



E-DATA & RESEARCH

Jaargang 11 | nummer 1

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: Centerdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het RIVM.

INHOUD

2

Verslagen van events in Gehoord en bijgewoond

3



Robert Erdmann over Jheronimus Bosch

4

Deze datasets zijn ook sinds kort beschikbaar

5

Ron Dekker (NWO) over de Open Acces Week

6

eHumanities.nl brengt partijen samen

7

Nieuw platform voor citizen science

8

Dilemma rondom de Auteurswet

**24 -30 oktober
Open Access week
openaccess.nl**

Lees ook het interview met Ron Dekker op blz. 5



Scan deze QR code met een smartphone om de website van E-data te bezoeken.
www.edata.nl

Nieuwe landelijke en Europese regels omtrent persoonsgegevens

‘Wees je bewust: een datalek ligt zo op de loer’

Zowel landelijke als Europese regels rondom persoonsgegevens zijn veranderd. Wat betekent dit voor wetenschappelijk onderzoek in Nederland? Het advies van twee deskundigen. *Marika de Bruijne*

“Per 1 januari 2016 is de Wet Bescherming Persoonsgegevens (WBP) aangescherpt. Er is onder meer een meldplicht gekomen in het geval dat persoonsgegevens in verkeerde handen komen of onrechtmatig worden verwerkt. Daarnaast zijn de sanctiemogelijkheden exponentieel in breedte en hoogte gestegen,” vertelt Moswa Herregodts, functionaris gegevensbescherming aan de Tilburg University.

In mei 2016 is bovendien de Europese Algemene Verordening Gegevensbescherming in werking getreden. “Het recht op privacy is een grondrecht van iedere burger,” legt Marlon Domingus het uitgangspunt van de verordening uit. Domingus is projectmanager research data management aan de Erasmus Universiteit en samen met Esther Hoorn van de RUG voorzitter van de juridische werkgroep van het Landelijk Coördinatiepunt Research Data Management (LCRDM).

Universiteitsbrede aanpak

Onderzoeksinstituten hebben tot mei 2018 om hun bedrijfsvoering in te richten volgens de Europese verordening. Per die datum moeten ze kunnen aantonen dat ze de nieuwe regelgeving volgen.

Tilburg University heeft een functionaris voor de gegevensbescherming aangesteld en een werkgroep voor privacy-gerelateerde vraagstukken in het leven geroepen die zich onder meer bezighoudt met datalekken. “Zo’n lek ligt snel op de loer. In het eerste half jaar hebben we al tientallen casussen onderzocht,” vertelt Herregodts. “Verder willen we dit jaar disk-encryptie universiteitsbreed implementeren bij laptops, later bij alle werkplekken. Een tweede speerpunt is bewustwording van de invloed van eigen gedrag, ook in wetenschappelijk onderzoek.”

Begin met beschikbare middelen

“De meldplicht datalekken houdt voor individuele onderzoekers in dat zij intern melding moeten maken van een (potentieel) datalek en volledige medewerking moeten verlenen aan het onderzoek naar het datalek,” legt Herregodts uit. “Maar belangrijker is uiteraard rechtmatige verwerking



De digitale sporen die wetenschappers achterlaten kunnen gevoelige persoonlijke informatie bevatten. Illustratie Rik van Schagen

van persoonsgegevens en het voorkomen dat de data terecht komen bij onbevoegden.”

Domingus adviseert: “Bedenk al bij de opzet van je onderzoek hoe je zorgvuldig omgaat met gegevens die herleidbaar zijn tot personen. Realiseer je wie toegang tot deze gegevens moet hebben, en wie absoluut niet.”

Tips van Herregodts

- Splits al tijdens het verzamelen de herleidbare gegevens ('het communicatiebestand') van onderzoeksdata en bewaar deze bestanden op gescheiden en veilige locaties.
- Bewaar geen gevoelige onderzoeksdata bij publieke cloud-opslagdiensten zoals Dropbox, GoogleDrive, OneDrive en Box. Gebruik hiervoor SURFdrive.
- Download niet klakkeloos een leuk tooltje van het internet. Laat je informeren welke toepassingen veilig gebruikt kunnen worden.
- Lock altijd de pc / laptop.
- Gebruik voor opslag bij voorkeur geen draagbare digitale opslagmedia zoals een laptop, plug-in hard drive, usb-stick, cd of dvd. De netwerkschijven op de universiteit zijn betrouwbaarder en veiliger.
- Bewaar je gevoelige onderzoeksdata toch op je laptop, zorg dan voor disk-encryptie via IT Support.
- Gebruik Secure FileSender voor het veilig verzenden van grote bestanden.
- Laat de pc / laptop niet onbeheerd achter, doe de deur van je kantoor op slot.

De instelling heeft de verantwoordelijkheid om hierbij te helpen “door heldere procedures te communiceren over de omgang met data, voorzieningen te bieden voor veilige opslag, encryptie, anonimiseren van data, enzovoorts. En door onderzoeksondersteuners aan te stellen die de onderzoeker kunnen helpen bij de opzet van een onderzoek,” geeft Domingus aan.

De hoge boetes dreigen echter te leiden tot een verkramping. “We zien dat veel betrokkenen neigen naar uitstelgedrag bij het oppakken van het onderwerp ‘hoe privacy goed te borgen’. Onze boodschap, die we ook uitdragen via de Werkgroep Juridische aspecten van het Landelijk Coördinatiepunt, is: begin gewoon met de middelen die er zijn. Vergroot de bekendheid door met elkaar in gesprek te raken en te blijven - en van elkaar te leren,” moedigt Domingus aan.

autoriteitpersoonsgegevens.nl/nl/melden/meldplicht-datalekken
lcrdm.nl

Parade van 12 panels
in het Trippenhuys

Marion Wittenberg

Hoe veranderen gedurende de levensloop facetten als inkomen, gezondheid, en normen en waarden? En hoe verhouden deze aspecten zich tot elkaar? Longitudinaal panelonderzoek maakt het mogelijk dit soort vragen te onderzoeken. Donderdag 8 september presenteerden een aantal van de belangrijkste Nederlandse longitudinale panels zich op de jaarbijeenkomst van het Nederlandstalig Platform voor Survey Onderzoek (NPSO).

In de keynote presenteerde Jon Burton (Universiteit van Essex) het Britse longitudinale panelonderzoek ‘Understanding Society’. Dit internationaal toonaangevende panelonderzoek volgt jaarlijks de leden van zo’n 40 duizend Britse huishoudens. Vervolgens kwamen in twee parades elf Nederlandse panels voor het voetlicht. Aan de sprekers was gevraagd om in hun presentatie in te gaan op een bepaald aantal aspecten, met name op de gebruiksmogelijkheden en de uitdagingen die je als panelonderzoeker tegenkomt. Het programma was met elf panels een beetje vol, maar juist het feit dat alle panels dezelfde issues behandelden, werd door de deelnemers aan deze dag zeer gewaardeerd. Na de panelpresentaties ging Peter Lugtig van de Universiteit van Utrecht in op het issue van paneluitval en besprak Jurjen Iedema (Sociaal en Cultureel Planbureau) hoe je dit soort data het beste kan analyseren.

Hoewel heel verschillend van doelgroep of onderwerp, lopen bijna alle panels tegen dezelfde zaken aan. Deze bijeenkomst gaf de mogelijkheid om ervaringen uit te wisselen en contacten voor mogelijke sa-



Krakau was gaststad van Digital Humanities 2016 foto Shutterstock

menwerking te leggen. Tijdens de discussie werd aan het NPSO gepocherd om een vervolgbijeenkomst te organiseren, waarin dieper op deelaspecten van panelonderzoek kan worden ingegaan.

npsa.net

Open digitaal erfgoed
heeft enorme reikwijdte

Olaf Janssen

Het digitaliseren van cultureel erfgoed in Europa verloopt volgens plan, aldus de EU. Eind juni kwamen onder de vlag van het Nederlands EU-voorzitterschap zo’n 250 mensen uit de Europese cultuur- en erfgoedsector bijeen om samen met eindgebruikers uit onder andere de creatieve industrie en onderzoekswereld bij te praten over het gebruik en de impact van deze digitale content.

De eerste dag van de conferentie bracht weinig schokkende of

nieuwe inzichten. Het was vooral een bevestiging van de mantra’s die de laatste jaren in digitaal erfgoedland te horen zijn. Neem zichtbaarheid, verspreiding en hergebruik van collecties als uitgangspunten, betrek vooral het publiek bij die missie en werk samen met andere instellingen. De enorme reikwijdte en impact van een goed uitgevoerd open collectie(beleid) werd niet alleen geïllustreerd door het immer inspirerende (én onvermijdelijke) Rijksmuseum, maar ook door cases uit Spanje, Noorwegen en Vlaanderen (debeeldcapsule.be).

Waar ik persoonlijk heel erg blij van werd, is de inmiddels brede erkenning van de rol die de Wikipedia-beweging speelt in dit geheel. Werd die een paar jaar geleden nog niet erkend of zelfs genegeerd, anno nu is Wikipedia niet meer weg te denken als het gaat om het breed zichtbaar en herbruikbaar maken van erf-

GEHOORD & BIJGEWOOND

goed. De tientallen succesvolle samenwerkingsverbanden tussen de Wikipedia-community en erfgoedinstellingen in heel Europa spreken wat dat betreft boekdelen.

Google timmert met zijn Cultural Institute lekker aan de weg, maar de ING-bank zou ik nou niet meteen met erfgoed associëren. Op dag twee moest ik dus alsnog even rechtop gaan zitten in de sessie over publiek-private samenwerking. Hun Next Rembrandt, een 3D-geprint schilderij gebaseerd op big data van Rembrandts oeuvre, werd voor de gelegenheid tentoongesteld, voorzien van bijbehorende presentatie over het culturele sponsorbeleid van de bank. De ochtendsessie werd afgesloten met een uitleg over PERSIST, het digitale duurzaamheidsprogramma van de Unesco. De middag had een werkbezoek aan het EYE-filminstituut in Noord in petto: qua architectuur, buzz en collectie een aanrader!

Ready to reach out: Connecting cultural heritage collections and serving wider audiences, 29 & 30 juni, Amsterdam

Groeidend DH2016
zoekt naar inclusiviteit

Steven Claeysens

Begin deze zomer troeften in Krakau een week lang een ongezien aantal onderzoekers samen voor de jaarlijkse hoogmis van de Digital Humanities. Met meer dan negenhonderd deelnemers was dit het drukstbezochte Digital Humanities-congres tot nu toe. De Poolse organisatie bleek echter op alles voorbereid en verzorgde een uitstekende vijfdaagse. Ook Nederland was

goed vertegenwoordigd en mocht aan het eind van het congres bij monde van Arjan van Hessen aankondigen dat DH2018 in Utrecht plaatsvindt.

De keynotes vielen tegen dit jaar. Agnieszka Zalewska kon niet inzichtelijk maken wat de digitale geesteswetenschappen moeten leren van CERN en Helen Agüera stelde teleur door niet veel verder te kijken dan een geschiedenis van het Text Encoding Initiative (TEI) in relatie tot haar eigen loopbaan. Enkel Claire Warwick wist enigszins te beklijven met haar reflecties op de verschillen tussen digitale en papieren leescultuur.

Gelukkig viel op het niveau van de gemiddelde lezing en poster minder aan te merken: voor elk wat wils, van experimenteel tot gevestigd, van multidisciplinair tot specialistisch. Bovendien hebben sommige deel terreinen intussen dusdanig veel aantrekkingskracht of maturiteit dat ze de hoofdmoot vormden van meerdere sessies en het daarmee mogelijk werd je een conferentie lang onder te dompelen in uitsluitend de stylometrie of enkel de ontwikkelingen op het audiovisuele vlak. Hierin schuilt wel een gevaar. Deze conferentie hoort ook een plaats te zijn waar onverwachte kruisbestuivingen tot stand komen en dat wordt bemoeilijkt met deze verkaveling.

Dit congres kent ook een sterke onderstroom die ijvert voor meer inclusiviteit. De ophef van vorig jaar in Sydney over de ‘mannenparade’ tijdens de opening kreeg zijn weerslag in verschillende sessies en paneldiscussies, maar ook, en niet onomstreden, in de aankondiging van een officiële Frans-Engelse editie in Montreal volgend jaar.

dh2016.adho.org

NWO heeft internationale primeur

Drie miljoen voor replicatiestudies

De Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) stelt drie miljoen euro beschikbaar voor een pilotprogramma Replicatiestudies. Marion Wittenberg

Wetenschappers kunnen hiermee onderzoek, dat door anderen is gedaan, opnieuw uitvoeren. De pilot richt zich op het repliceren van studies met een grote impact op wetenschap, overheidsbeleid of het publieke debat. Met het financieren van de replicatie van dergelijk ‘hoeksteenonderzoek’ wil NWO onderzoekers stimuleren replicatieonderzoek uit te voeren. Deze pilot bestrijkt het

terrein van de sociale wetenschappen, gezondheidsonderzoek en zorginnovatie. Op basis van de resultaten van de pilot bekijkt NWO hoe ze het replicatieonderzoek op kan nemen in al haar onderzoekprogramma’s.

Grote impact

Het pilotprogramma richt zich op twee vormen van onderzoek: reproductie en replicatie. Bij reproductie worden datasets van de oorspronkelijke studie opnieuw geanalyseerd, bij replicatie worden de data opnieuw verzameld op basis van het onderzoeksprotocol van de oorspronkelijke studie. Door dit soort onderzoek te stimuleren wil NWO

Brengt glimlach geluk?

Replicatie van hoeksteenonderzoek leidde in september tot veel media-aandacht. In 1988 publiceerde de Duitse psycholoog Fritz Strack de resultaten van psychologische experimenten waaruit bleek dat een mens gelukkiger wordt door de mond in een glimlach te plooiën. De studie werd de hoeksteen van de ‘gezichtsfeedback-hypothese’, de gedachte dat een emotionele gelaatsuitdrukking de stemming kan bepalen. Dit onderzoek is ruim 1400 keer geciteerd en wordt gebruikt door sommige therapeuten. Zeventien laboratoria uit de VS, Canada en Europa hebben onlangs de experimenten van Strack herhaald. Hiermee werd het onderzoek van Strack volledig onderuitgehaald, geen van de laboratoria kon enig effect vinden.

een bijdrage leveren aan het vergroten van transparantie van onderzoek en aan de kwaliteit en de volledigheid van het rapporteren van resultaten.

De pilot is met name gericht op studies waarvan de resultaten een grote impact hebben gehad. Bijvoorbeeld omdat ze veel geciteerd zijn, omdat ze in tekstboeken voor studenten zijn

opgenomen, onderzoek dat veel media-aandacht heeft gekregen en daarmee het publieke debat heeft beïnvloed, of onderzoeksresultaten op grond waarvan belangrijke beleidsbeslissingen zijn genomen. Het pilotprogramma Replicatiestudies heeft een looptijd van drie jaar (2017-2019), per jaar stelt NWO 1 miljoen euro beschikbaar. DANS is in het programma opgenomen als steun- en informatiepunt voor onderzoekers, en stelt haar data-archief EASY open voor het beschikbaar stellen van de basisbestanden en het duurzaam archiveren van de resultaten uit de replicatiestudies.

nwo.nl

COLOFON Uitgever: Stichting Uitgeverij E-data & Research Den Haag. Redactieadres: Postbus 93067, 2509 AB Den Haag, 070-3494450, edata@dans.knaw.nl, www.edata.nl.

Hoofd-/eindredacteur: Erica Renckens. Redactie: Marika de Bruijne, Steven Claeysens, Rutger Nugteren, Erica Renckens, Marieke Willems, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning.

Aan dit nummer werken mee: Heidi Berkhout, Elly Dijk, Valentijn Gilissen, Marjan Grootveld, Arjan van Hessen, Hella Hollander, Olaf Janssen, Jos Koldewij, Milo van de Pol, Nina Westzaan.

Opmaak: Colette Sloots, Haarlem. Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Oplage: 7500 papier, 1900 digitaal. ISSN: 1872-0374.

We hebben getracht alle belanghebbenden met betrekking tot het gebruikte beeldmateriaal te benaderen. Degenen die menen rechten te kunnen doen gelden, kunnen zich tot ons wenden. Toezending papieren en/of digitale versie is kosteloos aan relaties van de stakeholders en studenten in de alfa- en gamma-richtingen. Wilt u een artikel uit E-data overnemen? Neem dan contact op met de hoofdredacteur. Overname van artikelen geschiedt altijd onder vermelding van de bron E-data & Research en de naam van de auteur van het artikel.

Hoogleraar Visualizing Art History Robert Erdmann:

‘Je wilt niet verdrinken in de data, je wilt ze laten zingen’

Het Boschjaar 2016 is bijna ten einde. Prof. dr. Robert Erdmann ontwikkelde nieuwe computer- en visualisatietechnieken waarmee het oeuvre van Jheronimus Bosch onderzocht is.

Erica Renckens

“De mens heeft een beperkt visueel geheugen,” zegt Robert Erdmann terwijl hij switcht tussen twee afbeeldingen op zijn scherm. “Dat

maakt het maken van vergelijkingen heel lastig,” Erdmann werkt bij het Rijksmuseum en is als hoogleraar verbonden aan de Universiteit van Amsterdam en de Radboud Universiteit. Door een toevallige samenloop van omstandigheden belandde hij bij het Bosch Research and Conservation Project (BRCP), waarvoor hij beeldverwerkings- en visualisatietechnieken ontwikkelde.

“Als hoogleraar Material Science and Engineering in Arizona deed ik met NASA onderzoek naar de invloed van zwaartekracht op vloeistofstromingen in metaal. We lieten dunne staven metaal stollen in de ruimte, die we terug op aarde in honderden plakjes opdeelden. Van elk plakje maakten we honderden microscopische foto’s. Ik ontwikkelde technieken waarmee we die foto’s konden stitchen, naadloos op elkaar laten aansluiten.” De overstap naar de kunstgeschiedenis volgde na een rondleiding van een vriendin bij het Art Institute in Chicago. “Het conserveringsprogramma daar genereerde ook grote hoeveelheden foto’s die verwerkt moesten worden. Nadat ik daarover een lezing gaf bij het NIAS in Wassenaar raakte ik bij het Bosch-project betrokken.”

Stitchen en registreren

In het BRCP zijn de werken van Bosch op veel verschillende manieren gefotografeerd. “We hadden een prachtige Hasselblad-camera waarmee we soms wel 210 foto’s van 60 megapixels van één schilderij maakten. Met een speciaal filter maakten we met die camera ook infraroodfoto’s, waarop je door de oppervlaktelagen de onderliggende schildering of schets kunt zien. Dan hadden we nog een Osiris-camera, een speciale infraroodcamera met een nog langere golflengte. Soms maakten we opnames met UV-licht, waarmee bepaalde organische materialen oplichten. En musea beschikten soms over röntgenopnames.”

Per type opname moeten de beelden aan elkaar gesticht worden, als bij een panorama-foto. Erdmann: “Dat is extreem moeilijk om perfect te doen, zonder variatie in lichtinval en dergelijke.” Vervolgens moeten de verschil-

lende typen opnamen worden geregistreerd: precies op elkaar uitgelijnd worden. Maar stitchen geeft altijd een kleine afwijking en de afwijkingen in twee verschillende opnamen zullen nooit matchen. Erdmann ontwikkelde daarom een techniek waarin het stitchen en registreren gelijktijdig plaatsvindt en elkaar assisteert.

Alles in de cloud

“Maar nu hadden we een nieuw probleem gecreëerd: de bestanden zijn te groot om te openen in Photoshop. Ik zou ze bovendien niet makkelijke kunnen versturen naar de

projectleden, die over verschillende instituten verspreid zitten.” Uiteindelijk ontwikkelde Erdmann een complete online suite met visualisatietechnieken die de afbeeldingen instant toegankelijk maakt. “Iedereen kan er waar ook ter wereld van gebruikmaken, zolang je een

internetverbinding hebt. Alles staat in de cloud bij Amazon of op de server van het Bosch-project, in mijn appartement.”

Heb je dan geen megasnelle verbinding nodig? “We maken gebruik van een piramide van decompositie: het originele bestand is ook opgeslagen op halve resolutie, op een kwart, een zestiende, enzovoort. Zo zijn de verschillende zoomlevels alvast voorberekend. Alle afbeeldingen zijn opgeknipt in kleine stukjes en al die jpg’s hebben een eigen url. Zo vraagt de viewer steeds het minimale benodigde aantal data op van de server.”

Erdmann toont verschillende visualisatie-



Prof. dr. Erdmann belandde vanuit het onderzoek naar vloeistofstromingen in metaal in de kunstgeschiedenis. “Ook daar moeten grote hoeveelheden foto’s verwerkt worden.”

foto Herman Wouters

BoschDoc

In BoschDoc zijn bijna duizend schriftelijke documenten over Jheronimus Bosch, zijn naaste familie, zijn leven en zijn werk ontsloten. De databank bevat facsimiles, transcripties en vertalingen van handgeschreven en enkele gedrukte teksten tot 1800. Elk document is voorzien van een informerende tekst over de context, de bron en de relatie tot andere documenten in de databank. boschdoc.huygens.knaw.nl

technieken die hij ontwikkelde binnen het Bosch-project. In de Curtain Viewer staan vier typen opnamen van een doek twee bij twee op het scherm. Samen vormen ze één beeld, het pijltje van de muis is de kruising en

bepaalt zo welke opname het grootst is. “Zo kan je vergelijkingen maken binnen een schilderij, maar ook tussen schilderijen. Elke keer als je inzoomt of van plek verandert in de viewer, verandert de url direct mee. Als je die url naar je collega stuurt, krijgt hij precies te zien wat jij nu ziet.”

In de Fading Mode lopen twee opnamen via de muis geleidelijk in elkaar over en Rolling Registration is bedoeld voor het vergelijken van kopieën die onvoldoende matchen om volledig te kunnen registreren. De registratie is dan perfect op de plek van de muiscursor. In Morelli’s Vision zijn alle details op de doeken handmatig getagd. “Als ik op deze hand klik, toont hij alle handen uit de werken van Bosch, gesorteerd op visuele gelijkheid. Deze tools helpen om beter te kijken en om bewijzen te verzamelen om je punt te onderbouwen. Het is een overtuigendere, wetenschappelijkere aanpak dan zeggen: ‘Goh, ziet dit er niet anders uit?’ Je wilt alle data samenbrengen en er niet in verdrinken, maar ze laten zingen.” Dankzij deze technieken konden kunsthistorici vaststellen dat twee aan Bosch toegeschreven werken toch niet van zijn hand waren.

Observeren

Hoe wist je wat kunsthistorici en conservatoren nodig hadden? “Ik heb eerst heel lang geobserveerd hoe ze werken, waar frictie en frustratie was,” legt Erdmann uit. “Ik zag hele tafels bezaaid met afdrukken en boeken waar ze steeds tussen schakelden. Ik heb ook zelf geprobeerd met de opnamen te werken en veel gelezen over visualisatie. Ik bedacht veel ideeën die in de praktijk niet bleken te werken; het was een doorlopend proces.”

Erdmanns technieken zijn momenteel nog niet beschikbaar voor andere onderzoekers. “In de toekomst zal alles open source worden met een eenvoudige webinterface voor het stitchen en registreren. Maar tot dan zullen onderzoekers contact met mij moeten opnemen. Als zij dan de afbeeldingen naar mij opsturen, stuur ik de url’s terug. Dat doe ik graag.”

boschproject.org

INTERVIEW

‘Veel ideeën bleken in de praktijk niet te werken’



Uit de infraroodfoto van dit fragment van het schilderij ‘Johannes de Doper in de wilderness’ bleek dat Jheronimus Bosch oorspronkelijk een man had geschilderd onder de plant credits Hieronymus Bosch (circa 1450–1516) [Public domain], via Wikimedia Commons

SINDS KORT BESCHIKBAAR

Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij CentERdata, Data Archiving and Networked Services en Huygens ING.

CentERdata

• Wat doen Nederlanders in hun vrije tijd?
De nieuwste data van het onderzoek 'Social Integration and Leisure' zijn beschikbaar. De dataset maakt deel uit van de achtste meting van LISS Core Study en bevat gegevens van bijna 6000 respondenten. De data zijn verzameld in oktober en november 2015. Met dit longitudinale onderzoek is het mogelijk om de ontwikkelingen in tijd sinds 2008 te bestuderen. Het onderzoek gaat over tijdsbesteding aan onder meer vrijwilligerswerk, mantelzorg, sporten, media, hobby's, cultuur, online activiteiten en vakanties. Tevens worden de sociale contacten van de respondenten in kaart gebracht. De data zijn beschikbaar via LISS Data Archive.
lissdata.nl/dataarchive



Alkmaar foto Tasfoto/Shutterstock

Ook sinds kort beschikbaar:
Studies LISS panel
• Brummelman, E.; Thomaes, S.; Slagt, M.; Overbeek, G.; Orobio de Castro, B.; Bushman, B., februari 2013, On Parents Transferring Their Unfulfilled Ambitions Onto Their Child
• Gunnink, H.; Habraken, J.; Tranzo institute (Tilburg University), februari 2015, The opinion of the Dutch population about the intellectually disabled
• Naumann, E.; Stötzer, L., januari 2015, How does Immigration Influence Support for Redistribution?
• Vermeulen, F.; Soest, A. van, maart 2015, Time Use and Consumption - Wave 4



Deze bestanden zijn kosteloos beschikbaar via www.lissdata.nl/dataarchive. Bezoek deze site of scan de QR-code

DANS

• Rijke archeologische diversiteit in Vasse
RAAP Archeologisch Adviesbureau heeft in 2012 en 2014 opgravingen uitgevoerd op de locatie De Steenbrei 3 te Vasse in de gemeente Tubbergen. Daarbij zijn niet alleen palenrijen, -kransen en zeer bijzondere grafmonumenten met bijgiften uit de prehistorie aangetroffen, maar ook de resten



Opgraving van een grafveld en een herenhof in Vasse (Tubbergen) foto RAAP Archeologisch Adviesbureau

van een herenhof en andere bewoningssporen uit de Merovingische tijd (5e-8e eeuw). Het vondstmateriaal behorende bij de middeleeuwse bewoning toont dat het herenhof een hoge status had en intensieve contacten had met het noorden en het zuiden. De tabellen, digitale veldtekeningen en veldfoto's zijn via EASY beschikbaar gesteld.
dx.doi.org/10.17026/dans-xhq-etx5

Ook sinds kort beschikbaar:
• Weerden, drs. J.F. van der (BAAC bv) (2014): Een Spaans fort met een Frans jasje. Venlo, Kazerneterrein/MFC, opgraving. <http://dx.doi.org/10.17026/dans-xs6-kdtc>
• Kempenaar, G.F.H.M. (Bureau Oudheidkundig Onderzoek Rotterdam (BOOR)) (2015): Verslag van het documenteren van de funderingen van molen de Meerkoet in het Kralingse Bos in Rotterdam, vindplaats 06-47. <http://dx.doi.org/10.17026/dans-z7n-azrn>
• Loopik, J. (ADC ArcheoProjecten) (2015): Leven en dood aan het Schriksel - Venlo Q4. Een archeologische opgraving op de locaties Bergstraat-west en Maaskade-zuid. <http://dx.doi.org/10.17026/dans-x5b-rp4y>
• Meroño Peñuela, Dr A. (Vrije Universiteit Amsterdam); Ashkpour, A. (Internationaal Instituut voor Sociale Geschiedenis); Guéret, Dr C. (BBC) (2016): CEDAR RDF database, the Dutch historical censuses (1795-1971) as Linked Open Data. <http://dx.doi.org/10.17026/dans-xpk-wj5w>
• Dool, drs. R. van den (Mulier Instituut) (2016): Nationaal Sportonderzoek 2016. <http://dx.doi.org/10.17026/dans-266-pyjb>
• Shitova, MSc N. (RU Radboud Universiteit); Roelofs, Prof. Dr. A.P.A. (Donders Centre for Cognition) (2016): Locus of the Stroop-like interference in colour and picture naming. <http://dx.doi.org/10.17026/dans-z2r-kjs4>
• Sociaal en Cultureel Planbureau (SCP) (2016): Arbeidsaanbodpanel 1985 t/m 2014. <http://dx.doi.org/10.17026/dans-xum-7b2r>
• Biezen, MSc M van der (Radboudumc) (2015): Naar een optimale teamsamenstelling op de speedpost (towards an optimum ratio of NPs and GPs in out-of-hours primary care). <http://dx.doi.org/10.17026/dans-z2p-b85a>
• Centraal Bureau voor de Statistiek - CBS (2015): Zelfstandigen Enquête Arbeid 2015 - ZEA 2015. <http://dx.doi.org/10.17026/dans-zuz-hqbz>
• ROA, Research Centre for Education and the Labour Market, Maastricht University (2015): Project Onderwijs en Arbeidsmarkt tot 2020. <http://dx.doi.org/10.17026/dans-zjw-xkc5>
• Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS) (2016): Onderzoek Verplaatsingen in Nederland 2015 - OVIN

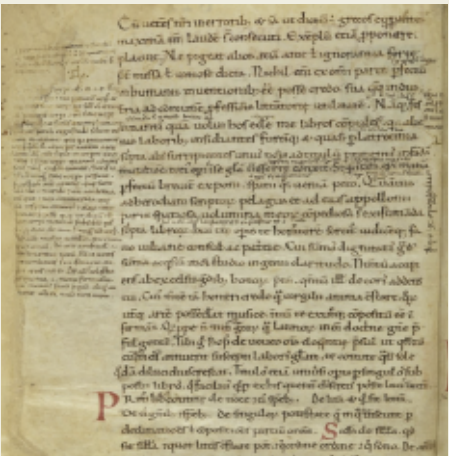
2015. <http://dx.doi.org/10.17026/dans-z38-prz4>



Via easy.dans.knaw.nl zijn deze bestanden beschikbaar. Bezoek deze site of scan de QR code.

Huygens ING

• Marginal Scholarship: speuren in de marge
In veel middeleeuwse manuscripten zijn de bladzijden gevuld met aantekeningen om de tekst heen en tussen de regels. Onderzoekers Mariken Teeuwen, Irene van Renswoude en Evina Steinová brachten de praktijken van het annoteren van teksten in de vroege Middeleeuwen in kaart. Het project Marginal Scholarship: The Practice of Learning in the Early Middle Ages (c. 800 - c. 1000) ontwikkelde een database met 350 geannoteerde vroegmiddeleeuwse handschriften. De zoekresultaten zijn te filteren op herkomst, periode, collectie in bibliotheek of archief en mate waarin de marges van de handschriften zijn gevuld.
marginalscholarship.nl



Manuscript van Ars Grammatica van Priscianus Caedariensis met aantekeningen in de kantlijn credits Bibliothèque Nationale de France

Ook sinds kort beschikbaar:
• Boschdoc: <http://boschdoc.huygens.knaw.nl/>
• Ystroom: <http://deystroom.huygens.knaw.nl/edition/>
• Briefwisselingen van de Hollandse en Friese Stadhoudersvrouwen (1605-1725): http://emlo-portal.bodleian.ox.ac.uk/collections/?page_id=2756



Deze publicaties zijn beschikbaar via huygens.knaw.nl/tools-en-data/. Bezoek deze site of scan de QR-code.

AGENDA

24 - 30 oktober • overal
Open Access Week
Internationale week om meer bekendheid te genereren voor open access in onderzoek. Dit jaar is het thema 'Open in Action'.
openaccessweek.org

7 november • Amsterdam
DH: Is the neutrality of technology possible or desirable?
Welke invloed heeft het gebruik van digitale technieken op de uitvoering van onderzoek? Korte presentaties uit de praktijk.
humanities.nl

9 november • Den Haag
Uitreiking Dataprijs
Wie wint de Nederlandse Dataprijs 2016? Kom naar de prijsuitreiking met interessante lezingen!
researchdata.nl/diensten/dataprijs

9 november • Amsterdam
Afsluiting COMMIT/
De publiek-private ICT-onderzoeksgemeenschap sluit haar activiteiten af in Club Panama.
commit-nl.nl

10 november • Nieuwegein
KNVI Jaarcongres
'Do it yourself!' Dat is het thema van het congres voor informatieprofessionals, georganiseerd door KNVI, SOD en Ngi-NGN.
congres.knvi.info

16 - 18 november • Wenen
Conference on Digital Heritage and New Technologies
Op deze 21e editie van CHNT staat de duurzaamheid en het hergebruik van archeologische en historische data centraal.
chnt.at

30 november • Londen
Uitreiking Digital Preservation Awards
De uitreiking van deze internationale prijzen vindt plaats tijdens de tweedaagse conferentie van het PERICLES-consortium.
dpconline.org

25 - 27 januari • Hilversum
Winter School for Audiovisual Archiving
Driedaagse training van het Nederlands Instituut voor Beeld en Geluid over het duurzaam bewaren van digitale audiovisuele collecties
beeldengeluid.nl/winterschool

4 februari • Utrecht
Grote Taaldag
Het grootste Nederlandse taalcongres wordt afgesloten met het Taalgala, waar o.a. de AVT/Anéla Dissertatieprijs wordt uitgereikt.
anela.nl

10 februari • Leuven
CLIN27
De jaarlijkse conferentie over computationele taalkunde in het Nederlandse taalgebied vindt dit jaar plaats in Leuven.
ccl.kuleuven.be/CLIN27

Ron Dekker (NWO) in aanloop naar Open Access Week:

‘Datamanagementplan hoort bij goede wetenschap’

Maandag 24 oktober gaat de internationale Open Access Week van start. Volgens Ron Dekker (NWO) gaat het in Nederland in rap tempo de goede kant op.

Erica Renckens

“Die data zijn niet van jou,” vermaande Dekker de onderzoekers nog in 2011 in dit blad. Sindsdien zijn onderzoekers zich meer bewust geworden van het nut en de noodzaak om data toegankelijk te maken, aldus Ron Dekker, die sinds september is gedetacheerd bij de Europese Commissie als nationaal expert op het gebied van open science. “Soms door negatieve ervaringen als datafraude en -manipulatie, waarbij de wetenschappelijke integriteit op het spel staat. Maar er zijn gelukkig ook positieve ontwikkelingen, zoals het Open Science-beleid van NWO en de NWO-pilot met een datamanagementplan (DMP) in het onderzoeksvoorstel.”

Data zijn een ‘asset’

Deze pilot is onlangs succesvol afgerond en is sinds oktober ingevoerd bij nieuwe programma's. Elke aanvraag bij NWO bevat voortaan een



Ron Dekker foto Monique van Zeijl

datamanagementparagraaf. Na toekenning moet de onderzoeker zo snel mogelijk een DMP opstellen. “Dat is ook in het belang van de onderzoeker,” aldus Dekker. “Het vooraf vastleggen en verzamelen van metadata die de data beschrijven, bespaart achteraf veel tijd en zorgt voor een hogere kwaliteit metadata. Het hoort in feite bij goed wetenschap doen.” Het opstellen van een DMP kost tijd, maar het hergebruik van data dat dit mogelijk maakt, bespaart ook weer tijd en geld. “Bovendien, als je weet dat jouw data beschikbaar zullen zijn voor derden, dan komt dat ook de wetenschappelijke integriteit ten goede.” Er moet nog wel een systeem van datacitatie

van de grond komen, beaamt Dekker. “Data zijn een ‘asset’ – onderzoekers hebben daar veel tijd en energie in gestoken, daar wil je op z’n minst de credits voor krijgen.”

Tijd voor actie

Dekker was projectleider Open Access bij het ministerie van OCW en speelde als adviseur van staatssecretaris Sander Dekker een belangrijke rol bij de voorbereiding van het Open Wetenschapsbeleid. In 2013 kondigde de staatssecretaris aan toe te willen werken naar 100% open toegankelijke publicaties in 2024. Onder zijn voorzitterschap vervroegde de Europese Raad voor Concurrentievermogen deze deadline zelfs naar 2020. Is dat niet wat ambitieus?

Volgens Dekker is het haalbaar: “De succesvolle onderhandelingen van de Nederlandse universiteiten laten zien dat grote uitgevers bereid zijn om én toegang tot hun tijdschriften te blijven verlenen én artikelen van Nederlandse onderzoekers direct via open access beschikbaar te maken. Dan gaat het snel met de percentages. Ook de Duitse Max Planck Gesellschaft pleit voor een switch naar open access-modellen voor publicaties in 2020. Er wordt al bijna 25 jaar over open access gepraat – nu is de tijd om daadwerkelijk tot actie over te gaan.”

openaccessweek.org

Corpora SoNaR en CGN voortaan gezamenlijk te doorzoeken

Niet meer verdwalen in tekst- en spraakwoud

In OpenSoNaR+ zijn het tekstcorpus SoNaR en het spraakcorpus CGN tegelijk doorzoekbaar. Er is in het systeem nog plaats voor andere corpora. Erica Renckens

“Wetenschappers die kijken naar taalgebruik zijn meestal geïnteresseerd in tekst én spraak,” aldus Nelleke Oostdijk, taaltechnoloog aan de Radboud Universiteit. “Neem nou voegwoorden: in spraak kom je bijna alleen ‘en’, ‘want’ en ‘dus’ tegen, maar in teksten zie je veel meer variatie met ook archaïsche voegwoorden zoals ‘desalniettemin’.” Het is slechts een van de onderzoeksvragen die gesteld kunnen worden met OpenSoNaR+, een interface waarmee twee grote dataverzamelingen gelijktijdig doorzocht kunnen worden.

In het nieuwe systeem zijn momenteel twee corpora opgenomen. SoNaR bestaat uit ruim 540 miljoen woorden aan moderne Nederlandse teksten die automatisch zijn voorzien van tags en lemma's. De teksten komen uit alle typen media en genre, die een gebalanceerde af-

spiegeling vormen van de hedendaagse schrijftaal. Het Corpus Gesproken Nederlands (CGN) bestaat uit 900 uur spraakopnames, gemaakt in een groot aantal verschillende situaties. De ruim 9 miljoen woorden zijn voorzien van transcripties, lemma's en woordsoortinformatie.

Schat aan data

Het CGN had al een eigen zoekinterface, COREX, maar deze bleek niet om te kunnen gaan met de grote hoeveelheden data uit SoNaR. “Tijdens de ontwikkeling van SoNaR was er helaas onvoldoende budget voor een exploitatie-omgeving,” herinnert Oostdijk zich. “Als je geen technische achtergrond had, verdwaalde je in al die data. Daardoor maakte maar een heel beperkte groep gebruik van deze schat aan data.”

“In een CLARIN-NL demonstratieproject is toen OpenSonar ontwikkeld, waarin het tekstcorpus voor iedereen doorzoekbaar werd, maar die interface was weer niet bruikbaar om ook spraak mee te doorzoeken,” aldus de projectleidster

van OpenSoNaR+, één van de laatste nog door CLARIN-NL gefinancierde projecten. “Met OpenSoNaR+ kan het nu eindelijk allemaal. Het systeem kan zelfs ook nog andere corpora inpassen, zoals het JASMIN-corpus met spraak van jongeren, anderstaligen en senioren.”

Snoepwinkel

Oostdijk is enthousiast over de mogelijkheden van de nieuwe zoekinterface: “Je kunt natuurlijk zoeken naar woorden, maar je kunt ook aanvullende eisen stellen, zoals ‘fiets’, maar alleen als werkwoord. Bij de resultaten zie je steeds een stukje context en kun je doorklikken om het hele fragment te lezen of te beluisteren. Het is een snoepwinkel voor onderzoekers.”

“We hopen dat het gebruik van de corpora zich nu meer zal gaan verspreiden. Communicatiewetenschappers wisten het CGN al wel te vinden, maar ook SoNaR is voor hen relevant. Dat corpus bevat ook teksten van sociale media, waar mensen vaak juist schrijven in spreektaal.”

opensonarplus.science.ru.nl



Oudezijds Achterburgwal 185, Amsterdam foto Geert Jan van Rooij

Nieuw: ‘KNAW Humanities Cluster’

Vanaf 1 oktober 2016 vormen het Meertens Instituut, Huygens ING en het Internationaal Instituut voor Sociale Geschiedenis (IISG) het KNAW Humanities Cluster. De instituten spelen hiermee in op internationale ontwikkelingen binnen het geesteswetenschappelijk onderzoek. Doel van de samenwerking is het versterken van vernieuwend, interdisciplinair, geesteswetenschappelijk onderzoek dat belangrijke wetenschappelijke en maatschappelijke thema's aansnijdt. Het KNAW Humanities Cluster bouwt ook aan een digitale infrastructuur om het geesteswetenschappelijk onderzoek methodologisch te ondersteunen. Datasets, digitale collecties en daarmee verbonden diensten worden in het KNAW Humanities Cluster gezamenlijk gevormd, ontwikkeld en beheerd als onderdeel van de nationale infrastructuur voor de geesteswetenschappen.

knaaw.nl

KORT

Nieuw dataplatform voor de sociale wetenschappen

Donderdag 27 oktober 2016 wordt in het Centraal Museum te Utrecht het Dataplatform voor de mens- en maatschappijwetenschappen gelanceerd. Binnen dit platform werken de Nederlandse universiteiten, NWO, CBS, DANS, Rijkskennisinstellingen, overheden en bedrijven samen om één netwerk met data ten behoeve van de mens- en maatschappijwetenschappen te creëren. Tijdens dit event wordt ook de naam van het platform, werktitel Nationale Data-infrastructuur voor de Sociale Wetenschappen (NDSW), bekendgemaakt. (MW) nwo.nl/onderzoek-en-resultaten/programmas/NDSW

Nieuwe site NWO toont grootschalige voorzieningen

Afgelopen zomer heeft de Permanente Commissie voor Grootschalige Wetenschappelijke Infrastructuur van NWO een nieuwe website gepresenteerd: onderzoeksfaciliteiten.nl. In deze online database kunnen onderzoekers, beleidsmakers en andere geïnteresseerden onderzoeksfaciliteiten vinden en mogelijkheden tot samenwerking verkennen. Alle vormen van grootschalige onderzoeksfaciliteiten uit de volle breedte van de Nederlandse wetenschap komen aan bod: van wetenschappelijke meetapparatuur tot databases en ICT-infrastructuren. (NWO) onderzoeksfaciliteiten.nl

Beeld en Geluid ontvangt keurmerk

Het Nederlands Instituut voor Beeld en Geluid heeft als eerste nationale audiovisuele archief het Data Seal of Approval (DSA) gekregen. Hiermee is het nationale omroeparchief gecertificeerd als 'Trusted Digital Repository' (TDR). Beeld en Geluid heeft volgens de beoordeling van het Data Seal of Approval aangetoond dat het de digitale duurzaamheid van de audiovisuele collecties kan waarborgen. Het instituut heeft met verschillende afdelingen meerdere jaren gewerkt aan het verkrijgen van deze certificering. Het ging hierbij zowel om het vastleggen van beleid en procedures voor digitale duurzaamheid van de audiovisuele collecties, als om het daadwerkelijk inrichten en aanpassen van de systemen, de processen en de werkwijzen.

beeldengeluid.nl

Archiefsysteem DANS uitgebreid

EASY toont zoekresultaten op de kaart

Het vinden van relevante archeologische datasets in het archiefsysteem van DANS is onlangs vereenvoudigd. Zoekresultaten worden op een kaart geprojecteerd.

Valentijn Gilissen

Via EASY, het archiefsysteem van DANS, zijn duizenden rapporten en grotere datasets van opgravingen uit de Nederlandse archeologie beschikbaar gesteld. Het zoeksysteem om deze data inzichtelijk te maken is eerder dit jaar uitgebreid met een kaartmodule.

Metadataveld

Stel dat een student onderzoek doet naar de archeologie van de binnenstad van Utrecht. Dan kan hij rechtstreeks zoeken op de tekst ‘Utrecht’, maar dat geeft te veel irrelevante datasets: de student is immers niet geïnteresseerd in een dataset over Kuinre, waarin melding wordt gemaakt van ‘de bisschop van Utrecht’. Beter kan hij gericht zoeken via *advanced search* in het metadataveld ‘coverage’. Maar een



Opgraving op het Vredenburg in Utrecht in het najaar van 2007 foto Projectorganisatie Stationsgebied en BAAC bv

dergelijke zoekopdracht naar ‘Utrecht’ geeft nog steeds veel meer dan enkel de datasets van projecten in de stad Utrecht: de zoekresultaten bevatten alle datasets uit de gehele provincie. Dit zijn meer dan 1500 resultaten, waarin hij de datasets die specifiek over de binnenstad gaan

niet zomaar kan terugvinden.

Onlangs is EASY daarom uitgebreid met een hulpmiddel dat het zoeken een stuk eenvoudiger en gebruiksvriendelijker maakt: zoekresultaten worden nu op een kaart geprojecteerd. Het archiefsysteem geeft in eerste instantie de zoekresultaten in

de vertrouwde lijst weer. Daarboven kan de gebruiker schakelen tussen de weergave van de treffers in een lijst of op de kaart.

Op deze manier komt de gebruiker snel bij de relevante resultaten. In eerste instantie toont EASY agglomeraties van resultaten. Bij inzoom-

men worden deze steeds nader verfijnd, totdat enkele resultaten worden gepresenteerd. Zo kan op de binnenstad van Utrecht worden ingezoomd, waardoor een gedetailleerd beeld ontstaat van de exacte locaties van de archeologische vindplaatsen in deze stad. Dit beeld kan vervolgens aangescherpt worden met aanvullende browse-opties en zoekopdrachten. De onderzoeker kan bijvoorbeeld op complextype en datering filteren en zo de zoekresultaten voor een specifiek gebied verder sorteren.

Verrijking

Momenteel gebruikt EASY voor de kaartweergave de Rijksdriehoekskoördinaten. De meeste archeologische datasets bevatten deze metadata, vandaar dat deze functionaliteit nu met name een verrijking biedt voor de archeologie. De kaartmodule is echter ook geschikt voor andere soorten datasets met een plaatsbepaling, ongeacht de discipline of het gebruikte meetsysteem. Deze uitbreiding van EASY is tot stand gekomen dankzij de steun van het Europese project ARIADNE (Advanced Research Infrastructure for Archaeological Dataset Networking in Europe).

easy.dans.knaw.nl

Sally Wyatt, programmaleider van de nieuwe, nationale eHumanities.NL:

‘You can’t do DH alone’

Op 20 mei ging in het Amsterdamse Trippenhuis, hoofdkwartier van de KNAW, eHumanities.NL officieel van start. We spraken met Sally Wyatt, een van de vier programmaleiders. Steven Claeysens



Sally Wyatt foto Moniek Wegdam

In de eerste plaats wil eHumanities.NL onderzoekers uit alle kamers van het academisch huis, maar ook erfgoedmensen en ICT-bedrijven samenbrengen. “Samen staan we sterker,” aldus Wyatt. “We zijn er voor iedereen die geïnteresseerd is in de nieuwe onderzoeksmogelijkheden die digitale technologie, digitale collecties en big data bieden. De concurrentie voor onderzoeksfinanciering is nationaal én internationaal groot en dus is samenwerking geboden. Dat geldt des te meer voor de Digital Humanities, waar een brede waaier aan competenties, vaardigheden en kennis vereist is. You can’t do DH alone.”

Interfacing is cruciaal

De Canadese Wyatt is senior onderzoeker bij het HuguensING en bijzonder hoogleraar Digital Cultures in Development aan de Universiteit Maastricht. Ze leidde eerder het ge-

lijksnamige vijfjarige eHumanities-programma van de KNAW. Gevraagd naar de concrete plannen van het nieuwe netwerk, benadrukt ze het belang van interfacing. “Zowel om samenwerking tussen alle disciplines te bevorderen, als om het bereik van de Digital Humanities te vergroten hebben we interfacing nodig – onderling, maar ook met andere groepen in de samenleving. Denk aan onderzoeksfinanciers, uitgeverij of het bedrijfsleven. We moeten ontmoeting en uitwisseling op alle niveaus bevorderen. Daarom gaan we bijvoorbeeld lezingen en cursussen organiseren, werken we

‘De concurrentie voor onderzoeksfinanciering is groot, dus samenwerking is geboden’

aan een overkoepelend sectorplan en zijn we in gesprek met een uitgever om een DH-reeks op te zetten.”

Drie speerpunten

Bij de eerste plannen voor eHumanities.nl, toen nog onder de naam Center for Humanities and Technology (CHAT), was ook IBM betrokken. Hoe staat het met de plannen om private partijen in het netwerk te betrekken? “We hebben drie speerpunten: fundamenteel DH-onderzoek stimuleren, de verdere ontwikkeling van een nationale DH-infrastructuur (CLARIAH) ondersteunen en waarborgen dat de Digital

Humanities daadwerkelijk impact hebben. Op al deze punten kan de private sector een stempel drukken. De meeste deelnemers aan het netwerk hebben vaak al goede banden met de industrie of de uitgeverij. We willen die banden zeker aanhalen.”

Public Humanities

“Vergeet ook niet de publieke functie van Digital Humanities. In Canada en de VS is het wellicht zichtbaarder dan in Europa, maar DH heeft ook een rol in het aanjagen van nieuwe vormen van Public Humanities: het publiek betrekken bij geesteswetenschappelijk onderzoek. Dat gebeurt niet alleen door een bredere verspreiding van onderzoeksresultaten via digitale wegen, maar ook door mensen volwaardig te betrekken bij de productie en creatie van kennis. Bekend zijn de vele crowd-initiatieven waarbij bijvoorbeeld getranscribeerd of geannoteerd wordt.”

Op 7 november komt eHumanities.nl in Amsterdam weer bijeen. Het middagprogramma, over de invloed van het ontwerp van digitale tools, algoritmes en software op het onderzoeksresultaat, is openbaar toegankelijk.

ehumanities.nl

Datacursus RDNL maakt kans op prijs

De cursus Essentials 4 Data Support van Research Data Netherlands (RDNL) maakt kans op een prestigieuze Digital Preservation Award op het gebied van Teaching and Communications. Deze internationale prijzen worden elke twee jaar uitgereikt aan mensen of organisaties die op een significante en innovatieve manier hebben bijgedragen aan het toegankelijk houden van ons wereldwijde digitale erfgoed. De prijsuitreiking vindt 30 november plaats in Londen.

De nieuwe ronde van de RDNL-cursus gaat in februari 2017 van start. De cursus is bedoeld voor data supporters: iedereen die onderzoekers wil ondersteunen bij het opslaan, beheren, archiveren en delen van hun onderzoeksdata. De twee groepsbijeenkomsten worden begeleid door coaches en omvatten presentaties van experts, online cursusmateriaal en wekelijkse opdrachten. RDNL wil met deze cursus een bijdrage leveren aan de professionalisering van en de afstemming tussen datasupporters.

Aanmelden kan tot 9 januari. (MG) datasupport.researchdata.nl dpconline.org

KORT

Eenvoudiger zoeken naar voorouders

De verschillende databases van het Centrum voor Familiegeschiedenis (CBG) zijn via een vernieuwde website tegelijk te doorzoeken. Voortaan krijgt de gebruiker met één zoekopdracht toegang tot archiefbronnen, stamboomgegevens, familienamen en familiewapens. Naast de collectie van het CBG zijn ook de ruim 43 miljoen aktes en 117 miljoen persoonsvermeldingen uit WieWasWie gelijktijdig te doorzoeken. WieWasWie is hét Nederlandse online platform voor familiegeschiedenis en biedt historische documenten die Nederlandse archiefinstellingen daar samen aanbieden voor historisch persoonsonderzoek.

cbg.nl

DH-congressen komen naar Utrecht

De komende jaren zullen twee grote congressen op het gebied van Digital Humanities plaatsvinden in Utrecht. Eerst zal van 3 tot en met 5 juli 2017 Digital Humanities Benelux neerstrijken in de stad. Dit congres, dat dan voor de vierde keer zal plaatsvinden, wordt afwisselend gehouden in Nederland, België en Luxemburg. Vervolgens komt in juli 2019 het prestigieuze Digital Humanities-congres naar Utrecht, mede dankzij een investering van CLARIAH.

dhbenelux.org
adho.org

Nieuwe richtlijnen data Horizon2020

De Europese Commissie heeft nieuwe richtlijnen gepubliceerd ten aanzien van onderzoeksdata voortkomend uit Horizon 2020-projecten. Deze richtlijnen gelden voor onderzoeksprojecten die onderdeel zijn van de Open Research Data Pilot (ORDP). Onderzoekers uit negen vastgestelde vakgebieden moeten hun onderzoeksdata als open data deponeren in een repository. Opting out behoort tot de mogelijkheden, bijvoorbeeld ter bescherming van de privacy. De onderzoekers zijn bovendien verplicht een Data Management Plan (DMP) in te leveren. Nieuw in deze richtlijnen is dat één DMP voor het hele onderzoeksproject moet worden geschreven en niet meer een DMP per dataset. Ook moeten onderzoekers werken volgens de zogenaamde FAIR data principles, oftewel de data moeten Findable, Accessible, Interoperable and Reusable zijn. De ORDP geldt vanaf 2017 voor alle H2020-projecten. (ED)

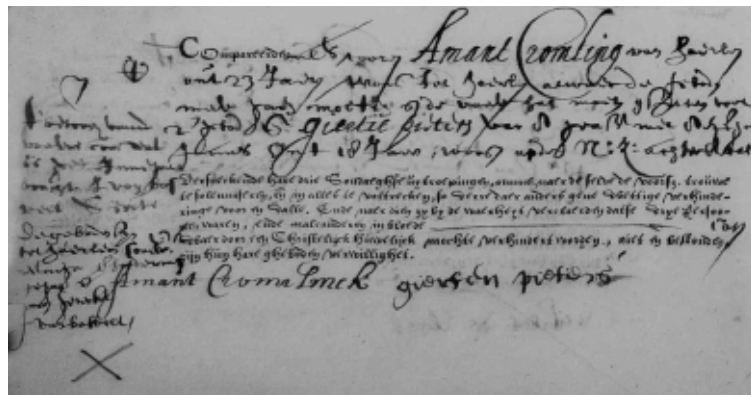
ec.europa.eu/programmes/horizon2020

Citizen science-projecten verzameld op platform

Vrijwilligers helpen mee

Het nieuwe platform IedereenWetenschap-per.nl verzamelt projecten waarin burgers de wetenschap een handje helpen. Prof. dr. Tine de Moor had in haar project ‘Ja, ik wil!’ veel aan deze hulp. *Erica Renckens*

Wetenschappers doen steeds vaker een beroep op burgers bij hun onderzoeksprojecten. Vrijwilligers helpen bij het digitaliseren van archieven, het tellen van bodemdierpjes of het meten van fijnstof. In 2011 lanceerde Stadsarchief Amsterdam al Vele-Handen.nl, een online platform waar musea en archieven hun scans en foto's aanbieden. Tine de Moor, hoogleraar Economische en sociale geschiedenis aan de Universiteit Utrecht, vertelt tijdens het KNAW-symposium Citizen Science (16 juni j.l.) over haar ervaringen hiermee. “In het project ‘Ja, ik wil!’ hielpen zo’n vijfhonderd vrijwilligers bij het



Ondertrouwinschrijving Amant Cromling en Giertie Pieters credit Ja, ik wil!, Universiteit Utrecht

digitaliseren van ondertrouwaktes vanaf de zestiende eeuw uit het archief van Amsterdam,” vertelt ze. “Een kleine harde kern van ongeveer dertig mensen heeft bijna 80 procent van het werk gedaan. Hoe ouder de akte, hoe moeilijker die te lezen is. Dus om vrijwilligers uit te dagen, nodigden we de meer ervaren mensen uit verder terug in de tijd te gaan.”

Goede communicatie is heel belangrijk, aldus De Moor. “Zo’n project is tweerichtingsverkeer. Je moet van tevoren duidelijk uitleg geven over de doelen, de beoogde resulta-

ten en de tijdsplanning. Daarnaast is persoonlijke aandacht belangrijk. Ga in op vragen en verzoeken, en organiseer rondleidingen, lezingen en cursussen.”

Vrijwilligers

De inzet en begeleiding kosten volgens de hoogleraar veel tijd, geld en energie, maar met die investering kan veel worden bereikt. “Het is nergens voor nodig om huiverig te zijn over de kwaliteit van de data uit zo’n project,” stelt De Moor. Elke akte uit ‘Ja, ik wil!’ is door twee verschillende vrijwilligers ingevoerd. Een

derde persoon vergeleek de ingevoerde teksten, waarna de onderzoekers uitsluitel boden voor de gemarkeerde verschillen. Met betaalde arbeidskrachten zou er onvoldoende budget voor een dergelijk project zijn geweest.

Tijdens het symposium presenteren Eos Wetenschap en de Vlaamse Jonge Academie IedereenWetenschappers.nl. Dit platform moet dé Nederlandse en Vlaamse ontmoetingsplek voor wetenschappers en burgers in citizen science-projecten worden. Inmiddels staan er al ruim tachtig projecten op de site. “Alle projecten voldoen aan de richtlijnen van de European Citizen Science Association (ECSA),” aldus Liesbeth Gijzel van Eos Wetenschap. “Een project moet tot een wetenschappelijk resultaat leiden en mag geen ‘klassiek’ onderzoek zijn, al is die grens soms moeilijk te trekken. Maar sociologisch onderzoek met enquêtes waar de deelnemers persoonlijk niks aan hebben, is geen citizen science. Via Eos Wetenschap brengen wij de projecten blijvend onder de aandacht.”

iedereenwetenschapper.nl

Van Dijk breidt kennis statistische methoden uit in Michigan, VS

‘Summerschool is een unieke ervaring’

Om de overlevingskansen van families te ontrafelen, duikt Ingrid van Dijk, PhD-student aan de Radboud Universiteit, in historische data. *Marion Wittenberg*

Bij de vakgroep Economische, Sociale en Demografische Geschiedenis in Nijmegen werkt sociologe Ingrid van Dijk aan een proefschrift over sterfte in families in de 19e eeuw. Hoe komt het dat in een tijd dat bijna de helft van de kinderen overleed voor het vijfde levensjaar, kindersterfte zich in bepaalde families concentreerde, terwijl andere families dit leed werd bespaard? “Voor mijn onderzoek gebruik ik het databestand LINKS van het Internationaal Instituut voor Sociale Geschiedenis,” vertelt Van Dijk. “Dit bestand bevat de geboorte-, huwelijks- en overlijdensaktes geregistreerd tussen 1812 en 1917 van miljoenen Nederlanders uit de provincies Zeeland, Drenthe, Groningen en Limburg. Deze certificaten zijn aan elkaar gelinkt op basis van de namen van de personen en hiermee zijn familieconstructies gemaakt.” Met behulp van deze data volgt Van Dijk families door de tijd

JONG TALENT

heen. Ze kijkt met name naar gezinnen met een hoge kinder- en zuigelingensterfte. “Ik bekijk de overlevingskansen van hun overlevende broertjes en zusjes op de lange termijn, en naar de sterfte in de eventuele volgende generatie.”

‘Aantonen hoe sterfte zich verspreide door familienetwerken’

Op het moment is ze vooral bezig met de datacleaning en -preparatie. De familieconstructies moeten afgemaakt en verbeterd worden, daarnaast maakt ze allerlei demografische variabelen. Van Dijk: “Doordat we werken met hele oude data die destijds ook niet voor dit doel zijn verzameld, weten we relatief weinig over de gezinnen die we volgen. Daardoor kunnen we een aantal interessante kenmerken van

mensen helaas niet in het onderzoek betrekken. Ik ga proberen om een aantal ideeën die ik niet kan toetsen op mijn eigen data op buitenlandse datasets te toetsen.”

Opgeleid als kwantitatief socioloog, met een researchmaster in vergelijkend onderzoek heeft Van Dijk al een stevige achtergrond in de statistiek. Toch had ze behoefte aan een update van haar kennis op het gebied van statistische methoden. Hiervoor heeft ze deze zomer cursussen gevolgd bij de ICPSR Summerschool in Ann Arbor, Michigan. Volgens haar ‘een unieke ervaring’ die ze sterk kan aanbevelen aan alle andere (beginnende) wetenschappers. “Met deze methoden hoop ik te kunnen aantonen hoe sterfte zich verspreide door familienetwerken, en welke karakteristieken een causale invloed uitoefenden op vroege sterfte in families.”

Het promotieonderzoek ‘Death and the family’ maakt deel uit van het NWO-project ‘Genes, Germs and Resources’.

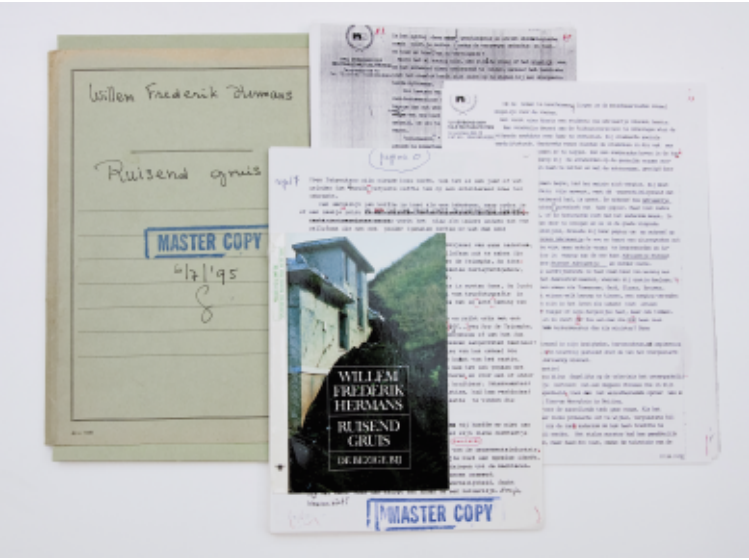
ru.nl/genesandgerms



Ingrid van Dijk
foto Karin van Dijk

ICPSR Summerschool

De ICPSR Summerschool is wereldwijd toonaangevend in cursussen op het gebied van kwantitatieve methoden van sociaalwetenschappelijk onderzoek. Studenten en onderzoekers uit verschillende disciplines en meer dan 40 verschillende landen volgen elke zomer onderwijs in een breed scala aan geavanceerde statistische technieken. DANS stelt elk jaar subsidie beschikbaar voor researchmasterstudenten en PhD's om hieraan deel te nemen. Momenteel onderzoekt DANS of deze subsidieregeling kan worden uitgebreid naar soortgelijke summerschools voor andere vakgebieden.



Bronnenmateriaal ‘Mandarijnen op zwavelzuur’, W.F. Hermans archief credits wfhermansvolledigewerken.nl



Nieuwe software helpt bij collationeren

CollateX vergelijkt automatisch teksten

Tientallen versies van een tekst met elkaar vergelijken is een arbeidsintensieve klus. Bij Huygens ING experimenteerden onderzoeker en ontwikkelaar met software die hierbij helpt.
Marieke Willems

Door verschillende versies van een tekst te vergelijken, kunnen onderzoekers te weten komen wat de oorspronkelijke brontekst van bijvoorbeeld de bijbel is. Ook voor onderzoek naar de totstandkoming van een tekst is dit ‘collationeren’ een geschikte methode, die letterkundigen bijvoorbeeld gebruikten in Volledige Werken, een Huygens ING-project over het werk van W.F.

Hermans. Daarnaast kan collationeren de kwaliteit van getranscribeerde handgeschreven brieven verbeteren wanneer verschillende transcripties elkaar aanvullen.

Basistekst kiezen

Wanneer een onderzoeker handmatig collationeert, kiest hij eerst een basistekst, een versie van de tekst waartegen hij alle andere versies vergelijkt. In het geval van honderden versies is dat een moeilijke keuze die een bias met zich meebrengt. Een internationaal team van ontwikkelaars ontwierp daarom CollateX, software die meerdere versies onderling vergelijkt zonder keuze van een basistekst. Een computer kan tientallen tekstversies met

elkaar vergelijken in enkele seconden of minuten en doet dat consistent en zonder bias. De gebruiker kan de keuze voor een basistekst zo uitstellen tot een later moment in het onderzoek of zelfs helemaal achterwege laten.

Training onderzoeker

Het is belangrijk dat onderzoekers begrijpen hoe de software werkt, stelt Ronald Haentjens Dekker, lead engineer in Huygens ING en ontwikkelaar van CollateX. “Indien de computer en de gebruikte tools een zogenaamde ‘blackbox’ blijven, is het onduidelijk welke aannamen gedaan zijn tijdens het onderzoek en hoe betrouwbaar het resultaat is.” Haentjens Dekker trainde daarom

onderzoekster Elli Bleeker van de Universiteit van Antwerpen. Bleeker onderzocht hoe de computer kan worden ingezet voor tekstgenetisch onderzoek, waarbij het schrijfsproces belangrijker is dan het eindproduct. Dit resulteert in zeer complexe transcripties met meerdere ‘schrijflagen’. Haentjens Dekker en Bleeker experimenteerden binnen CollateX met het collationeren van twee XML-files, inclusief alle auteurscorrecties, schrijflagen en andere tags. De resultaten waren veelbelovend en worden verder uitgewerkt. Bleeker concludeert: “Al met al heeft mijn training en onderzoek in CollateX zeer waardevolle inzichten opgeleverd. Niet alleen op het gebied van geautomatiseerde tekstcollatie, maar evengoed over het belang van het kennen én begrijpen van de software die je gebruikt, en over de ideale vorm van samenwerking tussen onderzoeker en ontwikkelaar.”

huygens.knaw.nl/collate-x
github.com/interedition/collatex

GELEZEN

Peer Review Quality and Transparency of the Peer-Review Process in Open Access and Subscription Journals,
J.M. Wicherts
Milo van de Pol

Veel wetenschappelijke Open Acces Journals nemen het niet zo nauw met het peer-review-proces. OA-uitgevers zouden economisch gewin een belangrijker afweging vinden dan wetenschappelijke betrouwbaarheid. Althans, dat is de mening van heel wat wetenschappers. Een berucht voorbeeld dat vaak wordt aangehaald is het hoax artikel van wetenschapsjournalist John Bohannon dat door bijna de helft van alle OA-journals zonder meer werd geaccepteerd. Jelte Wicherts, psycholoog en UD aan de Tilburg University, zocht uit of bovenstaande mening wel klopt. Hij ontwikkelde een vragenlijst waarmee de transparantie van het peer-reviewproces bij zowel OA als bij traditionele wetenschappelijke tijdschriften kan worden gemeten. Wicherts' onderzoek toont aan dat enerzijds veel OA-tijdschriften zich inderdaad schuldig maken aan een falend peer-reviewproces, anderzijds zijn er ook traditionele, subscription based Journals die er een potje van maken. In die laatste groep bevinden zich ook enkele High Impact bladen. De tool van Wicherts om de kwaliteit van peer review bij zowel Open Acces als trationele wetenschappelijke tijdschriften te meten is beschikbaar via: qoam.eu

COLUMN

#Digilemma

Ooit had ik een grote fysieke bibliotheek. Zo’n veertig kasten vol boeken en tijdschriften, vooral op het gebied van taal- en letterkunde. Maar die collectie werd me tot last en ik maakte er niet optimaal gebruik van. Daarom heb ik me erin verdiept hoe je boeken moet digitaliseren en indexeren, zodat je grote collecties in één keer kunt doorzoeken. Met vallen en opstaan heb ik dat onder de knie gekregen. Nu heb ik een grote digitale bibliotheek. Ruim honderd zestigduizend boeken en afleveringen van tijdschriften, het overgrote deel in pdf-formaat. Veel heb ik zelf gescand, of laten scannen door studenten, maar ik heb ook het nodige van internet geplukt. Met regelmaat word ik benaderd door wetenschappers met de vraag of ik ze bepaalde bronnen digitaal kan leveren. Of een corpus. Bij de Digitale Bibliotheek voor de Nederlandse Letteren zijn, vanwege de beperkingen van de Auteurswet, nauwelijks literaire bronnen van na 1950 voorhanden. Digitaliseren voor ‘thuis-

gebruik’ mag van die wet. Ik heb juist duizenden boeken van na 1950 gedigitaliseerd, want die zijn doorgaans makkelijker te scannen dan oude boeken.

Die verzoeken van wetenschappers stellen me voor een dilemma en ik hoor graag uw mening (het liefst via Twitter, @ewoudsanders #digilemma). Hoe moet ik op dergelijke verzoeken reageren?

Als ik me aan de Auteurswet wil houden, luidt het antwoord altijd: nee. Ja, ik heb alle spellinggidsen van het Nederlands gedigitaliseerd en ja, ik zou je aan een corpus van duizend romans uit de jaren zestig kunnen helpen. Maar nee, ik kan dat niet doen, want daarmee overtreed ik de Auteurswet.

Overigens valt het me op hoe ‘kleinschalig’ sommige verzoeken zijn. Er is mij weleens gevraagd of ik vijftientig titels kon leveren voor een onderzoek naar verfelstructuren in naoorlogse Nederlandse romans. Aangezien er sindsdien tienduizenden romans zijn gepubliceerd, snap ik niet hoe zo’n piepkleine selectie ooit representatief zou kunnen zijn, maar de details van dat onderzoek zijn me niet bijgebleven.



foto Leo van Velzen

Ik zou dus altijd nee moeten antwoorden, maar in de praktijk zeg ik vaak ja. Ik leef van mijn pen en ik begrijp het belang van de Auteurswet, maar mijn hart gaat uit naar de wetenschap. Dus toen ik onlangs werd benaderd door een wetenschapper die dolgraag samen met zijn studenten enkele naamkundige naslagwerken digitaal wilde kunnen doorzoeken, ben ik meteen gaan kijken of ik die toevallig al digitaal beschikbaar had.

Als ik iemand digitale bronnen lever, zeg ik er altijd bij dat ze alleen voor ‘thuisgebruik’ zijn en dat ze nooit zomaar op internet terecht mogen komen, maar ik begrijp goed dat ik daar geen controle over heb. Bij grote verzoeken (‘Het liefst alle duizend

streekromans!’) antwoord ik: je kunt ze bij mij thuis komen raadplegen. In feite vind ik dit gerommel in de marge. Het zou niet nodig moeten zijn dat wetenschappers zich wenden tot een goedbedoelende particulier. De belangrijkste standaardwerken op alle vakgebieden zouden onder bepaalde voorwaarden voor wetenschappers en studenten beschikbaar moeten worden gesteld, simpelweg omdat dit de wetenschap vooruit helpt. Hetzelfde geldt voor goed gestructureerde, representatieve literaire corpora. Ik hoor graag ideeën over hoe dat zou kunnen worden gerealiseerd, ondanks de beperkingen van de Auteurswet, die dateert uit 1912 en mijns inziens hoognodig aan een update toe is.

Ewoud Sanders
Taalhistoricus en journalist.
Sanders is vaste medewerker van onder meer NRC Handelsblad en Onze Taal.