanced students and researchers in the behavioral, social, and health sciences, this 'must have' resource will appeal to those interested in conducting or using survey data from anywhere in the world, especially those interested in comparing results across countries. The book also serves as a state-of-the-art text for graduate level courses and seminars on survey methodology. A companion website contains additional readings and examples.

Column

Michiel Leenaars

Dweilen onder de kenniskraan

Onze moderne wetenschap is het meest achtergebleven in de manier waarop we haar vastleggen. We maken grifgebruik van computers om te experimenteren, modelleren, visualiseren en via repositories beschikbaar te stellen. Maar daarna doen we het nog steeds zoals we honderden jaren geleden ook al deden: we schrijven jaar in jaar uit verslagjes van onze experimentjes en ons onderzoek, alsof het krantennieuws is. Het is allemaal zo comfortabel quasi-lineair, zo ad hoc, zo incrementeel archiefgericht – eigenlijk: zo geautomatiseerd zeventiende-eeuws. We hebben wat je wilt weten wel ergens hoor, ga maar eerst binnen het juiste vakgebied het juiste tijdschrift vinden en dan twintig jaargangen heen en weer bladeren om handmatig alle directe en indirecte referenties naar het oorspronkelijke artikel in alle andere wetenschappelijke tijdschriften na te lopen. De belangrijkste innovatie dankzij de computer: alles online, en dan op citaten kunnen klikken en full text search kunnen doen. Jippie.

Citeren kunnen we allemaal geweldig – en dat valt in ons te prijzen. Het was vier eeuwen terug ook een enorme sprong vooruit dat we leerden hoe systematisch naar elkaar te verwijzen. Maar is het geen tijd om in het tijdperk van een wereldwijd internet en dito web meer te willen? De boekhouding van de wetenschap is een constructie van lik-mijn-vestje. Als wetenschapper B vijftiengintig jaar na wetenschapper A bewijst dat iets onjuist is, hebben we geen mechanisme om al het onderzoek dat besmet is door de fouten van A tussen haakjes te zetten – en iedereen die patiënten onnodig behandelt met een gevaarlijk medicijn met een mailtje te kunnen waarschuwen. Is onze kennis echt niet zo vast te leggen dat we real-time overzicht hebben en dat we minder meta-onderzoek nodig hebben om te snappen wat de stand van zaken is? Bij gebrek aan beter, moeten we binnen de wetenschap uit pure armoede terugvallen op Wikipedia en haar talrijke kleinere zusters om onze kennis in niet-semantic hypertext te disambigueren en aan elkaar te linken, om discussie in real-time samen te vatten en te annoteren. De kenniskraan staat open, en wat we weten als individu hangt af van hoe goed we kunnen dweilen – maar weten hoe het opschiet doen we niet. We zetten geen ICT in om ons voortschrijdend inzicht zelf te mappen, en op een slimme manier te beschrijven hoe complex en genuanceerd ons totaalmodel van de werkelijkheid nu is. Dat model is er ook niet, tenminste niet buiten onze eigen hoofden. We hebben wel stapels en stapels met tienduizenden wetenschappelijke tijdschriften in onze bibliotheken en online. Die razendsnel kunnen doorzoeken maakt het alleen niet helderder hoe die informatie zich nu tot elkaar verhoudt, maar laat ook zien hoe pijnlijk de kloof is tussen letterlijk honderden miljoenen wetenschappelijke artikelen en de mens die kennis tot zich wil nemen. Een feitelijk onmogelijke taak. Data zijn beschikbaar, maar niet verwerkbaar. Waar blijft de ICT als je haar echt nodig hebt?

Michiel Leenaars is directeur van de vereniging Internet Society Nederland en adviseur van de NWO-stichting Nationale Computerfaciliteiten.

Klijn, Edwin en Yola de Lusenet: Tracking the reel world, a survey of audiovisual collections in Europe; European Commission on Preservation and Access (ECPA), Amsterdam, 2008 - ISBN 9789069845463

This report presents the results of a study on audiovisual collections undertaken in the framework of TAPE (Training for Audiovisual Preservation in Europe). Based on data, collected from hundreds of organizations all over Europe, most of them archives, libraries, museums and (research) institutes that hold audiovisual recordings as minority collections, this is the first extensive report on preservation and access of sound and moving image material outside the broadcasting sector and large audiovisual archives.



PAUL COERTEN