

Pagina 2 • Avatars ondervraagd •

Een enquête onder avatars was een van de opvallende bijdragen op de e-Social Science Conference die eind juni in Keulen werd gehouden. De respons bedroeg 13%.

Pagina 3 • Soundbites online •

Eind september gaat de website met 'spreekende kaart' van het Soundbites-project online. Het Meertens Instituut maakt met de kaart zo'n duizend uur geluidsopnames van gesprekken in dialect beschikbaar.



WIEBE KIESTRA

Pagina 5 • Karina van Dalen • Nu het kan, wil onderzoekster Karina van Dalen-Oskam van het Huygens Instituut een empirisch fundament onder de letterkunde leggen.

Pagina 6 • Pelgrimsroutes • Een internettoepassing met geo-technologie bracht het onderzoek naar pelgrimstochten in de late middeleeuwen in een nieuwe fase.

EN VERDER

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

Duurzame opslag data: weg is nog lang

In de wetenschap is digitale informatie verder doorgedrongen dan bij de overheid en in de sector cultureel erfgoed. Toch is juist daar nog het meeste onduidelijk over het duurzaam opslaan van gegevens en de selectie daarvan. Dat blijkt uit het interimrapport *Toekomst voor ons digitaal geheugen* van de Nationale Coalitie voor Digitale Duurzaamheid (NCDD).

Het rapport verscheen begin juli als voorlopige uitkomst van de Nationale Verkenning Digitale Duurzaamheid van de NCDD. Op 18 september is er in Den Haag een werkconferentie over met betrokken bestuurders en managers. De uitkomsten daarvan moeten van het interimrapport een definitief strategisch plan maken.

De verkenning maakt onderscheid tussen overheid, wetenschap en cultureel erfgoed. Bij alle drie is het bewustzijn van de noodzaak om digitale informatie duurzaam beschikbaar te houden nog onvoldoende. De verkenners adviseren dan ook aan de NCDD om 'voortdurend aandacht te vragen voor het langetermijnbelang: lokaal, regionaal, landelijk en in de diverse disciplines'. Vooral bij de overheid en in de wetenschap hebben de producenten van digitale informatie vaak andere prioriteiten dan het voor de toekomst veilig stellen daarvan. Heeft de overheid nog een Archiefwet die bepaalde verplichtingen meebrengt, voor de wetenschap geldt dat niet. 'Voor onderzoekers is de wetenschappelijke publicatie nu eenmaal het belangrijkste,' licht René van Horik toe, verantwoordelijk voor

de sector wetenschap binnen de verkenning. 'Zij erkennen best het belang van langetermijnbewaring, maar ze voelen zich er niet vanzelf verantwoordelijk voor.'

Onderzoekfinanciers zouden daarom een verplichting moeten kennen om duurzame toegankelijkheid van de data in projectvoorstellen op te nemen.

Ook wordt aandacht gevraagd voor de langetermijnfinanciering van data-projecten. Van Horik: 'Er zijn veel projecten die wetenschappelijk materiaal via internet toegankelijk maken. Maar als de financiering ophoudt, blijkt die toegankelijkheid ook vaak op te houden.' Geef instellingen die systematisch werken aan duurzame toegankelijkheid uitzicht op structurele financiering, aldus het rapport, dat ook aandringt op grootschalige opslagfaciliteiten. Die zijn per eenheid goedkoper dan kleine en om de kosten te bestrijden kunnen ze ge-

deeld worden. Verder zou er meer gedaan moeten worden aan het beschikbaar stellen van diensten, gereedschappen en kennis, vooral aan kleinere instellingen. Een landelijk kenniscentrum zou daar goed voor zijn, aldus het rapport.

Zo'n kenniscentrum lijkt ook beleidsmedewerker Louise Perbal van het financierende ministerie van OCW een goed idee. Zij is ook verder tevreden over de verkenning. 'Het is een goed doorwrocht geheel met nuchtere suggesties voor verbeteringen in de verschillende sectoren. We kunnen dit in de komende tijd goed gebruiken'. Ook Perbal wijst op de mindere belangstelling bij onderzoekers voor duurzame toegankelijkheid: 'Ze worden beoordeeld op hun publicaties, maar wat ze met hun data moeten doen staat vaak niet op het netvlies.' Dat moet anders: 'Volgens onze bewindsman zou het

OCW komt met nota Wetenschappelijke Informatievoorziening

Voor het eind van het jaar brengt het ministerie van OCW een nota uit over de omgang met digitale informatie in het wetenschappelijk onderzoek. In 2006 was dat ook al het plan, maar door een wisseling van bewindslieden en politieke prioriteiten kwam het concept in een la terecht. De huidige minister is volgens beleidsmedewerker Louise Perbal ook door de Open Access-gedachte gemotiveerd om het onderwerp weer op te pakken. Voor de nota zal worden geput uit het vroegere werk maar ook uit de Nationale Verkenning.

eigenlijk normaal moeten zijn dat wanneer je onderzoek wil doen, publicatie en de opslag van de data gewoon onderdeel uitmaken van het projectvoorstel'. (MdG)

© www.ncdd.nl www.ncdd.nl/werkconferentie2009.php

Boeken voor de nieuwe lichting



ANP KOEN SUYK

In het zicht van het nieuwe schooljaar werden vanuit Kampen de afgelopen weken zo'n 12 miljoen schoolboeken verstuurd naar 750.000 scholieren in heel Nederland. Vierduizend vakantiekrachten deden dat werk bij Van Dijk Educatie, de grootste schoolboekendistributeur van Nederland. Naar de cognitieve en sociaal-emotionele ontwikkeling van die scholieren, en naar de ontwikkeling van hun bur-

gerschapscompetenties, wordt intussen longitudinaal onderzoek gedaan binnen het grote cohortonderzoek COOL 5-18. Leerlingen van vijf tot achttien jaar worden in het onderzoek betrokken met drie rondes van toetsen en vragenlijsten. De Nederlandse Organisatie voor Wetenschappelijk Onderzoek NOW is opdrachtgever, sa-

men met het ministerie van Onderwijs, Cultuur en Wetenschap. Van het onderzoek, dat loopt van 2007 tot 2016, zijn afgelopen zomer de eerste resultaten gedeponereerd in het elektronisch archief EASY van DANS (Data Archiving and Networked Services) en daardoor beschikbaar gekomen voor wetenschappelijk gebruik.

© www.cool5-18.nl/ <http://easy.dans.knaw.nl> > search COOL 5-18

19-23 oktober: Open Access Week



In de derde week van oktober vindt de wereldwijde Open Access Week plaats. Ook aan Nederlandse universiteiten en onderzoekinstellingen zijn er dan activiteiten om de aandacht te vestigen op de gewenste vrije toegang tot wetenschappelijke informatie.

Internationaal wordt de week georganiseerd door instellingen en organisaties als PLoS (The Public Library of Science) en SPARC (the Scholarly Publishing and Academic Resources Coalition). In Nederland speelt de universitaire ict-organisatie SURFfoundation een coördinerende rol. De bedoeling is dat allerlei activiteiten worden georganiseerd om steun te werven voor Open Access. Zo houdt de Rijksuniversiteit Groningen op verschillende plaatsen en tijdstippen Open Access Lounges met korte toespraken. Het blad van de noordelijke universiteit komt met een special over het onderwerp. De Universiteit Utrecht organiseert op 23 oktober het symposium 'Read me, Cite me, Count me!'. Andere instellingen waren bij het sluiten van deze krant nog bezig met voorbereiden. (MdG)

© www.surfoundation.nl/nl/themas/openonderzoek/OpenAccess

AGENDA

7-9 september

Belfast, Queens University
Digital Resources for the Humanities & Arts 2009; Dynamic Networks of Knowledge and Practice: Contexts, Crises, Futures

The Digital Resources in the Humanities and Arts conference is held annually at various academic venues throughout the United Kingdom. The conference this year aims to promote discussion around dynamic networks of knowledge and practice, new digital communities of knowledge and practice, engaging users and digitisation of cultural heritage.

<http://dho.ie/drhaz2009/>

14-15 september

Den Haag, Museon

Europeana Plenary Conference 2009: Creation, Collaboration & Copyright

This conference will bring ideas and inspiration to the next level – it will explore different possibilities of content re-use, mash-up and APIs. The plenary will also address public domain and associated copyright issues. <http://version1.europeana.eu/web/guest/plenary/>

18 september

Den Haag, Koninklijke Bibliotheek

Toekomst voor ons digitaal geheugen
Werkconferentie van de Nationale Coalitie Digitale Duurzaamheid - NCDD over digitale duurzaamheid in Nederland
www.ncdd.nl/werkconferentie2009.php

25 september

Den Haag, Huygens Instituut KNAW

Presentatie van een elektronische editie van Van Deysse

De editie die hier wordt gepresenteerd is het resultaat van intensieve samenwerking tussen de ontwikkelaars van het Huygens Instituut en de editor, Jan-Willem van der Weij. Het gaat om een elektronische editie van Lodewijk van Deysse's proza-gedicht *Menschen en Bergen*. De editie presenteert de ontstaansgeschiedenis van de tekst en toont parallel de verschillende tekstversies.

www.huygensinstituut.knaw.nl/weblog/?p=740

12-13 oktober

Ljubljana, Slovenia

Training European Social Survey: Socio-Demographic Background Variables for Cross-National Comparative Survey

The objective of this course is to provide the participants with a thorough background on survey methods for comparative measurement of socio-demographic background variables.

www.europeansocialsurvey.org/

15 oktober

Den Haag, Data Archiving and Networked Services KNAW

Fedora op klompen

Bijeenkomst voor Nederlandse Fedora-gebruikers en informatiedag voor geïnteresseerden in de Fedora software van DuraSpace, het vroegere Fedora-commons. www.dans.knaw.nl/nl/dans_symposia/fedora_op_klompen_15_oktober_2009/

2-4 november

Leiden, Universiteit Leiden

Autumn School Tools voor onderzoekers – SURFfoundation

Er zijn veel (generieke) tools die handig kunnen zijn bij zowel het verrichten van onderzoek, als het communiceren over en publiceren van de resultaten. De Autumn School laat u kennis maken met deze tools. Diverse tools worden kort gepresenteerd en waar mogelijk wordt er in een proeftuin hands-on mee gewerkt. www.surfoundation.nl/nl/bijeenkomsten/Pages/AutumnSchool.aspx

12-13 november

Middelburg

Reuvsdagen, wetenschappelijk archeologisch congres

Als enig algemeen wetenschappelijk archeologisch congres in Nederland brengen de Reuvsdagen archeologen uit alle geledingen en specialismen samen. De dagen, georganiseerd door Erfgoed Nederland, bieden niet alleen een veelheid aan lezingen en workshops, maar ook kansen om contacten uit te bouwen en om het eigen archeologische instituut of bedrijf in de kijker te spelen. Jaarlijks wordt bovendien de tijd genomen om aandacht te besteden aan bijzondere personen. www.erfgoednederland.nl/

Bibliotheken zoeken digitale identiteit

Sinds digitale informatie direct van internet ingeburgerd is geraakt en vele zoekmachines beschikbaar zijn gekomen om die informatie ook te vinden, zijn de wetenschappelijke bibliotheken ietwat in een identiteitscrisis geraakt. Wat blijft er over van hun centrale rol in de informatiehuishouding van de universiteit als onderzoekers de bibliotheek omzeilen of niet eens meer zien dat het de bibliotheek is die toegang tot allerhande uitgeverplatforms vanaf hun eigen PC mogelijk maakt? Is gedrukte informatie nu het lot beschoren van verstoffen in het magazijn? Blijft alleen het gebouw in functie als een sociaal trefpunt op de campus? Dergelijke existentiële vragen voor wetenschappelijke bibliotheken stonden op het programma van de 38ste conferentie van LIBER, de *Association of European Research Libraries*, van 30 juni tot 3 juli in Toulouse.

Er blijken bibliotheken te zijn die voorsnog afwachten: ze zetten wat digitale informatie op het web en hopen dat de storm overwaait. Andere bibliotheken zien voor zichzelf een geheel nieuwe rol: ondersteuning van de onderzoeker door het aanbieden van allerhande diensten in

Aristoteles Latinus en digitale edities

The Letter before the Spirit: The Importance of Text Editions for the Study of the Reception of Aristotle. Dat congres vond van 2-5 juni 2009 in het Huygens Instituut in Den Haag plaats. De laatste dag was gewijd aan de toekomst van dergelijke tekstedities. Die zijn immers heel ingewikkeld door hun meertaligheid en door de combinatie van verschillende alfabetten. Afgezien van de uitzonderlijke combinatie van almaar zeldzamer wordende specialismen is het nog steeds een worsteling om links-rechts (Latijn, Grieks, moderne talen) en rechts-links alfabetten (Arabisch, Hebreeuws, Syrisch) in één document goed naast elkaar te krijgen.

Verscheidende inleiders lieten hun licht schijnen over de problemen van het maken van elektronische tekstedities, ieder vanuit hun eigen achtergrond en ervaring: Paul Tombeur (Brepols Publishers), George Kiraz (Gorgias Press), Thomas Milo (consultant voor Unicode voor Arabische schriften), Wido van Peursen (Universiteit Leiden) en Peter Boot, Mariken Teeuwen, Bas Doppen en Ronald Dekker (medewerkers van het Huygens Instituut). Tombeur benadrukte dat teksten simpelweg digitaliseren niet voldoende is. De complexiteit en gelaagdheid van teksten dreigt gemakkelijk verloren te gaan als zij zonder degelijk tekstkri-



Lars Björnshauge van DOAJ en Bas Savenije van de KB tekenen de overeenkomst voor de open access-tijdschriften van DOAJ.

datamanagement. Maar de moderne wetenschap speelt zich vooral af in samenwerkingsverbanden die de wereld omspannen, en bibliotheken hebben nog nauwelijks een traditie in het werken met interactiviteit. Tijdens de conferentie daagde Guus van den Brekel van de Groningse medische universiteitsbibliotheek zijn gehoor uit om zich actief te manifesteren op weblogs, Hyves en Twitter. Daar zijn de wetenschappers actief.

Aan de hand van door onderzoekers zelf ontworpen webdiensten als OpenWetWare, Science Commons, ProteoPedia en PloS liet Van den Brekel zien waar wetenschappers behoefte aan hebben.

Zijn de bibliotheken er al aan toe om zo innovatief te werk te gaan? Het publiek in Toulouse leek nog te aarzelen, ook omdat veel Oost- en Zuid-Europese bibliotheken worstelen met marginale budgetten. Het motto voor de conferentie was daarom *Innovation through cooperation*: een organisatie als LIBER kan ho-

pelijk genoeg kritische massa ontwikkelen om ook voor kleinere bibliotheken de interactieve digitale wereld nabij te brengen. In de tussentijd was het mooi om te zien dat deze LIBER-conferentie de achtergrond vormde voor de ondertekening door de *Directory of Open Access Journals* (DOAJ) en de Koninklijke Bibliotheek (KB) van een overeenkomst om tot duurzame archivering te komen van de vele wetenschappelijke tijdschriften die bij kleinere uitgeverijen in open access ter beschikking komen. (IA)

Ⓜ <http://liber2009.biu-toulouse.fr/>

Hoe trek je een steekproef van avatars?

Over een enquête onder avatars ging een van de opvallende bijdragen op de vijfde internationale e-Social Science Conferentie die van 24 tot 26 juni in Keulen werd gehouden. Organisatoren waren het National Centre for e-Social Science en GESIS, het Leibniz Instituut voor de Sociale Wetenschappen. De conferentie ging over allerlei nieuwe computer-toepassingen en ict-technieken van belang voor de sociale wetenschappen. Ook e-infrastructuren, zoals digitale datavoorzieningen, kregen ruim aandacht. Vanuit Nederland was de KNAW opvallend sterk vertegenwoordigd met presentaties van DANS, de Virtual Knowledge Studio en het Rathenau-Instituut.

De eerste dag was gewijd aan workshops en tutorials, door sommigen als het boeiendste onderdeel van de conferentie beschouwd. Van de presentaties was de meest memorabele die van Daniel Witte en Andreas Muchlichen, die een een survey hebben gehouden in de virtuele wereld van 'Second Life' en rapporteerden over hun methoden en bevindingen.

Hoe trek je een steekproef van avatars? Ze benaderden in totaal vierhonderd avatars in een virtueel winkelcentrum, waarvan bijna driekwart neutraal of positief stond tegenover een interview. Uiteindelijk maakte 13% het interview af. De onderzoekers probeerden erachter te komen in hoeverre uiterlijk, activiteiten en 'karakter' van de avatars verschilden van de zich daarachter verschuilende personen.

Natuurlijk waren er verschillen, maar de overeenkomsten waren toch opmerkelijker. De conclusie was in ieder geval dat meting van lifestyle in virtuele werelden mogelijk is. De beide heren hadden zelf verschillende avatars gemaakt (vrouwelijk en mannelijk, studentikoos en zakelijk) om interviewer-effecten na te gaan. Met hun probleemstelling en de aanpak daarvan komen ze zeker in aanmerking om te worden genomineerd voor een Ig Nobel Prize voor onwaarschijnlijk onderzoek. De conferentiewebsite bevat naast veel informatie ook blogs, twitters en discussies. (Peter Doorn)

Ⓜ <http://www.ncss.ac.uk/conference-09/>
<http://www.improb.com/ig/ig-pastwinners.html>

STCN bibliografie tot 1800 voltooid

Na 27 jaar arbeid heeft de Koninklijke Bibliotheek (KB) in juni het project 'Short-title Catalogue Netherlands' (STCN) voltooid verklaard.

190.000 retrospectieve titelbeschrijvingen in meer dan een half miljoen exemplaren van Nederlandse publicaties (en publicaties elders in het Nederlands) tussen 1540 en 1800 zijn bij elkaar gebracht en online gezet. Dat gebeurde met de exemplaren in de hand in circa vijfentwintig bibliotheken.

Het was de Koninklijke Nederlandse Akademie voor Wetenschappen (KNAW) die het project in 1980 van de grond tilde nadat de Rijkscommissie van Advies inzake het Bibliotheekwezen in 1969 de wenselijkheid had aangegeven van een volledige nationale bibliografie, vooral om historische onderzoekers te helpen hun bronnen sneller te vinden. In 1980 dacht men daarbij aan een boek met titelbeschrijvingen, maar gelukkig heeft het project lang genoeg geduurd om de waarde-



gen, maar gelukkig heeft het project lang genoeg geduurd om de waarde-

volle informatie nu als database aan te kunnen bieden. Vanaf 1997 zijn er trefwoorden toegevoegd aan de titels, en sinds 2005 worden scans van titelpagina, colofon en andere belangrijke bladzijden toegevoegd.

De STCN geeft antwoord op vele vragen. Hoeveel boeken werden in een bepaald jaar in een bepaalde stad gedrukt? Hoe verhielden de verschillende talen, lettertypes en formaten zich tot elkaar? Wat verscheen er in druk in het 'Rampjaar' 1672? Wat wist men rond 1700 van wiskunde, van medicijnen? Welke Engelse auteurs werden hier gelezen? Wat voor liedjes hoorde Hieronymus van Alphen op straat zingen? Welke bijbelvertalingen waren er in omloop? Wat leerden de kinderen op school?

Waaraan hadden de regenten van het weeshuis zich te houden?

Maar de STCN ligt ook ten grondslag aan grootschalige digitaliseringsprojecten zoals het project *Dutch Prints Online*, waarbinnen 1,3 miljoen pagina's uit de periode 1781-1800 gedigitaliseerd worden. De Short-title Catalogue levert de basis voor de selectie en de metadata. Vanaf 2010 zal de STCN-gebruiker steeds meer links naar volledige tekstbestanden vinden.

Binnenkort gaat ook een vernieuwde website de lucht in. Omdat er nog geregeld onbekende exemplaren opduiken, zal de KB zal de *Catalogue* als reguliere dienst blijven onderhouden en uitbreiden. (IA)

www.stcn.nl

Soundbites gaan online

Op 24 september wordt de website met de 'sprekende kaart' van het Soundbites-project voor het publiek opengesteld. In dat project worden authentieke gesprekken tussen dialectsprekers uit alle delen van Nederland voor onderzoek toegankelijk gemaakt.

Met de lancering van de website wordt het project 'Soundbites uit vervlogen tijden' van het Meertens Instituut en DANS (Data Archiving and Networked Services) afgerond precies een jaar nadat het van start ging. Soundbites is door de beide instellingen gefinancierd in het kader van het DANS-programma Kleine Dataprojecten. Het Meertens Instituut bezit ruim duizend

uur, ongeveer zes weken non-stop luisteren, aan geluidsopnames van gesprekken in dialect tussen twee of meer mensen, zonder inmenging van de onderzoeker. Ze werden verzameld tussen 1950 en de jaren tachtig, toen de opnames voldoende gespreid waren over Nederland. Tot nu toe was maar een heel kleine selectie van fragmenten op de website van het Meertens In-

stituut te vinden als 'de sprekende kaart'. Vanaf eind september is de gehele collectie op een duurzame manier voor onderzoek toegankelijk via internet. De sprekende kaart is daarbij een hulpmiddel waarbij de verschillende dialecten kunnen worden beluisterd door te klikken op een kaart van Nederland.

De verzameling geeft behalve van de verschillende dialecten ook een uniek beeld van het vroegere Nederland. De gesprekken gaan over zware armoede, de werkverschaffing in de crisisjaren, over de oorlogstijd, sociale tegenstellingen, lokale gebruiken, het drogen van wedstrijdduiven met hennep, het plukken van brandnetels voor de varkens, de slechte omstandigheden op vissersboten waar het drinkwater uit de ton nagenoeg zwart was.

Voor de verschillende dialecten bestaat een groeiende wetenschappelijke belangstelling. Bij het Meertens Instituut worden plannen ontwikkeld om de verzameling op te nemen in de Europese infrastructuur CLARIN (Common Language Resources and Technology Infrastructure). De opnames kunnen dan, met behulp van hulpmiddelen die via die infrastructuur beschikbaar komen, met andere opnames vergeleken worden. Daarmee kunnen onderzoekers ook mogelijkheden verkennen die buiten het traditionele dialectonderzoek liggen. (Douwe Zeldenrust)

www.meertens.knaw.nl/soundbites vanaf 24/9

Online varianteneditie van Van Deysse

Op 25 september presenteert het Huygens Instituut in Den Haag een elektronische editie van *Menschen en Bergen*, een prozagedicht van Lodewijk van Deysse. *Menschen en Bergen* is volgens het instituut een sleuteltekst in de ontwikkeling van het Nederlands proza en heeft een fascinerende ontstaansgeschiedenis. Het handschrift, de drukproeven (met verbeteringen) en de verschillende drukken tonen hoe Van Deysse streefde naar een tot het uiterste geïntensiveerd proza.

Onderzoeker Jan-Willem van der Weij promoveerde op het Nederlandse prozagedicht en werkte het afgelopen jaar bij het Huygens Instituut aan een online editie die de totstandkoming van *Menschen en Bergen* inzichtelijk maakt. Samen onder meer Walter Ravenek, Bas Doppen en Ronald Dekker werd op het editieplatform *eLaborate* een editie gemaakt waar de verschillende tekstgetuigen afzonderlijk en in samenhang kunnen worden geraadpleegd. Vanuit elke tekstversie kan het volledige variantenapparaat worden geraadpleegd. Een en ander is ingebed in een toelichting op de geschiedenis van de tekst en het literaire belang daarvan.

Bij de presentatie belicht Jan Willem van der Weij de keuzes die gemaakt zijn bij het produceren van de editie. Prof. Elisabeth Leijne spreekt over *Menschen en Bergen* in de context van zijn tijd, en actrice en voordrachtskunstenares Mieke Lelyveld draagt fragmenten uit de tekst voor. De editie wordt officieel in gebruik genomen door Robert Alberdingk Thijm, achterkleinzoon van Van Deysse. (PB)

www.huygensinstituut.knaw.nl/weblog/?p=740

Handel met de West

Het elektronisch archief EASY van DANS is aangevuld met gegevens over Nederlandse scheepvaart in het Atlantisch gebied van 1600 tot 1800. Meer data uit hetzelfde project worden nog toegevoegd, samen met gegevens die in het verleden al waren verzameld. De data komen uit het project 'Dutch connections: the circulation of people, goods and ideas in the Atlantic world, 1680-1795' van de Universiteit Leiden, dat de handel met de West voor het voetlicht wil halen. Die heeft een veel grotere bijdrage aan de welvaart van de Nederlandse Republiek geleverd dan tot nu toe altijd werd gedacht.

www.kitlv.nl/home/Projects?id=19

Dataforum opgericht

Organisaties in het wetenschappelijk onderzoek hebben een *Onderzoeksdata Forum* opgericht dat ervoor gaat zorgen dat kennis en deskundigheid over duurzaam opslaan en uitwisselen van onderzoeksdata worden uitgewisseld. SURFFoundation, de organisatie voor universitaire ict-ontwikkeling, nam met DANS (Data Archiving and Networked Services) het initiatief. Deelnemers zijn de universiteitsbibliotheken die digitale data in repositories opslaan of dat willen gaan doen. De Nationale Coalitie Digitale Duurzaamheid (NCDD) coördineert de activiteiten met andere sectoren uit de publieke sector als cultureel erfgoed en archieven. Er zijn werkgroepen voor technische en infrastructuur, voor juridische en organisatorische aspecten en het bevorderen van kennis en expertise. Het project valt binnen het SURFshare programma en heeft in eerste instantie een looptijd van twee jaar. (IA)

ICPSR op YouTube

Data-archief ICPSR plaatste onlangs een aantal handleidingen voor het gebruik van zijn data op YouTube. Het Inter-university Consortium for Political and Social Research in het Amerikaanse Michigan is het grootste sociaalwetenschappelijke data-archief. De handleidingen bieden praktische hulp bij het zelf oplossen van vaak voorkomende problemen bij het werken met ICPSR-data. Zo is er een handleiding voor het inlezen van files in de statistische pakketten SPSS of STATA met behulp van een set-up file. Ook is er informatie te vinden over het gebruik van de SDA online analyse tool. Zonder SPSS, STATA of SAS is het met deze tool toch mogelijk om op specifieke ICPSR datasets statistische analyses uit te voeren. Zoeken op YouTube met *ICPSR*. www.youtube.com/results?search_type=&search_query=icpsr

Model voor digitalisering

Digitaal Erfgoed Nederland (DEN) ontwikkelt in het kader van het project Meer Digitale Feiten een rekenmodel voor digitaliseringskosten. Gemikt wordt op een eerste versie in december van dit jaar. Er wordt gewerkt met een eerste versie die het Gelders Archief vorig jaar ontwikkelde voor intern gebruik. Bijzonder is dat het model ook de kosten berekent die na het afronden van het project worden gemaakt, zoals opslag- en beheerskosten.

DEN wil het model uitbouwen tot algemeen bruikbaar rekenmodel in de vorm van een belastingformulier met een stapsgewijze aanpak, gepubliceerd als webtoepassing.

www.erfgoednederland.nl

'Scientwists': politiek achter mooi plaatje



Je zou denken dat wetenschap een medium vereist dat iets minder gehaast is dan Twitter. Toch houdt de weblog Sciencebase een lijst bij van inmiddels 522 twitterende wetenschappers ('scientwists'). Ook op Twitter zijn er rangen en standen, en een maat voor je sociale positie op Twitter is het aantal mensen dat je berichten volgt. Andrew Maynard maakte daarvan een interactieve

bubble chart met behulp van Many Eyes (zie *e-data@research* september 2007).

Even interessant als de *bubble chart* is het gevecht om wie op de lijst mogen staan. Een wetenschapsjournalist is geen wetenschapper, merkt de één op. En een ander is 'maar een dokter'. Achter elke mooi plaatje liggen politieke beslissingen. (PB)

www.meertens.knaw.nl/soundbites vanaf 24/9

Duurzame toegang tot onderzoeksdata nog zeer fragmentarisch

INGE ANGEVAARE

Duurzaam archiveren en toegankelijk houden van wetenschappelijke data is nog steeds meer uitzondering dan regel, zo blijkt uit het onlangs verschenen interimrapport *Toekomst voor ons digitaal geheugen: duurzame toegang tot informatie in Nederland*. Dit in tegenstelling tot wetenschappelijke publicaties, die via de digitale archieven van de universiteiten duurzaam worden opgeslagen in het e-Depot van de Koninklijke Bibliotheek.

In de Nationale Verkenning Digitale Duurzaamheid, die ten grondslag ligt aan het rapport, werd het onderscheid tussen digitale publicaties en onderzoeksdata al snel duidelijk. Voor de eerste categorie bestond er in het analoge tijdperk een gevestigde praktijk. De informatie werd door de auteur naar een uitgever gestuurd, na *peer review* gepubliceerd in een vaktijdschrift, en vervolgens duurzaam opgeslagen in (nationale) bibliotheken. Die praktijk kon in het digitale tijdperk grotendeels worden voortgezet.

Complicerende factoren

Digitale onderzoeksdata vormen een heel nieuwe categorie, waarvan de verkenning uitwees dat voor duurzame opslag en hergebruik nog maar incidenteel voorzieningen zijn

getroffen. Daar zijn een aantal oorzaken voor. Onderzoeksdata zijn er in ontelbare vormen en formaten, en ze bevinden zich overal en nergens – op de eigen PC van een onderzoeker, op servers van instellingen, in grote internationale datacentra. Ze bewegen zich tussen al die netwerken en worden steeds opnieuw bewerkt. Dat alles maakt het lastig om grip te krijgen op die informatiestromen. Zeker internationaal is het immers niet gemakkelijk om partijen aan te wijzen die verantwoordelijk zijn voor de organisatie van duurzame toegankelijkheid én voor de financiering daarvan.

Een andere complicerende factor is het feit dat de onderzoeker die de data produceert het belang van duurzame toegankelijkheid vaak niet als het zijne ervaart. Uit onderzoek in

het kader van andere projecten, zoals PARSE.insight, blijkt ook dat onderzoekers wel data van anderen willen gebruiken, maar zelf veel minder bereid zijn om hun data beschikbaar te stellen aan derden. Ze vrezen dat anderen misbruik zullen maken van het door hen verrichte werk en ervaren de extra inspanning die van hen wordt gevraagd als lastig en tijdrovend. Het belang van de samenleving, die als geheel wel baat heeft bij het hergebruiken van onderzoeksdata, staat verder af van de individuele onderzoeker.

Selectie nodig

En dan is er de grote vraag wat wel en niet moet worden opgeslagen voor hergebruik. Bij wetenschappelijke publicaties fungeert het uitgeverbedrijf met behulp van de *peer review* als

selectiemechanisme. Een dergelijk mechanisme ontbreekt voor onderzoeksdata, terwijl de hoeveelheden veel groter zijn en de verschijningsvormen (ook in technische zin) veel gevarieerder en vaak ook ingewikkelder dan bij publicaties, waar pdf de boventoon voert. Er zijn ook grote verschillen tussen disciplines: daar waar onderzoek onmogelijk is zonder longitudinale data, bijvoorbeeld in de meteorologie, is duurzame opslag al snel tot stand gekomen. Ook wettelijke regels kunnen helpen, zoals in de archeologie waar het Verdrag van Malta dat deponeren van onderzoeksgegevens verplicht stelde. Archeologische data vormen de grootste categorie in de archieven van Data Archiving and Networked Services (DANS).

Buiten DANS zijn in Nederland

maar enkele wetenschappelijke data-archieven bekend die echt beleid en organisatie hebben afgestemd op toegankelijkheid voor de lange termijn. Voorbeelden zijn het Instituut voor Nederlandse Lexicologie, het Max-Planck-Instituut for Psycholinguistics, het Internationaal Instituut voor Sociale Geschiedenis en het KNMI. Een datacentrum van de drie technische universiteiten (3TU) verkeert in de opbouwfase. Verreweg de meeste onderzoeksdata vallen momenteel nog buiten enig duurzaamheidsregime.

Het ontwikkelen van selectiemechanismen is een belangrijke richting waarin volgens het rapport gewerkt zou moeten worden om duurzame toegankelijkheid te bevorderen. Ook zouden onderzoekers die hun data ter beschikking stellen op de een of andere manier beloofd moeten worden voor hun inspanningen. Financiers zouden deponeren verplicht kunnen stellen, zoals NWO met zijn programma Middelgroot al doet. Datacentra, ten slotte, kunnen en moeten onderzoekers helpen bij het organiseren en beheren van hun data, opdat die na afloop van het onderzoek geschikt zijn voor hergebruik.

www.ncdd.nl

Er zit toekomst in het verleden

MARJOLIJN KRUIP

Middeleeuwse pelgrimsroutes op de kaart gezet

Op de afdeling Kunstgeschiedenis van de Radboud Universiteit wordt al jaren onderzoek verricht naar pelgrimage in de late middeleeuwen. Dit onderzoek heeft een nieuwe impuls gekregen door de ontwikkeling van een internettoepassing met geo-technologie.

Pelgrims die een bedevaartsoord hadden bereikt, kochten daar in de late middeleeuwen vaak insignes of ampullen. Deze religieuze objecten herinnerden de gelovigen door de afbeelding aan het bezochte heiligdom; daarnaast boden ze bescherming, weerden ze het kwaad af en brachten ze geluk. In de onderzoeksgroep Kunera, onder leiding van professor Jos Koldewij, wordt al ruim tien jaar gewerkt aan een database van zulke laatmiddeleeuwse pelgrimssouvenirs.

Souvenirs die in de ene plaats werden gemaakt en in de andere als bodemvondst opduiken, geven inzicht in het reisgedrag van laatmiddeleeuwse pelgrims. Daarom, en om de Kunera-database met op

dit moment vijftienduizend religieuze en profane objecten op lange termijn niet te laten verworden tot een inactieve digitale kaartenbak, is hij recentelijk ontsloten met een Geografisch Informatiesysteem (GIS). Dat was mogelijk dankzij een subsidie van de Stichting Reinier Post. Door het gebruik van een nieuwe GIS-applicatie kunnen de geografische patronen uit het insigne- en ampullenmateriaal op kaarten van Google Maps worden geanalyseerd (www.kunera.nl).

Pilots

De nieuwe impuls die zo aan het kunsthistorisch onderzoek kon worden gegeven, kreeg in het najaar van 2008 een vervolg met twee Kleine Data-projecten van DANS: 'Alfa-Geo', mede mogelijk gemaakt door het programma Ruimte voor Geo-informatie. Het doel van die pilots was het onderzoek naar middeleeuwse pelgrimsroutes te verbreden door het insignemateriaal binnen de nieuwe internettoepassing samen te brengen met geografische patronen uit heel ander bronmateriaal. Daartoe zijn twee schriftelijke bronnen

ontsloten: het *Mirakelboek van Onze Lieve Vrouw van 's-Hertogenbosch* en de Brugse Wegwijzer.

Alle 481 wonderverhalen uit het *Bossche Mirakelboek* (1382-1603), zijn aan de hand van de herkomstplaatsen van de pelgrims op een kaart gezet. Het patroon dat daaruit naar voren komt, is vergeleken met de patronen van Bossche insignes. Die vergelijking bevestigt de toepasbaarheid van geo-technologie in insigneonderzoek, maar ook het nut van de confrontatie met ander historisch bronmateriaal. Het mirakelboek-patroon vult het patroon van de uit 's Hertogenbosch meegenomen insignes aan: de kwetsbare metalen objecten worden namelijk enkel opgegraven in gebieden met een vochtige, luchtafsluitende bodem. Het patroon van de insignes die in 's-Hertogenbosch zijn gevonden, brengt een afwijking aan het licht: het verspreidingspatroon is veel groter dan dat van de eerdergenoemde insignes. De afstand die de reizende Bosschenaar wilde afleggen was klaarblijkelijk groter dan de afstand die de niet-Bosschenaar voor de Bossche Maria wilde overbruggen.



Insigne OLV van 's-Hertogenbosch, vindplaats: Nieuwlande. Bron: R.M. van Heeringen, A.M. Koldewij en A.A.G. Gaalman, Heiligen uit de modder. In Zeeland gevonden pelgrimstekens (Clavis Kunsthistorische Monografieën 4), Utrecht en Zutphen 1987, pp. 78-79, nr. 11.2

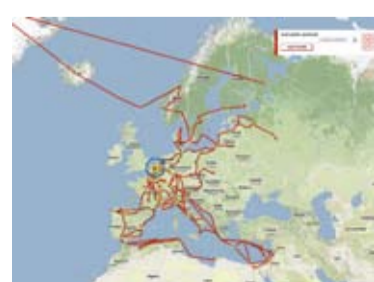
Middeleeuwse TomTom

Een nog concretere bron voor middeleeuwse reisroutes is de *Brugse Wegwijzer* (1370-1400). De ontsluiting van deze onderbelichte, maar uiterst waardevolle bron met een Geografisch Informatiesysteem is

spectaculair. De routes uit de geschreven bron zijn omgezet in een digitale wegenkaart, waardoor het totaalconcept van deze middeleeuwse routeplanner voor het eerst helder in beeld is gebracht. Daarnaast is de tekst gedigitaliseerd en toegankelijk gemaakt. De geografische patronen van pelgrimssouvenirs met herkomst- en vindplaats Brugge kunnen over het patroon van de wegenkaart worden geplaatst. Vergelijking van beide soorten bronmateriaal heeft ook hier interessante resultaten aan het licht gebracht.

Door de lancering van de internettoepassing Kunera 2.0 is bestaand abstract onderzoeksresultaat zichtbaar gemaakt met behulp van een Geografisch Informatiesysteem. Daardoor is het toegankelijk geworden voor een breder historisch en kunsthistorisch publiek. Het gebruik van geo-technologie is heel waardevol gebleken voor het ontsluiten en vergelijken van historisch bronmateriaal. Kortom, er zit toekomst in het onderzoek naar het verleden.

Op de website www.kunera.nl kunnen de pilot-projecten geraadpleegd worden. Ga via 'direct naar de database' naar het kopje 'Steden in perspectief'.



Brugse Wegwijzer gevisualiseerd met een Geografisch Informatiesysteem

Kwantitatief georiënteerde Neerlandica Karina van Dalen-Oskam:

MARTIJN DE GROOT

‘Vechten voor het belang van empirie’

Binnen de letterkunde loopt een scheidslijn tussen de mensen die hun vakgebied vooral als een intellectuele en kunstzinnige discipline beschouwen en degenen die er een empirisch fundament onder willen leggen. Die laatste groep heeft het niet makkelijk, vindt letterkundige Karina van Dalen-Oskam, onderzoekster bij het Haagse Huygens Instituut.

‘Het onymisch landschap’. Een onderzoeker die bij die woorden de ogen wat toeknijpt en verzucht: ‘Dat vind ik nou een mooie term’, dat zou wel eens een letterkundige kunnen zijn. Karina van Dalen is zo iemand. Zij heeft al vanaf haar schooltijd belangstelling voor namen, en de laatste tijd in het bijzonder voor het geheel aan namen van personages dat je in een roman kan tegenkomen. *Het onymisch landschap*. Gevoel voor schoonheid hoort ook bij het vak. ‘Letterkundigen schrijven zelf ook heel mooie teksten waarvoor je de AKO-literatuurprijs kunt krijgen.’

‘Letterkundigen schrijven zelf ook heel mooie teksten waarvoor je de AKO-literatuurprijs kunt krijgen’

Maar er is een *maar*. En die tekent zich af als Van Dalen het woordje ‘traditioneel’ aan haar betoog toevoegt. De *traditionele* letterkundigen hebben wel een heel intieme relatie met hun teksten maar geen waardering voor de empirie en dat is een bedreiging voor het vak. ‘De Neerlandistiek is in gevaar’ is dan ook de openingzin van een column die Van Dalen recent schreef in het *Tijdschrift voor Nederlandse Taal- en Letterkunde*, ter gelegenheid van de 125-ste jaargang van dat tijdschrift. Het gevaar bleek zich vooral aan de letterkundige kant van het continuüm taalkunde-letterkunde te bevinden. ‘Taalkunde is van huis uit al empirisch georiënteerd,’ legt de Neerlandica uit. Met de groeiende rol van ict dienen zich ook voor de letterkunde steeds meer kansen aan om elektronisch gereedschap in hun onderzoek te gebruiken. Die kansen maken een meer empirische benadering van de letterkunde mogelijk. Maar de ‘traditionele letterkundigen’ laten ze aan zich voorbij gaan, constateert Van Dalen. Sterker nog, de minderheid die de empirie wel in het letterkundig onderzoek wil betrekken, ‘wordt uit het centrum weggeslingerd en zoekt zijn heil in het buitenland,’ aldus de onderzoekster ict en Teksten van het Haagse Huygens Instituut. ‘Empirie wordt niet gewaardeerd in de Nederlandse letterkunde. Je moet echt heel hard vechten om aan te tonen dat het nuttig is.’

Matrixprintertje

Karina van Dalen-Oskam is een echte alfa. ‘Meer alfa kan je ze niet krijgen,’ zegt ze zelf in haar kamer met uitzicht op het ministerie van



Karina van Dalen-Oskam: ‘De minderheid die wel empirisch wil werken, wordt uit het centrum weggeslingerd en zoekt zijn heil in het buitenland’

Buitenlandse Zaken. Het Huygens Instituut is ingekwartierd bij de Koninklijke Bibliotheek die zelf weer onderdeel is van het cluster kantoor-kolossen dat zich aan het Haagse Centraal Station lijkt te hebben vastgeklit. Maar die alfa-oriëntatie heeft haar niet kunnen weerhouden van een vroege belangstelling voor computers en wat je daarmee kan doen. Haar vriend verdient daarvoor krediet. ‘Hij studeerde natuurkunde en door hem maakte ik kennis met computers. Ik schreef mijn scriptie op de computer, wat toen nog niet gebruikelijk was, en ik printte hem uit op een laserprinter in plaats van zo’n matrixprintertje dat in die tijd de norm was.’ En ook nu nog maakt Van Dalen graag gebruik van de praktische aanwijzingen van de vriend die al weer heel lang haar echtgenoot is. ‘Hij kijkt vaak over je schouder mee en dan zegt hij bijvoorbeeld: ‘Dat kan toch veel eenvoudiger!’

‘Je merkt dat traditionele letterkundigen vaak heel enthousiast zijn over alles wat er digitaal verschijnt’

De drijvende kracht achter Van Dalen’s betrokkenheid bij elektronische data is echter niet haar eigen talent voor ict of dat van haar man, zo blijkt bij nadere beschouwing. Dat is haar ‘empirische oriëntatie’. ‘Als ik bepaalde dingen zie of conclusies wil trekken, dan wil ik een onderbouwing zien. Dan zoek ik naar getallen om zo’n conclusie mee te ondersteunen.’

Tekstgerelateerd

Nadat Van Dalen in de jaren tachtig Nederlands studeerde in Utrecht, kreeg ze haar eerste baan bij het Instituut voor Nederlandse Lexicologie in Leiden. Daar was ze een aantal jaren betrokken bij de samenstelling van het *Vroegmiddelnederlands Woordenboek*. Dat werk beantwoordde heel goed aan haar drijfveren want het project moest in alle fases met behulp van de computer worden uitgevoerd – ‘dat was nieuw in die tijd’ – maar het was ook empirisch georiënteerd. Empirisch? ‘Ja, we betrokken in dat project bijvoorbeeld alleen teksten waarvan we zeker wisten die dat uit de dertiende eeuw kwamen. Dus niet teksten die ook wel in een latere tijd konden zijn gekopieerd. Alleen op die manier kan je zo’n woordenboek later gebruiken om vergelijkingen te maken of om andere teksten op hun historische komaf te beoordelen. Letterkundigen begrepen die aanpak maar moeilijk. Mij heeft de lexicografie juist bevestigd in mijn empirische oriëntatie.’

Was taalkunde, waar een kwantitatieve benadering al heel normaal is, dan eigenlijk niet meer haar richting geweest? ‘Nee,’ antwoordt ze beslist, ‘Want de vragen die ik stel zijn uiteindelijk tekstgerelateerd. Hoe heeft een auteur zijn woorden in een tekst gebracht, welke stilistische elementen zijn daarin te herkennen en wat is de betekenis daarvan?’ En bovendien gaat het haar niet om haar eigen voorkeur maar om de toekomst van de letterkunde, zo maakt Van Dalen duidelijk. Niet dat er geen plaats meer is voor

de opvatting waarin ‘veel lezen en selectief observaties doen op basis van het geheugen van de onderzoeker’ het hoogste goed is. Als er maar meer ruimte komt voor het streven om waarnemingen met empirische observaties te schragen.

Wiki-principe

Pessimistisch is de Huygens-onderzoekster trouwens niet: ‘Je merkt dat traditionele letterkundigen vaak heel enthousiast zijn over alles wat er digitaal verschijnt. Hoe groter die massa van beschikbare werken wordt, des te meer behoefte zal er zijn aan gereedschappen om dingen op te zoeken. Daardoor hoop ik dan dat men zich het belang van de empirie zal realiseren.’ Die hoop baseert de onderzoekster goeddeels op haar eigen ervaring. In 2002 verruilde ze haar werkplek bij het INL voor één als hoofd van de afdeling Neerlandistiek bij het toenmalige Nederlands Instituut voor Wetenschappelijke Informatiediensten NIWI. Behalve het werk aan de *Bibliografie der Nederlandse Taal- en Letterkunde* kreeg ze daar de kans

Karina van Dalen-Oskam studeerde in de jaren tachtig Neerlandistiek in Utrecht, met als specialisaties Middelnederlandse Letterkunde, Oudgermaanse Taal- en Letterkunde en Mediëvistiek. Ze werkte in de jaren negentig als redacteur op het Nederlands Instituut voor Lexicologie in Leiden en kwam aan het begin van deze eeuw bij het wetenschappelijk informatie-instituut NIWI terecht als

om, samen met onder meer Joris van Zundert, veel energie te steken in de ontwikkeling van de online werkomgeving voor tekstonderzoekers *eLaborate*. ‘Het schoot toen niet op met het digitaliseren van belangrijke tijdschriften uit het verleden. Het was gewoon ook teveel om allemaal over te typen. Toen zeiden we tegen elkaar: waarom alles overtypen als je ook scans kan maken en die met automatische tekenherkenning, OCR, beschikbaar kan maken voor iedereen om ze te verbeteren? OCR’s leveren immers bepaald geen foutloos resultaat op. Maar het effect daarvan zouden we met een soort Wiki-principe kunnen neutraliseren, dat was de basisgedachte.’

Open Source aanbieden

In de loop van de jaren, ook nadat Van Dalen met haar afdeling naar het Huygens Instituut verhuisde toen het NIWI werd opgeheven, is vervolgens stap voor stap verder gewerkt en verbeterd aan de online werkplaats die het uiteindelijk werd. Onderzoekers maken er steeds meer gebruik van, verzekert Van Dalen. ‘Dat komt doordat het systeem toegankelijk is en het gebruik eenvoudig. Dan gaan de mensen de voordelen er wel van zien!’ Hulpmiddelen als deze zijn niet uniek, geeft Van Dalen toe. ‘Maar wij waren wel de eersten die het online hadden, en waar het niet voor één maar voor heel veel projecten kan werken. Dat heeft ons de voorsprong gegeven waarmee we nu kunnen kijken hoe we het verder kunnen ontwikkelen, zodat het ook voor andere talen en in andere landen gebruikt kan worden. Ik hoop dat we het dan Open Source aan kunnen bieden. Dan heb je kans dat anderen er op een gegeven moment hun krachten aan gaan geven.’ Van Dalen zelf heeft dan wel weer nieuwe projecten die haar aandacht zullen opeisen. Zoals auteursherkenning. ‘Daar zijn steeds meer moderne technieken voor en ik merk dat er ook maatschappelijk belangstelling voor is.’ Of namen. ‘Ik wil in de loop van dit jaar twintig Nederlandse romans aan een analyse onderwerpen van hun *onymisch landschap*. Er lijken bepaalde regels te zijn die het gebruik van soorten namen verklaren. Maar die kan ik pas hard maken als ze na bestudering van die twintig overeind blijven.’

hoofd van de afdeling Neerlandistiek. Toen was Van Dalen al in Leiden gepromoveerd (1997) op taal- en letterkundige ‘Studies over Jacob van Maerlands Rijmbijbel’. In 2005 werd ze onderzoekster op het Huygens Instituut voor literatuur- en wetenschapsgeschiedenis, waar ze verantwoordelijk is voor het snelgroeiende terrein van ict-toepassingen in het onderzoek.

WIEBE KLESTRA

FOCUS

ERCOMER

THIJS HERMSEN

Het bevorderen van vergelijkend, kwantitatief onderzoek naar migratie en ethnische relaties in Europa, dat is volgens directeur prof. dr. Louk Hagendoorn het doel van het European Research Centre on Migration and Ethnic Relations (ERCOMER), dat in november zijn vijftienjarig bestaan viert.



Directeur van ERCOMER Louk Hagendoorn: 'Het is een grote wens om een goed data-archief te maken'

ERCOMER is een interuniversitair onderzoekcentrum, gevestigd in Utrecht met afdelingen in Nijmegen en Rotterdam. Het is 1994 vanuit de faculteit Sociale Wetenschappen van de Universiteit van Utrecht opgericht, door een driemanschap waarvan behalve de huidige directeur ook Han Entzinger en Malcolm Cross deel uitmaakten. Doel was om vergelijkend, kwantitatief onderzoek te bevorderen op het gebied van (internationale) migratie en ethnische relaties. Hagendoorn: 'Wij hadden een grote wens om buitenlandse expertise te verwerven en ons eigen onderzoek uit te breiden. Om dat te bewerkstelligen wilden wij post-docs uit Europa naar Utrecht halen en dat financieren met behulp van Marie Curie Fellowships. Voor die beurzen kwamen toen alleen erkende onderzoekcentra in aanmerking. We hadden dus geen keus en hebben eerst maar een onderzoekcentrum gestart. Toen we aan alle eisen hadden voldaan, konden we meedingen en werd de eerste aanvraag meteen gehonoreerd.'

'Alles wat met migratie in Europa te maken heeft valt in ons blikveld,' aldus Hagendoorn nu. 'Onze vier hoofdthema's zijn: migratie en migranten, immigranten en minderheden in de nieuwe samenleving, culturele verschillen en reacties in de samenleving en, als laatste, nationalisme en ethnische conflicten.' Het merendeel van de onderzoekers van ERCOMER doet zijn werk in Utrecht: 'Daar werken 32 vaste onderzoekers, en er zijn altijd enkele visiting scholars. Er is ook een staf van vijf medewerkers. Bij de afdelingen van het instituut in Nijmegen en Rotterdam werken in totaal vijf onderzoekers.'

ERCOMER is niet het enige Nederlandse universitaire onderzoekcentrum op het gebied van migratie. 'Het Institute for Migration and Ethnic Studies (IMES) in Amsterdam houdt zich bezig met dezelfde problematiek, maar doet vooral kwalitatief onderzoek terwijl ERCOMER kwantitatief onderzoek verricht.'

ERCOMER is bij veel grote internationale empirische onderzoeken betrokken. Het instituut vormt daartoe netwerken met allerlei andere instituten, afdelingen en faculteiten in binnen- en buitenland en de organisaties die alles financieren om steeds weer nieuw, innovatief onderzoek van de grond te krijgen. Hagendoorn: 'De belangrijkste partners in Nederland zijn de afdeling Sociologie van de Radboud Universiteit in Nijmegen en de faculteit Sociale Wetenschappen en de afdeling Economie van de Erasmus Universiteit in Rotterdam. Binnenkort hopen we ook het contact met de afdeling Politieke Wetenschappen van de Universiteit Leiden te formaliseren. Zij houden zich bezig met de politieke gevolgen van migratie.' Ook met het Sociaal Cultureel Planbureau in Den Haag onderhoudt ERCOMER goede banden. Buiten Nederland werkt het centrum samen met universiteiten in België, Hongarije, Zweden, Italië, de Verenigde Staten, Oekraïne en Rusland. 'Goed onderzoek heeft baat bij uitstekende contacten,' aldus de directeur. 'Die komen echter niet uit de lucht vallen. Samen onderzoek verrichten én dat publiceren en op die manier langdurige en plezierige contacten op te bouwen en te onderhouden was en is het credo.'

ERCOMER heeft meer taken dan alleen onderzoek: 'Wij leiden

Gereedschap voor de computational humanities

PETER BOOT

Softwaregereedschappen voor de humaniora moeten beter onderling samenwerken, meer soorten inhoud aan kunnen en beter te vinden zijn, volgens een rapport over 'Tools for Data-driven Scholarship' dat afgelopen voorjaar in Amerika verscheen. *e-data&research* verdiepte zich erin. Het rapport vertoont paar blinde vlekken maar ook hoopvolle overeenkomst met de doelstellingen van het Nederlandse Alfalab.

Empirisch onderzoek in de geesteswetenschappen vraagt om gereedschappen voor het bewerken en analyseren van data. Over de vraag hoe zulke tools ontwikkeld kunnen worden en wat er nodig is voor het succesvol gebruik en voortbestaan ervan wordt al lang gediscussieerd. In 2003 constateerde de Amerikaanse anglist en informatiewetenschapper John Unsworth al dat *'[the] sense of déjà vu is particularly keen when it comes to tool-building for humanities computing.'*

Veel ontwikkelwerk aan tools is mogelijk dankzij subsidies, en de subsidiërende instellingen zien natuurlijk graag dat hun projecten resulteren in producten die breed worden gebruikt en na een periode van financiële ondersteuning op eigen benen verder kunnen. Dat blijkt vaak niet het geval, en drie grote Amerikaanse funders (National Science Foundation, National Endowment for the Humanities, Institute of Museum and Library Services) organiseerden daarom in oktober 2008 de workshop *Tools for Data-Driven Scholarship: Past, Present, Future*. Het eindrapport kwam uit in maart 2009, en is verplichte

ook studenten van de Universiteit Utrecht op.' Daarnaast levert ERCOMER op verzoek informatie aan allerlei organisaties die in migratieproblematiek zijn geïnteresseerd, waaronder de overheid en scholen. Hagendoorn: 'Vroeger besteedden we daar meer tijd aan maar we zijn tot de conclusie gekomen dat de primaire taak, het onderzoek, in het gedrang kwam.'

Op de website van het instituut zijn abstracts van alle projecten te vinden en de titels en meestal ook abstracts van publicaties. De data-sets waarmee het kwantitatieve onderzoek dat door de onderzoekers van ERCOMER wordt verricht zijn veelal niet beschikbaar. Hagendoorn: 'Sommige sets die met publieksgeld zijn gemaakt zijn in principe voor iedereen toegankelijk maar veel sets zijn van onderzoekers zelf en daardoor niet toegankelijk.' Intern zijn de sets beschikbaar maar niet via een gecentraliseerde data-opslag. 'Helaas, een goed data-archief maken is lastig en daar zijn wij bij ERCOMER nog niet in geslaagd. Het is wel nog een grote wens van mij om dat nog een keer te realiseren.'

kost voor toolontwikkelaars en hun subsidiegevers.

Institutionele focus

Volgens het rapport vertonen bestaande gereedschappen drie soorten tekortkomingen: ze werken niet goed genoeg samen met andere tools, ze werken niet goed genoeg met verschillende soorten inhoud, en het is voor potentiële gebruikers te moeilijk om ze te vinden. In verschillende fasen van het bestaan van een tool zijn verschillende aspecten van die problemen belangrijk. In de fase van het eerste ontwerp spelen vragen over het beoogde bereik van de tool en de samenwerking met andere hulpmiddelen een rol. Op het moment dat een project de bouwfase in gaat, gaan projectmanagement en planning een vooraanstaande rol spelen. Wanneer de tool een bruikbare staat heeft bereikt, wordt het belangrijk gebruikers aan te trekken en vast te houden.

Het rapport heeft een duidelijk institutionele focus, wat gegeven de achtergronden niet verrassend is. Het veronderstelt dat gereedschappen worden ontwikkeld in teams die gemanaged moeten worden. Daarmee negeert het zowel de succesvolle tools die worden ontwikkeld door één of enkele onderzoeker(s) voor het eigen onderzoek, als de grote niet aan een instituut gebonden open source projecten.

Een daarmee samenhangende aanname is dat het ontwikkelen van gereedschappen moet resulteren in (her)bruikbare tools die zijn gebouwd volgens de ontwikkel- en managementmethoden van professionele ontwikkelaars. Die aanname laat echter buiten beschouwing dat dat ontwikkelen ook de uitwerking van onderzoeksideeën is, en dat de vruchtbaarheid van een idee voor onderzoek zich niet laat voorspellen zonder het uit te werken.

Professioneler tools

De aanbevelingen van het rapport liggen op drie terreinen: bevorderen van de interoperabiliteit – het vermogen van tools om met elkaar samen te werken –, het creëren van infrastructuur en het ontwikkelen van tools zichtbaarder waarderen. Deaanbevelingen voor de infrastructuur lijken het verste te gaan: er moet gestreefd worden naar een gedeelde



Informatiewetenschapper John Unsworth: 'sense of déjà-vu'

omgeving voor ontwikkelaars, waar verschillende ontwikkelhulpmiddelen in samenhang worden aangeboden. Ontwikkelaars in kleinere instellingen zouden zo toegang krijgen tot een professionele ontwikkelomgeving. Een gedeeld en professioneel ontwikkelproces leidt wellicht tot professioneler tools en maakt het uitwisselen van kennis makkelijker. Een andere nieuw onderdeel van de infrastructuur zou een site zijn die repositoryfuncties biedt voor tools, een tijdschrift en reviews, en die bijdraagt aan de zichtbaarheid van tools voor gebruikers en collega-ontwikkelaars. Essentieel is dat de tools gereviewed zouden moeten worden, net als de tijdschriftartikelen, zodat de bezoeker kan vertrouwen op de kwaliteit van het gebodene. Tegelijkertijd biedt een review van tools ook de mogelijkheid om ontwikkelwerk academisch te waarderen. Een andere aanbeveling is dan ook afzonderlijke discussies te wijden aan mechanismen voor peer review van wetenschappelijke gereedschappen.

Betere zichtbaarheid

In veel opzichten herinneren de aanbevelingen aan de doelstellingen van het Nederlandse Alfalab (zie *e-data&research* december 2008). De interoperabiliteit van tools en data is daarvan een belangrijk onderdeel, net als een betere zichtbaarheid van tools (Alfalab's 'disseminarium'). Laten we hopen dat dankzij de resultaten van deze workshop en het werk aan Alfalab en gerelateerde projecten Unsworth's gevoel van *déjà-vu* spoedig tot het verleden zal behoren.

Wat kost duurzaam archiveren?

Vervolg op digitalisering Volkstellingen



Palaiologk presenteert haar conclusies aan de medewerkers van DANS. Links DANS-directeur Peter Doorn.

Econome Anna Palaiologk uit Griekenland heeft de afgelopen maanden bij Data Archiving and Networked Services (DANS) een onderzoek uitgevoerd naar de kosten van duurzame toegankelijkheid van onderzoekgegevens. Zoals ook al bleek uit het onlangs in concept verschenen rapport *Toekomst voor ons digitaal geheugen* (zie p. 1 en 4) is daarover momenteel nog te wei-

nig bekend om financiers en data-archieven in staat te stellen gedegen afwegingen te maken tussen de kosten en de baten.

Palaiologk bestudeerde de beschikbare literatuur, met name de samenvatting van onderzoek tot nu toe in het Brits-Amerikaanse rapport *Sustaining the digital investment*, en voerde uitgebreide interviews met DANS-medewerkers om

erachter te komen welke werkzaamheden komen kijken bij het archiveren van onderzoekgegevens. Dit zou de sector in staat moeten stellen om het *activity-based costing model* toe te passen dat bekend is uit andere sectoren. Maar daarvoor is nog het nodige voorbereidende werk nodig, concludeerde Palaiologk. Zo ontbreken ervaringscijfers op basis waarvan begrotingen gemaakt kunnen worden omdat duurzame archivering een erg jong vakgebied is. Dit gemis zal ertoe leiden dat begrotingen de eerstkomende tijd een zeer grote onzekerheidsfactor hebben.

Om die factor te verkleinen is het nodig om de juiste vragen te gaan stellen, en juist daar heeft de Griekse zich in deze fase op geconcentreerd. Welke werkzaamheden reken je toe aan digitale archivering? Wat zijn de 'producten' die digitale archivering oplevert? Wat zijn de variabelen waar men mee te maken krijgt? De antwoorden op deze vragen kunnen verschillen per organisatie of zelfs per afdeling; ze zijn afhankelijk van de strategie die een organisatie vaststelt.

DANS is door het onderzoek van Palaiologk aan het denken gezet in de goede richting, concludeerde DANS-directeur Peter Doorn bij het afscheid van de econome, maar het laatste woord over wat duurzame archivering kost is nog lang niet gezegd. (IA)

http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf



Het Centraal Bureau voor de Statistiek (CBS) en Data Archiving and Networked Services (DANS) bereiden een vervolg voor op het grote Volkstellingenproject dat werd uitgevoerd tussen 1997 en 2006. In enkele nieuwe projecten worden nu gegevens gedigitaliseerd van een aantal bedrijfstellingen, volks-, beroeps- en woningtellingen, deels op wijk- en buurtniveau. In totaal gaat het om circa 350 duizend pagina's.

Ook wordt de website www.volks-tellingen.nl, die maandelijks zo'n 3.500 bezoekers trekt, gerenoveerd. Onderzocht wordt of die kan worden uitgebouwd tot een 'portal historische tellingen' met toegang tot andere websites met historisch statistisch materiaal, zoals de website van het CBS, de Hub for Aggregated Social History (HASH), NL-GIS en de Historische Databank Nederlandse Gemeenten (HDNG).

In het oorspronkelijke Volkstellingenproject werden de publicaties

van de Nederlandse volkstellingen van 1795 tot en met 1971 gedigitaliseerd. Daarbij kwamen rond 42 duizend pagina's aan data beschikbaar voor nieuw onderzoek, niet alleen als afbeelding maar ook in de vorm van elektronisch verwerkbaar gegevens. Het beoogde vervolg gaat niet alleen om het digitaliseren van nog meer gegevens, maar ook om voorzieningen die nodig zijn om de grotere hoeveelheid materiaal beter toegankelijk en bruikbaar te maken.

In een *expert meeting* medio juni zijn daarvoor ervaringen en suggesties uitgewisseld tussen een twintigtal onderzoekers die in de afgelopen jaren bij het Volkstellingenproject betrokken waren. De deelnemers waren afkomstig van het CBS, DANS en verschillende universitaire instituten. In korte presentaties gevolgd door discussie werden nieuwe mogelijkheden voor vormgeving en presentatie van gedigitaliseerde historische tellingen verkend. (Jan Jonker)

Sinds kort beschikbaar

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

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

Centraal Bureau voor de Statistiek	Periode
Algemeen Bedrijfsregister (ABR-GPSKADER)	2006-2008
Bedrijfsfraude (BFS)	2005-2007
Bijstandsfraude en Bijstandsdebiteurenstatistiek (BFS-BDS)	2008
Bijstandsuitkeringenstatistiek (BUS-TRANS)	2008
Doodsoorzakenstatistiek (DO)	2006-2007
Productiestatistiek Industrie	1993-1999
Statistiek re-integratie gemeenten (SRG)	2008
Beschikbaar via DANS EASY	
<i>Archeologie</i>	
Zandmaas – Stuwpannd Lith zomerbedverdieping km 176-181 (Projectteam Archeologie Maaswerken, Alterra)	2004
Schipluiden – Harnaspolder, campagne 03HP (Archeologisch projectbureau ADC)	2006
Mierlo – 't Karrewiel (Archeologisch projectbureau ADC)	2007
Plangebied Bedrijventerrein Avelingen te Gorinchem (Projectbureau RAAP)	2008
<i>Geschiedenis</i>	
Dutch Atlantic connections: the circulation of people, goods and ideas in the Atlantic world (Universiteit Leiden)	1680-1795
Erfgoed van de Oorlog; interviews afgenomen in het kader van het project Getuigen Verhalen:	1940-1945
<ul style="list-style-type: none"> • Anna's stille strijd: doven in de oorlog (AVA filmproducties) • Het vergeten bombardement, Rotterdam 31 maart 1943 (Stichting Mondelinge Geschiedenisprojecten) • Retourtje Sachsenhausen (Nationaal Monument Kamp Vught) • Negen vrouwen ontsnapt uit dodenmars 1945 (Ange Wieberdink Producties; restricted access) • Vluchten uit Overloon op last van de bezetter (Liberty Park Overloon; restricted access) • Boerinnen en boerendochters in de Tweede Wereldoorlog (Meertens Instituut; restricted access) • NSB vrouwen in de oorlog (IIAV; restricted access) 	
<i>Sociale wetenschappen</i>	
Kwaliteitskaart Voortgezet Onderwijs (Inspectie van het Onderwijs)	2009
Cohortonderzoek Onderwijsloopbanen van 5-18 jaar COOL 5-18 (ITS Radboud Universiteit, SCO Kohnstamm Instituut)	2007
Bont en blauw, een onderzoek naar de behandeling van geweldszaken tegen politieambtenaren (Ministerie van Justitie, WODC)	2008
Sociale uitsluiting (Sociaal en Cultureel Planbureau)	2003
Centerdata millennium survey (Centerdata, Universiteit Tilburg)	1999

Symposium bij 45 jaar e-archief

Begin december organiseert DANS (Data Archiving and Networked Services) het symposium 'Gedreven door Data' over nieuwe kansen voor onderzoek in de alfa- en gammawetenschappen door de groeiende beschikbaarheid van elektronisch gearchiveerde data.

Aanleiding is het samenvallen van een aantal lustrums in 2009. Zo is het vijftienjarige jubileum van het eerste elektronisch data-archief in Nederland werd opgericht: het Steinmetzarchief voor de sociale wetenschappen. Daarnaast is

het twintig jaar geleden dat het Nederlands Historisch Data Archief in het leven werd geroepen, terwijl vijftien jaar geleden NWO en het CBS besloten om het Wetenschappelijk Statistisch Agentschap op te richten, dat zou gaan zorgen voor de verspreiding van CBS-data voor wetenschappelijke doelen. Enten slotte is 2009 het vijfde bestaansjaar van het EDNA, het e-depot van de Nederlandse archeologie.

Op het symposium 'Gedreven door Data' wordt aandacht besteed aan de ontwikkeling van de rol van

elektronische data tot nu toe, maar vooral aan de nieuwe perspectieven die zich aandienen voor het onderzoek in de alfa- en gammawetenschappen door de steeds grotere beschikbaarheid van data. Er wordt gerekend op sprekers uit uiteenlopende vakgebieden als demografie, politicologie, economische geschiedenis, ornithologie, methodologie en criminologie. De toegang is gratis, informatie kan worden gevraagd bij info@dans.knaw.nl. (MdG)

COLOFON

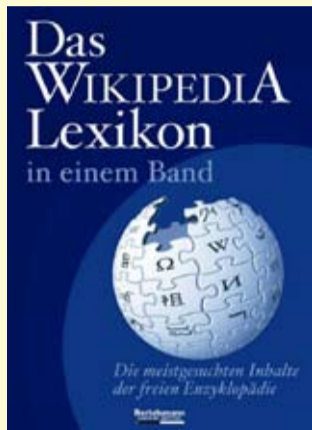
e-data@research is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van DANS, het Huygensinstituut, het Internationaal Instituut voor Sociale Geschiedenis, het Centraal Bureau voor de Statistiek, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen. Oplage: 8800. *e-data@research* is online te raadplegen op www.edata.nl

Uitgever: Stichting Uitgeverij *e-data@research*, Postbus 93067, 2509 AB Den Haag
Redactieadres: Postbus 93067, 2509 AB Den Haag; t (070)3494450 f (070)3494451 e edata@dans.knaw.nl
Redactie: Inge Angevaere, Ronald van der Bie, Peter Boot, Martijn de Groot (hoofd/eindredacteur), Thijs Hermsen, Jetske van der Schaaf
Aan dit nummer werkten mee: Peter Doorn, Jan Jonker, Marjolijn

Kruip, Hans van Maanen, Mariken Teeuwen, Douwe Zeldenrust.
Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf
Vormgeving en opmaak: Ellen Bouma
Productie: Uitgeverij Aksant, Amsterdam
Druk: Thieme Almere
ISSN: 1872-0374

Printpedia?

Een beetje lacherig en ongeloofig was de algemene reactie toen het Duitse mediabedrijf Bertelsmann in 2008 een papieren versie van de Wikipedia aankondigde. Inmiddels is het boek verschenen. Al zijn er signalen dat het bij deze eerste uitgave blijft, het fenomeen is bijzonder genoeg om nader in ogenschouw te nemen.



Eerst wat feiten. Wat staat erin? Het boek bevat twintigduizend artikelen op 992 bladzijden; dat wil zeggen dat er ongeveer twintig artikelen op een pagina staan. De inhoud van de Duitstalige Wikipedia, die aan de tekst ten grondslag ligt, is dus in twee opzichten radicaal bekort: het aantal artikelen online (nu 940.000) is met ruim 97% teruggebracht; het aantal woorden per artikel is nog eens gereduceerd met meer dan 95 procent, zo leert een kleine steekproef. De koper koopt dus niet de encyclopedie, en zo wordt het boek dan ook niet verkocht. De titel luidt *Das Wikipedia Lexikon in einem Band. Ausgabe 2009*. Het bekorten van de artikelen is volgens een eenvoudige procedure gebeurd. Wikipedia-artikelen beginnen met een inleiding, vaak niet meer dan één of enkele zinnen. De uitgever heeft het

Wikipedia Lexikon samengesteld op basis van deze inleidingen. De geselecteerde ingangen zijn die welke online het vaakst worden opgevraagd. Het voorwoord van het *Lexikon* gebruikt dat als verkoopargument: het boek bevat niet alleen de begrippen uit een overgeleverde canon, maar ook films, popsterren en andere begrippen uit de cultuur van alledag.

Mag dat zomaar, het werk van vrijwilligers overnemen en daar winst mee maken? Het antwoord is duidelijk: ja, dat mag. Wikipedia-teksten kunnen door iedereen worden gebruikt, ook in aangepaste vorm, ook voor commerciële doeleinden, mits de originele auteur de credits krijgt voor zijn werk en het aangepaste product zonder beperkende copyrightbepalingen aan derden ter beschikking wordt gesteld. Deze voorwaarden leiden tot wat zonder twijfel de meest curieuze dertig pagina's uit het boek zijn: de pagina's waar in een zeer klein lettertype de in totaal ruim



99000 auteurs worden genoemd die aan de Wikipedia artikelen hebben bijgedragen. Wellicht in een poging om eventuele critici de wind uit de zeilen te nemen, kondigt de uitgever in de inleiding aan de Duitse Wikipedia financieel te zullen ondersteunen. Want sommige van de auteurs voelen zich wel degelijk gebruikt, zo

blijkt uit levendig webdebat.

Het *Lexikon* is dus legaal. Maar is het ook handig? Het lijkt onwaarschijnlijk dat iemand met toegang tot het Internet het boek zal raadplegen (want: niet doorzoekbaar, geen aanklikbare verwijzingen, artikel wellicht gedateerd). Misschien als de computer in een andere kamer staat, of er te lang over doet om op te starten? Het *Lexikon* zou dan concurreren met het netbook (kleine laptop, vooral voor huiskamergebruik). Een reden om het *Lexikon* aan te schaffen kan ook zijn dat onbewust een boek toch iets betrouwbaarder lijkt dan een website.

Heeft de redactie een betrouwbaarder boek willen maken dan de Wikipedia? Men geeft aan 'alleen zeer behoedzaam' de teksten voor het boek te hebben aangepast, maar geeft daarvoor tamelijk wijde criteria (bijvoorbeeld: belang van het onderwerp blijkt niet uit de tekst). Er wordt niet aangegeven in welke artikelen wijzigingen zijn aangebracht. Voor Wikipedia berust betrouwbaarheid juist mede op het fundamentele uitgangspunt dat precies kan worden nagegaan wie welke wijzigingen aanbrengt. De uitgever lijkt hier meer op zijn gezag te vertrouwen. Een uiting van dit autoriteitsdenken is ook dat men de aangebrachte veranderingen niet zelf in de Wikipedia-artikelen heeft willen doorvoeren maar ze 'ter controle aan de Wikipedia-gemeenschap ter beschikking gesteld'. Uit het woord 'controle' blijkt de achtergrond van de traditionele maker van naslagwerken.

Peter Boot

Column

Hans van Maanen

Enquetekeurmerk

Ook voor nieuws, en zeker voor nieuws over opiniepeilingen, geldt de wet van Gresham: slecht werk verdrijft goed werk. Dat ligt natuurlijk aan de journalisten, die goed werk niet van slecht werk kunnen onderscheiden, geen tijd hebben om de merites van een onderzoek te bepalen, en liever scoren dan horen en wederhoren. Of het ligt aan de opiniepeilers, die in hun streven naar media-aandacht niet altijd alle zwakke plekken en losse schroeven in hun onderzoek benadrukken.



BARBARA THIEL

Over deze kwestie kan men symposiaal lang discussiëren, en dat gebeurt dan ook regelmatig, maar mijn vraag is: kunnen bonafide onderzoekers wetenschapsjournalisten niet een beetje helpen het onderscheid tussen goed en slecht surveyonderzoek te maken? Zodat de onzin uit de krant blijft, en het goed onderbouwde nieuws een betere kans krijgt?

*Wat is het belangrijkste verschil tussen wetenschap en opinieonderzoek? Als het over wetenschap gaat, weet elke journalist – en een beetje ervaren lezer – wel dat een verhaal in *Nature* een andere status heeft dan een verhaal in de *Acta Zoologica Bulgarica*. Het ene blad is het andere niet, en dat werkt als een soort keurmerk. Een dergelijk filter heeft het opinieonderzoek in Nederland niet. Er is al vaak over keurmerken gepraat, maar nog steeds mag rijp en groen tot advertentie/persbericht worden voorgekookt en aan de kranten worden opgediend.*

Hier zie ik een mooie rol voor het nieuwe Nederlandstalig Platform voor Survey-Onderzoek. Als de leden van het platform nu eens voortaan in elk persbericht in de derde zin iets zetten als: 'Dit onderzoek is, naar de maatstaven van de branche-vereniging, redelijk betrouwbaar.' Of, als het beter is uitgevoerd, 'betrouwbaar'. Of, als er bijna niets op is af te dingen: 'zeer betrouwbaar'. In een voetnoot zou dan kunnen worden uitgelegd wat de criteria precies inhielden: was het een online-panelonderzoek of een echt deur-tot-deuronderzoek, is het betrouwbaarheidsinterval redelijk of smal, is het representatief, valide, dat soort dingen.

De meeste journalisten zetten zo'n persbericht toch zo in de krant, dus dan gaat het meteen goed, en van degenen die wat kritischer zijn, wordt eerst de nieuwsgierigheid geprikkeld en vervolgens nemen zij het waarschijnlijk als handzame informatie ook gewoon over. Het is een soort keurmerk, door de branche-organisatie zelf in het bericht gezet.

Het aardige is natuurlijk, dat onderzoek dat niet aan die minimumeis van 'redelijk betrouwbaar' voldoet – we weten allemaal over welke bureaus we het hebben – een dergelijke frase niet mee krijgt, en dus na verloop van tijd vanzelf door de mand valt. 'Hé', denkt de journalist, en de lezer, 'er staat niet bij of het betrouwbaar is, dan zal het wel niet deugen.'

Het lijkt mij, als wetenschapsjournalist, wel een mooi idee. De uitwerking ervan – wat de criteria dan worden, en wie dat gaat bepalen en controleren – laat ik uiteraard graag aan het platform over.

Hans van Maanen is freelance wetenschapsjournalist.

Gelezen

Madeleine de Cock Buning, Allard Ringnalda, Tina van der Linden (CIER): *De juridische status van ruwe data; een wegwijzer voor de onderzoekspraktijk*; Utrecht, SURFfoundation, juli 2009

Rapport om duidelijkheid te verschaffen over de juridische bescherming van onderzoeksdata, voor de onderzoeker die wil weten wat hij mag doen met de data van anderen. De mogelijkheden van ict en internet maken de toegankelijkheid van onderzoekresultaten en -data steeds groter en opener. De mogelijkheden groeien om aanvullend aan tekst, publicaties te verrijken met o.a. de onderzoeksdata. Door deze toegankelijk te maken, wordt het makkelijker om de resultaten van onderzoek te verifiëren, te reproduceren en te hergebruiken. Maar met hergebruik van de (ruwe) onderzoeksdata is het belangrijk te weten wat de juridische status is van het materiaal. Soms is toestemming van de maker vereist maar sommige handelingen kunnen ook zonder toestemming worden verricht. SURFdirect wil inzicht verschaffen in de juridische bescherming van onderzoeksdata, voor de onderzoeker die wil weten wat hij mag doen met de data van anderen. Met deze informatie kan de onderzoeker ook bepa-

len hoe wenselijk bescherming van de eigen onderzoeksdata is. www.surfoundation.nl/nl/publicaties/Pages/Default.aspx



Inge Angevaere (red.): *Toekomst voor ons digitaal geheugen, duurzame toegang tot informatie in Nederland*; Den Haag, Nationale Coalitie Digitale Duurzaamheid (NCDD), juli 2009

Het is voor het eerst dat in Nederland de risico's die digitale bestanden lopen én de maatregelen die nodig zijn om die risico's te beheersen over de

hele breedte van de publieke sector (wetenschap, overheid, cultureel erfgoed) in kaart zijn gebracht. Het rapport besteedt vooral aandacht aan de organisatorische en beleidsmatige maatregelen die nodig zijn om te voorkomen dat ons digitaal geheugen onbereikbaar wordt omdat nieuwe computers de oude bestanden niet meer kunnen lezen. Het nu gepubliceerde document is een interimrapportage die besproken zal worden tijdens de nationale conferentie die de NCDD op 18 september 2009 organiseert. Na die consultatieronde zal de NCDD haar strategische plan vaststellen voor het realiseren van een duurzame infrastructuur voor toegang tot digitale informatie in Nederland. Het rapport werd geschreven door een projectteam bestaande uit René van Horik (Data Archiving and Networked Services, DANS: wetenschap), Annelies van Nispen (Digitaal Erfgoed Nederland, DEN: erfgoed), Ingrid Dillo (Koninklijke Bibliotheek, KB: projectmanagement), Inge Angevaere (NCDD: projectmanagement), Petra Helwig (Nationaal Archief, NA: overheid/archieven).

www.ncdd.nl/activiteiten-natverkenning.php

Paul Brous: *Managing metadata for geographic information*; Delft, TU Delft, GIMA Masters Thesis, 2009

Our knowledge of the real forces and mechanisms behind metadata production is still very limited. The reason for this difficulty might lie in the complex, dynamic and multifaceted interaction of organisations, the metadata producers themselves, the available technology, the spatial data and the metadata itself during the metadata production process. In order to be able to produce appropriate metadata we must have a proper understanding of what the main barriers and issues surrounding metadata production are and how these is-

sues affect the appropriateness of the metadata. This research describes a new approach to understanding the issues preventing appropriate metadata production by investigating them through the dimensions of a Geographic Information System (GIS) as described by Traub (1998) and comparing them to the categories of data quality as described by Strong et al. (1997).

www.msc-gima.nl/index.php/news