

Pagina 3 • UNESCO • In Amsterdam spraken op uitnodiging van UNESCO wetenschappers van over de hele wereld over Open Access in ontwikkelingslanden.

Pagina 4 • Historisch GIS • Het project HISGIS van de Fryske Akademy heeft een basissysteem voor historische geo-informatie in een viertal provincies, en wil dat graag over heel Nederland uitrollen.



Pagina 5 • Esther Jansma • Begin januari ging de elektronische archiefwerkplaats voor dendrochronologen DCCD open voor gebruikers. Boomtjidskundigen uit alle windrichtingen kunnen hier hun bevindingen opslaan en vergelijken. *e-data@research* sprak met initiatiefneemster Esther Jansma.

Pagina 6 • Meertens Instituut • De rubriek Focus richt zijn lens op een Amsterdams instituut dat Nederlandse taal en cultuur als werkgebied heeft.

Pagina 6 • Muziekgelijkenis • Aan de Universiteit Utrecht wordt gezocht naar manieren om muziekfragmenten automatisch te doen herkennen.

EN VERDER

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

KNAW: programma e-Humanities van start

Bij de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) is in februari het programma e-Humanities van start gegaan. Kennis opbouwen van informatietechnologie voor de alfawetenschappen en onderzoek op dat terrein bijeenbrengen, is het doel. Er wordt gestreefd naar samenwerking met universitaire onderzoekers. Het programma loopt tot 2015 en heeft twee thema's: *computational humanities*, dat rond de omgang met onderzoeksgegevens draait, en *cultures of e-humanities*, waarin het effect centraal staat van innovatie op de manier van onderzoek doen. Bij het programma zijn verschillende medewerkers van de voormalige Virtual Knowledge Studio betrokken.

© <http://ehumanities.nl/>

Nieuwe SURF-projecten helpen onderzoeker

Onder de vleugels van SURFshare is onlangs een aantal projecten van start gegaan om te bevorderen dat onderzoekers in de geestes- en sociale wetenschappen de digitale mogelijkheden ten volle benutten. De subsidieronde Verrijkte Publicaties van het SURFshare programma heeft zes projecten opgeleverd die half januari van start zijn gegaan. Het CARDS-project (controlled storage of and access to research data) is bedoeld om onderzoekers te helpen hun data efficiënt en doeltreffend te beheren.

Uit internationaal onderzoek, zoals PARSE.Insight (zie e 5-2) is gebleken dat veel wetenschappers best bereid zijn om hun gegevens ter beschikking te stellen aan anderen, als ze maar de controle kunnen houden over wie er op welk moment en onder welke voorwaarden gebruik van mag maken. Ook kwamen grote verschillen naar voren tussen de werkomgeving in diverse disciplines. Daarom wil het CARDSproject specifieke, lokale ondersteuning bieden aan onderzoekers vanuit de universitaire repository's en vanuit data-archieven als DANS (Data Archiving and Networked Services) en het 3TU.Datacentrum. Daarin spelen accountmanagers een belangrijke rol. Het is niet de bedoeling, aldus de SURF-website, 'dat de accountmanager binnenkomt met een map

vol adviezen, procedures en regels'. Daarentegen speelt hij 'al luisterend in op de problemen en biedt oplossingen en advies'. Daarnaast zullen de ervaringen uit de projecten leiden tot een digitaal informatiepakket voor grotere groepen onderzoekers. Welke onderzoeksgroepen deelnemen aan het project was bij het ter perse gaan van deze *e-data@research* nog niet bekendgemaakt.

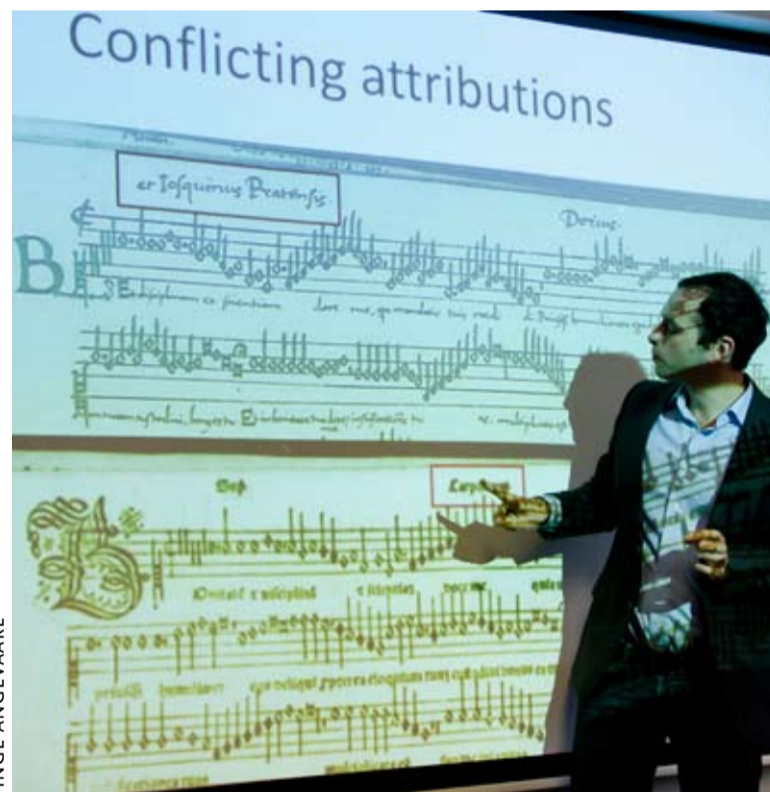
De verrijkte publicatie staat centraal in een zestal andere SURFshareprojecten: Open Data and Publications (economie), Lenguas de Bolivia en Enhanced NIAS Publications (taalwetenschappen), TheOther Josquin (muziekwetenschappen), Enhancing Scholarly Publishing in the Humanities and Social Sciences (communicatiewetenschappen) en VP cross (geowetenschappen). In

alle gevallen bieden SURF, DANS en de universitaire repository's de onderzoekers hulp bij het verrijken van hun publicaties met allerhande online bronnen. Linked Data technieken (semantisch web) worden ingezet om die bronnen ook goed

doorzoekbaar te maken. DANS en de KB doen op basis van deze voorbeelden concreet onderzoek naar de duurzaamheid van verrijkte publicaties en hun vindbaarheid op de lange termijn.

© www.surfoundation.nl/nl/actueel/Pages/Onderzoekersverrijkenhunpublicaties.aspx

Betwiste composities belicht



INGE ANGEVAARE

Musicoloog Theodor Dumitrescu van de Universiteit Utrecht presenteert een onderdeel van de website die hij met hulp van het SURFshare programma bouwt; een site met informatie over alle betwiste composities van de zestiende-eeuwse componist Josquin des Prez – ter verrijking van de gedrukte publicatie *The New Josquin Edition*, die alleen door de samenstellers als authentiek beoordeelde composities bevat. Nu kunnen muziekwetenschappers van over de hele wereld zelf hun oordeel vellen over de keuzes die de samenstellers hebben gemaakt. (IA)

CLARIN-NL deelnemer in e-data&research

Sinds 1 januari is het infrastructuurproject CLARIN-NL als *stakeholder* aan *e-data@research* verbonden. CLARIN-NL (Common Language Resources and Technology Infrastructure), dat is geworteld in de taalwetenschappen, is de Nederlandse pijler van een Europees project om een elektronische onderzoek-infrastructuur tot stand te brengen voor onderzoekers in de geesteswetenschappen.

CLARIN-NL is na CentERdata de tweede nieuwe deelnemer in korte tijd die zich aansluit bij het

Centraal Bureau voor de Statistiek, DANS, het Huygens Instituut voor Nederlandse Geschiedenis, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Deze partners maken het met hun financiële en redactionele bijdrage mogelijk dat *e-data@research* wordt verspreid onder meer dan achtduizend onderzoekers in de alfa- en gammawetenschappen.

© www.clarin.nl

Google Award voor Nederlandse onderzoekster

Suzan Verberne, taaltechnoloog aan de Radboud Universiteit, heeft van Google vijftigduizend dollar ontvangen om haar projectvoorstel *Extracting Factoids from Dutch Texts* uit te voeren. In het project werken taalkundigen en informatici samen aan software die feiten uit een tekst extraheert. Met deze software kunnen grote tekstcollecties makkelijker gebruikt worden voor onderzoek binnen de geesteswetenschappen.

Een gewone zoekmachine signaleert alleen het voorkomen van de zoekterm in een tekstbestand. De onderzoeker moet dan zelf op zoek naar de informatie die er in de tekst over het zoekonderwerp te vinden is.

Verbernes software moet syntactische en semantische rollen in de tekst onderscheiden en labelen. Zo krijgt de onderzoeker direct toegang tot de feiten over het zoekonderwerp die liggen opgesloten in de brontekst. De onderzoekster verwacht dat vooral historici en letterkundigen veel aan de software zullen hebben.

Het project is het enige Nederlandse voorstel dat binnen Google's eerste Europese *Digital Humanities Research Awards* is gehonoreerd. Met deze Awards hoopt het zoekbedrijf de data uit haar GoogleBooks-project beter beschikbaar te maken voor onderzoekers uit de geestes- en sociale wetenschappen. (ER)

Beschermheilige van het internet



Eind vorig jaar lanceerde het Meertens Instituut samen met DANS een website met beschermheiligen, gebaseerd op de Volkskundige Atlas van Nederland en Vlaanderen uit de vorige eeuw. De site biedt informatie over de heiligen en links naar diverse relevante bronnen, zoals de Databank Bedevaart en Bedevaartplaatsen. De heilige Isidorus van Sevilla (ca. 560-636, foto) heeft in die databank een bijzondere plek gekregen. Hij werd in ons land vooral vereerd om zijn vermogen paarden te genezen, maar is volgens het instituut inmiddels de 'officiële patroonheilige van het internet' geworden.

© www.meertens.knaw.nl/beschermheiligen
 © www.meertens.knaw.nl/bedevaart

AGENDA

14 en 15 maart

Hilversum – Beeld en Geluid

Screening the Future – New Strategies and Challenges in Audiovisual Archiving
Participants will help set the agenda for AV preservation in Europe, and benefit from interacting with leading institutions, funders, vendors, and policy makers. Participants include managers from broadcast archives, film archives, audio archives, government archives, corporate archives, libraries, museums, universities, and commercial organisations, as well as funders, educators, technologists and film makers. <http://screeningthefuture.eventbrite.com/>

3-8 april

Vienna – Austria Center Vienna

Data publishing: paving the way to data-intensive science; 2011 General Assembly European Geosciences Union
Topics: Data publication workflows (incl. cross-linking issues between journals & data archives), trustworthiness of data archives (certification and accreditation), peer review models and QA/QC procedures, persistent identifier & catalogues, bibliometrics, data citation issues (syntax, granularity etc.), intellectual property rights & licensing, metadata & data exchange standards, editorial systems for data, business models, user experiences, expectations, and needs, new concepts (linked data, open data, clouds). <http://meetings.copernicus.org/egu2011/home.html>

<http://meetings.copernicus.org/egu2011/home.html>

12 en 13 april

Amsterdam – Science Park Amsterdam

2011 User Forum European Persistent Identifier Consortium (EPIC)

Het programma omvat een technisch deel en een deel over beleidsontwikkeling. EPIC is een samenwerkingsverband van CSC – IT Center for Science Ltd (Helsinki), GWDG – Gesellschaft für wissenschaftliche Datenverarbeitung mbH (Göttingen) en SARA Reken- en Netwerkdiensten (Amsterdam) om aan Europese wetenschappers een Persistent Identifier (PID)-infrastructuur aan te bieden. www.sara.nl/news/recent/20101216/EPIC_User_Forum_2011_ned.html

9 mei

Den Haag – Koninklijke Bibliotheek

Out of the Box ; Building and Using Web Archive Collections

This public event, presented by the International Internet Preservation Consortium (IIPC) and the National Library of the Netherlands, will offer an opportunity for curators and creators of web archives to show off the contents of their collections through stories, illustrations and demonstrations and discuss their collection policies. Researchers and users of web archives will be able to present and discuss their needs and expectations when approaching a collection of archival web sites. <http://netpreserve.org/events/2011GAoutofthebox.php>

11-13 mei

Kopenhagen

Click-on-Knowledge – web based knowledge and contemporary scholarship

The conference will discuss a broad range of issues, all concerning the interactions between research and study in the humanities and so-called web-based knowledge. Topics are the epistemologies of knowledge, issues concerning legal copyrights, the construction of national legacy on the web, the development of web-based learning techniques, and the historical perspectives on modern-day IT culture and the dissemination of knowledge. <http://engerom.ku.dk/clickonknowledge/about/>

12 en 13 mei

Capri (It.), ISTI-CNR, University of Athens

Workshop on Global scientific data infrastructures, the big data challenges

The next generation of research data infrastructures is facing two main challenges: to effectively and efficiently support data-intensive science and to effectively and efficiently support multidisciplinary/interdisciplinary science. In order to develop such data infrastructures several data, applications, systems, and organizational as well as policy challenges must be successfully tackled. The workshop will address the main data challenges. They include data modelling, data management, and data services challenges. <http://datachallenges.isti.cnr.it/>

De rol van het boek in nieuwe kennisomgevingen

Boeken zijn de dragers van kennis, tradities en cultuur. Wanneer we boeken vervangen door websites en e-readers, dan is de vraag hoe die dragende rol van het boek in nieuwe omgevingen zeker gesteld kan worden. Het beantwoorden van die vraag is de hoofddoelstelling van het Canadese INKE project ('Implementing New Knowledge Environments'). Het project, geleid door Ray Siemens, organiseerde op 15 december een symposium in de Koninklijke Bibliotheek in Den Haag, in samenwerking met de afdeling Book and Digital Media Studies van de Universiteit van Leiden.

Zulke vragen kunnen alleen beantwoord worden in een dialoog van historici, technici, interface ontwerpers en specialisten in mediastudies. Historici rapporteerden over het online naslagwerk *The Architectures of the Book* – de faciliteiten die het

boek in de loop van de eeuwen heeft ontwikkeld voor de efficiënte overdracht van kennis (Cunningham). Een voorbeeld van zulke faciliteiten zijn uitgebreide systemen voor thematische kruisverwijzingen in sommige Bijbels (Nelson). Andere onderzoekers richtten zich juist op de faciliteiten van elektronische apparatuur. Welke tekstsoorten komen bijvoorbeeld tot hun recht op welke soorten beeldschermen (Petty)? Wat voor hulpmiddelen voor navigatie zijn het best geschikt om toegang te bieden tot de grote tekstcollecties van de toekomst (Sondheim)? Hoe kunnen we navigatiehulpmiddelen (Giacometti)?

Culturele tradities verschillen in hun omgang met de digitale mogelijkheden. In digitale edities van Homerus worden tegenwoordig de afzonderlijke handschriften ge-



ORR9220

waardeerd om hun verschillen. In de studie van het Nieuwe Testament is men daarentegen nog op zoek naar de uiteindelijke 'ware' tekst (Clivaz). In de studie van moderne literatuur

wordt, vaak nog voorzichtig, geprobeerd digitale technologie in te zetten voor interpretatie (Howell).

Het spreekt trouwens niet voor zich dat we de waarden van de boekcultuur (zoals stabiliteit en gezag) in de elektronische omgeving moeten proberen te repliceren. Er is juist ook ruimte voor experimenten met meer 'vloeiende', doorlopende en collaboratieve publicatievormen (Adema). Dialoog blijft geboden, was de conclusie van Adriaan van der Weel (Leiden), mede-organisator van de bijeenkomst. (PB)

DDI wordt volwassen

Onder het motto 'DDI – The Basis of Managing the Data Life Cycle' vond op 8 en 9 december de tweede European DDI Users Group meeting plaats bij SurfFoundation in Utrecht. Duidelijk werd dat die basis steeds breder wordt en dat het niet langer de vraag is of, maar op welke wijze de DDI standaard moet worden geadopteerd en geïmplementeerd.

Peter Wittenburg van het Max Planck Institute of Psycholinguistics, en een van de auteurs van het EU-rapport *Riding the Wave* over een Europese infrastructuur, was meteen de eerste om de groeiende gebruikersgemeenschap voor te houden dat we er nog lang niet zijn. In zijn bevolgen keynote speech kwamen de verschillende uitdagingen aan bod die in de data-intensieve wetenschap nog voor ons liggen. Gedurende de bijeenkomst werden vervolgens vele initiatieven gepresenteerd om deze uitdagingen het hoofd te bieden.

'Data Documentation Initiative' (DDI), een XML-standaard voor het documenteren van data en metadata in de Sociale Wetenschappen, ging bijna veertig jaar geleden van start. Sinds een kleine drie jaar is versie 3.x in gebruik. Nieuw in deze versie is dat de volledige Data Life Cycle wordt ondersteund. Ook kunnen nu longitudinale data worden gedocumenteerd. Omdat versie 3 nog maar relatief kort beschikbaar is, was er op de bijeenkomst veel belangstelling voor toepassingen en gereedschappen voor DDI 3.x en voor ervaringen met het omzetten van DDI versie 2 naar 3.x.

De jaarlijkse European DDI Users Group meeting is bedoeld als platform voor de Europese DDI gebruikersgemeenschap. Dit jaar werd de bijeenkomst georganiseerd door de bibliotheek van de Universiteit van Tilburg, DANS, CentERdata en Surf Foundation. Volgend jaar zal Swedish National Data Service (SND) in Göteborg de bijeenkomst herbergen. (EB)

www.iza.org/conference_files/eddio

Verdere dialoog nodig voor Biografisch Portaal

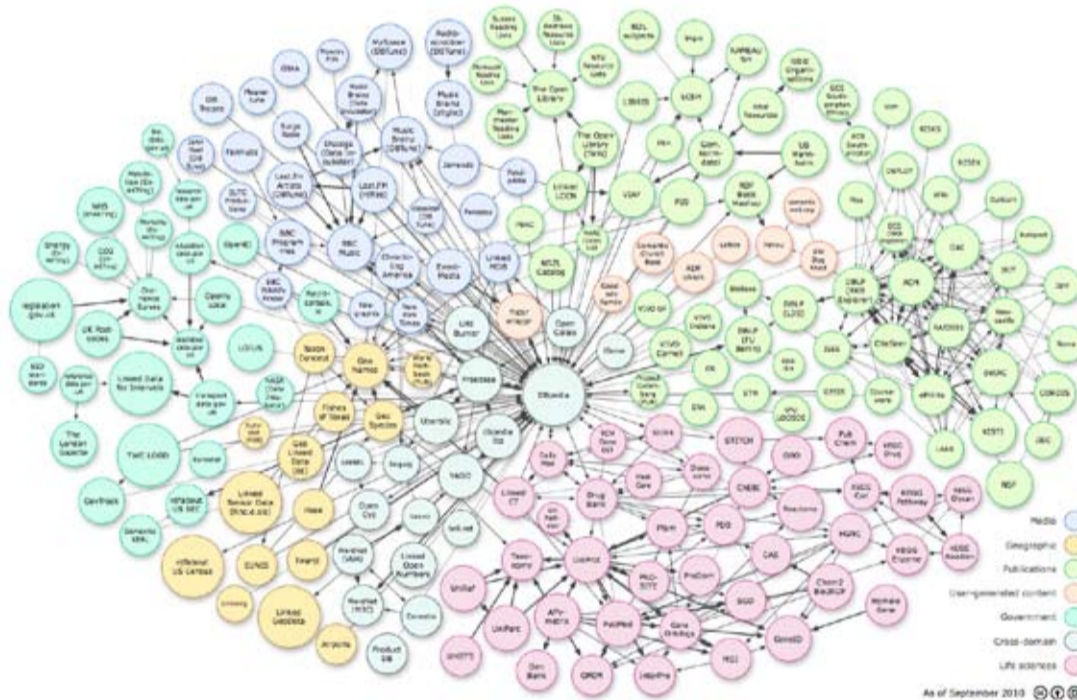


Diagram onderhouden door Richard Cyganiak (DERI, NUI Galway) en Anja Jentzsch (Freie Universität Berlin) waarin datasets worden getoond – in de e-versie als weblinks – die zijn gepubliceerd in het Linked Data format.

De dialoog tussen onderzoekers en informatiekundigen moet worden voortgezet om de kwaliteit en bruikbaarheid van het Biografisch Portaal te laten groeien. Dat bleek op 10 december in Den Haag, bij een werkconferentie over de mogelijkheden van het portaal als instrument voor wetenschappelijk onderzoek. Organisator was het portaal zelf, dat als taak heeft toegang te bieden tot biografisch materiaal over kopstukken en andere opzienbarende figuren uit de Nederlandse geschiedenis.

Dat wetenschappelijk onderzoekers het biografisch materiaal intensief gebruiken, bleek uit de bijdragen over de Historische Steekproef Nederlandse Bevolking en twee projecten op het gebied van collectief biografisch onderzoek (het Ecartico-project over de Amsterdamse kunst-

handel in de Gouden Eeuw en een prosopografisch onderzoekproject naar de Bourgondische adel). De presentaties maakten duidelijk dat verschillende projecten verschillende eisen stellen aan biografische gegevens. Waar voor het ene onderzoek globale kenmerken van een persoon voldoende, is er voor het andere een volledige biografie noodzakelijk. Het is een aardige uitdaging voor het Biografisch Portaal om aan alle wensen tegemoet te komen.

De bijdragen uit de informatiekundige hoek gaven een aanzet tot een oplossing. Onder de noemer van het Semantic Web komen er steeds meer diensten beschikbaar om informatie te verrijken, te koppelen en te delen. Zogenaamde *Linked Data* technieken zorgen er voor dat gestructureerde datacollecties met elkaar kunnen worden verbonden.

Steeds meer databestanden komen beschikbaar als *Linked Open Data*, waarvan de DBpedia, de database van WIKIpedia, wel de bekendste is. Door de biografische databestanden als *Linked Open Data* aan elkaar te koppelen ontstaat er een groot digitaal biografisch apparaat en het Biografisch Portaal lijkt de aangewezen organisatie om de kwaliteit van de biografische informatie te verzekeren.

De voor het onderzoek vereiste eenduidige koppelingen tussen persoonsgegevens vereisen de beschikbaarheid van een unieke persoonscode. De totstandkoming van dit 'historisch burgerservicenummer' vergt nader overleg en ook hier ligt een rol voor het Portaal. (René van Horik)

www.biografischportaal.nl
www.biografischportaal.nl/blog
(verslag)



INCE ANCEVAARE

Online panelonderzoek naar kwaliteit van leven (ex-)kankerpatiënten

Vorig najaar is onder de naam Profiel begonnen met het online verzamelen van gegevens over de kwaliteit van leven van (ex-) kankerpatiënten in de regio Noord-Brabant en Noord-Limburg. De studie, met als officiële titel *Patient Reported Outcomes Following Initial treatment and Evaluation of Long term Survivorship (PROFILES)*, wordt mogelijk gemaakt door een investeringssubsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). De looptijd is in eerste instantie vijf jaar, maar volgens hoofdonderzoeker Lonneke van de Poll-Franse is alles erop gericht om het onderzoek na die vijf jaar te continueren.

Profiel is een samenwerkingsverband tussen onderzoekers van het Integraal Kankercentrum Zuid, de Universiteit van Tilburg en medisch specialisten van de ziekenhuizen in Noord-Brabant en Noord-Limburg. Sinds 2004 doen deze partners al onderzoek naar de psychosociale en lichamelijke gevolgen van kanker en de behandelwijzen. Vorig jaar is daar een longitudinale en een online component aan toegevoegd. Met het door

CentERdata ontwikkelde digitaal patiëntenvolgsysteem kunnen mensen die kanker hebben of hebben gehad periodiek via het internet worden ondervraagd. Van de Poll-Franse, verantwoordelijk voor het uitwerken van nieuwe onderzoeksvragen en het coördineren van de opzet en uitvoer van de studie: 'De techniek van ondervraging via het internet in plaats van op papier maakt ons onderzoek naar mensen die kanker overleefden flexibeler, betrouwbaarder en betaalbaar'.

De data van de studie zullen beschikbaar zijn voor wetenschappelijk onderzoek binnen de geldende privacyregels. Profiel volgt hierbij het Data Seal of Approval, zoals opgesteld door DANS (Data Archiving and Networked Services). In de loop van dit jaar zullen de data via internet beschikbaar worden gesteld. (EB)



www.profielstudie.nl

Unesco: open access voor het Zuiden

A *Global View on Open Access* was het thema van een bijeenkomst die de Nationale Unesco Commissie op 20 januari organiseerde in het statige Trippenhuis in Amsterdam. Onder voorzitterschap van Jos Engelen (NWO) en Robbert Dijkgraaf (KNAW) discussieerde een grote groep betrokkenen van over de hele wereld over de problematiek van *open access* tot wetenschappelijke kennis in ontwikkelingslanden. Waar de moderne technologie goede mogelijkheden zou kunnen bieden om wetenschappelijke kennis breed te verspreiden, blijkt diezelfde digitale technologie vaak een struikelblok in landen waar de ICT-infrastructuur slecht ontwikkeld is. Een bibliothecaris uit Kameroen illustreerde dit aan de hand van haar eigen praktijk: haar bibliotheek heeft continu last van een onbetrouwbare elektriciteitsvoorziening. Ook het 'author pays'-model is voor ondermaats gesubsidieerde wetenschappers in het zuiden geen optie.

Tijdens de bijeenkomst werd overigens duidelijk dat 'het' zuiden niet bestaat: er zijn grote verschillen tussen continenten, landen en ook regio's. In Azië wordt al 75% van de wetenschappelijke tijdschriften in open access gepubliceerd, tegen slechts 20% in Afrika.

Unesco wil met de uitgevers en met beleidsmakers overleggen over mogelijkheden om open access in te

zetten als instrument om de ontwikkeling van de wetenschap in het zuiden te bevorderen – en daarbij gaat het niet alleen om het beschikbaar stellen van westerse kennis, maar ook om het streven wetenschappelijke kennis uit het zuiden meer aandacht en aanzien te geven in het noorden. Want momenteel publiceren wetenschappers uit het zuiden toch het liefst in westerse tijdschriften met een hoge impactfactor, hetgeen de ontwikkeling van de eigen wetenschappelijke traditie zeker niet ten goede komt. En dat leidt weer tot een gebrek aan primaire data waar beleidsbeslissingen op gebaseerd kunnen worden. (IA)

www.unesco.nl/

DARIAH loopt warm voor bouwfase

De voorbereidende fase van het Europese infrastructuurproject DARIAH (Digital Research Infrastructure for the Arts and Humanities) loopt op 1 maart ten einde. Medewerkenden aan het project zoals de DANS-medewerkster Milena Piccoli bereiden de fase voor, waarin de infrastructuur daadwerkelijk moet worden opgebouwd. Die zal ongeveer vijf jaar duren en gaat officieel begin 2012 van start wanneer nog een aantal procedurele stappen zijn doorlopen rond de deelname van verschillende landen. 'Anders dan in de voorbe-

reidende fase waarin instituten de deelnemers waren,' aldus Piccoli, 'zijn het in de uitvoering de landen en hun ministeries van wetenschappen die die rol vervullen.'

DARIAH, waarin vanaf het begin het Nederlandse DANS (Data Archiving and Networked Services) een coördinerend aandeel had, streeft naar een Europese infrastructuur die met ICT-gerelateerde methoden en gereedschappen onderzoekers helpt om nieuwe onderzoeksvragen te beantwoorden en oude tot nieuwe oplossingen te brengen. Niet alleen de

financiering van het vier miljoen euro vergende project, ook de organisatie wordt anders dan tot nu toe. De Nederlandse rol wordt kleiner; Frankrijk en Duitsland gaan samen de coördinatie doen. Binnen het project, komen vier zogenaamde *Virtual Competence Centres*. Piccoli: 'Eén voor de infrastructuur waarin DANS een belangrijke rol op het tweede plan speelt; één voor het *content management* waarin Nederland samen met Frankrijk het voortouw heeft. Een derde competentiecentrum met Denemarken en Ierland als hoofdpartners houdt zich bezig met onderwijs en onderzoek, en het vierde centrum is voor communicatie en verspreiding van de resultaten.' Dat wordt volgens de coördinator vooreerst ondergebracht bij de centrale organisatie in Frankrijk.

Nederland, Ierland, Griekenland en Kroatië hebben al getekend voor deelname aan de bouwfase. Van Oostenrijk-Denemarken wordt diestap op korte termijn verwacht. Naast het volledig partnerschap van deze landen, is er ook een 'samenwerkend partnerschap' mogelijk. Landen als Zwitserland, Spanje, Noorwegen en Litouwen die daarvoor kiezen, betalen niet maar houden door deelname op een laag pitje wel hun opties open om alsnog op de trein te springen. Van het Verenigd Koninkrijk, dat in de voorbereidende fase nog een vooraanstaande rol speelde, is de deelname aan DARIAH onzeker geworden door de bezuinigingspolitiek van de regering-Cameron. (MdG)

Vrouwenlexicon

Ruim duizend levensbeschrijvingen van opmerkelijke vrouwen uit de Nederlandse geschiedenis tot 1850 zijn voor iedereen te raadplegen op de website van het Digitaal Vrouwenlexicon van Nederland. Onder leiding van historica Els Kloek werd op het Huygens ING meer dan vijf jaar gewerkt aan de verzameling beschrijvingen van opmerkelijke vrouwen die soms 'beroemd of berucht waren, maar vaak in de vergetelheid zijn geraakt'. Het gaat om 'kunstenaresen, nonnen, moordenaressen, prinsessen, oplichtsters, weldoeners, onderneemsters, avonturiersters, professen en vele anderen,' aldus het instituut.

Nieuwe website NPSO

Het Nederlandstalig Platform voor Surveyonderzoek NPSO heeft zijn website opgeschud. De site biedt nu informatie over en van het platform, en over de inmiddels tot 11 leden gegroeide reeks instellingen die het steunen. Onder die leden bevinden zich de universiteiten van Tilburg, Utrecht en Nijmegen en een drietal stakeholders van e-data&research: het CBS, CentERdata en DANS. Er is ook een forum waar zich discussies kunnen ontvouwen over surveygerelateerde onderwerpen of interne onderwerpen. Daarvoor is lidmaatschap vereist, dat ter plaatse kan worden aangevraagd en verkregen.

NOD in NARCIS bij DANS

De Nederlandse Onderzoek Databank (NOD) wordt vanaf 1 mei onderdeel van de wetenschapsportal NARCIS. Beide websites worden beheerd door de voormalige afdeling Onderzoek Informatie van de KNAW, die sinds februari is ondergebracht bij DANS. Informatie en functies van de NOD werden de afgelopen tijd al steeds meer in NARCIS geïntegreerd. Vanaf mei is er geen reden meer om de NOD als aparte dienst naast NARCIS te onderhouden, aldus DANS in een persbericht. Ook de Nieuwsbrief Onderzoek Informatie houdt op te bestaan. Informatie over NARCIS wordt voortaan door DANS verspreid via zijn eigen kanalen.

Data in beeld



Het is heel goed mogelijk om speciale visualisaties te ontwikkelen bij de data die door DANS worden opgeslagen en beschikbaar gesteld. Dat blijkt uit een experiment dat het data-instituut heeft uitgevoerd. Er is steeds meer belangstelling voor visualisatie als middel om beter inzicht te krijgen in onderzoekdata, denkt DANS. In het experiment zijn de data van de beroepentelling 1899 gebruikt voor twee onderling verbonden presentatievormen, een dynamische treemap en een dynamische bevolkingspiramide. Er kwam een toepassing uit die relaties als die tussen geslacht, huwelijkse staat, regio, beroepsklasse, beroep en leeftijdsklasse kan verkennen.

Vriendschappen in kaart gebracht



Vijfhonderd miljoen vrienden telt de sociale netwerksite Facebook inmiddels volgens eigen zeggen. Over de invloed van de onderlinge contacten wordt de laatste maanden veel gespeculeerd. Nadat de sociale-netwerksite eerder belangrijk was in de volksofstand tegen de uitslag van de Iraanse presidentsverkiezingen, werd er in

de afgelopen maanden – vaak in één adem genoemd met Twitter – een doorslaggevende rol aan toegeschreven in de maatschappelijke omwentelingen in Noord-Afrika. De kaart hierboven, gebaseerd op een steekproef van zo'n tien miljoen vriendenparen, laat de wereldwijde relaties zien tussen steden, waarbij de kleur en zicht-

baarheid van de lijntjes varieert met de afstand en het aantal vrienden. De maker, Paul Butler, werkt bij het *Data Infrastructure Research Team* van Facebook. Hij postte de kaart op 14 december. Bij het ter perse gaan van deze *e-data&research* vonden 2263 vrienden dit leuk.

www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919

Esther Jansma, datapionier van de Nederlandse dendrochronologie:

‘Ik wil echt dat jonge onderzoekers straks verder kunnen met deze data’

MARTIJN DE GROOT

Begin januari ging de elektronische archiefwerkplaats voor dendrochronologen DCCD open voor gebruikers. Onderzoekers uit alle windrichtingen kunnen hier hun bevindingen op het gebied van de boomtjkdkunde opslaan en vergelijken. *e-data&research* sprak met initiatiefneemster Esther Jansma. De datadrijfveren van een dendrochronologe.

‘Ik geef je een simpel voorbeeldje.’ Esther Jansma zegt het aan het eind van het interview, met een hand aan de deurknop. Het gesprek is dan gekomen op het wezen van haar vak, de dendrochronologie. Is het wetenschap of is het ambacht? Ergens ertussenin, meent Jansma, maar dat is juist het mooie. En als ze dan een simpel voorbeeldje belooft, weet je intussen, komt er weer zo’n intrigerend verhaal dat het verleden tot leven brengt: ‘In de jaren negentig werden opgravingen gedaan naar een brug over de Maas bij Cuijk, gebouwd in de nadagen van het Romeinse Rijk. Over de bouw hebben ze toen wel een paar jaar gedaan, dat kon je zien aan de kapdata van het hout dat voor de fundamenten werd gebruikt. Die bomen haalden ze uit een bosje ergens in de buurt. Later werden er waterputten geslagen bij een paar boerderijtjes bij Gennep, niet ver van daar. Het hout dat daarvoor werd gebruikt kwam uit hetzelfde bosje als dat van die brug, maar de jaarringpatronen vertelden nog meer over hoe mensen in die tijd met houtteelt omgingen. Want op de plaats waar de bomen voor de brug waren gekapt, kregen de boompjes die overbleven opeens veel meer licht en die gingen dus veel sneller groeien. En juist dát hout werd geselecteerd voor die waterputten. Het lijkt er dus op dat de mensen in die tijd heel goed wisten welk hout ze wel en niet konden gebruiken. Nou, is zulk jaarringonderzoek wetenschappelijk baanbrekend? Niet echt. Maar maakt het wetenschappelijke inzichten levend? Ja, en dat is het mooie van dit vak.’

‘Is zulk onderzoek baanbrekend? Niet echt. Maakt het wetenschappelijke inzichten levend? Ja, en dat is het mooie van dit vak.’

Esther Jansma heeft het in dit gesprek al meer dan eens gezegd. Dendrochronologie, boomtjkdkunde, is geen ‘hoog-wetenschappelijke’ discipline. Maar het is wél een kunde die op belangrijke punten een kwalitatieve draai kan geven aan onderzoek op aangrenzende gebieden zoals archeologie en geschiedenis. ‘Landschapsreconstructie is zo’n activiteit die beter mogelijk wordt als je de dateringen van hout en bomen tot je beschikking hebt’. En dat is ook een van de redenen waarom Jansma zich als geen ander inspant om de

steeds groeiende hoeveelheden data die haar vak oplevert toegankelijk te maken: ‘Ik wil de deuren openzetten naar andere disciplines’.

‘Ik wil de deuren openzetten naar andere disciplines’

Maar de *tour d’horizon* van de dendrochronologie is nog niet klaar. We gaan nu nog een verdieping naar beneden in het van daglicht doortrokken nieuwe gebouw van de Rijksdienst voor het Cultureel Erfgoed (RCE). Onderweg naar het laboratorium van de stichting RING, het Nederlands Centrum voor Dendrochronologie dat onder het dak van de RCE is gehuisvest, blijkt de deur van Jansma’s kamer een van de weinige te zijn geweest die we in dit gebouw passeren. Terwijl ze haar bewondering uitspreekt voor collega’s die ‘het nieuwe werken’ in de praktijk brengen in gezamenlijke kantoorruimtes, naderen we het laboratorium dat je ook wel gewoon een werkplaats zou mogen noemen. Twee studenten zitten achter microscopen. Overal hout. Snippers, schuursel, brokken, tot enorme boomstamschijven aan toe. Hier wordt een gedeelte van het geld verdiend dat nodig is om de relatief kleine onderzoekdiscipline van de boomtjkdkunde in Nederland op de been te houden, aldus Jansma die het centrum begin jaren negentig oprichtte. Vanuit de archeologie en de bouwhistorie is veel vraag naar dateringen van hout. En het is een plek waar studenten de kans krijgen om zich in het vakgebied te bekwamen.

Zo’n soort plek was er niet toen de huidige bijzonder hoogleraar zélf aan het begin van haar wetenschappelijke loopbaan stond. ‘Ik wilde eigenlijk geschiedenis gaan studeren,’ had ze eerder verteld, ‘maar daar werd ik uitgeloot. Toen werd het wijsbegeerte met als bijvak archeologie. Ik kwam terecht in de opgraving van een Romeins fort in Velsen en daar zat ik als student dus al snel in stukjes van één bij één meter oude sedimenten door allerlei hout heen te troffelen. Dat blijft onder water lang bewaard. Je vond daar van alles: hout was eigenlijk het plastic van vroeger.’ Jansma’s belangstelling was gewekt en zou niet meer wijken. Hout was een gebied waarvoor in Nederland maar beperkte aandacht bestond, en toen haar bedoelingen serieus bleken mocht ze naar Zwitserland om zich verder



WIEBE KIESTRA

Esther Jansma: ‘Onderzoek en spannende artikelen leveren meer eer en roem op dan het bijeenbrengen van data’

te bekwamen in het determineren en dateren van hout. ‘Ze zagen me daar komen, een meisje van net in de twintig, en dachten ‘die zullen we eens laten zien wat dendrochronologie is’. Ik mocht op opgravingen rondlopen met hout van bomen die in bijvoorbeeld 3620 voor Christus waren gekapt en ik dacht: dat wil ik ook.’ Jansma ging terug naar Nederland, leerde een computertaal, vroeg en kreeg geld om een meettafel te laten maken in de medische instrumentmakerij van het Amsterdams Medisch Centrum, en studeerde in 1985 af op de dendrochronologische datering van het Romeinse fort aan de IJmond waar ze haar eerste schreden op het vakgebied had gezet.

‘Je kunt met computeranalyse allerlei regelmatigigheden uit de groeipulsen van bomen halen.’

Het woord computertaal is gevallen, en daarmee is de enorme verandering geïntroduceerd die er in het vakgebied bezig was, en waar Jansma niet alleen aansluiting bij vond maar waar ze uiteindelijk ook zelf een belangrijk aandeel in zou hebben. ‘Het gaat in ons vak om kalenders, die – kort gezegd – over een reeks van jaren

de breedte van de jaarringen aangeven. De variatie van die ringbreedtes komt tot stand door veranderingen in de omstandigheden: het weer, overstromingen, branden, vulkaanuitbarstingen. Die omstandigheden hebben bomen in grotere gebieden met elkaar gemeen, dus als je van één boomsoort in een gebied of streek de kalender hebt kan je daarmee andere stukken soortgelijk hout uit dezelfde periode in dat gebied thuisbrengen.’ Het zoeken naar dergelijke overeenkomsten is een klus waarbij de computer een enorm verschil kan maken, zag Jansma in en ze mobiliseerde de hulp van een collega-archeoloog, Hans Karmers, die met een sterke Amsterdams computerclub om zich heen kon helpen met programmeren. ‘Die mensen vonden dendro heel leuk; dat was in de jaren tachtig echt nieuw’. De mogelijkheden groeiden nu snel. ‘Als je in staat bent om intelligente tijdreeksanalyses te maken, komen daar allerlei signalen uit die je kunt gebruiken. De groei van een boom in een bepaald jaar heeft bijvoorbeeld ook invloed op de groei in het daaropvolgende jaar. De jaarringen veranderen ook naar gelang de leeftijd van de boom. Dus er zijn allerlei regelmatigheden die je met computeranalyse uit de groeipulsen van bomen kan halen.’

Er kwamen in de loop der jaren, niet in de laatste plaats door toedoen van Jansma, steeds meer kalenders beschikbaar, ‘zo ontzettend veel dat het moeilijk te overzien is.’ Hetzelfde gebeurde in het buitenland en zo dienden zich wéér grote nieuwe mogelijkheden aan die schreeuwden om te worden benut. Dat laatste is precies het doel van Jansma’s laatste grote verworvenheid, de *Digital Laboratory for Cultural Dendrochronology (DCCD)*: een digitaal archief van jaarringgegevens dat het mogelijk maakt dat onderzoekdata meer dan eens, op grotere schaal en door meer onderzoekers worden gebruikt. De DCCD werkt op basis van de Tree-Ring Data Standard TRiDaS waar de naam van Esther Jansma ook al overheen staat geschreven.

‘Het wordt straks veel makkelijker voor historici, voor klimatologen, voor conservatoren van musea.’

Onder de weinige dendrochronologen die ons land rijk is, en trouwens ook internationaal, valt Jansma op omdat ze zich onvermoeibaar sterk maakt voor digitale opslag en uitwisseling. Waarom? ‘Ja,’ geeft ze wel toe, ‘onderzoek en spannende artikelen leveren meer eer en roem op dan het bijeenbrengen van data. Dit is meer dienstverlenend, voorwaardenscheppend. Zorgen dat het straks mogelijk is, dat vind ik belangrijk. We beschikken intussen over geweldige hoeveelheden data maar er gebeurt veel te weinig mee omdat er geen overzicht is. Dat is omdat we ons als vakgebied niet goed hebben georganiseerd. Je moet eigenlijk kunnen zeggen, en dat kan straks ook, geef me de meetreeksen van al het hout uit alle boerderijen in dat-en-dat gebied in die-en-die periode. Dat soort vragen. Als je die kunt stellen en beantwoorden, dan kan je studenten heel mooie datasets aanleveren en prachtig promotieonderzoek laten doen. En niet alleen in dit vak maar juist in aangrenzende gebieden. Het wordt straks veel makkelijker voor historici, voor klimatologen, voor conservatoren van musea. En voor jonge onderzoekers. Ik wil echt dat die er later mee verder kunnen!’

Esther Jansma studeerde in 1985 in Leiden af als archeologe. Ze promoveerde in 1995 cum laude op een proefschrift op basis van haar bijdrage aan de Nederlandse dendro, waar het oprichten van het Nederlands Centrum voor Dendrochronologie (stichting RING) onderdeel van was. Ze is nu directeur van dat centrum, onderzoeker bij de Rijksdienst voor het Cultureel Erfgoed en bijzonder hoogleraar aan de Universiteit Utrecht. Jansma is ook dichter. Ze publiceert sinds 1988 regelmatig dichtbundels en won ook op dit gebied verschillende prijzen.

Focus

Meertens Instituut

Het Amsterdamse Meertens Instituut, een onderdeel van de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW), onderzoekt en documenteert alledaagse verschijnselen in de Nederlandse taal en cultuur. *e-data* sprak met Sjef Barbiers, onderzoeksleider van de afdeling variatielinguïstiek, over een instituut in beweging.

Een gesprek over het Meertens Instituut kan niet anders dan beginnen met een vraag naar Voskuil, die met zijn zevendelige romancyclus *Het bureau* het Meertens Instituut bij velen bekend maakte. Barbiers herkent zich niet in de instelling zoals Voskuil die beschreef. 'Integendeel. Voskuil beschreef een in zichzelf gekeerde wereld, die niet deel nam aan het grotere wetenschappelijke debat. Onderzoekers van het huidige Meertens Instituut staan middenin de internationale academische wereld.'

Inhoudelijk is er natuurlijk wel sprake van continuïteit: het Instituut houdt zich nog steeds bezig met variatie in cultuur en taal. Maar waar vroeger papieren vragenlijsten naar alle hoeken van het land werden gestuurd, gingen voor de recente Syntactische Atlas van de Nederlandse Dialecten veldwerkers naar 267 plaatsen in het Nederlands taalgebied. Barbiers: 'Methodologisch was de grote vernieuwing dat we zoveel mogelijk met paren van lokale interviewers werkten. Daardoor treedt minder vervuiling op ten gevolge van het taalgebruik van de onderzoeker. Bovendien zijn de geluidsopnamen on-line beschikbaar, opgelijnd met transcripties en verrijkt met woordsoortinformatie.' De gegevens zijn zo veel makkelijker doorzoekbaar dan voorheen. De resultaten van zoekvragen kunnen op kaarten worden gepresenteerd die inzicht geven in de verspreiding van dialectverschijnselen. Barbiers: 'Al onze bestanden komen op deze manier ter beschikking van onderzoekers en het bredere publiek.' Onder de databases van het Instituut zijn een paar zeer populaire bestanden: de database van Nederlandse familienamen bijvoorbeeld, maar ook de Volksverhalendatabank, de Liederenbank en Soundbites, een collectie van dialectopnamen. Onderzoek en collectievorming gaan voor het Meertens hand in hand. Barbiers verklaart: 'Het één kan niet zonder het ander. Dankzij grote digitale collecties ontstaan er nieuwe onderzoeksmogelijkheden. We verzamelen niet om het verzamelen.' Voor zover er nog vragenlijsten worden verzonden, loopt dat tegenwoordig via het web. Via het Meertens Panel worden regelmatig vragenlijsten voorgelegd aan meer dan 4500 respondenten. Ook is sinds vorig jaar het Meldpunt Taal online waar mensen interessante taalverschijnselen kunnen melden.



DIEDERIK VAN DER LAAN

Onderzoekleider Sjef Barbiers: 'Onze onderzoekers staan middenin de internationale academische wereld'

Het Meertens Instituut wil zich ontwikkelen tot centrum voor technologische innovatie binnen de geesteswetenschappen. Barbiers: 'Het is beleid van de KNAW, maar we waren zelf al hard op weg.' Het Meertens wordt één van de expertisecentra binnen de infrastructuur voor de geesteswetenschappen die het project CLARIN tot stand wil brengen. Er staat een groot aantal computationele taalkundige projecten op stapel, zoals bijvoorbeeld INPOLDER, voor het annoteren van historisch taalmateriaal – nu vaak onder andere vanwege afwijkende spelling nog een probleem in tekstgeoriënteerd onderzoek. Barbiers: 'De programmatuur die het resultaat is van dit soort projecten willen we via web services aan andere onderzoekers aanbieden. Het zijn trouwens niet alleen de taalkundigen die met computationele technieken werken: onderzoeker Theo Meder gaat bijvoorbeeld langs computationele weg de volksverhalen onderzoeken op motieven en thema's.'

Wenkend perspectief, voor zowel taalkundigen als etnologen, is de vraag naar het computationeel modelleren van menselijke cognitie. Barbiers: 'De musicologen willen weten hoe melodieën worden herkend. Taalkundigen zijn altijd al geïnteresseerd geweest in de representatie van taal in het brein. Digitale bestanden en onderzoeksmethoden kunnen een antwoord op zulke vragen dichterbij brengen.'

MESS migrantenpanel nu open voor wetenschappers

Een evaluatiecommissie van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO) oordeelde vorig jaar niet alleen gunstig over het verloop van het MESS-project (Measurement and Experimentation in the Social Sciences) van CentERdata. Op advies van dezelfde commissie heeft NWO nu 750.000 euro extra toegekend aan MESS.

Het eerste oordeel van de commissie werd gevraagd in het kader van een tussentijdse evaluatie van de vijf grootschalige onderzoeksfaciliteiten die vanuit BIG-programma van NWO zijn gefinancierd. MESS is een van die vijf, met als kern het LISS-panel (Longitudinal Internet Studies for the Social sciences).

Projectleider Annette Scherpenzeel van MESS is dubbel blij: 'De evaluatie van de commissie is natuurlijk een opsteker, maar we verheugen ons er vooral over dat we met de extra middelen weer nieuwe activiteiten kunnen opzetten die inspelen op recente wetenschappelijke en maatschappelijke ontwikkelingen.' Een deel van de extra subsidie zet CentERdata in om zogenoemde biomarkers in het LISS panel te ge-

bruiken: innovatieve apparaten en zelftests waarmee gezondheidsaspecten objectief gemeten kunnen worden. Op dit moment hebben duizend huishoudens in het LISS-panel bijvoorbeeld een geavanceerde weegschaal in gebruik die gewicht en vetpercentage meet en via internet aan CentERdata doorstuurt. Andere mogelijkheden zijn bloeddrukmeters en bewegingssensoren.

Met het tweede deel van de extra subsidie wordt het migrantenpanel, het tweede online panel binnen MESS, opengesteld voor wetenschappers. Dat panel omvat een extra steekproef van de verschillende migrantengroepen in Nederland die, net als de steekproef voor het LISS panel, getrokken is uit de Gemeentelijke Basis Administratie. Door de extra subsidie kunnen wetenschappers nu onderzoekvoorstellen indienen om gratis eigen data te verzamelen in het migrantenpanel. Ze zullen bovendien kosteloos over alle data kunnen beschikken die in het migrantenpanel verzameld zijn, net als bij het LISS panel. (EB)

 www.lissdata.nl

Op zoek naar gelijkennis in muziek

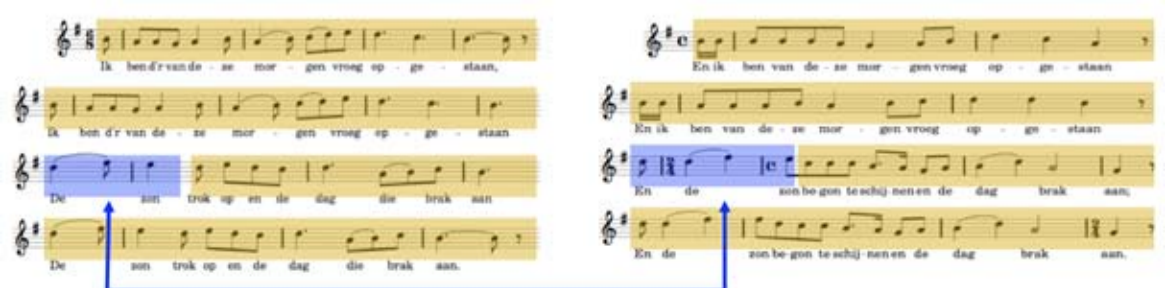
Mensen kunnen overeenkomsten tussen muziekfragmenten herkennen. Maar kunnen computers dat ook? Als het antwoord op die vraag bevestigend is levert dat nieuwe mogelijkheden op voor het elektronisch zoeken en vinden van muziek. Aan de universiteit Utrecht wordt naar dat antwoord gezocht, met een Vidi-subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek NWO.

Het herkennen van gelijkennis is cruciaal voor ons begrip van de wereld. In de cognitiewetenschap zijn formele modellen ontwikkeld die gelijkennis theoretisch beschrijven en empirisch meetbaar maken. Maar er bestaat nog geen algemeen model van gelijkennis in muziek. Dit is een cruciaal probleem in het jonge, snel groeiende vakgebied van *music information retrieval* (MIR). MIR onderzoekt methoden die het mogelijk maken om op basis van gelijkennis muziek te vinden in grote digitale muziekcollecties. Omdat er geen algemeen model voor gelijkennis in muziek bestaat, is het een fun-

damenteel probleem in MIR hoe de informatie die uit een digitaal muziekdocument wordt gehaald, gerelateerd kan worden aan de ervaring van gelijkennis door de gebruikers. In het project van de Universiteit Utrecht wordt een computationeel model ontwikkeld dat beschrijft hoe luisteraars gelijkennis in muziek ervaren. Het probleem wordt benaderd door een belangrijk principe in muziek, het variatieprincipe, als basis te nemen voor het modelleren van gelijkennis. De afbeelding toont een voorbeeld voor twee volksliederen uit de Nederlandse Liederenbank van het Meertens Instituut. De blauw gekleurde motiefjes tonen prominente patronen die door variatie met elkaar verbonden zijn (de toonhoogtes zijn dezelfde, maar de ritmische structuur verschilt). Omdat bij het luisteren naar één van de liedjes aan het gerelateerde motiefje van het andere liedje gedacht wordt, worden de twee liedjes als gelijkend beoordeeld. Op basis van dit fenomeen wordt de wisselwerking tussen lokale en globale kenmerken van

muziek (zoals melodie, ritme, harmonie) in een model ondergebracht om gelijkennis te bepalen. Het herkennen van de blauwe patronen tijdens het luisterproces is namelijk mede afhankelijk van de globale context (in de afbeelding in geel), waarin de patronen zijn ingebed.

Het project zal een model voortbrengen van gelijkennis voor drie muziekstijlen: klassiek, populair en volksmuziek. Hiervoor wordt gestreefd naar samenwerking tussen *music information retrieval*, muziekwetenschap en cognitiewetenschap. Niet alleen wordt een theoretisch kader ontwikkeld voor gelijkennis in muziek, ook worden computermethoden toegepast die het mogelijk maken gelijkende muziekdocumenten te vinden op basis van variatie. Daarmee zet het project een beslissende stap op weg naar een algemeen model van gelijkennis in muziek. Dat vormt dan weer de basis voor methoden om succesvol naar muziek te zoeken in online digitale collecties. (Anja Volk)



KB digitaliseert door ondanks minder middelen

Eind 2009 zette de Koninklijke Bibliotheek (KB) het ambitieuze beleidsvoornemen op papier om alle publicaties van en over Nederland vanaf 1470 digitaal beschikbaar te stellen. Het initiatief ontmoette waardering maar ook twijfel over de uitvoerbaarheid, zeker in een tijd van bezuinigingen. Toch blijkt de KB nog geen anderhalf jaar later goed op weg te zijn. Om met minder middelen toch veel materiaal online te kunnen aanbieden werd samenwerking gezocht met private partijen. Zo worden 160.000 boeken uit de achttiende en negentiende eeuw gedigitaliseerd door Google en 50.000 bijzondere oudere drukken door uitgever ProQuest. In het conserveringsprogramma Metamor-

foze werkt de KB samen met het Nationaal Archief verder aan de digitalisering van boeken, kranten en tijdschriften van na 1840, waaronder twee miljoen krantenpagina's uit de periode 1850-1950. Astrid Verheusen, programmamanager Digitale Bibliotheek van de KB: 'De eerste fase van onze beleidsdoelstelling, tien procent online in 2013, gaan we daarmee zeker halen.'

Achter de schermen werkt de KB ook hard aan een betere beschikbaarheid van het materiaal. Verheusen: 'In de loop der jaren hebben veel digitaliseringsprojecten eigen websites ontwikkeld. Historisch gezien is dat goed te verklaren, maar we merken nu dat gebruikers moeite hebben om snel te vinden wat ze



KONINKLIJKE BIBLIOTHEEK

zoeken. Daarom streven we naar een platform waarop je alles kunt vinden wat de KB aanbiedt. En misschien

wel meer, want we willen ook samenwerkingsmogelijkheden onderzoeken met andere bibliotheken.' (IA)

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

SINDS KORT BESCHIKBAAR

Het overzicht toont databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS, CentERdata, Huygens ING en DANS. Een volledig overzicht van de CBS-bestanden is te vinden op www.cbs.nl/microdata. De LISS panel bestanden van CentERdata zijn kosteloos beschikbaar via het

LISS Data Archief www.lissdata.nl/dataarchive. De bestanden van Huygens ING zijn beschikbaar via www.inghist.nl. De DANS databestanden komen van diverse andere onderzoeksinstituten. Deze kunnen kosteloos worden gedownload vanuit DANS EASY: <https://easy.dans.knaw.nl>

Centraal Bureau voor de Statistiek	Periode
Productiestatistieken:	
Zorg en Welzijn	2003-2007
Detailhandel	1993-2008
Transport	2000-2008
Rijksfinanciën	2007
Investeringen in materiële vaste activa	2008
Enquete ICT-gebruik bedrijven	1995-2008
Woonruimteregister verrijkt met Voorraad woonruimte per adres	2008
CentERdata LISS Data Archive	
Ambiguity Aversion and Household Portfolio Choice (Dimmock, S.; Kowenbergh R.; Wakker, P.)	2010
Restraint Eating vs. Food Worry and Power of Food as Competing Explanations of Overeating and Overweight: Part 1 and 2 (Ridder, D. de)	2010
Explaining Public Support for Vigilantism: Part 1 and 2 (Haas, N.)	2009
LISS Core Study – Economic Situation: Housing, Wave 3 (CentERdata)	2010
LISS Core Study – Economic Situation: Income, Wave 3 (CentERdata)	2010
LISS Core Study – Economic Situation: Assets, Wave 2 (CentERdata)	2010
LISS Core Study – Work and schooling, Wave 3 (CentERdata)	2010
Perceived Diversity Effects on Organizational Outcomes (Garib, Y.R.)	2010
Waist Circumference (CentERdata)	2010
Commercial Opportunities (Block, J.; Gruber, M.; Petty, J.S.)	2010
Huygens ING	
Nieuwe brieven van Willem van Oranje	1572-1573
Digitale gids Grondwetcommissies	1883-1983
Registers van de graven van Holland in de Henegouwse periode	1299-1345
Namenindex op databases en gedigitaliseerde publicaties Huygens ING	2010
Kroniek van Kattendijke – Rijks Geschiedkundige Publicatiën	1490
Brieven en andere bescheiden Daniel van der Meulen – Rijks Geschiedkundige Publicatiën	1584-1600
Bronnen betreffende de Midden-Molukken – Rijks Geschiedkundige Publicatiën	1900-1942
Bronnen tot de geschiedenis der Leidsche universiteit – Rijks Geschiedkundige Publicatiën	1574-1811
Via DANS EASY	
Historische Wetenschappen	
Interviews Veteranen Instituut – IPNV	2009
Archeologie	
Een archeologische begeleiding (protocol opgraven) aan de Eexterweg te Gieten, gemeente Aa en Hunze (Dr) (ARC)	2009
Westland Poeldijk De Kreken fase II opgraving (ADC ArchoProjecten)	2010
Life on the watershed. Reconstructing subsistence in a steppe region using archaeological survey: a diachronic perspective on habitation in the Jordan Valley (Eva Kaptijn)	2009
Noordwijk, 't Pleintje, Gemeente Noordwijk (Becker en Van de Graaf b.v.)	2009
Visresten van Albrandswaard-Portland. Archeozoologisch onderzoek van de laat-neolithische vindplaats 20-134 (4145 ± 45 BP) (BOOR)	2008
Sociale Wetenschappen	
Arbeidsvraagpanel (voorheen OSA Arbeidsvraagpanel) (A.M. de Voogd-Hamelink, Sociaal en Cultureel Planbureau, Den Haag)	1991-2008
International Comparative Study of Ethno-Cultural Youth (ICSEY team; Paul Vedder, Rommert Casimir Institute, Leiden University)	2006
Mobiliteitsonderzoek Nederland – MON (Ministerie van Infrastructuur en Milieu – Rijkswaterstaat)	2009
Tijdsknelpunten in de samenleving – tijd op orde (Sociaal Cultureel Planbureau, Den Haag)	2009

DANS wil na onderzoek data gaan reviewen

DANS (Data Archiving and Networked Services) gaat onderzoeken of het in zijn archiefsysteem EASY een standaardvraag aan afnemers om feedback kan opnemen. De antwoorden zouden vervolgens aan andere EASY-gebruikers meer informatie verschaffen over de betreffende dataset.

Eind vorig jaar voerde het instituut een pilotonderzoek uit om het reviewen van onderzoeksdata te stimuleren. 140 mensen beoordeelden de kwaliteit van de dataset(s) die zij uit het EASY-archief hadden gedownload. 92% van de respondenten zou de dataset aan anderen aanbevelen, zo bleek. De gemiddelde scores voor aspecten als de kwaliteit van de data en de documentatie en structurering binnen de dataset lagen rond de 4 op een schaal van 1 tot 5. 'Kwaliteit van de data' stond bovenaan met 4,12. De dataset helpt in 70% van de gevallen om onderzoeksvragen te beantwoorden. Overigens downloadt niet iedereen vanwege een onderzoeksvraag: ook 'pure interesse' wordt als reden genoemd. DANS vindt deze uitkomsten waardevol voor zichzelf, maar ook voor de onderzoekers die de datasets hebben gedeponeerd.

Respondenten kenden in het kader van de pilot ook tags toe aan de dataset die ze beoordeelden. Deze trefwoorden bleken overwegend inhoudelijk, bijvoorbeeld 'Romeinse tijd' of 'woningmarkt', maar ook meningen zoals 'weinig toegankelijk' en 'handig' komen voor. Voorbeelden van gereviewde datasets zijn te vinden op de DANS-website.

In december verscheen ook een verkenning van kwaliteitsbeoordeling van onderzoeksdata in het kader van SURFshare. Volgens dat rapport bestaan er voor datasets die door onderzoeksfaciliteiten en overheidsorganisaties worden geprodu-

ceerd, veel mechanismen om kwaliteit te monitoren en te verbeteren. *Supplementary data* – behorend bij publicaties – worden echter in praktijk slechts beperkt betrokken in de peer review. Ook is onderzocht op welke manieren de kwaliteit van datasets te stimuleren valt. Hiertoe kreeg een landelijke steekproef van hoogleraren en hoofddocenten negen opties voorgelegd. Drie opties blijken populair in alle disciplines: het opzetten van tijdschriften voor datapublicaties, het citeren van datasets en hergebruikers commentaar laten geven op de kwaliteit. Ook deze uitkomst is volgens DANS een stimulans om een review-mogelijkheid aan EASY toe te voegen.

Marjan Grootveld

@ <http://datareviews.dans.knaw.nl/index.php>
www.surffoundation.nl/nl/publicaties/Pages/Verkenningonderzoek.aspx

Negen nieuwe CLARINprojecten van start

Begin 2011 gaan negen nieuwe projecten van start die zijn gehonoreerd binnen de tweede call van CLARIN-NL, het infrastructuurproject voor de geesteswetenschappen. De projectuitvoerders zullen bestaande data en analyseinstrumenten geschikt maken om in de CLARIN-infrastructuur te worden opgenomen. Het gaat hier onder andere om de Nederlandse Liederbank, een historische verzameling van artikelen uit vijf grote Nederlandse kranten en een PoS-tagger. In de huidige call ligt de nadruk op data en instrumenten uit de letteren, geschiedenis, taalverwerving en historische taalkunde. 'Wij hopen hiermee over de gehele breedte van de geesteswetenschappen onderzoekers enthousiast te maken over de voordelen die de infrastructuur hun bij de uitvoering van hun onderzoek kan bieden', zegt CLARIN-NL-directeur Jan Odijk.

CLARIN is een Europees initiatief om e-science mogelijk te maken binnen de geesteswetenschappen en de sociale wetenschappen. Ruim dertig instituten uit 25 verschillende landen werken als partners samen aan een infrastructuur waarbinnen onderzoekers vrij toegang krijgen tot de data die zij voor hun onderzoek nodig hebben. Bovendien kunnen onderzoekers gebruik maken van speciale tools om relevante data te vinden en de gevonden data te analyseren en te verrijken, zoals automatische zinsontleders en textmining-software. (ER)



@ www.clarin.nl

COLOFON

e-data@research is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van CentERdata, CLARIN-NL, DANS, het Huygens ING, het Centraal Bureau voor de Statistiek, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen. Oplage: 9100. *e-data@research* is online te raadplegen op www.edata.nl.

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

Redactieadres: Postbus 93067, 2509 AB Den Haag; t (070)3446484 f (070)3494451 e edata@dans.knaw.nl

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

Aan dit nummer werkten mee: Onno Crasborn, Marjan Grootveld, René van Horik, Jeroen Sondervan, Anja Volk.

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Vormgeving en opmaak: Ellen Bouma, Alkmaar

Productie: Amsterdam University Press

Druk: Ten Brink, Meppel

ISSN: 1872-0374

Culturomics: een nieuw vakgebied?

Begin december lanceerde Google de *Ngram Viewer*, een website waarop een selectie van ruim vijf miljoen boeken uit de Google Books-collectie doorzocht kan worden op reeksen van maximaal vijf woorden (n-grammen). De resultaten worden gepresenteerd in een grafiek die het relatieve voorkomen van de zoektermen of reeksen per jaar weergeeft. Tegelijk met de lancering publiceerde *Science* een artikel van de ontwikkelaars waarin zij uitgebreid ingaan op de mogelijkheden van de Ngram Viewer. Daarnaast presenteerden zij Culturomics, een nieuw vakgebied waarbinnen taal en cultuur bestudeerd worden aan de hand van grootschalige data-analyse.

De *Science*-publicatie zorgde voor opschudding onder met name geesteswetenschappers. Is deze *culturomics* een arrogante poging uit de bètahoek om de alfa's en de gamma's uit de brand te helpen, een brand die er helemaal niet is? Taal en cultuur worden toch al tijden onderzocht aan de hand van de analyse van data? Daarnaast is de Ngram Viewer nog helemaal niet zo bruikbaar voor onderzoek als de ontwikkelaars in hun artikel doen voorkomen.

Reden voor de Linguistic Society of America (LSA) om tijdens haar jaarlijkse bijeenkomst een extra sessie in te lassen met Jean-Baptiste Michel en Erez Lieberman Aiden, de hoofdauteurs van de *Science*-publi-

catie. Antal van den Bosch, hoogleraar Geheugen, taal en betekenis aan de Universiteit van Tilburg, was erbij.

'Het was erg duidelijk dat de ontwikkelaars zelf geen taalkundigen zijn; ze weten zelf amper wat ze doen', aldus Van de Bosch. 'Maar ze staan wel erg open voor suggesties om hun toepassing te verbeteren.' In de eerste week na de publicatie heeft deze houding al tot enkele essentiële verbeteringen geleid. Zo is er een dubbele indexering uitgevoerd, waardoor woorden als 'can't' niet meer alleen als een bigram ('can' en 't'), maar ook als unigram ('can't') vindbaar zijn. Uit de kritieken uit de zaal tijdens de LSA-bijeenkomst sprak vooral de wens van een breder aanbod, zowel van cijfers als van metadata. Ook hier werken de ontwikkelaars graag aan mee – voor zover Google bereid is de benodigde informatie vrij te geven.

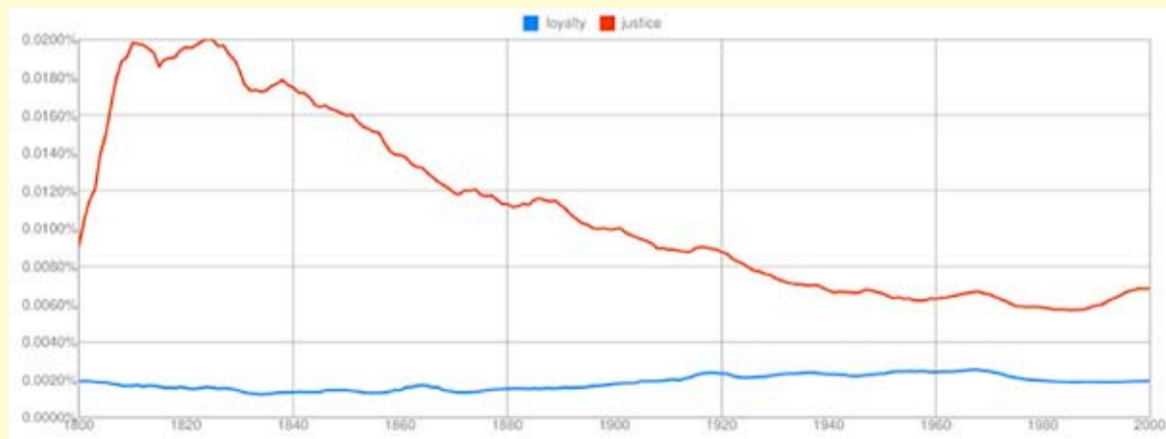
Momenteel geeft de Viewer als enige uitkomst het relatieve voorkomen van een woord per jaar ten opzichte van het totaal aantal woorden in dat jaar. Wetenschappers hopen binnenkort ook te beschikken over de absolute aantallen, zowel per token als per boek. Daarnaast zou een uitbreiding van de metadata erg waardevol zijn. In de huidige vorm is alleen een grove categorisering aangebracht tussen bijvoorbeeld fictie en nonfictie. 'Iemand die onderzoek doet naar poëzie uit Zuid-

Frankrijk uit de tweede helft van de negentiende eeuw zal nu nog handmatig moeten zoeken', aldus Van den Bosch. 'Dat moet veel makkelijker kunnen.' Ten slotte zullen de publicaties van vóór 1900, die niet meer auteursrechtelijk beschermd zijn, binnen afzienbare tijd als *full text* gepubliceerd worden.

Dan blijven er nog twee grote struikelblokken over waar de ontwikkelaars weinig aan zullen kunnen veranderen. Ten eerste de kwaliteit van het toegepaste OCR-proces: die is nog altijd verre van perfect waardoor de resultaten niet altijd even betrouwbaar zijn. En ten tweede de eenzijdige bron van boeken. 'Boeken representeren de tijdsgeest maar ten dele,' legt Van den Bosch uit. 'Ze zijn representatief voor de mensen doorwie en voorwie boeken werden geschreven. In de 19^e eeuw lag dat heel anders dan nu.' Maar de Ngram Viewer is nu eenmaal gebaseerd op het GoogleBooks-project.

En dan die term, culturomics, zal die beklijven? Van den Bosch verwacht van niet, maar wellicht kunnen we het over een paar jaar objectief bekijken met een verbeterde versie van de Ngram Viewer. Eentje waarin wellicht ook websites en tijdschriften zijn opgenomen, want zoals Van den Bosch het zegt, 'wat staat er tegenwoordig nog in boeken?'

Erica Renckens



Uitkomst van de Ngram Viewer voor het voorkomen van de woorden 'loyalty' en 'justice' in de periode van 1800 tot 2000.

Gelezen

Heiko Tjalsma, Jeroen Rombouts and An-nemiek van der Kuil (eds): *Selection of Research Data: Guidelines for appraising and selecting research data, a report by DANS and 3TU.Datacentrum*. Utrecht, Stichting SURF, 2010.

Er zijn algemene richtlijnen opgesteld voor het beoordelen en selecteren van onderzoeksdata om te bewaren. Die zijn te gebruiken door iedereen die iets met de beoordeling en selectie te maken heeft. Het rapport bevat de laatste stand van zaken op het gebied van het selecteren van onderzoeksdata, gebaseerd op literatuuronderzoek, een aantal interviews met belangrijke personen en de ervaringen van DANS en het 3TU Data centrum. Hoewel onderzoeksgegevens veelal bewaard moeten worden voor gebruik of hergebruik, of ter validatie van onderzoeksresultaten, geldt dit niet voor alle gegevens. Om te bepalen welke data waardevol zijn als (bron)materiaal voor onderzoek zijn er een aantal praktische richtlijnen opgesteld en samengebracht in een checklist. www.surfoundation.nl/nl/publicaties/



Martijn de Groot en Marion Wittenberg (eds): *Driven by data: Exploring the research horizon*. Amsterdam, Pallas Publications, Amsterdam University Press, 2010; ISBN: 978 90 8555 038 9

Op 2 December 2009, the symposium 'Driven by data' took place in The Hague to celebrate the anniversaries of a number of archiving institutions that are now all part of DANS (Data Archiving and Networked

Services). It was a lively meeting, in which attention was paid to various surprising and creative possibilities associated with the reuse of electronic data. DANS decided to publish this jubilee volume to give a larger public a taste of this versatility and creativity. The contributions cover topics as diverse as dendrochronology, politics, shipping traffic history, bird migration, philosophy of science and statistics. This publication can be downloaded free of charge from the DANS website: www.dans.knaw.nl

Stromen van Kennis. Utrecht, SURFshare, 2011

In deze publicatie geven lectoren, docenten en studenten hun visie op het toegankelijk maken en delen van onderzoekresultaten van hogescholen. Dit is belangrijk omdat de HBO-raad in 2009 de 'Berlin Declaration on Open Access' heeft getekend. Daarmee geven de hogescholen aan zich te zullen inzetten voor het vrij beschikbaar stellen van hun onderzoeksresultaten. De behoefte aan vrije toegang tot kennis is

Column

Onno Crasborn

Onverwachte gasten

In 2008 publiceerden we het *Corpus Nederlandse Gebarentaal (NGT)*. Zeventig uur video-opnames van dialogen tussen doven. Het leeuwendeel is openbaar, en dus door elke onderzoeker te gebruiken. 'Corpus' is een groot woord voor onze verzameling, leerden we steeds meer inzien: zonder uitgebreide annotaties is een video niet alleen voor gebarentaal-leken onbegrijpelijk, maar ook voor taalkundigen betekenisloos.



In de twee jaar die de kleine investerings-subsidie van NWO ons eraan liet werken, was er nauwelijks tijd over om gebaren en zinnen te vertalen. Laat staan dat we een fonetische transcriptie hadden gemaakt van alle details. Vingerbewegingen, gezichtsuitdrukkingen, hoofdknikken, leunen naar links of naar achteren... je bent zo een hele dag bezig met een enkel minuutje video.

Gelukkig wisten we van tevoren dat de verzameling niet alleen interessant zou zijn voor kwantitatief onderzoek. Alleen al er naar kunnen kijken maakt de filmpjes bijzonder. Er is namelijk niet zoveel gebarentaal te zien geweest in het verleden. We hebben geen dovenomroep, er zijn geen dove speelfilms (met ondertiteling voor horenden), en er is geen *Sesamstraat* met een gebarende Pino. Afen toe is er een tolk op televisie. Wel zijn er twee planken met videobanden gemaakt sinds de jaren tachtig, en later nog een hele plank dvd's, maar je kunt er nog geen hoekje van een bibliotheek mee vullen. Het *Corpus NGT* bevat discussies die voor veel doven interessant zijn. Hebben dove ouders het recht om via IVF een dove embryo te selecteren? Is het positief of negatief dat er regionale variatie is in gebarentaal? Het vormt een aanzienlijke verrijking van het dove medialandschap, dachten wij zelf met gepaste trots.

En inderdaad, op de publiekssite www.corpusngt.nl kijken zowel doven als horende tolken en gebarentaaldocenten regelmatig rond. Maar het mailtje van collega's uit Hamburg hadden we niet verwacht. Of de publiekssite niet ook in het Duits vertaald kon worden, om de filmpjes te kunnen bekijken. Hè? Maar de Duitse Gebarentaal en de Nederlandse zijn toch twee verschillende talen? De dove oosterburen gaan de Nederlandse Gebarentaal toch niet begrijpen? Nee, dat was waar. Maar ze wilden het gewoon zien. Kunnen kijken. Er is geen vreemdetalenonderwijs voor gebarentaal, niet in Nederland en ook niet in Duitsland. Ontwikkeling, zo denkt men in het dovenonderwijs nog altijd, is gekoppeld aan gesproken taal. En daarom hebben de gebarentaalcorpora die nu overal gemaakt worden een bijzonder en onverwachte functie erbij gekregen: ze maken een blik op een andere taal mogelijk, voor doven uit allerlei landen.

Onno Crasborn is gebarentaaldeskundige en senior onderzoeker Taalwetenschap aan de Radboud Universiteit.

groot. Zowel omdat praktijkgericht onderzoek een directe link heeft met de samenleving, alsook omdat, met de komst van lectoraten en kenniskringen, hogescholen steeds meer onderzoek doen. Het delen van de stroom kennis die momenteel vanuit het hbo los komt en waarmee hogescholen zich sterker willen profileren, vraagt om een gedegen kennisinfrastructuur. SURFfoundation ondersteunt dit proces, onder meer door het helpen realiseren van repositories, een onmisbaar onderdeel van die infrastructuur. www.surfoundation.nl/nl/publicaties/



e-data&research
wil graag iets van u weten

Vul de bijgesloten antwoordkaart in en stuur hem op of ga naar www.edata.nl