

Pagina 3 • Open Access • Op de Frankfurter Buchmesse werd de digitale bibliotheek OAPEN gepresenteerd, die al ruim 700 wetenschappelijke boeken vrij beschikbaar heeft.

Pagina 4 • Registerdata • Het gebruik van registerdata in sociaalwetenschappelijk onderzoek biedt zo veel nieuwe mogelijkheden dat het de moeite waard is de nog bestaande hindernissen te overwinnen.



Pagina 5 • Neelie Kroes • De eurocommissaris voor de Digitale Agenda toont zich in een interview enthousiast over het voorstel van experts voor een Europese digitale infrastructuur. Maar ze vindt het ook belangrijk om 'realistisch en praktisch te werk te gaan om de klus te klaren'.

Pagina 6 • Gouden Eeuw • Onderzoekproject ECARTICO moet inzicht opleveren in het netwerk van kunstschilders en handelaars in het Amsterdam van de Gouden Eeuw.

Pagina 8 • UNESCO • De VN-organisatie voor onderwijs, wetenschap en cultuur gaat voor duurzame toegang tot informatie.

EN VERDER

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

Holocaustproject brengt archieven samen

Twintig partners uit dertien Europese landen gaan samenwerken om een database tot stand te brengen die materiaal over de Holocaust samenbrengt dat nu verspreid ligt over archieven in heel Europa. Dat wordt mogelijk gemaakt door de Europese Commissie, die zeven miljoen euro beschikbaar stelde. De leiding van de European Holocaust Research Infrastructure (EHRI) zoals het project heet, is in handen van het Nederlands Instituut voor Oorlogsdocumentatie (NIOD). De officiële lancering vond plaats op 16 november in Brussel. 'Door deze gezamenlijke inspanning kunnen zowel onderzoekers als het grote publiek profiteren van en bijdragen aan meer kennis en bewustzijn over de Holocaust,' zei EHRI directeur Conny Kristel bij die gelegenheid.

www.ehri-project.eu

Eigen computer populairste plek voor data

De eigen computer is nog steeds de populairste plek om data uit wetenschappelijk onderzoek op te slaan. Ruim zeventig procent van de onderzoekers in Nederland is gewend de eigen computer voor dit doel als archief te gebruiken. Dat blijkt uit de voorlopige resultaten van een enquête van DANS onder onderzoekers in alle disciplines.

Tegenover het hoge aandeel 'thuisbewaring' staat een kleine minderheid van 18% die een externe opslagvoorziening buiten de eigen afdeling of het eigen instituut noemt als plek om de onderzoekgegevens op te slaan. Externe harde schijven of backupmedia worden door 46% genoemd, terwijl 63% zegt gebruik te maken van een netwerkschijf van eigen afdeling of instituut.

Data-instituut DANS (Data Archiving and Networked Services) deed het onderzoek als onderdeel van een verkenning om de uitbreiding van zijn activiteiten naar de bètawetenschappen voor te bereiden. De webenquête werd ingevuld door 556 onderzoekers uit een steekproef uit alle onderzoekrichtingen in de Nederlandse wetenschap.

Zoals verwacht mocht worden werken de ondervraagden met veel verschillende soorten data. Onder beoefenaars van de bètawetenschappen zijn spreadsheets (69%) en instrumentdata (61%) het meest populair, in de alfa- en gammawetenschappen wordt het meest gewerkt met statistische data (68%) en spreadsheets (53%).

Experts willen open Europese infrastructuur

Er moet een internationale *Collaborative Data Infrastructure* komen waaraan zowel bedrijven als instituten, universiteiten, overheden en individuen kunnen deelnemen. Dat zegt een groep experts in een advies aan de Europese Commissie.

De groep, die door commissaris Neelie Kroes van de Digitale Agenda aan het werk was gezet, pleit daarnaast voor het ontwikkelen van manieren om de waarde van data te bepalen en van aparte universitaire opleidingen voor datawetenschappers. Ook wil ze maatregelen om het energieverbruik van de snel groeiende e-sector het hoofd te bieden. De Europese Commissie moet verder het initiatief nemen om een internationale groep samen te stellen die zich buigt over de *global governance* van een wetenschappelijke e-infrastructuur, schrijven de deskundigen

http://cordis.europa.eu/fp7/ict/e-infrastructure/high-level-group_en.html

Een grote meerderheid van de onderzoekers blijkt – meestal naast zelf verzamelde gegevens – gebruik te maken van data die door anderen bijeen zijn gebracht. Slechts dertig procent doet zijn of haar onderzoek alleen op basis van eigen data, maar in deze groep geven weer zes van de tien te kennen wel eens behoefte te hebben aan data uit andere dan eigen bron. Op de vraag of er voldoende mogelijkheden zijn om gebruik te maken van gegevens van anderen antwoordt 42% van de ondervraagden bevestigend, terwijl een kwart die mogelijkheden onvoldoende vindt en nog eens een kwart het niet weet. Al zijn er volgens vier van de tien onderzoekers voldoende mogelijkheden, dat neemt niet weg dat 58% van alle ondervraagden hindernissen en bezwaren meldt bij pogingen om er gebruik van te maken. 'Geen of beperkte toegang' wordt door bijna driekwart van hen genoemd en scoort daarmee het hoogst, gevolgd door problemen met de kwaliteit van de gewenste data (de helft), privacy-verwegingen (41%), ontbrekende metadata en onvoldoende controle op de betrouw-

baarheid (beide 36%).

De definitieve rapportage over de enquête wordt, als onderdeel van een

publicatie over de gehele verkenning, in de komende maanden verwacht. (MdG)

Europa en de cultuur



HELLA HOLLANDER

Achter de vlag van de Europese Unie voor het universiteitsgebouw 'La Scuola Normale Superiore' gaat het wapenschild schuil van de familie De' Medici, geflankeerd door allegorische figuren. Sinds 1563 siert deze timpaan de hoofdingang van het voormalige Paleis van de Ridders in Pisa, in welke stad bijna vierhonderdvijftig jaar later de deelnemers van het Europees erfgoedproject CARARE elkaar zouden treffen. Meer over deze bijeenkomst op pagina 7. Europa is goed vertegenwoordigd in deze *e-data&research*. Op deze pagina treft u een bericht over de rol van de EU bij het maximaal benutten van wetenschappelijke data, en op pagina 5 een interview daarover met Eurocommissaris Neelie Kroes. (Hella Hollander/MdG)

CentERdata nieuwe deelnemer in e-data&research

Met ingang van deze aflevering van *e-data&research* is het Tilburgse instituut voor dataverzameling en onderzoek CentERdata als stakeholder aan het blad verbonden. CentERdata is gespecialiseerd in het verzamelen, analyseren en ter beschikking stellen van (panel)data. Het instituut is onder meer de bedenker en beheerder van het grote LISS-panel (Longitudinal Internet Studies for the Social sciences), waaraan vijfduizend Nederlandse huishoudens deelnemen.

Door zijn toetreding sluit CentERdata zich aan bij het rijtje dat tot nu toe bestond uit het Centraal Bureau voor de Statistiek, DANS, het Huygens Instituut, 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.

Zie ook pagina 5.

AGENDA

2 december
Leiden – LUMC

Closing the Deal – Conferentie Groot-schalige Onderzoeksfaciliteiten
Dit initiatief van de Taskforce Stimulering Groot-schalige Onderzoeksfaciliteiten biedt ontmoetingsmogelijkheden met verschillende organisaties, presenteert succesverhalen en laat zien wat er gedaan moet worden om een project binnen de kaders van groot-schalige onderzoeksfaciliteiten te laten slagen.
www.closingthedeal.nl/

6-8 december

Chicago – Chicago Mart Plaza
6th International Digital Curation Conference – Participation and Practice – Growing the Curation Community through the Data Decade
Digital curation manages, maintains, preserves, and adds value to digital data throughout the lifecycle. This year's Conference (IDCC) will be presented jointly by the Digital Curation Centre, UK and the Graduate School of Library and Information Science, the University of Illinois at Urbana-Champaign, in partnership with the Coalition for Networked Information (CNI). www.dcc.ac.uk/events/conferences/6th-international-digital-curation-conference

7 en 8 december

Rotterdam – De Doelen
Digitaal Erfgoed Conferentie – 'Lang Leve Lerer!'
Centraal staan de digitale ontwikkelingen rond erfgoed en educatie, van onderwijs tot publieksbegeleiding en van kennisoverdracht tot 'lifelong learning'.
www.deconferentie2010.nl/

8 december

Utrecht
DDI (Data Documentation Initiative) User Group Meeting – The Basis of Managing the Data Life Cycle
Een consortium van de Library and IT Services van Universiteit Tilburg, DANS, SURF en CentERdata slaat de handen ineen met GESIS – Leibniz Instituut voor Sociale Wetenschappen en het International Data Service Center (IDSC) van het Instituut voor de Studie naar Arbeid (IZA) in Bonn om de toepassing en de verspreiding van DDI te bevorderen. Zij organiseren gezamenlijk deze tweede Europese DDI User Group Meeting (EDDI).
www.dans.knaw.nl/content/symposia

10 december

Den Haag – Koninklijke Bibliotheek
Werkconferentie Biografisch Portaal
Conferentie georganiseerd door het Instituut voor Nederlandse Geschiedenis. Het Biografisch Portaal van Nederland verzamelt zoveel mogelijk biografisch materiaal dat betrekking heeft op de geschiedenis van Nederland. Technisch en inhoudelijk deskundigen, wetenschappers en andere belanghebbenden kunnen hier met elkaar van gedachten wisselen over de perspectieven van het gebruik van het Portaal als onderzoeksinstrument. www.inghist.nl/Nieuws/Actueel/expertmeeting

15 december

Den Haag – Koninklijke Bibliotheek
Research Foundations for Understanding Books and Reading in the Digital Age; Textual Methodologies & Exemplars
Digital technology is fundamentally altering the way we relate to writing, reading, and the human record itself, thus creating a gap between core social and cultural practices that depend on stable reading and writing environments, and the new kinds of digital artefacts that must sustain those practices now and into the future. This one-day gathering explores research foundations pertinent to understanding those new practices and emerging media. www.ucl.ac.uk/dh-blog/2010/08/23/books-and-reading-in-the-digital-age/

16 en 17 december

Den Haag – Koninklijke Bibliotheek
Texts and Literacy in the Digital Age
Jointly organised by the Leiden University (Book and Digital Media Studies), LIBER, Tiele-Stichting and the National Library of the Netherlands, this conference will bring prominent and inspiring speakers together in order to address the changes in scholarly communication within the humanities and social sciences, and more importantly, their meaning to authors, intermediaries and users.

De kosten van duurzaam beheer

De fase waarin gegevens in een archief worden opgenomen lijkt het grootste deel van de kosten van duurzaam digitaal beheer op te slokken. De grootste onzekerheid rond de totale omvang van die kosten zit in de vraag hoe vaak de data in de toekomst gemigreerd zullen moeten worden naar nieuwe bestandsformaten. Dat bleek op een studiebijeenkomst in Den Haag.

Duurzaam beheer van digitale gegevens opdat ze hergebruikt kunnen worden in de toekomst kost geld. Maar op de vraag van financiers *hoeveel* geld dat beheer gaat kosten, hebben digitale archieven nog geen antwoord. Ook is het nog zoeken naar beleidsbeslissingen die het kostenniveau kunnen beïnvloeden.

De Nationale Coalitie Digitale Duurzaamheid (NCDD), Digitaal Erfgoed Nederland (DEN) en de Koninklijke Bibliotheek organiseerden daarom op 16 september een internationale bijeenkomst voor deelnemers aan alle bekende onderzoeksprojecten over de kosten van duurzame toegankelijkheid. Doel: kennis uitwisselen en onderzoeken of het mogelijk is om tot één internationaal kostenmodel te komen. Aanwezig waren het Engelse KRDS-project (Keeping Research Data Safe), het Deense CMDP-project (Costing Models for Digital Preservation), de DANS Cost Model Methodology (Nederland), het Testbedmodel van het Nederlandse Nationaal Archief en het LIFE3-project (Life-cycle Information for E-Literature) uit Engeland.



INGE ANGEVAARE

Het was de eerste keer dat al deze projecten om één tafel zaten, en daarom werd vooraleerst veel kennis uitgewisseld. Er bleken flink wat overeenkomsten te zijn tussen de modellen, maar op belangrijke punten waren er even duidelijke verschillen. Het blijkt lang niet gemakkelijk om alle relevante factoren in een schema onder te brengen dat zowel handzaam als informatief is. Voorlopig lijkt het erop dat de *ingest*-fase (opname in het archief) de duurste fase is, die zo'n vijftig procent van de kosten bepaalt. Dan immers moeten de databestanden van wetenschappers geschikt gemaakt

worden voor opname in het archief. De grote vraag die niemand nog kon beantwoorden is hoe vaak alle gegevens in een archief *gemigreerd* zullen moeten worden naar een nieuw bestandsformaat. David Rosenthal, van het LOCKSS-archief, poneerde de stelling dat bestandsformaten veel robuuster blijken te zijn dan we eerder dachten, en dat we daarom voorlopig maar niet moeten migreren. De meeste gesprekspartners wilden dat risico voorlopig niet nemen, en normaliseren de bestanden bij ontvangst in het archief. (IA)

<http://tinyurl.com/376lry4>

Internet Research 11.0

Duurzaamheid en participatie waren de centrale thema's van de elfde conferentie van de Association of Internet Research (AoIR) in oktober Göteborg, maar uiteindelijk was er vooral veel aandacht voor Facebook en Twitter.

Op de conferentie onder de noemer 'IR 11.0: Sustainability, Participation, Practice' kwamen deze enigszins losjes samenhangende thema's op diverse momenten en manieren terug, maar ze drukten uiteindelijk geen zware stempel op de inhoud.

In hun poging een duurzame conferentie te organiseren gaven de organisatoren zelf ook aan dat het aantal door de deelnemers afgelegde vliegekilometers elke ambitie op dat vlak enigszins in de schaduw stelde. Toch waren er linnen tassen, programma's die alleen op aanvraag werden

geprint, ecologisch eten en drinken, en in plaats van plastic waterflessen kannen met water. Ook ging de *keynote* van milieu-econoom Peter Arnfalk uitgebreid in op de impact van ict op het milieu. Zo bleek onder meer dat de CO₂-uitstoot van alle e-mails samen (ongeveer tachtig miljard per jaar) zeven tot twintig keer hoger is dan die van het traditionele postverkeer.

In de overige bijdragen aan de conferentie bleek duurzaamheid uiteindelijk een redelijk marginaal thema. Tijdens deze conferentie was er na-

melijk, naast *games* en *Second Life*, vooral veel aandacht voor Web 2.0 en sociale netwerksites. Met name Facebook werd uitgebreid besproken, bejubeld en, in sommige gevallen, kritisch bekeken. Daarnaast was er veel aandacht voor Twitter. Een groot deel van de ruim 250 deelnemers bleek zelf actief op dat platform, ook tijdens presentaties, waarbij *keynote speaker* Nancy Baym als twitteraar (intussen meer dan 10.000 tweets) waarschijnlijk de kroon spant. Baym's verhaal ging echter niet over Twitter, maar over de relatie tussen de muziekindustrie en het internet. Haar verhaal was niet alleen erg goed gepresenteerd, maar ook juist hier in Zweden, het land van *Pirate Bay* en *Spotify*, erg op z'n plek.

Vanuit Nederland was alleen de Virtual Knowledge Studio op de conferentie vertegenwoordigd, maar hun inbreng was met de organisatie van twee rondetafelgesprekken en twee paneldiscussies wel groot. (Stefan Dormans)

<http://ir11.aoir.org/>

'ICT-projecten missen vaak aansluiting op praktijk kunsthistoricus'

Resultaten van ICT-projecten in de kunstgeschiedenis blijven vaak ongebruikt omdat ze de aansluiting missen op het werk van de beoogde gebruiker. Dat was een terugkerende constatering op de conferentie 'Networked Humanities: Art History in the Web' van 9 t/m 14 oktober in het Italiaanse Acquafredda di Maratea. Het initiatief voor de bijeenkomst kwam van de European Science Foundation en de Cooperation in Science and Technology (ESF-COST).

ESF-COST-conferenties zijn relatief kleinschalig, met veel ruimte voor discussie en informele contacten. De dagen in Acquafredda waren dan ook lang. Er waren sessies gewijd aan wetenschappelijk publiceren, onderzoeksondersteuning, *social tagging*, en de erfgoedsector als beheerder van het onderzoeksmateriaal. Maar ook was er aandacht voor nieuwe kunst- en cultuurvormen als internet art, interactieve (digitale) kunst en *games* als objecten van wetenschappelijke studie.

De klassieke discipline is aan ten minste drie excentrische krachten onderhevig, bleek uit de discussies die geregeld raakten aan het structureren van onderzoeksgegevens (*linked data*), het analyseren van grote dataverzamelingen (*data mining*) en de kennis die nodig is om participatie van professionals en leken te werkstelligen (*social media*).

Informatici zorgden voor deskundige inbreng inzake *linked data* en *data mining*. De technologie biedt ongekende mogelijkheden en lijkt niet belemmerend. Veel belangstelling was er voor Muruca, een *semantic digital library framework* voor de erfgoedsector. Over de vraag hoe gecompliceerde online toepassingen aansluiten op de werkwijze van de kunsthistoricus moet echter nog worden nagedacht. Het initiatief ligt bij vernieuwende projecten in de regel buiten de kunsthistorische wetenschap en het gebruik van de resultaten blijft vaak achter: annotatiefuncties die niet worden benut, onderzoeksomgevingen waar de onderzoekers bij wegblijven, *crowdsourcing* dat alleen werkt als een project net is gelanceerd. Te vaak wordt over het hoofd gezien dat de voordelen van het werken met nieuw ontwikkelde toepassingen voor de kunsthistorische gebruiker duidelijk moeten zijn. Het zou goed zijn wanneer in onderzoeksvoorstellen een paragraaf wordt gewijd aan een onderbouwde beantwoording van de vraag: waarom gaat de kunsthistoricus/kunstliefhebber hier iets mee doen? (Gerhard Jan Nauta)

www.esf.org/activities/esf-conferences/details/2010/confdetail342.html



MATHIAS KLANG www.flickr.com/photos/wrote/5101629251/

Open Access project OAPEN van start in Frankfurt

Half oktober werd op de Frankfurter Buchmesse de digitale bibliotheek OAPEN gepresenteerd: Open Access Publishing in European Networks. OAPEN wil een infrastructuur helpen ontwikkelen voor Open Access publicatie van monografieën, inclusief een online bibliotheek van OA gepubliceerde werken in de geestes- en sociale wetenschappen. Daarmee moet het project bijdragen aan de zichtbaarheid en geloofwaardigheid van het OA publicatiemodel voor auteurs, uitgevers, bibliotheken en subsidiegevers.

De Open Access (OA) beweging heeft zich in het verleden vooral geconcentreerd op wetenschappelijke artikelen, de vorm waarin het leeuwendeel van het wetenschappelijke werk wordt gepubliceerd. In de geesteswetenschappen speelt het boek echter nog steeds een belangrijke rol als medium voor de verbreiding van wetenschappelijke kennis. Een vorm van boekpublicatie waarbij vrije online toegankelijkheid van de inhoud is gewaarborgd is dus gewenst, beseffen ook wetenschappelijke uitgevers. Het was reden voor Amsterdam Uni-

versity Press (AUP) om het initiatief te nemen tot OAPEN.

Er zijn nu ruim zevenhonderd werken beschikbaar van ongeveer vijftien uitgevers. De boeken zijn als pdf te downloaden, volledig doorzoekbaar en zonder beperkingen (tekst kan worden gekopieerd en geprint). Er zijn links aanwezig naar boekhandels waar het boek kan worden aangeschaft. Desgevraagd meldt Eelco Ferwerda, coördinator OAPEN bij AUP, nog flinke groei te voorzien in zowel het aantal betrokken uitgevers als het aantal monografieën. AUP zal voortaan in principe alle monografieën opnemen in de OAPEN Library. In verband met rechten van derdengeldt dat niet voor boeken die veel illustraties bevatten; het geldt ook niet voor naslagwerken en leerboeken.

Eén van de veronderstellingen achter Open Access is dat het de zichtbaarheid van wetenschappelijke publicaties vergroot en dat dus het aantal verwijzingen ernaar zal toenemen. Onderdeel van het OAPEN initiatief is systematisch onderzoek naar deze veronderstellingen. Eer-



FRANKFURTER BUCHMESSE

De Education Hotspot op de Frankfurter Buchmesse: plaats van handeling

ste experimenten door Ronald Snijder van AUP leren dat de verkoop niet lijdt onder OA publicatie, dat de zichtbaarheid inderdaad toeneemt, maar het aantal citaten nog niet. Dit kan een gevolg zijn van de beperkte tijdsduur van het experiment. Nieuwe pilots, in Engeland en in Nederland, moeten meer helderheid verschaffen.

Onder de deelnemende uitgeverijen zijn ook een aantal instituten van de KNAW. Zo stelt het Huygens Instituut in samenwerking met Brill

een groot aantal delen van het volledige werk van Erasmus ter beschikking. 'Open Access is voor ons een principieel uitgangspunt', zegt directeur Henk Wals van het Huygens Instituut, die in Frankfurt de OAPEN library mede ten doop hield. 'Het is een business model dat even grote gevolgen zal hebben als de uitvinding van de drukpers vijfhonderd jaar geleden'. (PB)

@ www.oapen.org
<http://project.oapen.org/>

NCDD wordt landelijk dekkend netwerk

Op 1 november heeft de Nationale Coalitie Digitale Duurzaamheid (NCDD) de 'Culturele Coalitie Digitale Duurzaamheid' verwelkomd als lid. Daarmee is volgens de NCDD een belangrijke lacune in de organisatie opgevuld. Tot voor kort was de cultureel-erfgoedsector zo gefragmenteerd dat het niet mogelijk was een partner op landelijk niveau te benoemen om de belangen van die sector te vertegenwoordigen. Op initiatief van Digitaal Erfgoed Nederland en het sectorinstituut voor e-cultuur Virtueel Platform heeft de sector zich nu georganiseerd in een netwerkverband. Daaraan nemen onder meer deel het Rijksmuseum, de Nederlandse Museumvereni-

ging, het Nederlands Architectuur Instituut, het Nederlands Instituut voor Mediakunst, en het nieuwe sectorinstituut EYE Film Instituut Nederland.

De NCDD koos ook een nieuw dagelijks bestuur. Het werk aan een nationaal plan van aanpak voor duurzame toegankelijkheid van digitale gegevens kan nu worden aangestuurd door Sandra den Hamer (Culturele Coalitie), Bas Savenije (Koninklijke Bibliotheek, voorzitter), Jan Müller (Beeld en Geluid, vice-voorzitter), Peter Doorn (Data Archiving and Networked Services, penningmeester) en Martin Berendse (Nationaal Archief, secretaris) (op de foto van links naar rechts).



INGE ANGEVAARE

Het nieuwe dagelijks bestuur van de NCDD

Grondwetscommissies

Het Instituut voor Nederlandse Geschiedenis (ING) heeft een grote hoeveelheid gegevens bijeengebracht over de zeventien grondwetscommissies die Nederland sinds 1883 kende. Dat gebeurde in de digitale onderzoeksgids *Grondwetscommissies 1883-1983*, die op 15 november in Den Haag werd gepresenteerd. Informatie over hun opdracht, hun leden en hun archieven is samengebracht in een webpublicatie die onder meer ruim tienduizend pagina's notulen en verslagen biedt, en aldus een inzicht in discussies over onderwerpen die ook nu nog vaak actueel zijn, aldus het ING. www.inghist.nl

Datakeurmerk doe-het-zelf

Sinds 1 december kunnen datarepositories zelf een vragenlijst opvragen en invullen om het Data Seal of Approval (DSA) te krijgen. Daarmee kunnen ze aan de buitenwereld laten zien dat ze voldoen aan een aantal eisen om zich een duurzaam en *trusted* repository te mogen noemen. Het DSA-keurmerk komt voort uit een internationale samenwerking tussen gezaghebbende datainstellingen.

Belangstellenden gaan naar de website www.datasealofapproval.org en vragen een zogenaamd self-assessment aan. Als ze die hebben ingevuld wordt hij gereviewd door het DSA-bestuur. Na goedkeuring mag de betreffende repository het DSA-zegel op zijn website tonen. De webapplicatie werd gemaakt door DANS, de Nederlandse deelnemer in het project. (MdG)

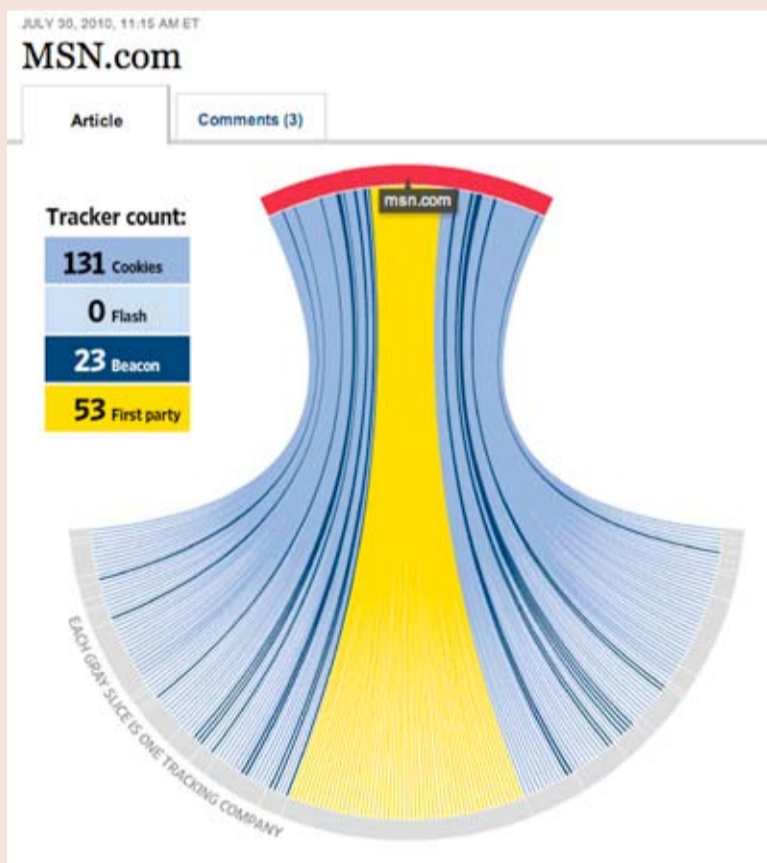
Zesdeklklassers 1952-2010

Een bestand met gegevens van Brabantse schoolkinderen uit 1952 is dit jaar voor de vierde keer aangevuld met nieuwe data van dezelfde personen. Doordat in de loop der tijd deelnemers uit het zicht verdwenen of niet meer wilden meewerken, zijn van de oorspronkelijke 5800 kinderen in de versie van 2010 circa 3000 overgebleven. Behalve schoolprestaties in 1952 bevat het bestand nu gegevens over opleiding, marktpositie en (in tweede instantie) ondernemerschap in 1983 en 1993 en over de levensduur in 2001 en 2010. Verantwoordelijk voor de laatste twee aanvullingen is prof J.S. Cramer van de faculteit Economie en Bedrijfskunde UvA, die zegt te hopen dat 'over vijf of tien jaar de operatie wordt herhaald'. Het bestand is te vinden in het EASY-archief van DANS onder de naam 'Brabantse zesdeklklassers 1952-2010'. <https://easy.dans.knaw.nl>

Gildenwebsite

De Universiteit Utrecht lanceerde in september een nieuwe website over collectieve instituties zoals gilden, gemene gronden en waterschappen: 'Institutions for Collective Action'. De site gaat grote hoeveelheden informatie aanbieden over dergelijke instituties, zoals woordenlijsten, bronbeschrijvingen, bronnenmateriaal, bibliografieën, onderwijsmateriaal en beschrijvingen van case studies. De onderzoekers onder leiding van prof. Tine de Moor hopen daarmee onderzoek en onderwijs rond collectieve instituties te stimuleren. De site biedt ook toegang tot een GIS-module, door DANS geleverd, waarmee historische data kunnen worden getoond op historische kaarten. www.collective-action.info

'Wat zij weten'



The *Wall Street Journal* publiceert op internet een serie verhalen onder de titel 'What They Know' met fraaie webgrafieken waarin wordt getoond hoe marktonderzoekers gegevens verzamelen over internetgebruikers. 'Ze observeren en onthouden hun clicks, en bouwen en verkopen gedetailleerde dossiers over hun activiteiten en belangstelling', schrijft de krant. Van de vijftig populairste Amerikaanse websites werden de *tracking files* geanalyseerd, kleine bestandjes die op de computer van de gebruiker worden geïnstalleerd. Soms om het gebruik van die site makkelijker te maken, vaak ook om

marketeers en bedrijven die data verzamelen te informeren over de acties van de gebruiker op het web. De grafiek toont de *tracker files* van msn.com (bovenin de cirkel) de soort *tracker file* (geel is voor het gemak van de gebruiker, blauw is voor gebruik door derden) en de eindbestemming van de informatie (elk grijs stukje onderin de cirkel is één bedrijf). De website toont ook een hitparade van de websites met het grootste aantal *tracker files*, die wordt aangevoerd door *dictionary.com* met 234 van zulke spionnetjes. (PB/MdG)

@ <http://blogs.wsj.com/wtk/>

ERIC BALSTER EN ROB GRIM

Vruchtbare ontmoeting statistiek en Linked Data

Een ontmoeting tussen enerzijds statistici en anderzijds Linked Data adepten die bij voorkeur alle mogelijke databronnen aan elkaar koppelen. Dat was het doel van de bijeenkomst op 7 en 8 juli in Tilburg van de Open Data Foundation (ODaF), een drijvende kracht achter de data-infrastructuur die door de OESO, WHO, IMF en Verenigde Naties wordt gebruikt.

De oproep van Tim Berners Lee, 'Bring your data to the web' zal de lezer van dit blad niet ontgaan zijn. Sir Tim doelde daarbij op Linked Data, als opvolger van het Semantisch Web tijdperk waarin belofte en verwachting blijvend botsten. Het begrip Linked Data verwijst naar een verzameling van internettechnologieën waarmee op eenvoudige wijze relaties kunnen worden gelegd en beheerd tussen brokjes informatie. Linked Data technologie ligt bijvoorbeeld vaak aan de basis van *augmented reality* toepassingen. Een voorbeeld van zo'n toepassing is dat je als pizzaliefhebber in een onbekende stad op je mobieltje geattendeerd wordt op een sympathieke Italiaan in de buurt. Overtuigende toepassingen van Linked Data in de wetenschap zien we voornamelijk in de bio-medische wetenschappen. Maar op de ODaF-bijeenkomst werd duidelijk dat ook toepassingen in de sociale wetenschappen voor de hand liggen.

Voordeel wordt nadeel

Ook voor Linked Data geldt de wijsheid van Johan Cruijff: 'Ieder voordeel heeft z'n nadeel'. Het voordeel van Linked Data is dat we een klein stukje informatie als een eenheid kunnen publiceren (nanopublicatie). Denk aan een eiwit met een streepjescode waarin is opgeslagen wat de bron is van het eiwit en hoe het eiwit in elkaar steekt. Dit voordeel, een microbestandje met contextinformatie, kent ook een nadeel. Het gevaar zit hem erin dat de brokjes makkelijk kunnen worden samengevoegd tot bijvoorbeeld een nieuw eiwit of bestand. Linked Data kan strenge regels opleggen voor de manier waarop de nanopublicaties opgaan in een groter geheel, maar is ook een vrijplaats voor het creëren van bestanden zonder regels. De computer ziet er bijvoorbeeld geen probleem in om een relatie aan te maken waarin mannen kinderen kunnen krijgen. Wat eerst een voordeel was – het behoud van

context – is nu een nadeel geworden: de semantiek gaat verloren als het verband tussen de nanopublicaties onderling niet is vastgelegd. Hoe het automatisch koppelen mis kan gaan werd duidelijk tijdens de presentatie van Dan Gillman, waarvan de video-opname (net als die van de andere presentaties van de bijeenkomst) is te vinden op de website van de ODaF.

Controleerbaar en betrouwbaar

Open Data en Linked Data worden vaak in één adem genoemd. Een voorbeeld is het Open Data project dat vorige maand is ingezet door E-overheid voor Burgers, een initiatief om overheidsinformatie via webapplicaties te ontsluiten en vindbaar te maken. Toch is Open Data niet hetzelfde als Linked Data! Open Data omvat meer dan de geschetste technologie alleen en eist dat data ook controleerbaar en betrouwbaar zijn.

http://odaf.org/events/odaf_europe_2010.php

In een tijd waarin databestanden op het Web automatisch gekoppeld kunnen worden is het belangrijk dat de betrouwbaarheid van gegevens is gewaarborgd.

De ontmoeting tussen statistici en Linked Data adepten heeft de partijen ontegenzeggelijk dichterbij elkaar gebracht. De belangstelling voor Linked Data is groot, er is sterke behoefte aan kennisoverdracht en er zijn dan ook veel vervolgspraken gemaakt. Een van de afspraken is om de samenwerking tussen Nederland en het Verenigd Koninkrijk voor Linked Open Data toepassingen voor statistische gegevens te versterken. In december verschijnt het rapport waarin de case studies van de ODaF-bijeenkomst worden besproken. De bijeenkomst werd gesponsord door de Universiteit van Tilburg, DANS, SURF en de VU.



Registers en sociaalwetenschappelijk onderzoek: een geslaagde combinatie?

BART F.M. BAKKER EN
LÉANDER KUIJVENHOVEN

Het gebruik van registerdata in sociaalwetenschappelijk onderzoek levert veel nieuwe mogelijkheden op, maar ook nieuwe hindernissen. Het is de moeite waard om die laatste te slechten, bleek op een symposium eerder dit jaar, waarvan de bijdragen nu zijn gebundeld.

Steeds vaker worden administratieve gegevens gebruikt in sociaalwetenschappelijk onderzoek. Dat vraagt echter om nieuwe methoden en technieken en die zijn nog lang niet altijd voorhanden. Wel is er al vrij veel onderzoek gedaan dat kan helpen deze lacune te dichten. Een binnenkort te publiceren symposiumbundel* brengt daarvan een aantal voorbeelden samen. Nieuwe mogelijkheden zijn er zo veel, blijkt daaruit, dat het ontlopen van registerdata en de daarbij behorende methodologische hindernissen allang geen optie meer is.

Kwaliteit

Een van de factoren die voor het gebruik van registerdata pleit is dat de kwaliteit van informatie uit enquêtes is afgenomen, vooral door een steeds grotere non-response die ook steeds selectiever is geworden. Registerdata hebben daar geen last van. Een andere reden is dat enquêtes in toenemende mate duurder zijn dan het verwerven van registraties, die immers toch al om andere redenen worden bijgehouden. Ook komen er in dit e-tijdperk steeds meer registraties beschikbaar, door digitalisering of compleet nieuw ontworpen. In Nederland zorgt de steeds bredere toepassing van het BurgerService-Nummer voor meer mogelijkheden om onderling te koppelen.

Mogelijkheden

Een belangrijke winst ligt ook in de nieuwe onderzoeksmogelijkheden. Omdat registraties vaak de totale populatie beschrijven kunnen bijvoorbeeld heel kleine subpopulaties of zeldzame fenomenen worden onderzocht. Ook bieden registerdata meer mogelijkheden voor onder-

zoek tussen generaties en longitudinaal onderzoek. Bovendien kan de vertekening worden uitgeschakeld die enquêtes vaak opleveren door bijvoorbeeld sociaalwenselijk antwoordgedrag of geheugeneffecten – factoren die bij registraties een veel kleinere rol spelen. Natuurlijk kunnen niet altijd alle relevante vragen worden beantwoord met alleen registerdata, zodat het nodig kan zijn de ontbrekende variabelen met een enquête binnen te halen. Daarbij kunnen de registraties dan weer – heel efficiënt – als steekproefkader dienen. En ten slotte kunnen ad-hoc-vragen vaak eenvoudig worden beantwoord op basis van registraties, ook als ze gaan over variabelen die men niet snel samen in een enquête tegenkomt.

Obstakels

Er valt dus veel te winnen bij het gebruik van registerdata. Maar er is ook nog veel te overwinnen. Beide aspecten kwamen aan de orde op het symposium waarvan de bijdragen nu zijn gebundeld. Er zijn bijdragen over de grote waarde van register-

gegevens voor criminologisch onderzoek, over de relatief lage kosten van longitudinaal onderzoek als registraties over loopbaan- en ontslaggegevens worden gecombineerd, en over de groei van het onderzoek over volksgezondheid door het koppelen van registraties. Het gebruik van btw-gegevens krijgt aandacht in een vergelijkende studie over Nederland en Groot-Brittannië. Aan de obstakels wordt ook ruime aandacht besteed. Zo is er aandacht voor de validiteit van data uit enquêtes en registers (geven ze weer wat we willen meten?), voor aanpakken om die te vergelijken en voor het gebruik van registerdata als standaard om de kwaliteit van enquêtegegevens te beoordelen.

Oplossingen

Mogelijkheden om registerdata zo goed mogelijk te benutten en de beperkingen te overwinnen komen eveneens aan de orde. Manieren bijvoorbeeld om het ontbreken van variabelen in administratieve bronnen te overwinnen, uitvoeringsproblemen bij het gebruik van registerdata in de geestelijke gezondheidszorg,

een raamwerk om de statistische bruikbaarheid van administratieve en andere secundaire bronnen in kaart te brengen, gebruik van schattingen om tegenstrijdigheden te overwinnen tussen registerdata en enquêtegegevens. Het CBS komt ten slotte met een onderzoekprogramma op het terrein van registers voor de toekomst. Dit programma levert niet alleen interessante onderzoeksvragen op voor het CBS zelf, maar ook voor de sociale wetenschap.

Inzicht

Voor veel van de opdoemende problemen kunnen oplossingen worden gevonden, zo lijkt het. Eenvoudig zal het echter niet zijn. Er is nog veel meer inzicht nodig. Inzicht in de belangen van registratiehouders en geregistreerden, in manieren om dreigende vertekeningen en mogelijke lacunes in de gegevens te overwinnen, en om de omvang van zeer grote bestanden terug te brengen. Terwijl we de noodzaak van nieuwe methoden en technieken onderkennen, kan evenwel het huwelijk tussen registerdata en sociaalwetenschappelijk onderzoek alvast aangekondigd worden. De mogelijkheden die deze data bieden zijn te talrijk en te veelbelovend om daar geen gebruik van te maken.

* Registers in sociaalwetenschappelijk onderzoek: Mogelijkheden en valkuilen; informatie: www.cbs.nl/infoservice.



...combineren van loopbaan- en ontslaggegevens...

Europees commissaris voor de Digitale Agenda Neelie Kroes:

MARTIJN DE GROOT

‘Wetenschap is altijd grenzeloos geweest’

Begin oktober werd in Brussel het rapport gepresenteerd van een groep ‘high-level’-adviseurs over het belang van wetenschappelijke data en infrastructuren daarvoor: *Riding the Wave: How Europe can gain from the rising tide of scientific data*. Commissaris Neelie Kroes van de Digitale Agenda nam het in ontvangst en sprak lovende woorden. *e-data&research* legde haar een aantal vragen voor over haar visie op de inhoud.

e: Over het pleidooi voor een open infrastructuur zei u ‘Ik zal dat met vreugde steunen’. Zijn er ook onderdelen van het rapport waarvoor uw enthousiasme gematigder is?

‘Het rapport gaat gedetailleerd en overtuigend in op de enorme uitdagingen waar wij de komende twintig jaar mee te maken krijgen. Ik heb het met plezier gelezen, als burger die getuige is van de spannende vooruitzichten in een snel veranderende wereld en als verantwoordelijke commissaris voor de Digitale Agenda. De geschetste scenario’s scheppen een visionair toekomstbeeld en laten zien hoe technologie en informatie alomtegenwoordig zullen zijn, en hoe samenwerking en sociale interacties kunnen veranderen. Bij de presentatie heb ik gezegd dat het verwezenlijken van de visie moeilijker is dan het misschien lijkt. Dat is echt waar. Ik ben erg enthousiast, maar als een typisch pragmatische Nederlander wil ik toch benadrukken dat het belangrijk is om realistisch en praktisch te werk te gaan om de klus te klaren.’

e: Welke partij moet het voortouw nemen bij het tot stand brengen van de gewenste infrastructuur?

‘Alle betrokken partijen moeten in harmonie samenwerken. De Commissie is bereid om een leidende rol te spelen op Europees niveau. Nationale regeringen en financiers hebben echter een specifieke verantwoordelijkheid om op nationaal niveau te handelen en partners te mobiliseren. Ons gemeenschappelijke doel moet het maximaliseren van de socio-economische impact van onderzoek zijn. Daarnaast moeten we ons richten op het positioneren van Europa als verzamelplaats van de wetenschappelijke top en als wetenschappelijk partner op wereldniveau. Europese onderzoekers, docenten en studenten zouden bovendien een belangrijke rol moeten spelen in de verdere ontwikkeling van het potentieel van e-Science.’

e: De High-Level Group dringt aan op een infrastructuur die grenzen overschrijdt van disciplines, maatschappelijke sectoren, landen en mondiale regio’s. Schuilt daarin niet het gevaar van stroperigheid en vertraging?

‘Europawilenvoorteenwetenschappelijke partner op wereldniveau zijn. De grote uitdagingen van de 21e eeuw, zoals klimaatverandering en energievoorziening, moeten geza-



‘Ik ben enthousiast, maar we moeten realistisch en praktisch te werk te gaan om de klus te klaren.’

menlijk worden aangepakt. Wetenschap is altijd grenzeloos geweest; iets wat we in ons achterhoofd moeten houden bij het bepalen van onze strategie. Tegelijkertijd kunnen we niet wachten tot alle problemen op een rijtje zijn gezet. Integendeel, het bouwen van een open, interdisciplinaire en internationale infrastructuur moet steunen op regionale, nationale en institutionele toepassingen. Zoals ik hiervoor al zei: we moeten realistisch en praktisch te werk gaan. En hoewel de doelstellingen ambitieus zijn, wil ik dat met de Europese Digitale Agenda alle inspanningen bijeen worden gebracht en dat alle doelstellingen worden verwezenlijkt.’

e: De High-Level Group onderstreept de logica en de noodzaak van steun uit publieke bronnen voor zijn ambities, en wijst daarbij op het Structuurfonds van de Europese Unie. Is de Commissie bereid om de financiële inspanningen voor dit doel te verhogen?

‘Structuurfondsen zorgen al voor een belangrijke bijdrage aan onderzoek in het algemeen en voor het onderzoek naar infrastructuren, waaronder e-infrastructuren, in het bijzonder. Van de 49,8 miljard euro die is bedoeld voor *Research, Technology, Development & Innovation* van 2007 tot 2013, is 9,8 miljard bestemd voor infrastructuren en kenniscentra. Met het oog op de toekomst zullen kennisinfrastructuren steeds belangrijker worden. Vergeet echter niet dat Europese Structuurfondsen onderverantwoordelijkheid van de lidstaten zelf vallen. Het is daarom aan de nationale regeringen om onderzoek en innovatie in het al-

gemeen, en e-infrastructuren in het bijzonder, voorrang te geven bij het toekennen van Structuurfondsen.’

e: Welke andere initiatieven kunnen we van de Commissie verwachten, behalve de Mededeling die u bij de presentatie van het rapport aankondigde?

‘Die Mededeling zal een belangrijke mijlpaal zijn voor de Digitale Agenda. De Commissie heeft sinds 2007 45 miljoen euro beschikbaar gesteld voor vijftien initiatieven op het gebied van ontwikkeling en implementatie. Daarbij zijn vele wetenschappelijke disciplines en aanbieders van e-infrastructuren betrokken. De initiatieven lopen uiteen van biodiversiteit en mariene systemen tot astronomie en geo-data. Daarnaast heeft de Commissie dit jaar een nieuwe *call for proposals* uitgezet ter waarde van 43 miljoen euro om de stap te zetten van ad hoc implementatie naar meer strategische initiatieven.

Wat betreft het toegankelijker maken van wetenschappelijke informatie heeft de Commissie in het Zevende Kaderprogramma de *Open Access Pilot* opgezet. Een e-infrastructuur die de Pilot ondersteunt, OpenAIRE, wordt in december 2010 gelanceerd. De Commissie financiert bovendien onderzoek naar duurzame toegang tot digitale objecten, waaronder wetenschappelijke gegevens. Projecten zoals CASPAR hebben het kader en de middelen gecreëerd om digitale data op de lange termijn te beheren. Nieuw onderzoek zal zich richten op de werkprocessen en links die tot de resultaten hebben geleid. Dit is een duidelijk voorbeeld van de steeds rijkere en complexere aard van digitale wetenschap.’

e: De Commissie heeft in 2008 een pilot gelanceerd om data uit onderzoek dat door haar is meebetaald standaard openbaar te maken. Zou dat beleid niet moeten worden uitgebreid naar alle publiek gefinancierd onderzoek en wat wilt u daaraan doen?

‘We kunnen gegevens niet achterhouden, omdat dit de impact van onderzoek op innovatie vermindert, met name in een kenniseconomie. Op dit moment is er echter geen formele verplichting om gegevens die voortkomen uit door de Commissie gefinancierd onderzoek, algemeen toegankelijk te maken. We gaan bepalen hoe we deze situatie in het volgende Kaderprogramma kunnen veranderen en dit zal deel uitmaken van onze besprekingen over de komende Mededeling. Daarnaast zullen we zelf suggesties doen en ontwikkelingen in de gaten houden ten aanzien van door de gemeenschap gefinancierd onderzoek op nationaal niveau. Dit blijft echter wel een zaak waarover de lidstaten zelf moeten beslissen.’

e: De open infrastructuur waarop de High-Level Group doelt zou ter beschikking moeten staan van wetenschap, ondernemers, maatschappelijke organisaties en overheden. Moet er van al die partijen ook inbreng gevraagd worden?

‘Het is niet alleen een kwestie van raadplegen, maar ook van het daadwerkelijk betrekken van de wetenschap, de private sector, NGO’s en regeringen bij de ontwikkeling van de infrastructuur. Daar zijn we al mee begonnen door middel van de projecten die ik eerder heb genoemd. Er moet natuurlijk nog veel meer gebeuren, met name op het gebied van financieringsmodellen, milieu en duurzaamheid en de rol van de private sector.

Ter verduidelijking: we hebben geen infrastructuur voor ogen die van hogerhand wordt opgelegd. De infrastructuur zal zich eerder ontwikkelen door een reeks niet-hiërarchische initiatieven – al of niet door de Commissie ondersteund – en structurering en begeleiding van bovenaf, om te zorgen dat de verschillende initiatieven uiteindelijk een naadloze infrastructuur vormen, in plaats van te eindigen als geïsoleerde eilandjes.’

e: Heeft de Commissie nog speciale initiatieven in petto om data-georiënteerd onderzoek te bevorderen in de geesteswetenschappen, in aanmerking genomen de huidige achterstand op dat gebied?

‘Naast de nieuwe oproep die we dit jaar deden voor voorstellen voor data-infrastructuren, waarvoor geesteswetenschappen van harte worden uitgenodigd zich in te schrijven, zijn er ook andere oproepen die vallen onder het *Research Infrastructures* programma. Deze richten zich specifiek op sociale en geesteswetenschappen. Onder de doelgebieden vallen bijvoorbeeld taalmiddelen, socio-economisch gedrag van individuen en gezinnen, en het integreren van digitale archieven over middeleeuwse en moderne Europese geschiedenis. Toekomstig onderzoek naar cultureel erfgoed zal zich bovendien richten op nieuwe technologieën en middelen om de toegang tot en het gebruik van culturele bronnen voor onderzoekers te ondersteunen.’

Wood en Kroes

Het rapport *Riding the Wave: How Europe can gain from the rising tide of scientific data* was het eindverslag van een *High-Level group on Scientific Data* die in opdracht van de Europese Commissie een ‘visie 2030’ op tafel moest leggen over infrastructuren voor wetenschappelijke data. Voorzitter was de Britse hoogleraar John Wood die ook leiding geeft aan het European Strategy Forum on Research Infrastructures (ESFRI). Zie ook pagina 1.

Van de zevenentwintig leden van de Europese Commissie, het ‘dagelijks bestuur’ van de Europese Unie, zijn er twee die e-science tot hun gebied mogen rekenen: de Ierse commissaris Máire Geoghegan-Quinn van Onderzoek, innovatie en wetenschap en de Nederlandse vicepresident van de Commissie Neelie Kroes, commissaris van de Digitale Agenda. Kroes kreeg na een lange politieke carrière in ons land in de vorige commissie de portefeuille Mededinging. In de commissie-Barroso II, die vorig jaar aantrad, kwam ze op haar huidige verantwoordelijkheidsgebied terecht.

UNESCO voor Open en Permanent Access

UNESCO in Nederland bereidt initiatieven voor om zowel in ons land als internationaal de aandacht te vestigen op de noden van minder geprievilegieerde landen op het gebied van duurzame toegang tot informatie. Dat blijkt uit het werkprogramma van de Nationale UNESCO-Commissie, dat deze zomer werd gepubliceerd.

Informatie- en communicatietechnologieën (icts) kunnen volgens de organisatie veel van haar doelen helpen realiseren, zoals democratisering, terugdringen van armoede, internationale samenwerking en wereldwijde deelname aan onderwijs en cultuur. In de praktijk zijn

er echter veel obstakels. Op sommige terreinen, zoals *standard setting*, het agenderen van vraagstukken en het uitwisselen van informatie kan UNESCO een goede rol spelen. Toegang tot informatie is een van de grondslagen van een democratische samenleving, en het waarborgen daarvan behoort tot haar kerntaken.

De organisatie wil daarom in de eerste plaats een Nederlandse expertmeeting organiseren over Open Access, UNESCO en de wetenschap in minder geprievilegieerde landen. De resultaten daarvan zouden internationaal verspreid kunnen worden en mogelijk tot een UNESCO-bijeen-

komst in Parijs kunnen leiden. Voor het probleem van *permanent access*, dus duurzame toegang, in de ontwikkelingslanden wil UNESCO een speciale bijeenkomst organiseren op de wereldconferentie 'Memory of the World', volgend voorjaar in Warschau. Dat moet leiden tot een ahangsel bij de *Guidelines for the Preservation of the Digital Heritage* die UNESCO in 2003 publiceerde, en in een later stadium tot voorstellen aan UNESCO wereldwijd om nieuwe standaarden op het gebied van duurzaam bewaren te bevorderen. Initiatieven op het gebied van permanent access zijn vooral nodig omdat er veel wordt gedigitaliseerd zonder oog voor de toekomstige bewaring, aldus UNESCO. Ook moeten er oplossingen komen voor de nationale en internationale werkverdeling en voor de vraag of en wanneer data uit de openbaarheid gehouden mogen worden.

UNESCO zoekt voor beide activiteiten samenwerking met een reeks internationale organisaties, met Nederlandse organisaties als SURF, DANS en de KNAW en met buitenlandse zusterorganisaties daarvan (MdG)

Samenwerking Europese data-infrastructuurprojecten

Supporting Digital Humanities (SDH) 2010 was de naam van een conferentie die in oktober in Wenen werd georganiseerd door DARIAH (Digital Research Infrastructure for the Arts and Humanities) en CLARIN (Common Language Resources and Technologies Initiative). Deze eerste SDH-conferentie, die werd gecombineerd met een bijeenkomst van het Networking Event for European Research Infrastructures (NEERI 2010) stond vooral in het teken van de verbreding van data-infrastructuren en de uitbreiding van onderzoekactiviteiten.

De sessies hadden onder meer betrekking op manuscriptstudies, musicologie, archeologie en sociaal-economische geschiedenis. De discussies gingen vooral over de volgen-

CARARE: Connecting Archaeology and Architecture in Europeana

In de nazomer van dit jaar, zes maanden na de start van het Europese cultureel-erfgoedproject CARARE, werden de resultaten van de projectpartners in Italië gepresenteerd op een steenworp afstand van de toren van Pisa. Een inspirerende plek waar archeologie, architectuur en toerisme onlosmakelijk met elkaar verbonden zijn.

Om cultureel erfgoed voor een groot publiek toegankelijk te maken werd begin dit jaar het Europese samenwerkingsverband CARARE opgericht. Het project zorgt ervoor dat vanuit het werkterrein van de archeologie en de architectuur digitale objecten worden toegevoegd aan de Europese cultuurwebsite Europeana (www.europeana.eu).

DANS (Data Archiving and Networked Services) levert via het e-depot voor de Nederlandse archeologie een bijdrage aan CARARE. Dat gebeurt selectief, want uitgebreide wetenschappelijke datasets zijn niet direct gericht op een breder publiek en er moet ook rekening worden gehouden met privacy en auteursrechten. Erfgoed Nederland coördineert



HELLA HOLLANDER

Op het Plein van de Wonderen in Pisa: connecting Archaeology and Architecture.

de bijdrage van Nederlandse organisaties rond cultureel erfgoed, waaronder musea. De Rijksdienst voor Cultureel Erfgoed leidt de organisatie en planning van het oogsten van de metadata. Nieuwe partners op het gebied van cultureel erfgoed zijn van harte welkom.

In Pisa keken de partners welke data ze beschikbaar kunnen stellen en in hoeverre ze daar technisch toe in staat zijn. De erfgoedinstellingen KUAS (Denemarken) en ADS (Engeland) hebben een metadata-schema opgesteld dat grotendeels is gebaseerd op het al bestaande Midas-model. De technische realisatie van CARARE en het oogsten hiervan voor Europeana ligt bij het Griekse Athena-project (DCU).

De verschillende thematische collecties, waaronder dus CARARE, vormen de verzamelingen binnen Europeana en worden de komende jaren toegankelijk gemaakt. Bij CARARE zijn veel wetenschappelijke instellingen betrokken. Europeana heeft zich tot nu toe vooral via musea en bibliotheken op een breed publiek gericht. Er is een focus op het type 3D en VR (Virtual Reality) bestandsformaten. 'Als het doel en het publiek met foto's bereikt kunnen worden dan gaat hier de voorkeur naar uit,' benadrukte de Belgische deelnemer Daniel Pletinckx. Een goed advies om op te volgen. (Hella Hollander)

www.unesco.nl/documents/documenten-natcom/
www.unesco.org/new/en/unesco/

SINDS KORT BESCHIKBAAR

Het overzicht toont een aantal databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS, CentERdata 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 bij DANS beschikbare databestanden komen van diverse andere onderzoeksinstellingen. Deze kunnen kosteloos gedownload worden vanuit DANS EASY: <https://easy.dans.knaw.nl>

Centraal Bureau voor de Statistiek	Periode
Bijstandsuitkeringenstatistiek (transactiebasis)	2009
Enquete ICT-gebruik bedrijven	2006-2008
Integrale veiligheidsmonitor	2008-2009
Woning Onderzoek Nederland	2006
Statistiek reïntegratie door gemeenten	2009
<i>Uit het Sociaal Statistisch Bestand (SSB)</i>	
- Banenbestanden	2008
- Overige arbeidsbestanden	2008
- Overige uitkeringen	2008
- Demografische persoonsgegevens	2008
- Uitkeringen buitenland	2008
- Werkloosheidswetbestanden	2008
- Zelfstandigen	2008
CentERdata LISS Data Archive	
<i>Ambiguity Aversion and Household Portfolio Choice</i> (Dimmock, S., Kouwenberg, R., Wakker)	
	2010
<i>Decreasing Trust in Democratic Governments, Wave 3</i> (Wisse, E.)	
	2010
<i>Tilburg Consumer Outlook Monitor</i> (Yabar, J., Pieters, R.)	
	2010
<i>Holiday Study: Wave 1 and 2</i> (Statistics Netherlands)	
	2010
<i>Organizations and older workers – Part 1 and 2</i> (Karpinska, K., Henkens, K., Schippers, J.)	
	2010
<i>LISS Core Study – Personality, Wave 3</i> (CentERdata)	
	2010
<i>Measuring Higher Order Risk Attitudes of the General Population</i> (Kuilen, G. van de, Trautmann, S.T.)	
	2009
<i>Sexuality and media use: Wave 1 and 2</i> (Peter, J.)	
	2010
Via DANS EASY	
<i>Historische Wetenschappen</i>	
<i>Bijbelvertalingen, digitale tekstituitgaven 2008-2010</i>	
(Nicoline van der Sijs, Stichting Vrijwilligersnetwerk Nederlandse Taal)	1749-1794
<i>Playlists and Repertoire of 18th-C The Hague Theatres mainly of the French Comedy</i> (Aldo Lieferring, Universiteit Utrecht – Muziekwetenschap)	
	1724-1795
<i>Eighteenth-Century Music and theatre advertisements from the 's-Gravenhaagsche Courant and Gazette de La Haye</i> (Aldo Lieferring, Universiteit Utrecht – Muziekwetenschap)	
<i>Interviews met Veteranen, Modern History en (IPNV – Veteranen Instituut) Contemporary History</i>	
<i>Archeologie</i>	
<i>The Džarylgač Survey Project</i> (GIA – Groninger Instituut voor Archeologie)	
	2010
<i>Breda Taxandrialaan (BAAC)</i>	
	2010
<i>Een villaterrein bij het Kloosterraderplein te Kerkrade (ADC Archeo-Projecten)</i>	
	2010
<i>Veldtekeningen van de opgravingen te Goedereede, Ouddorp (Provincie Zuid Holland 1982/1983)</i>	
	1959-1960
<i>IVO-O rioolpersleiding Woerden (Oranjewoud bv)</i>	
	2009
<i>Plangebied A12 BRAVO Molendijk (Gemeente Bodegraven) (RAAP)</i>	
	2010
<i>Sociale Wetenschappen</i>	
<i>Netherlands Longitudinal Lifecourse Study – NELLS – First Wave</i> (P.M. de Graaf, M. Kalmijn, G. Kraaykamp, C. Monden)	
	2009
<i>Brabantse zesdeklassers, persoonsgegevens (J.S.Cramer – Universiteit van Amsterdam)</i>	
	1994-2010
<i>Gebruik kinderopvang (Sociaal en Cultureel Planbureau)</i>	
	2004
<i>HEGESCO – Higher education as a generator of strategic competences (ROA – Universiteit Maastricht)</i>	
	2008
<i>REFLEX – REsearch into employment and professional FLExibility (ROA – Universiteit Maastricht)</i>	
	2005-2006
<i>Monitor Hoger Beroeps Onderwijs (ROA – Universiteit Maastricht)</i>	
	2009
<i>Registratie uitstroom en bestemming schoolverlaters – BVE Monitor (ROA – Universiteit Maastricht)</i>	
	2009
<i>God in Nederland (Radboud Universiteit)</i>	
	2006

COLOFON

e-data@research is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van CentERdata, DANS, het Huygens Instituut, 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)3494450 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 Hermesen, Jetske van der Schaaf
Aan dit nummer werkten mee: Bart Bakker, Marten Jan Bok, Jos Engelen, Rob Grim, Stefan

Dormans, Hella Hollander, Léander Kuijvenhoven, Gerhard Jan Nauta, Harm Nijboer.
Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf
Vormgeving en opmaak: Ellen Bouma, Alkmaar
Productie: Amsterdam University Press
Druk: Ten Brink, Meppel
ISSN: 1872-0374

Internetpanels niet betrouwbaar, of toch wel?

Internetbureaus die voorafgaand aan de Tweede-Kamerverkiezingen van 9 juni de burgers naar hun stemgedrag vroegen, bleken de politieke aardverschuivingen die de verkiezingen markeerden in de versterkte niet te hebben voorzien. Het bureau Peil.nl van Maurice de Hond zat er maar liefst achttien zetels naast. Het bureau Synovate deed het maar iets minder slecht. Het grote aantal zetels dat de Partij voor de Vrijheid (PVV) wist te winnen, had geen van beide bureaus ook maar bij benadering voorspeld. Eén van de redenen die de peilers noemden was het feit dat PVV'ers niet meededen aan het onderzoek, omdat die niet in de panels zaten.

Digitale kloof

Het gebruik van internet heeft het peilen zeker een stuk gemakkelijker gemaakt. Het maakt het mogelijk om heel snel en goedkoop heel veel gegevens te verzamelen bij heel veel mensen, die je niet eens hoeft te zoeken, want ze melden zich vaak spontaan. Die zogeheten zelfselectie is één van de grote problemen van het internetpeilen en van de zogeheten Accesspanels: deelnemers zijn niet willekeurig geloot uit een grotere groep. Onbekend blijft vaak wat de respons is en hoe die is verdeeld over bevol-

kingsgroepen. Peilingen zijn dan nauwelijks meer een objectief meetinstrument en de uitkomsten van dergelijke peilingen zijn dan ook niet erg betrouwbaar.

Voor de internetpanels worden mensen die geen internetaansluiting hebben, zoals veel ouderen, laagopgeleiden en allochtonen, in elk geval niet bereikt. Dit wordt wel de digitale kloof genoemd. Aan noodzakelijke voorwaarden om de kwaliteit van een onderzoek te beoordelen, zoals het kunnen vaststellen van responscijfers en van de over- en ondervertegenwoordiging van bevolkingsgroepen in de steekproef, wordt niet voldaan.

Onderzoeksverantwoording

Eigenlijk zou van iedere uitgevoerde enquête een adequate onderzoeksverantwoording moeten worden opgesteld, vindt prof. Jelke Bethlehem van het Centraal Bureau voor de Statistiek (CBS) en de Universiteit van Amsterdam. In zo'n verantwoording zou precies moeten staan wat de doelgroep is, hoe de steekproef getrokken is – geloot of via zelfselectie – wat de omvang van de steekproef en de non-respons is en wat de onzekerheidsmarges zijn. Investeren in arbeidsintensieve en dure één-op-één gesprekken met je respondenten, van

wie je de sociale achtergrondkenmerken kent, is volgens de CBS-onderzoeker de enige oplossing om betrouwbare enquêteresultaten te krijgen.

Interviewdruk

Positieve geluiden over het *on line* enquêteren ontbreken in de discussie evenwel niet. Zo schreef Edith de Leeuw, bijzonder hoogleraar Kwaliteit en Survey Onderzoek aan de Universiteit Utrecht onlangs in het statistiektijdschrift *STATOR* dat betrouwbaar en valide meten via het internet wel degelijk kan.

Enquêteren via internet heeft volgens De Leeuw vele voordelen. Internetenquêtes zijn snel, goedkoop en actueel. Zeker zo belangrijk is dat de interviewdruk ontbreekt. Zonder interviewer aan tafel worden er waarschijnlijk minder sociaal wenselijke antwoorden gegeven. In enquêtes zonder interviewer geven mensen normoverschrijdend gedrag vaker toe: ze drinken dan meer, frauderen meer en overtreden vaker de wet. Een ander voordeel is dat mensen die de vragenlijsten op de computer doorneemen geen vragen kunnen overslaan. De computer herinnert ze eraan als een antwoord op een vraag ontbreekt. *Missing data* komen om die reden niet voor.

Beter meten

Voorwaarde is wel dat de metingen 'accuraat' zijn. Het gaat dan om wat zij noemde de vier hoekstenen van goed survey-onderzoek: een hoge dekkingsgraad, een vakkundige steekproeftrekking, een hoge respons en een goede meting. Is dat gewaarborgd, dan vindt De Leeuw dat met internetenquêtes zelfs beter kan worden gemeten dan met traditionele methoden.

Ronald van der Bie

Zie ook pagina 6: Focus

Column

Jos Engelen

Open Access moet sneller

Open Access Publishing, ook in Nederland de meest gangbare benaming van Vrij Toegankelijk Publiceren, heeft met de opkomst van het World Wide Web een sterke impuls gekregen. Wetenschappelijke zoekmachines over de grenzen van disciplines heen, en doorklikken naar onderliggende onderzoeksdata, naar verwante publicaties en noem maar op: het wordt allemaal mogelijk. En in principe is het technisch nu allemaal al beschikbaar. Toch is Open Access publiceren nog altijd meer uitzondering dan regel, zoals in de Open Access Week van 11 tot 15 oktober weer duidelijk onder de aandacht kwam.



HARRY MEIJER

Hoe komt dat? Laten we het publiceren van wetenschappelijke artikelen nader onder de loep nemen. Ten eerste is het wezenlijke kenmerk van Open Access Publishing het 'business model' (wellicht het best vertaald als verdienmodel?). Heel in het kort: de kosten voor publicatie worden niet meer betaald uit abonnementsgelden, maar door de auteurs zelf. In principe is de overgang van het klassieke 'business model' naar Open Access kostenneutraal: abonnementskosten worden vervangen door publicatiekosten, die de auteurs zelf betalen. Publiceren is immers niet gratis: het organiseren van een excellente 'peer review' en de werkzaamheden van een 'editorial board' moeten bijvoorbeeld betaald worden. Dat geldt vanzelfsprekend ook voor de kosten voor de verspreiding en het duurzaam archiveren van publicaties en onderzoeksdata. De expertise die DANS heeft ontwikkeld voor het duurzaam beschikbaar stellen en archiveren van onderzoeksdata is hier erg waardevol.

Ten tweede hebben de gevestigde tijdschriften, vaak terecht, een grote reputatie die tot uitdrukking komt in een impactfactor. Deze factor is gelijk aan het gemiddeld aantal citaties per artikel dat door het tijdschrift is gepubliceerd. Onderzoekers willen daardoor graag publiceren in tijdschriften met een hoge impactfactor. In het beoordelingssysteem van de kwaliteit van onderzoekers houden organisaties, zoals NWO, hier rekening mee. Uitgevers die hun gevestigde tijdschrift niet Open Access willen maken, staan daarom sterk. En nieuw op te richten tijdschriften beginnen met impactfactor nul. Daarom stel ik voor om nieuwe Open-Access-tijdschriften een initiële impactfactor toe te kennen, gebaseerd op onder meer de kwaliteit van de redacteurs en de kwaliteit van het redactionele proces, inclusief 'peer review'. Alleen zo kunnen we de vicieuze cirkel doorbreken.

Ikzelf ben groot voorstander van Open Access. Als voorzitter van onderzoeksfinancier NWO heb ik de mogelijkheid om bij te dragen aan de realisatie van het Vrij Toegankelijk Publiceren. Zo heeft NWO een jaar geleden vijf miljoen euro ter beschikking gesteld voor Open Access. Dit is een eerste stap en nog lang niet voldoende. De voordelen en mogelijkheden van Open Access zijn immers zo enorm dat de ontwikkeling ervan niet te stoppen is. Het proces kan en moet alleen sneller.

Prof. dr. J.J. Engelen is voorzitter van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).



NATIONALE BEELDBANK

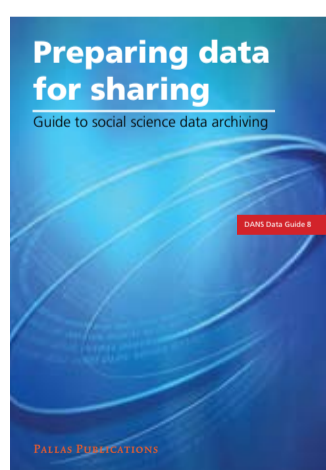
Traditionele enquêtes kennen meer interviewdruk

Gelezen

Harry van den Berg, Stef Scagliola, Fred Wester (red.): *Wat Veteranen Vertellen; Verschillende perspectieven op biografische interviews over ervaringen tijdens militaire operaties*. Amsterdam, Pallas Publications, 2010; ISBN 978-90-8555-034-1. Prijs € 39,95.

Publicatie, gebaseerd op biografische interviews met veteranen uit het Interviewproject Nederlandse Veteranen, een initiatief van het Veteraneninstituut in Doorn. Ondanks de grote verschillen tussen militaire operaties komen in de interviews telkens dezelfde dilemma's en trauma's terug. Bijdragen van onderzoekers vanuit verschillende disciplines geven een veelzijdig en openhartig beeld van wat veteranen meemaken.

Verrijkte publicatie: de digitale versie van het boek bevat links naar audiofragmenten van de interviews. Lezen van een artikel kan dus gecombineerd worden met luisteren naar de veteranen. Direct beluisteren kan ook op www.watveteranenvertellen.nl. Alle gebruikte interviews zijn gedeponneerd bij DANS en komen op termijn beschikbaar voor geregistreerde gebruikers. De andere interviews zijn alleen toegankelijk voor wetenschappelijk onderzoek met toestemming van het Veteraneninstituut.



Preparing data for sharing; Guide to social science data archiving. Amsterdam, Pallas Publications, 2010; DANS Data Guide 8; ISBN 978-90-8555039-6

Aimed at those engaged in the cycle of social science research, from applying for a research grant, through the data collection phase, and ultimately to the preparation of the data for deposit in the DANS archive or any other data repository. Adaption of the 4th edition of the *Guide to social science*

data preparation and archiving of 2009, published by the US based Inter-university Consortium for Political and Social Research (ICPSR). Intended to help researchers manage and document their data to prepare them for archival deposit, as well as think more broadly about the types of digital content that should be deposited in an archive.



Marcel Das, Peter Ester and Lars Kazmirek (eds.): *Social and Behavioral Re-*

search and the Internet; Advances in Applied Methods and Research Strategies. Oxford, Taylor & Francis Group, October 2010. ISBN: 978-1-84872-817-2

Highlighting the progress in using web-based surveys for data collection, this timely volume summarizes the experiences of leading behavioral and social scientists from Europe and the USA who collected data using the internet. Some chapters present theory, methodology, design and implementation, while others focus on best practices or issues such as data quality and understanding paradata. A number of contributors applied innovative web-based research methods to the LISS panel of CentERdata collected from over 5,000 Dutch households. Their findings are presented in the book. Some of the data are available on the book website. The book addresses practical issues such as data quality, how to reach difficult target groups, how to design a survey to maximize response, and ethical issues that

need to be considered. Innovative applications such as the use of biomarkers and eye-tracking techniques are also explored. See also <http://pub.centerdata.nl/>

Nico Bakker, Leen Breure, Wilko Quak (red.): *Geo-informatie kent geen tijd? Delft, Nederlandse Commissie voor Geodesie*, 2010. ISBN: 978-90-6132-325-9

In geografische bestanden is tijd vaak moeilijk weer te geven. Reeksen van oude en nieuwere kaarten laten ontwikkelingen zien, maar het juiste moment van wijziging is daarbij vaak niet meer te achterhalen. In het huidige digitale tijdperk is het echter mogelijk om zowel het moment van ontstaan van een object als het tijdstip waarop het in een database wordt opgenomen, direct vast te leggen. De publicatie vormt de neerslag van de presentaties van een studiemiddag over dit onderwerp. Verkrijgbaar als pdf:

www.ncg.knaw.nl/Publicaties/Groen/50BakkerGeoTijd.html