



E-DATA & RESEARCH

Jaargang 9 | 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

3 **Twitter: hype of here to stay voor de wetenschap?**

4 **Deze datasets zijn ook sinds kort beschikbaar**

5 **CLARIAH werkt aan een digitale infrastructuur**



6 **Lokale kennis als data mee in onderzoek**

6 **Deuren open voor virtueel onderzoek**

7 **RDA-week in beeld**

7 **KB: toegang tot ieder wetenschappelijk artikel**

8 **Hay Kranen: 'Wikipedia goudmijn voor onderzoek'**

8 **De bomen en het bos binnen de e-humanities**



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

Online toegankelijk maken en delen van data bekroond en beloond

Liederenbank en DINED winnaars Dataprijs 2014

De Nederlandse Dataprijs voor het toegankelijk maken en delen van onderzoeksdata gaat dit jaar naar de Nederlandse Liederenbank en DINED. *Heidi Berkhout*

Op 24 oktober jl. werd voor de derde keer de Nederlandse Dataprijs uitgereikt. € 7.500, een beeld en een diner voor de onderzoeksgroep, met die prijzenpot gingen de Nederlandse Liederenbank (winnaar Dataprijs humaniora en sociale wetenschappen) en DINED (winnaar Dataprijs exacte en technische wetenschappen) naar huis.

Nog lang niet klaar

"We zijn heel erg blij met de Dataprijs!", vertelt Martine de Bruin, lid van het projectteam van de Nederlandse Liederenbank. "De jury noemt onze database een 'grote aanwinst voor zowel het Nederlands academisch als cultureel erfgoed'. Dat is een bevestiging dat we op het goede spoor zitten. Het prijzengeld besteden we aan onze melodieënzoekmachine. Hiermee zijn beschreven melodieën te vergelijken en kunnen titelloze of gevarieerde melodieën worden geïdentificeerd. En daarvoor is veel data nodig. Het prijzengeld besteden aan een pilotproject lijkt ons een goede voorbereiding op een grotere subsidieaanvraag. Maar misschien wordt het ook nog iets anders. Er zijn nog zoveel verrijkingen en verdiepingen mogelijk: de Liederenbank is nog lang niet klaar!" Juryvoorzitter prof. Kees Aarts: "In de praktijk wordt de databank al door veel wetenschappers binnen de digital humanities gebruikt voor literair-, muziek-, godsdienst-, algemeen historisch en etnologisch onderzoek. Ook journalisten, musici en middelbare scholieren gebruiken de databank veel. Deze goed gestructureerde databank is daarom een grote aanwinst voor zowel het Nederlands academisch als cultureel erfgoed."

Begrijpelijke data

Ook Johan Molenbroek en Marijke Dekker, de mensen achter DINED, zijn verrast en blij: "Ons doel is altijd geweest om onbegrijpelijke data



Prof. Louis Grijp (Meertens Instituut) ontving de Dataprijs voor de Nederlandse Liederenbank, een database van 170.000 Nederlandse liederen van de middeleeuwen tot de 20e eeuw. Ir. Marijke Dekker is namens dr.ir. Johan Molenbroek (TU Delft) voor DINED aanwezig, een database met meetgegevens over de variatie van menselijke vorm en afmetingen foto Annemiek van der Kuil

begrijpelijk te maken. En dat lukt: met onze interactieve tool wordt de variatie in de afmetingen van het menselijk lichaam duidelijk, zonder dat de gebruiker veel kennis over anatomie en statistiek nodig heeft. Naast studenten gebruiken ook bedrijven en ziekenhuizen onze database. De vele vragen geven steeds weer aanleiding om verbeteringen aan te brengen en nieuwe data toe te voegen, zoals menselijke krachten. Onze faculteit Industrieel Ontwerpen van de TUDelft stelt ons steeds weer in de gelegenheid om het benodigde programmeerwerk door een student-assistent te laten uitvoeren. En door het winnen van de Dataprijs kunnen we nu ook een paar grotere, al langer gewenste verbeteringen maken."

Uitnodigend en innovatief

Juryvoorzitter prof. Karel Luyben benadrukte namens de andere juryleden de praktische kanten en de mooie interface van DINED. "De dataset is

heel praktisch opgebouwd waardoor het maatschappelijk nut groot is. Het innovatieve karakter van de set zal veel wetenschappers uitnodigen voor verder onderzoek. Door veel aandacht te schenken aan zowel de verzameling, beschrijving en het beschikbaar stellen van de set, is deze voor een groot publiek zeer toegankelijk."

Meer datasets

Naast de winnaars waren vier andere datasets genomineerd. Al zijn deze datasets niet bekroond, alle inzendingen waren mooie en inspirerende voorbeelden van het toegankelijk maken en delen van onderzoeksdata. Initiatiefnemer Research Data Netherlands is voornemens over twee jaar de prijs weer uit te reiken.

liederenbank.nl
dined.io.tudelft.nl
researchdata.nl

KB ResearchLab digitale speeltuin voor wetenschappers

Online aan de slag met data en tools

Wetenschappers kunnen in het KB ResearchLab zelf experimenteren met data van de KB en open-source tools. *Inge Angevaere*

De Koninklijke Bibliotheek (KB) heeft ruim 9 miljoen krantenpagina's, 1,5 miljoen tijdschriftpagina's en zo'n 90.000 andere publicaties gedigitaliseerd. Al die bestanden staan online op delpher.nl, waar iedere Nederlander ze kan doorzoeken. Maar onderzoekers willen vaak

meer dan alleen maar zoeken. Daarom biedt de KB sinds kort een omgeving aan waar wetenschappers met data en tools kunnen experimenteren: het KB ResearchLab.

Resultaten clusteren

Willem Jan Faber, manager van het Lab bij de KB: "Niet alleen de wetenschappers maar ook wijzelf staan nog aan het begin van een ontdekkingsstocht in Big Data. Om het maximale rendement uit die data te halen, werken we samen met onder-

zoekers. We bieden tools en data aan en nodigen wetenschappers uit om ermee te gaan werken." Ontwikkelt de KB die tools ook? Faber: "De KB is natuurlijk vooral een leverancier van content. Maar uit onze samenwerking met onderzoekers ontstaan soms ook tools die de content beter doorzoekbaar maken of die helpen om resultaten te clusteren. Die kunnen we dan weer aanbieden aan andere wetenschappers. De samenwerking met onderzoekers is trouwens niet helemaal nieuw. De

afgelopen jaren hebben we al wat ervaring opgedaan, bijvoorbeeld met taalkundige Marc van Oostendorp (zie E-data februari 2014, red), met wie we een 'scandeermachine' bouwden. Samen met het Instituut voor Nederlandse Lexicologie hebben we gewerkt aan een tool voor het taggen van 'named entities' (eigenamen). Nu zijn we zo ver dat we de mogelijkheden van ons lab breder kunnen aanbieden, aan alle onderzoekers die willen experimenteren."

VERVOLG OP PAGINA 2

GEHOORD & BIJGEWOOND

DH2014: de volle breedte van de Digital Humanities

Steven Claeysens

Het grootste Digital Humanities-congres vond dit jaar plaats in Lausanne. Meer dan 700 deelnemers verzamelden zich nabij het meer van Genève voor een volle week presentaties, paneldiscussies, postersessies en praktische workshops. Geesteswetenschappers uit alle windstreken bespraken nieuwe en oude digitale onderzoeksmethoden en ontvouwden hun onderzoeksplannen. Uit Nederland kwamen onder meer aansprekende bijdragen van het Translantis-team (translantis.nl), het Riddle of Literary Quality-project (literaryquality.huysens.knaw.nl) en het PoliMedia-project (polimedia.nl).

De breedte van het congres werd aangezet tijdens keynotes van Bethany Nowviski en Bruno Latour. Nowviski maakte een rondgang langs lopende onderzoeks- en kunstenaarsprojecten die de geschiedenis van de mensheid als maat nemen en hield zo een pleidooi om in de Digital Humanities de Longue Durée in ere te herstellen. Latour stelde het ‘digitale’ in de Digital Humanities ter discussie. Hij benadrukte de materialiteit van ‘alles digitaal’ en wees op de toenemende gelijkens tussen het boek en het web. Vroeger schreven mensen in scriptoria, vandaag schrijven ze in screentoria, aldus Latour. Volgend jaar is het congres in Sydney.

dh2014.org

ICPSR summerschool: verdieping en verbreding

Wouter Quite en Nina Conkova

Ieder jaar organiseert het Interuniversity Consortium for Political and Social Research (ICPSR) een summerschool met allerhande cursussen op het gebied van sociaalwetenschappelijke onderzoeksmethoden. Wouter Quite en Nina Conkova waren erbij, deels gesubsidieerd door



ICPSR summerschool leidt tot nieuwe vriendschappen voor Nina (links) en Wouter (vierde van links) foto Marcelo Lima

DANS. Wouter: “Ik had het geluk in het kader van mijn promotie-onderzoek naar de summerschool te mogen gaan. Het bezoek aan Ann Arbor was een zeer leerzame ervaring, waar ik naast de verdieping en verbreding van mijn methodologische kennis ook vele nieuwe contacten heb opgedaan. Een ervaring die ik niet had willen missen.” Nina Conkova vult Wouter aan: “Door deze summerschool heb ik mijn kennis in statistische methoden op kunnen frissen en mijn analytische vaardigheden kunnen uitbreiden. Sterker nog, ik heb genoten van het grote gezelschap van collega-promovendi van over de hele wereld en van de stukjes van de Amerikaanse cultuur!”

icpsr.umich.edu

Referentieculturen vinden in taal met online tools

Pim Huijnen

Kunnen historici gebruik maken van digitale krantenarchieven en digitale methoden en technieken? Die vraag stond centraal tijdens de conferentie ‘Reference Cultures and Imagined Empires from a Western Perspective, 1850-2000’ die 11, 12 en 13 juni plaatsvond in Utrecht. De conferentie was georganiseerd ter uitwerking van het begrip ‘reference

cultures’, dat als theoretisch fundament dient voor twee onderzoeksprojecten die vanuit de Universiteit Utrecht worden gecoördineerd: het NWO-project Translantis en het door HERA gefinancierde Asymenc. Deelnemers vonden elkaar in het idee dat ‘reference cultures’ in de taal moeten worden gezocht: de wijze waarop mensen andere culturen als referentie gebruiken (en veranderingen daarin: van ‘Amerikaanse droom’ naar ‘Amerikaanse toestanden’) zegt iets over hun cultuur of mentaliteit. Diverse sprekers toonden hoe digitale methoden en technieken kunnen bijdragen aan de studie hiervan. Zo liet Bob Nicholson van Edge Hill University op basis van zijn onderzoek naar digitale krantenarchieven zien hoe modewoorden (‘skedaddle’) en grappen in de negentiende eeuw vanuit de Verenigde Staten in Engeland terecht kwamen (zoek op Youtube op ‘Skeddadlemania’ voor een impressie). Evenzo inspirerend was de presentatie van Matthias Lemke van de Helmut-Schmidt-Universität Hamburg. Met behulp van een zelf ontwikkelde tekstanalyse-tool kon hij de stelling bewijzen dat Duitse kranten de afgelopen decennia in toenemende mate gebruik maken van neoliberale argumentatie om

politiek beleid te beschrijven – en daarmee de vraag die centraal stond in de conferentie bevestigend beantwoorden.

translantis.wp.hum.uu.nl/conferences

DHBenelux 2014: digitale wetenschappers verbinden

Joris van Zundert

De DHBenelux is een nieuwe conferentie met als onderwerp: digitale geesteswetenschappen (DH). Tijdens deze conferentie kunnen professionals die digitale componenten gebruiken in geesteswetenschappelijk onderzoek hun werk presenteren. Het doel is om (nader) kennis te maken met digitale geesteswetenschappen en in contact te komen met collega’s die digitale methoden en technieken gebruiken. De eerste editie van de conferentie vond plaats op 12 en 13 juni jongstleden bij het Huygens ING en de Koninklijke Bibliotheek in Den Haag. In haar keynote wenste professor Melissa Terras van het University College London Centre for Digital Humanities de organisatoren veel succes met het verbinden van de community van digitale geesteswetenschappers in de Benelux. De conferentie bood een breed overzicht van projecten, groen en rijp, waarin de bezoeker naar hartelust kon grasduinen: van concrete projecten die de historie van de bevrijding van Nijmegen in kaart brengen tot meer reflectieve bijdragen over de rol van digitale methoden in filosofie. Het afsluitende discussiepanel belichtte een aantal *hot topics* in de digitale geesteswetenschappen, zoals de rol van digitale methoden in het onderzoek, en de mate waarin deze onderdeel moeten zijn van het curriculum. Verslagen zijn door de organisatie en deelnemers in ware ‘DH-stijl’ online bij elkaar gebracht. dhbenelux.org/#Blogs-and-Reviews

Wrapper

KORT

Reis van de Razzia zeventig jaar later

Op 9 november 1944 legde het Duitse leger een cordon rond Rotterdam en Schiedam. In de twee daaropvolgende dagen werden ruim 52.000 Rotterdammers en Schiedammers tussen zeventien en veertig jaar oud opgepakt en afgevoerd richting Oost-Nederland en Duitsland om daar dwangarbeid te verrichten. Ten gevolge van ziekte, ondervoeding, vluchtpogingen, bombardementen en andere oorlogshandelingen zijn duizenden van deze mannen omgekomen. Filmmaker Erik de Jager heeft, 70 jaar later, met het project ‘Reis van de Razzia’ de verhalen van 75 overlevenden vastgelegd voor de toekomst. De interviews zijn online beschikbaar via EASY van DANS. (MP) stichtingreisvanderazzia.nl



Fremdenpas van dwangarbeider Jo Vermeulen, nummer 8416 credit Stichting Reis van de Razzia

SURFsara versterkt Research Data NL

Research Data Netherlands (RDNL) is een partner rijker: SURFsara. Met de komst van deze nieuwe partner kan RDNL meer wetenschappelijke instellingen vanuit een gezamenlijke aanpak bedienen bij het duurzaam opslaan en delen van onderzoeksdata. SURFsara is sterk op het gebied van nationale reken- en datadiensten, waaronder High Performance Computing, en vormt daarmee een aanvulling op het bestaande werkteerrein van RDNL: de technische wetenschappen (via 3TU.Data-centrum) en de sociale en geesteswetenschappen (via DANS). De officiële toetreding van SURFsara vond plaats tijdens de uitreiking van de Nederlandse Dataprijs. Axel Berg, adjunct-directeur SURFsara: “We zijn blij dat deze strategische samenwerking is bekrachtigd. Met de samenwerking verwachten we onderzoekers in Nederland nog beter en effectiever te kunnen ondersteunen, met een brede en gezamenlijke catalogus van diensten, expertise en innovaties.” researchdata.nl

VERVOLG VAN PAGINA 1

Experimenteren met data en tools

Permanente bèta-versie

Is het ResearchLab nu af? Faber: “Nee, en het zal ook nooit af zijn. We voorzien een permanente bèta-versie, waar steeds aan gewerkt en gebouwd wordt. We groeien mee met de technische mogelijkheden en de wensen van de onderzoekers. We zien trouwens het online ResearchLab niet als het enige resultaat van onze samenwerking met onderzoekers. We gaan ook bijeenkomsten organiseren, zoals hackatons waar onderzoekers en ontwikke-

laars met elkaar kunnen experimenteren in één ruimte.”

Ideeën en wensen welkom

Hoe kunnen onderzoekers het lab vinden? Faber: “Op kb.nl/lab. We nodigen wetenschappers van harte uit om daar te komen rondkijken en te experimenteren. Ook ideeën en wensen zijn welkom.”

kb.nl/lab

Speciale editie van E-data



Data archivering bestaat alweer 50 jaar in Nederland. In een speciale editie van E-data & Research halen (internationale) collega’s hun herinneringen op uit het verleden en delen ideeën voor de toekomst.

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:** Heidi Berkhout. **Redactie:** Marika de Bruijne, Steven Claeysens, René van Horik, Ilja Nieuwland, Rutger Nugteren, Erica Renckens. **Redactiesecretariaat:** Lucas Pasteuning. **Aan dit nummer werkten mee:** Inge Angevaere, Nina Conkova, Hella Hollander, Pim Huijnen, Erik de Jager, Emilie Kraaikamp, Annemiek van der Kuil, Ida Nijenhuis, Marieke Polhout, Wouter Quite, Renske Siemens, Bart van Vliet, Lotte Wilms, Carolien van Zuilekom, Joris van Zundert. **Opmaak:** Colette Sloots, Haarlem. **Productie:** Amsterdam University Press. **Druk:** Ten Brink, Meppel. **Webmaster:** Sonja Duijkers **Opplage:** 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.

Twitter onder de loep

Het is verleidelijk om social media in te zetten voor onderzoeksdoeleinden, maar hoe doe je dat op een verantwoorde manier? Lees mee met twee ervaringsdeskundigen.

Marika de Bruijne en Erica Renckens

Mick Couper,
Research Professor aan Michigan University:

‘Minder hype en meer hypothesen’

Mick Couper, Research Professor aan Michigan University en bekend surveymethodoloog, vertelt hoe social media en surveydata elkaar kunnen ondersteunen.

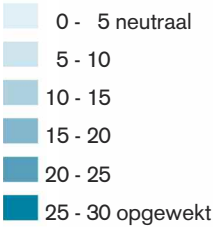
Mick Couper: “Social media en surveys kunnen worden gekoppeld op drie niveau’s: ge-aggregeerd, conceptueel of individueel. De meesten koppelen op geaggregeerd niveau en laten verbanden zien tussen surveyresultaten en social media.” Bekende voorbeelden zijn analyses op basis van Twitter-trends. Tijdens een griepseizoen wordt veel over griep getweet. Door te analyseren hoe vaak bepaalde woorden voorkomen, wordt een schatting gemaakt van de verspreiding van griep door het land. “Mijn zorg hierbij is dat surveys vaak worden gebruikt als maatstaf. Hierbij zijn de algoritmen voor het identificeren van trends in social media gekozen doordat ze passen bij de surveydata. Wat nodig zou zijn, is een onafhankelijke evaluatie van trends.”

Hypothesen testen

Couper vervolgt: “Met het koppelen op conceptueel niveau bedoel ik het testen van hypothesen in beide databronnen (surveys en social media). Bijvoorbeeld: is een relatie tussen houding en gedrag hetzelfde wanneer deze wordt gemeten in een survey als wanneer deze wordt gemeten in social media? Ook kan er niet alleen naar trends van losse indicators in social media worden gekeken, maar naar associaties zoals: sturen mensen die over een bepaald onderwerp tweeten ook tweets over een ander onderwerp?”

Beperkte groep Twitteraars

Couper: “Surveyresponses zijn ook op individueel niveau te koppelen aan persoonlijke berichten op social media. Maar dat is niet makkelijk. Op basis van surveydata weten we wie Twitter-gebruikers zijn, maar niet wat ze op Twitter posten. En binnen een groep respondenten maakt vaak slechts een beperkte groep gebruik van Twitter.” Zo blijkt bijvoorbeeld uit het LISS panel van Centerdata dat de Twitter-populatie verschilt van de gehele populatie en jonger is dan gemiddeld. Verder stuurt slechts een deel van de mensen op Twitter zelf actief tweets en heeft men te maken met automatische tweets, Twitterbots, retweets en bedrijfstweets.



Trends oppikken Voor de Volkskrant bracht Tjong Kim Sang met behulp van Twiqs het humeur van de Nederlandse twitteraar per gemeente en per provincie in kaart. Hiervoor maakte hij gebruik van een eigenhandig samengestelde lijst met positieve en negatieve woorden. Deze afbeelding laat het gemiddelde sentiment van tweets per gemeente zien voor november 2013. Tjong Kim Sang liep ook hier tegen een beperking aan van het gebruik van tweets als databron: Limburg en Vlaanderen kwamen opvallend somber uit de bus - waarschijnlijk door het ontbreken van typische dialectwoorden in de lijst. bron de Volkskrant

Manieren zoeken om onderzoeksmethoden te combineren of om ze samen te gebruiken en deze nog verder te ontwikkelen, is de boodschap. “Er worden al veel studies gepubliceerd, maar bij veel daarvan heb ik mijn twijfels. Social media bieden een kans voor onderzoekers en naarmate social media analytics zich ontwikkelen, zullen we meer studies zien van goede kwaliteit.” Of, zoals Couper concludeert: “We hebben minder hype en meer hypothesen nodig.” Het artikel *Is the Sky Falling? New Technology, Changing Media, and the Future of Surveys* uit Journal of the European Survey Research Association (volume 7, nummer 3, 2013) omschrijft meer bevindingen van Mick Couper. (MdB) <https://ojs.ub.uni-konstanz.de/srm/article/view/5751>

10 mln voor nieuw MIT-lab Twitter stimuleert ook zelf actief het gebruik van tweets voor wetenschappelijk onderzoek. Op 1 oktober maakte het bedrijf bekend 10 miljoen dollar te investeren in een nieuw lab aan MIT, het Laboratory of Social Machines. Onderzoekers in het lab krijgen volledige toegang tot de real-time stroom aan tweets en tot het gehele archief vanaf de begindagen van Twitter. Op basis daarvan zullen zij nieuwe technologieën ontwikkelen die wijs kunnen worden uit de semantische en sociale patronen in de online mediaberichten op Twitter en andere media.

Erik Tjong Kim Sang,
onderzoeker bij het Meertens Instituut:

‘Interessant om trends op te pikken’

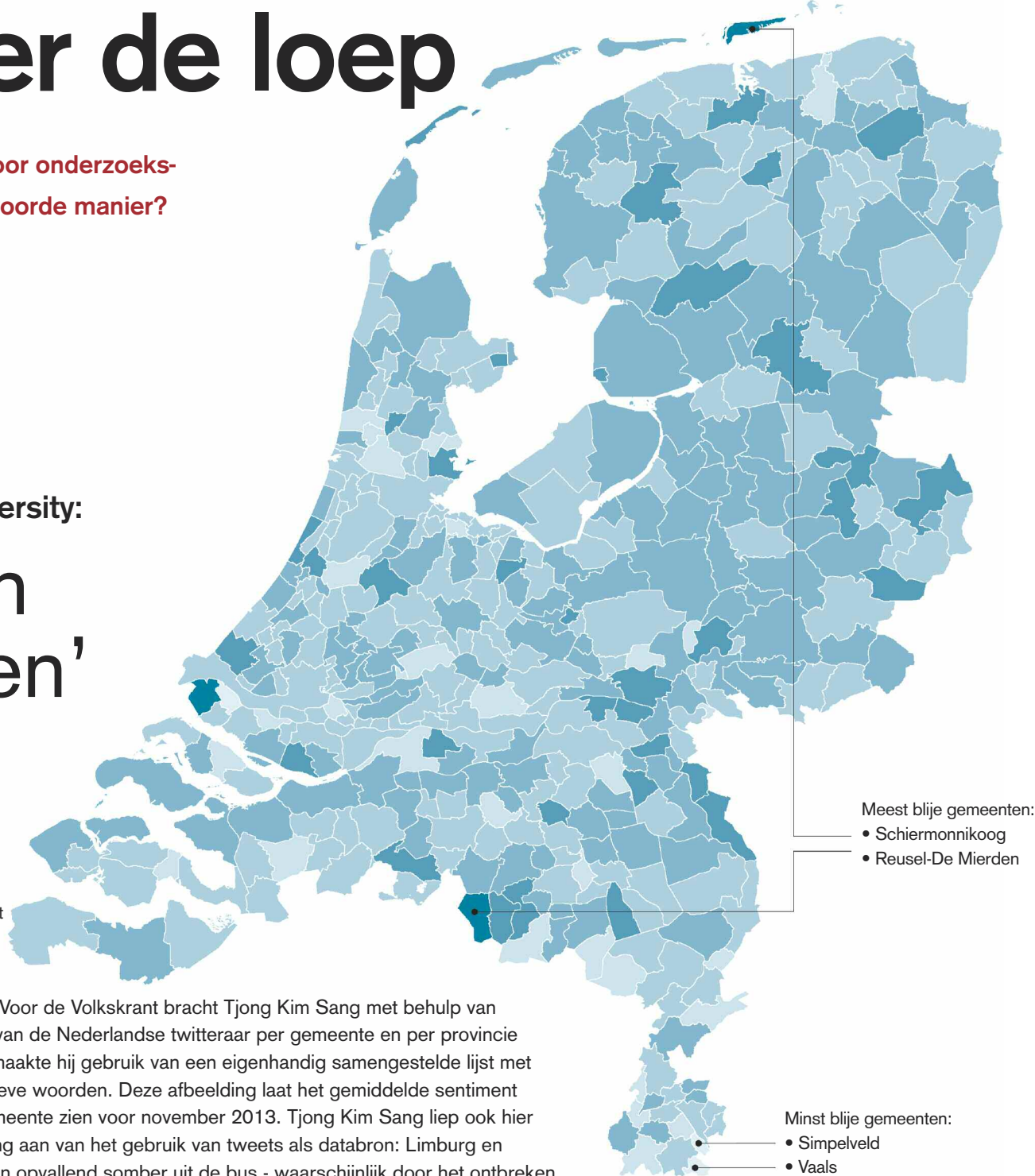
Erik Tjong Kim Sang ontwikkelde Twiqs, een zoekmachine waarmee onderzoekers, studenten en andere geïnteresseerden Nederlandse tweets kunnen analyseren.

Twiqs is het resultaat van het TwiNL, een project van het Netherlands eScience Center. “Het hoofddoel van dit project was om Nederlandse tweets te verzamelen en deze op een gemakkelijk doorzoekbare manier aan te bieden aan andere onderzoekers,” aldus Tjong Kim Sang, als onderzoeker verbonden aan het Meertens Instituut. Tweets bevatten naast de feitelijke inhoud ook metadata die waardevol kunnen zijn voor onderzoek, zoals dag en tijdstip van versturen, informatie over de afzender en hashtags, die een aanwijzing kunnen geven over het onderwerp.

Twitter-data voor onderzoek

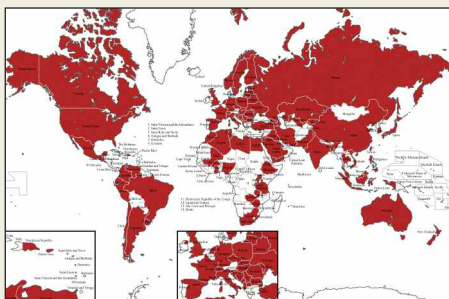
“Zelf hebben we ook gezocht naar informatie die uit deze tweets te halen was, bijvoorbeeld een voorspelling van een verkiezingsuitslag. De resultaten daarvan waren acceptabel, maar tweets zijn geen goede afspiegeling van de

Nederlandse bevolking.” In 2013 hadden ongeveer 3,3 miljoen Nederlanders een Twitter-account, van wie minder dan de helft dagelijks tuitert. De gebruikers zijn overwegend jong, hoogopgeleid en van het mannelijk geslacht. Deze niet-representatieve steekproef uit de Nederlandse bevolking is volgens Tjong Kim Sang niet het enige probleem bij het gebruik van twitter-data voor onderzoek: “Als je inzoomt op een bepaald onderwerp op Twitter, blijven er meestal weinig bruikbare tweets over.” De onderzoeker raadt zijn collega’s daarom niet direct aan om twitter te gebruiken als bron van data voor wetenschappelijk onderzoek. “Social media lijken mij niet geschikt om statistisch relevante informatie over de hele bevolking uit af te leiden,” zegt Tjong Kim Sang. “Ze zijn wel interessant om trends op te pikken. Het Instituut voor Nederlandse Lexicologie heeft bijvoorbeeld tweets gebruikt om nieuwe Nederlandse woorden te vinden. Dit levert geen compleet plaatje op – niet alle nieuwe woorden worden gebruikt op Twitter – maar wel diverse interessante vondsten.” (ER) twiqs.nl



SINDS KORT BESCHIKBAAR

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



Landen waar de WVS ooit is afgenomen
credit WVS

CentERdata

• World Values Survey

In december 2012 is de Nederlandse versie van de World Values Survey (WVS) afgenomen in het LISS panel, in opdracht van Bi Puranen, secretaris-generaal van de World Values Survey Associatie. Het hoofdonderzoek van de WVS vindt wereldwijd plaats in een groot aantal landen. Deze longitudinale studie maakt het mogelijk om de waarden en normen tussen landen te vergelijken en om de verandering daarvan te meten. De uitgebreide dataset bevat onderwerpen zoals sociaal leven, maatschappelijke waarden en vertrouwen, religie, media-gebruik, politiek, opvoeding, gezinsleven en werk. De data van de Nederlandse versie zijn beschikbaar via het LISS panel.
worldvaluessurvey.org

Ook sinds kort beschikbaar:

Studies LISS panel

- Jonge, J. de, juni 2013, Consumer heterogeneity with respect to morality in consumption decisions and perceptions of animal welfare - Part 3
- Prather, L.R., december 2013, Fighting Poverty at Home and Abroad: Explaining Attitudes Towards Poverty Assistance Programs
- Meuleman, R.; Lubbers, M., september 2013, Nationalism and the national dimension of cultural consumption - Wave 2
- Zijlstra, R.; Delbaere, K.; Kempen, R., maart, april/mei, juni 2013, Validation of three measures for fall prevention research among older people - Wave 1-2
- Hosch-Dayican, B.; Aarts, K.; Dassen, A.; Amrit, C., oktober 2013, Calibrating Twitter Data: Issue Salience and Issue Ownership in Social Media and in Surveys
- Velden, P.G. van der; Bosmans, M.W.G.; Knaap, L. van der, april/augustus/december 2012, Coping self-efficacy after potential traumatic events - Wave 1-3
- Kalmijn, M., juli/augustus 2013, Life History Questionnaire - Wave 2
- Pieters, R.; Leenheer, J., juni/september/december 2013, maart 2014, Tilburg Consumer Outlook Monitor
- CentERdata, februari/maart 2013, Social Integration and Leisure - Wave 6
- CentERdata, december 2013, januari 2014, Politics and Values - Wave 7
- CentERdata, november/december 2013, Health - Wave 7

Studies Immigrant panel

- Stupar, S.; Vijver, F. van de, augustus 2013, A Study on Underlying Motivation of Emotional Suppression in Immigrants and Mainstreamers in the Netherlands
- Bekhuis, H.; Lubbers, M., september 2013, Migrants National Identification and the National Dimension of Cultural Consumption
- Lelkes, Y.; Iyengar, S., januari 2014, Understanding the underpinnings of immigrant

attitudes towards immigration: A vignette experiment on visa applications

- Adams, B.G.; Vijver, F. van de, oktober 2013, The Association between Relational Orientation, Sources of identification, Identity, Personality, Culture and Psychological Well-Being in the Netherlands
- CentERdata, mei 2014, Personality - Wave 2
- CentERdata, juni 2012, Work and Schooling - Wave 1



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



Baksteenfundering van de toren en de westmuur van de middeleeuwse kerk credits gemeente Dordrecht, bureau Monumentenzorg en Archeologie

DANS

• Zonnen op Gods akker

De gemeente Dordrecht heeft archeologisch onderzoek gedaan naar een laatmiddeleeuws nederzettingsterrein. Het gaat om de vermoedelijke resten van het dorp Kruiskerke/Wolbrandskerke dat op de zuidoever van de Dubbel lag en tijdens de St. Elisabethsvloed van 1421 is verlaten en onder water kwam te staan. De resultaten van dit opgravingsproject zijn nu beschikbaar in het online archiveringssysteem EASY. Zonnen op Gods akker, archeologisch onderzoek van een laatmiddeleeuws nederzettingsterrein. Plangebied Gezondheidspark, Gemeente Dordrecht, T.H.L. Hos en M.C. Dorst, 2010.

Ook sinds kort beschikbaar:

Sociale Wetenschappen

- Heijting, Dr.ir. S. (Wageningen UR (WUR)) (2013-06-21), Weed count data Heijting et al.
[urn:nbn:nl:ui:13-gx2r-8l](http://nbn.nl:ui:13-gx2r-8l)
- Veldwijk, Drs. J. (2014-09-03), The effect of including an opt-out option in discrete choice experiments.
[urn:nbn:nl:ui:13-5ovx-16](http://nbn.nl:ui:13-5ovx-16)
- Hillebrink, C. (NWO-MaGW) (2013-03-03), Onderzoek Publicatieculturen NWO-MaGW. Onderzoek naar de verschillen in publicatieculturen tussen de disciplines van het NWO gebied Maatschappij- en Gedragswetenschappen.
[urn:nbn:nl:ui:13-0j3x-bl](http://nbn.nl:ui:13-0j3x-bl)
- Historische Wetenschappen
- Historical Sample of the Netherlands (2013-08-31), Data Set Linking Historical Lives (LHL)

Release 2013.01.

[urn:nbn:nl:ui:13-iy19-ze](http://nbn.nl:ui:13-iy19-ze)

- Sijs, Prof. Dr. N. Van der (Meertens Institute) (2014-05-01), Collectie kaarten behorende bij de Kaartenbank.
[urn:nbn:nl:ui:13-an8o-3t](http://nbn.nl:ui:13-an8o-3t)

Archeologie

- Bakker, A.M. (Archeodienst BV) (2014-04-01), Plangebied Vindicat, Grote Markt te Groningen.
[urn:nbn:nl:ui:13-87jq-el](http://nbn.nl:ui:13-87jq-el)
- Kremer, drs. H. (Synthegra BV); Leuving, drs. J.H.F. (Synthegra BV) (2014-07-08), Natuurontwikkeling IJzerhorst te Vierakker; Bureauonderzoek en inventariserend veldonderzoek, karterend booronderzoek.
[urn:nbn:nl:ui:13-ozqz-ww](http://nbn.nl:ui:13-ozqz-ww)

Ruimtelijke Wetenschappen

- Kadaster (2013-11-30), TOP25Raster, digitaal topografisch bestand - versie november 2013. [urn:nbn:nl:ui:13-o0n5-f2](http://nbn.nl:ui:13-o0n5-f2)
- NIROV (2008), Verwijzing naar de data van: Nieuwe Kaart van Nederland.
[urn:nbn:nl:ui:13-ucm-7lj](http://nbn.nl:ui:13-ucm-7lj)



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

Huygens ING

• Naoorlogs emigratiebeleid in Nederland 1945-1967

Deze online onderzoeksgids toont personen, overheidsinstellingen en particuliere organisaties die betrokken zijn bij zowel het Nederlandse emigratiebestel als bij de (beleids)processen rondom emigratie en de daaropvolgende immigratie in het land van bestemming. Via 367 gestandaardiseerde beschrijvingen zijn de taken, onderlinge verhoudingen en archieven geanalyseerd van instellingen en personen in Nederland, Australië en Canada en van enkele internationale organisaties. In 225 gevallen zijn de archivalia nader toegankelijk gemaakt.
resources.huygens.knaw.nl/emigratie



Achterblijvers op de kade zwaaien naar de vertrekkende familie op de boot credit Spaarnestad Foto

Ook sinds kort beschikbaar:

- Dagboekjaren 1831-1844 van Willem de Clercq. De digitale scans van deze dagboeken werden door de Universiteitsbibliotheek van Amsterdam voor dit project beschikbaar gesteld.
<http://resources.huygens.knaw.nl/dagboekdeclercq>
- Website VOC Kenniscentrum. De website informeert een ruim publiek op hoofdlijnen over allerlei aspecten van de VOC. Het KITLV heeft besloten de hosting en het beheer van de content aan het Huygens ING over te doen.
huygens.knaw.nl/website-voc-kenniscentrum-naar-huygens-ing



Deze publicaties zijn beschikbaar via www.historici.nl. Bezoek deze site of scan de QR code.

AGENDA

11 - 15 oktober • Girona

Archives and Cultural Industries

Deze conferentie is een bundeling van bijeenkomsten, waaronder de Europese conferentie van de International Council on Archives (ICA) en een gerenommeerd seminar op het gebied van beeldcollecties.

girona.cat/web/ica2014/eng

22 - 23 oktober • Brussel

Conferentie Alliance of Permanent Access (APA)

Tijdens deze conferentie wordt het APA-'Centre of Excellence' (CoE) op het gebied van digitale duurzaamheid gelanceerd. De diensten en tools die onderdeel zijn van het CoE zijn ontwikkeld in EU-projecten, zoals APARSEN en SCIDIP-ES.

alliancepermanentaccess.org

30 oktober • Hilversum

Studiedag Webarchivering in Nederland

Wat wordt er verzameld en door wie? Welke tools worden gebruikt en zijn de gearcheerde websites ook toegankelijk voor onderzoekers? Deze vragen komen aan bod tijdens deze studiedag.
ncdd.nl/studiedagwebarchivering

3 - 8 november • Limassol

International Conference on Cultural Heritage

Conferentie die zich richt op onderzoek op het gebied van restauratie, conservering en toegang tot cultureel erfgoed, ook in digitale vorm.
culturalheritage2014.eu

5 - 7 november • Berlijn

EVA: Electronic Media and Visual Arts

Al meer dan 20 jaar worden de EVA-conferenties georganiseerd. Ze richten zich voornamelijk op het gebruik van multimediatechnieken in onderzoek en analyse en de uitwisseling van kennis tussen erfgoedinstellingen en ontwikkelaars van technologie.

eva-berlin.de/en/eva-berlin.html

8 - 9 december • Den Haag

Exploring historical sources with language technology

Tijdens deze workshop staat het toepassen van taaltechnologie bij het analyseren van historische tekstuele bronnen centraal. Huygens ING is gastheer van deze workshop.
clarin.eu/events/2014

28 - 29 januari • Utrecht

THATCamp Utrecht

The Humanities and Technology Camp is een 'unconference'. Het is een open en informele bijeenkomst waarin geesteswetenschappers en ICT-ers samen komen en bespreken hoe technologie ingezet kan worden om onderzoek te ondersteunen.
utrecht.thatcamp.org/

29 januari en 12 maart • online, Delft en Den Haag

Cursus Essentials 4 Data Support

Essentials 4 Data Support is een introductiecursus voor diegenen die onderzoekers (willen) ondersteunen bij het opslaan, beheren, archiveren en delen van hun onderzoeksdata.

datasupport.researchdata.nl

Prof. dr. Lex Heerma van Voss, directeur Huygens ING en hoofdaanvrager CLARIAH:

‘CLARIAH biedt structuur aan de digitale onderzoeksrevolutie’

Vanaf 2015 werkt CLARIAH aan de ontwikkeling van een digitale infrastructuur voor geesteswetenschappers. De voorbereidingen daarvoor zijn momenteel in volle gang.

Erica Renckens

“Vanaf 1 januari mogen we geld uitgeven, maar voor het zover is, moet er nog heel wat overlegd worden over hoe we dat precies gaan doen,” aldus prof. dr. Lex Heerma van Voss, die als directeur van het Huygens ING hoofdaanvrager is van CLARIAH. Begin deze zomer kreeg CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) een NWO-subsidie van twaalf miljoen euro voor het ontwikkelen van een digitale infrastructuur waarin geesteswetenschappers grote databestanden kunnen interpreteren en ontsluiten.

Voor alles van nut

Het is de tweede keer dat het consortium van geesteswetenschappelijke onderzoeksinstituten de CLARIAH-aanvraag indiende. “De eerste keer kregen we als commentaar dat onze inzet op de gehele geesteswetenschappen te breed was,” vertelt Heerma van Voss. “Daarom hebben we nu de focus gelegd op drie deelgebieden: taalkunde met overwegend tekstuele data, mediastudies met audiovisuele bronnen en sociaal-economische geschiedenis met gestructureerde databestanden, databases.” Alle drie de vakgebieden zijn al voorloper in de Digital Humanities – een weloverwogen keuze. “Deze typen bronnen worden ook door andere geesteswetenschappelijke disciplines gebruikt. We hebben de verantwoordelijkheid voor de hele breedte van de geesteswetenschappen, dus wat we gaan maken, moet in principe voor alle disciplines van nut zijn.”

“Taalkundigen kunnen bijvoorbeeld al heel goed teksten automatisch parsen,” geeft Heerma van Voss als voorbeeld van een toepassing die breed inzetbaar is. “Nu willen historici of letterkundigen meestal niet speciaal weten of iets een lidwoord of een werkwoord is, maar zo’n techniek is wel een middel om grip te krijgen op een tekst. Je ziet steeds meer onderzoeken waar een parser wordt gebruikt om historische of letterkundige vragen te kunnen beantwoorden.”

Lab voor geesteswetenschap

“Net als in CLARIN zullen we bij CLARIAH calls uitschrijven waarin onderzoekers aanvragen kunnen indienen. We willen weten waar de geesteswetenschappers behoefte aan

hebben, dus we vragen mensen om zelf met voorstellen te komen voor nieuwe datasets of nieuwe tools om databestanden mee te ontsluiten.” Toch zullen er ook verschillen zijn met CLARIN. “In CLARIN zijn voornamelijk losse demonstrators gemaakt,” constateert Heerma van Voss. “Ik denk dat wij meer opschuiven naar een labsituatie, een volledige productieomgeving voor onderzoekers.”

In de naam CLARIAH is niet alleen ‘CLARIN’, maar ook ‘DARIAH’ te herkennen, beide grote Europese onderzoeksinfrastructuren – het eerste voor taalkundigen, het andere voor historici. “Nederland heeft een voortrekkersrol in CLARIN. In DARIAH minder, al is Nederland ook daar actief in. CLARIAH zal zowel in CLARIN als DARIAH de Nederlandse bijdrage leveren.” Daarnaast verbindt CLARIAH ook Nederlandse deelprojecten, als CLIO-Infra, Nederlab en EU Screen.

Eisen aan metadata

Goede afspraken omtrent metadata zijn voor zo’n brede samenwerking onontbeerlijk. “Daar zullen we het dit najaar – en ongetwijfeld gedurende het hele project – veel over moeten hebben,” beaamt Heerma van Voss. “Het is best ingewikkeld om tussen al die geesteswetenschappelijke disciplines te communiceren. CLARIN heeft eisen gesteld aan hoe de metadata eruit moeten zien. We zijn er nog niet over uit of het zin heeft die regels geesteswetenschappelijk breed voor te schrijven. In ieder geval streven we de interoperabiliteit ook internationaal na.”

“De demonstrators van CLARIN toonden aan waar onderzoekers in het veld behoefte aan hadden; daar werden standaarden voor ontwikkeld. Die standaarden zijn nog altijd volop in ontwikkeling. Deze revolutie zal ook doorzetten als CLARIAH niet gefinancierd zou zijn, maar nu hebben we in Nederland de mogelijkheid hier goed mee om te gaan en daar een structuur voor aan te bieden.”

Een grote vlucht

“Ik denk dat over vier jaar elke student een cursus Digital Humanities volgt,” voorspelt Heerma van Voss. “Niet omdat wij dat willen

Moral panic over prostitutie

Ook voor zijn eigen onderzoek ziet Prof. dr. Lex Heerma van Voss de mogelijkheden van de CLARIAH-infrastructuur. Hij onderzoekt arbeidsomstandigheden wereldwijd. “Ik heb bijvoorbeeld een artikel geschreven over prostitutie. Rond 1900 en rond 2000 zie je een vergelijkbare omslag in het denken over prostitutie: van gereguleerd en geaccepteerd naar verbieden vanwege dwang en slavernij. Inmiddels weten we uit literatuurstudies dat de moral panic van een eeuw geleden overdreven was, maar daaruit kun je natuurlijk niet concluderen dat dat nu weer zo is. Er blijken veel variabelen een rol te spelen in dit verhaal, meer dan ik als mens kan overzien. Met de data die we nu in CLARIAH verzamelen, heb ik een veel betere set in handen waarmee ik variabelen als ‘mate van migratie’ en ‘vrijheid van de vrouw op plek van herkomst’ beter in de vingers kan krijgen.”



Prof. dr. Lex Heerma van Voss foto Bart van Vliet

of geven, maar omdat steeds meer mensen er de mogelijkheden van inzien. We hebben verschillende onderzoeksscholen onze aanvraag laten lezen en gevraagd hoeveel van hun promovendi gebruik zouden maken van een dergelijke infrastructuur. Momenteel zou dat zo’n veertig procent zijn, maar ze gaven allemaal aan dat dit op korte termijn een nog veel grotere vlucht zal nemen.”

Bij het onderzoeksinstituut van Heerma van Voss zelf, het Huygens ING, is de geesteswetenschappelijke revolutie al voelbaar. “Wij

houden ons veel bezig met het maken van tekstedities. We wisten precies hoe dat moest; daar waren gewoon regels voor. Maar de digitale revolutie maakt dat alle basale handelingen voor geesteswetenschappers in beweging zijn: welke informatie is beschikbaar, hoe vind ik die, hoe bevraag ik de data, hoe laat ik de resultaten zien, hoe zorg ik dat alles teruggevonden kan worden. Ook al heb je een briljant idee, je kunt niet zeggen ‘Voortaan gaan we het zó doen.’ Alle relevante spelers werken mee aan CLARIAH, zodat we precies weten waar onderzoekers behoefte aan hebben.”

CLARIAH heeft voor vier jaar subsidie gekregen. Wanneer kan gesproken worden over een geslaagd project? “We hebben natuurlijk beloofd een infrastructuur te ontwikkelen die in de drie gebieden relevant is,” aldus Heerma van Voss. “Maar ik zou het echt een succes vinden als we iets maken waardoor ook andere geesteswetenschappelijke disciplines zeggen: ‘Oké, dit is zo interessant, dit ga ik ook gebruiken.’”

clariah.nl

Fort, kasteel of verdedigingswerk?

Hoe zijn diverse soorten data zo te combineren, dat dit tot interessante nieuwe observaties leidt? Carlijn Hageraats zocht het uit.

Heidi Berkhout

“Archeologie is in Udhruh prominent aanwezig in het landschap. De lokale bevolking weet wat er te vinden is en waarvoor de objecten in het verleden gediend hebben. Voor het Udhruh Oral History-project nam ik interviews af met de lokale bevolking.

Terug in Nederland ging Carlijn aan de slag. “Ik stelde woordenlijsten samen: één met informatie uit bestaand archeologische onderzoek en één met informatie uit de interviews. In Simple Knowledge Organisation System (SKOS), een data-model om kennissystemen te koppelen, kende ik labels toe aan individuele concepten: een preferred label voor de voorkeursterm (bijvoorbeeld ‘Romeins legionair fort’), alternative labels voor alternatieve benamingen (zoals ‘fort’ of ‘verdedigingswerk’) en hidden labels voor woorden die liever niet gezien worden, zoals woorden die verkeerd gespeld zijn.”

Close match

Vervolgens legde Carlijn met SKOS verbanden tussen de verschillende concepten. “Met broader en narrower worden hiërarchische relaties



Stagiaire Carlijn Hageraats (links) aan het werk bij het Ottoman-fort foto G. Gazenbeek

aangeduid, bijvoorbeeld tussen het fort en de objecten waar dit uit bestaat. Het fort is groter dan de stenen waar het uit bestaat, en een steen is kleiner dan het hele bouwwerk. Met related worden associatieve, niet-hiërarchische relaties aangeduid, bijvoorbeeld tussen het fort en de wachttorens op nabijgelegen heuvels. Uiteindelijk krijgen archeologie en oral history ieder hun eigen conceptschema met centrale concepten en onderlinge relaties.” Vervolgens worden beide conceptschema’s met elkaar ver-

bonden door de relaties tussen de onderliggende concepten in kaart te brengen. De laatste stap is het aanmaken van een semantisch repository, waarin de woordenlijst en een lijst met getagde documenten geüpload en doorzoekbaar gemaakt worden.

Interdisciplinair zoeken

Met deze manier van werken wordt bereikt dat bestanden doorzocht kunnen worden vanuit twee disciplinaire invalshoeken. Carlijn sluit af: “Een zoekopdracht op ‘fort’ of ‘verdedigingswerk’, leidt tot docu-

menten uit de archeologie die hierop betrekking hebben, maar ook tot interviews met de lokale bevolking die hierover gaan en waarin het concept met andere bewoording wordt aangeduid. Hierdoor kunnen er verbanden gelegd worden tussen twee ‘verschillende talen’; de lokale kennis en de wetenschappelijke kennis, wat het begrip en het samenwerken tussen beide discipli-

nes vergemakkelijkt. Ik verwacht dat deze methode verder te ontwikkelen is tot een algemeen toepasbaar concept in de combinatie van oral history en archeologie, wat het aantrekkelijker maakt om oral history frequenter te gebruiken samen met archeologisch onderzoek.”

projects.commonsites.net/nl/project/677/

JONG TALENT

Carlijn Hageraats

Carlijn Hageraats, studente Erfgoedstudies (Universiteit van Amsterdam), Archeologie (Universiteit Leiden) en stagiaire bij DANS, onderzocht de combinatie van oral history en archeologisch onderzoek binnen het Udhruh Oral History-project.

Dit project is onderdeel van het in 2011 gestarte Udhruh Archaeological-project, een samenwerkingsverband tussen de Universiteit Leiden en de Al-Hussein Bin Talal University in Jordanië.

Het resultaat staat in haar masterscriptie: ‘Who says myths are not real? Looking at archaeology and oral history as two complementary sources of data.’

‘Lokale kennis meegenomen in archeologisch onderzoek’

Augmented reality lab versmelt fysieke en virtuele wereld

Levensecht onderzoeken in Tilburg University

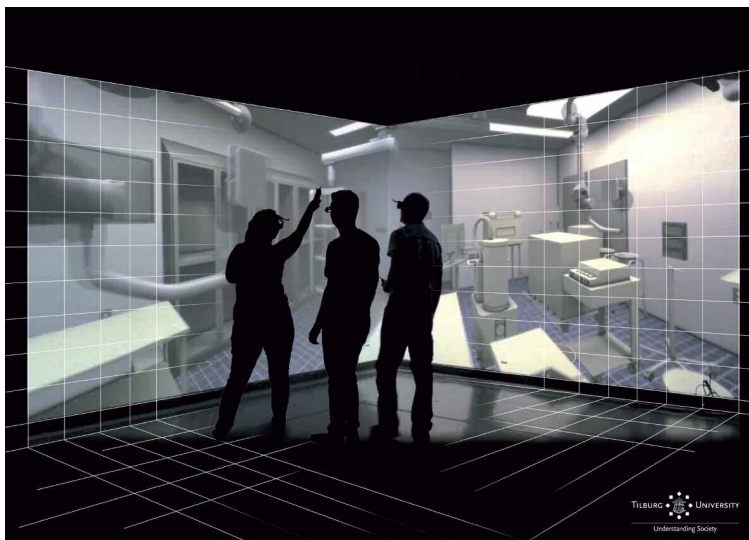
Tilburg University maakt onderzoeken levensecht met een augmented reality lab. Een bijzondere eigentijdse onderzoeksmethode voor.

Marike de Bruijne

“Om interactief onderwijs aan te bieden en nieuwe onderzoeksideeën te stimuleren, is een augmented reality lab ideaal”, zegt hoogleraar Cognitieve Psychologie en Kunstmatige Intelligentie en bedenker van het initiatief Max Louwerse van het Tilburg Center for Cognition and Communication (TiCC).

Toverstaf

Een augmented reality lab combineert de voordelen van een in-situ ervaring met die van een laboratorium. “Ons DAF Technology Lab brengt de werkelijkheid naar het laboratorium. In een ruimte van negen bij negen meter projecteren we een virtuele wereld op de vier wan-



De fysieke en de virtuele wereld versmelten. Studenten staan eerst in een virtuele rechtszaal die vervolgens in een operatiezaal wordt omgetoverd credit WorldViz/Tilburg University

den. Door een 3D-bril te dragen, verdwijnen de grenzen tussen de fysieke en virtuele wereld en ontstaat een ‘augmented reality’. Maar ook zonder zo’n bril wordt een vergelijkbare ervaring gecreëerd. Met een

speciale ‘toverstaf’ kan de gebruiker vervolgens interacteren met de virtuele wereld. Zijn of haar gedrag wordt opgenomen door verschillende camera’s en microfoons. Naast de 3D-brillen beschikken we ook

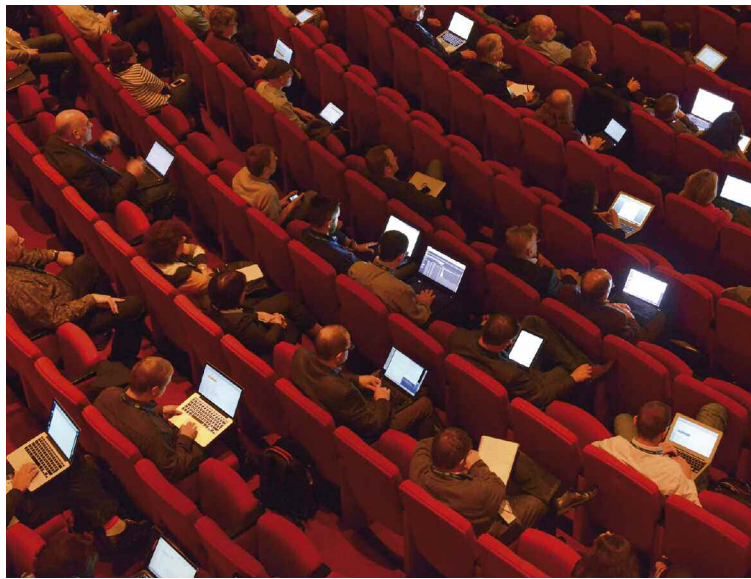
over geavanceerdere virtual reality-brillen. Deze zogenoemde Oculus Rift Head-Mounted Displays zorgen ervoor dat gebruikers zich volledig opgenomen voelen in een virtuele wereld en daarin vrij bewegen.”

Onbeperkt

Louwerse vertelt verder: “Het idee voor het lab ontstond mede door de wens om een virtuele rechtszaal te bouwen. Maar evengoed kan het lab gebruikt worden door andere sociale, medische of geesteswetenschappers. Bij economie kun je denken aan een onderzoeksvraag als ‘hoe meet je stress in directiekamers?’, bij psychologie aan ‘hoe creëer je een zo kindvriendelijk mogelijke onderzoeksomgeving?’ en bij medische psychologie kun je onderzoeken hoe dokter-patiënt interacties in ziekenhuizen werken. De toepassingsmogelijkheden zijn onbeperkt.” Kennis over het meten van het gedrag is op zijn universiteit al op

grote schaal aanwezig, stelt Louwerse. “Die is ook nodig om de technologie van het lab goed te kunnen benutten”, voegt hij toe. Wat hij graag meer wil zien, is interdisciplinaire samenwerking, iets wat het lab hopelijk aan zal wakkeren. “Het lab combineert onderwijs en onderzoek door research-based learning, verhoogt interdisciplinair onderzoek en onderwijs, biedt iets anders aan dan de standaard hoorcolleges en maakt de link naar technologie”, somt hij enthousiast op. Hij ziet uitbreidingen naar meer wearable computing, sensormetingen en andere toepassingen al voor zich.

Het DAF Technology Lab opent zijn deuren in 2015. Onderzoekers die van het lab gebruik willen maken, kunnen contact opnemen met projectcoördinator Rebecca Scholte of hoogleraar Max Louwerse van Tilburg University. tilburguniversity.edu

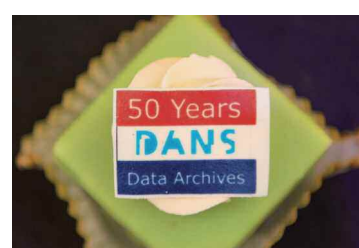


Nederland week lang internationaal toneel voor onderzoeksdata

4e plenaire bijeenkomst Research Data Alliance



500 Wetenschappers en data-experts uit vele landen en zo mogelijk nog meer disciplines kwamen van 22 tot 24 september naar Amsterdam voor de vierde plenaire bijeenkomst van de Research Data Alliance (RDA). Een keynotespreker suggereerde dat binnenkort een wetenschappelijk artikel een bijlage zal zijn bij de onderzoeksdata in plaats van andersom. Maar technisch, organisatorisch en juridisch moet er nog veel gebeuren – dus ook in de vele werkgroepen van de RDA. De resultaten van de conferentie staan op rd-alliance.org. tekst en foto's Inge Angevaere



Oorkonden en resoluties uit 15e en 16e eeuw online

“Materialen uitgeven op een manier die nu relevant is. Dat is waar het om gaat.” Lex Heerma van Voss vertelt over het ontsluiten van Nederlandse databronnen. Ilja Nieuwland

Samen met het Nationaal Archief ontsluit het Huygens ING de komende vijf jaar drie belangrijke Nederlandse databronnen online: de Nederlandse oorkonden van vóór 1500, de Generale Missiven (jaarlijkse rapporten vanuit Batavia aan het bestuur van de Verenigde Oost-Indische Compagnie) en de Resoluties van de Staten-Generaal (de verslaglegging van behandelde zaken tijdens de vergaderingen van 1576 tot 1796).

Heerma van Voss, directeur van het Huygens ING: “Wat ik boeiend vind,

is dat ik als student geschiedenis regels moest leren die uitontwikkeld waren, die al tachtig jaar als norm voor historici werden gehanteerd. Nu is dat allemaal in beweging; het is belangrijk om heel erg interessante materialen uit te geven op een manier die nu relevant is, die het stellen van nieuwe vragen toelaat. In eerste instantie worden scans van de databronnen online gezet. Dat verhoogt de toegankelijkheid al enorm. Vervolgens worden, naast andere technieken, vrijwilligers ingezet bij de basistranscriptie van de Generale Missiven en de Resoluties. Dit schrift is goed leesbaar. Je kunt je voorstellen dat zij voor het slapen even een Resolutie transcriberen.”

Echte uitdaging

De grootste mogelijkheid tot innovatie ziet Heerma van Voss in de

ontsluiting van de Resoluties. Voorlopers van het instituut zijn al voor 1910 aan de editie begonnen. Inmiddels is de periode tot 1630 ontsloten, met behulp van nieuwe technieken kan dat tempo behoorlijk worden opgeschoefd. In eerste instantie gebeurt dat door gescande pagina's ter beschikking te stellen aan onderzoekers.

Editeerwerk

Maar de echte uitdaging ligt in het editeerwerk. Huygens ING-senior-onderzoeker Ida Nijenhuis: “Het Nationaal Archief verstrekt de gescande pagina's van de Resoluties; wij zorgen voor verdere ontsluiting. In eerste instantie vinden twee pilots plaats: één met handschriftherkenning, waarvoor de samenwerking met de Rijksuniversiteit Groningen is gezocht, en één waarin, in samen-



Pagina uit gedrukte Resoluties van de Hoogh Mogende Heeren der Vereenighde Nederlanden, 1740 credit Nationaal Archief

werking met partners in Polen, getracht wordt karakterherkenning toe te passen op de gedrukte uitgaven van de Resoluties. We zijn nog niet zover, maar uiteindelijk willen we alles online ontsluiten.”

resources.huygens.knaw.nl

KORT

CLAPOP toont resultaten CLARIN-NL

CLARIN-NL heeft alle data en tools die de afgelopen jaren binnen het project zijn ontwikkeld, samengebracht in CLAPOP. Via deze portal hebben onderzoekers met één login toegang tot alle resultaten van het project dat komend voorjaar zal worden afgerond. De gebruiker kan filteren op onder andere vakgebied, taal en functionaliteit, zodat relevante data en tools worden voorgesteld. Ingepaste gebruikers hebben toegang tot deze bronnen en kunnen in contact komen met collega's die ook gebruikmaken van de CLARIN-infrastructuur. (ER) dev.clarin.nl

Digital Humanities cursussen op een rij

Eén interactief overzicht met alle cursussen Digital Humanities die aan Europese universiteiten worden gegeven. Dat is DODH (Dutch Overview of Digital Humanities), een online applicatie ontwikkeld door CLARIAH en DARIAH EU. Geïnteresseerden kunnen de database doorzoeken op locatie en opleidingsniveau, waarna de cursussen op de kaart worden weergegeven. Momenteel is het overzicht voor Nederland, Vlaanderen en Duitsland zo goed als compleet; gebruikers kunnen zelf ontbrekende cursussen toevoegen. Binnenkort verschijnt er ook een overzicht van lopende en afgeronde projecten op het gebied van Digital Humanities in Europa. (ER) clariah.nl/en/dodh

DSA-seal voor DNB Household Survey

Het internationale datakeurmerk Data Seal of Approval (DSA) is toegekend aan de DNB Household Survey (DHS). De DHS is een longitudinaal onderzoek over psychologische en economische aspecten van spaargedrag en wordt sinds 1993 afgenomen in het CentERpanel van CentERdata. Naast het bestuderen van financieel gedrag, worden de data ook gebruikt om andere studies te verrijken. De datasets bevatten informatie over werk, pensioen, wonen, hypotheek, inkomen, bezittingen, leningen, gezondheid, economische en psychologische concepten en persoonlijke eigenschappen. Het keurmerk geeft aan dat de DHS-data op een betrouwbare wijze worden ontsloten en bewaard. Om de langetermijnsopslag te garanderen, wordt samengewerkt met DANS. (MdB) dhsdata.nl

Wikipedian-in-Residence Hay Kranen:

‘Wikidata kan goudmijn zijn voor onderzoekers’

Hoe haal je de banden tussen Wikipedia en de Nederlandse erfgoedwereld aan? Wikipedia's Hay Kranen trok een jaar geleden in bij het Nationaal Archief en de Koninklijke Bibliotheek. En nu zwaait deze eerste Nederlandse Wikipedian-in-Residence weer af.
Steven Claeysens

Hij bracht Wikipedianen naar bibliotheek en archief en gaf aan archivissen en bibliothecarissen meer dan tien Wikipedia-workshops. Maar hij deed nog veel meer. Kranen is trots op wat er is bereikt. “Gaandeweg ontdekte ik dat ik er op een tweeledige manier in stond: een evangeliserende en een pragmatische. Het eerste is beter geslaagd, dan het laatste.”

90.000 views

“We hebben veel gedigitaliseerd Publiek Domein-materiaal van het NA en de KB toegevoegd aan Wikimedia Commons.” En uit alles blijkt dat van de vele scans van Nederlands historisch erfgoed duchtig gebruik gemaakt wordt. Enthousiast: “Het interessantst vond ik het VOC-materiaal uit de Atlas of Mutual Heritage. Een kaart uit deze collectie sierde intussen al de hoofdpagina van de Duitse Wikipedia en kreeg daar in één dag maar liefst 90.000 views.”



Hay Kranen foto KB

FOCUS

‘Een harde boodschap, maar wel de realiteit’

Dat enorme bereik raakt aan de kern van Kranens evangeliserende missie: “In elke lezing of cursus voor medewerkers wees ik erop dat maar één procent van de Nederlandse bevolking gebruikmaakt van een bibliotheek voor het opzoeken van feitelijke informatie. Mensen zoe-

ken informatie op Wikipedia of andere digital native-projecten, niet in de bibliotheek. Die boodschap kwam bij sommigen toch nog hard aan, maar het is wel de realiteit.” Die digitale realiteit is volgens Kranen nu wel doorgedrongen. Daarnaast zijn er ook de nodige plannen gesmeed om de verse banden tussen Wikipedia en de bibliotheek en het archief te bestendigen. Bovendien gingen intussen ook bij andere Nederlandse erfgoedinstellingen Wikipedian-in-Residence aan de slag.

Interessante dingen

Wat wetenschappers met de Wiki-projecten kunnen? “Veel. Ik ontmoet regelmatig onderzoekers die

interessante dingen doen op basis van ons materiaal. Alle teksten en data hebben een open licentie. Wikimedia Commons, bijvoorbeeld, is een gigantische, constant groeiende verzameling van meer dan 22 miljoen encyclopedisch relevante afbeeldingen en andere media. De metadata is jammer genoeg nog niet gestructureerd, maar daar gaat in de toekomst verandering in komen.”

15 miljoen gegevens

Vooral over de jongste telg van de Wiki-familie raakt Kranen moeilijk uitgepraat. “Wikidata kan een goudmijn zijn voor onderzoekers. In een paar jaar tijd hebben we al meer dan 15 miljoen gegevens verzameld en dat aantal zal de komende jaren erg hard groeien. Wikidata bestaat volledig uit gestructureerde data. Daar kan je ontelbare queries op loslaten. Mijn favoriete voorbeeld is de vraag naar alle burgemeesters ter wereld die niet mens zijn. Dan leer je dat in een klein plaatsje in Alaska een kat de burgemeesterssjerp draagt. Waar anders ontdek je dit?” Ook complete datasets kunnen sinds kort batch-gewijs worden toegevoegd. “Onlangs zijn zo alle Nederlandse rijksmonumenten mét geo-coördinaten geüpload. Het zou toch mooi zijn als we ook de volledige nationale bibliografie kunnen toevoegen en deze dan met crowd sourcing kunnen verbeteren en verrijken.” Andersom kan ook Wikidata veel hebben aan onderzoekers, geeft Kranen tot slot mee. “Als onderzoekers interessante datasets hebben die passen in het Wikidata-project, laat ze zich vooral melden. We voegen de data graag toe.”

www.wikidata.org
commons.wikimedia.org

GELEZEN

*‘Using Linked Data Effectively’, eBook, diverse auteurs.
Uitgever: The Open University
2013. Gratis via Apple iTunes en euclid-project.eu*
René van Horik



Er is groeiende belangstelling voor het semantisch web, waarin het ‘web van documenten’ transformeert naar een ‘web van data’. De onderliggende standaarden en technieken die voor deze transformatie nodig zijn, zijn bekend onder de naam ‘Linked Data’. Recentelijk verscheen het elektronische boek ‘Using Linked Data Effectively’. Dit boek geeft onderzoekers een overzicht van de mogelijkheden van de Linked Data standaarden en technieken. Het boek bestaat uit zes hoofdstukken over aspecten zoals het creëren, het gebruiken en het bouwen van applicaties op basis van Linked Data. Het bevat filmpjes met achtergrondinformatie bij de informatie. Het boek is ontwikkeld door EUCLID (EdUcational Curriculum for the usage of Linked Data), een Europees project waarbinnen kennisoverdracht aan ‘data-beoefenaars’ centraal staat. De visie van EUCLID is dat het gebruik van Linked Data het dagelijks werk van onderzoekers vergemakkelijkt. EUCLID biedt een combinatie van leermiddelen en activiteiten zoals eBooks, webinars en trainingen. Op de website van EUCLID staat meer informatie. euclid-project.eu

COLUMN

In de e-humanities ziet niemand door de bomen het bos meer

Onlangs bezocht ik een bijeenkomst van de Universiteitsbibliotheek Leiden over ‘Digital Scholarship and the Role of the Library’.

Een van de conclusies van de bijeenkomst was dat academische bibliotheken studenten en onderzoekers in de e-humanities beter moeten leren omgaan met digitale hulpmiddelen om patronen in grote datacollecties te ontdekken. De toverwoorden die hierbij horen zullen de meeste lezers van dit blad bekend zijn. Men heeft het niet over computerprogramma’s of software, maar over tools. Die moet je loslaten op een dataset. De meeste datasets zijn zo groot dat onderzoekers ze onmogelijk grondig kunnen bekijken. Maar gelukkig kunnen tools automatisch patronen uit die data vissen – wat distant reading wordt genoemd. Het liefst vindt men onbekende of nieuwe patronen. Alle aanwezigen, een internationaal publiek van zo’n zestig mensen, konden zich goed in die conclusie vinden. Wereldwijd stelt men vast dat veel studenten en

onderzoekers slimmer gebruik willen maken van datasets, maar welke tools werken het best voor bepaalde taken? En hoeveel tools zijn er eigenlijk?

Dat weet niemand. Nergens is een overzicht te vinden van alle tools die zijn of worden ontwikkeld – althans niet voor de geesteswetenschappen. Instituten krijgen nauwelijks geld voor het aanleggen van goede datasets, want subsidiegevers beschouwen dat niet als wetenschappelijk werk. Wel worden er bakken geld gegeven aan het bouwen van nieuwe tools, want daarmee be-geef je je volgens de subsidiegevers juist in de frontlinie van de wetenschap. Of er elders in de wereld al een

vergelijkbare tool is ontwikkeld, doet niet ter zake. Leiden heeft onlangs een eigen instituut voor e-humanities opgezet, maar werkt niet samen met Amsterdam of Utrecht, waar ze ook dergelijke centra hebben. In de frontlinie van de geesteswetenschappen knokt ieder voor zich – dat is nou eenmaal de traditie.

Of de tools vaak worden gebruikt, lijkt men niet belangrijk te vinden



foto Leo van Velzen

– het gaat erom dat ze er zijn. Men ziet het ook niet als taak om ze te onderhouden. Ze zijn er, wie ze verder wil ontwikkelen kan dat doen. Dat er soms heel weinig data aan de tool gekoppeld zijn, beschouwt men niet als probleem. In principe zou je er veel data aan kunnen hangen.

In de zaal had dit alles, als ik het goed heb gezien, een enigszins verflamend effect. Vele van de gedemonstreerde tools waren de aanwezigen onbekend. Ze stelden vragen als: ‘Waarom laat je ons speciaal deze tool zien voor deze functie? Zijn er ook andere beschikbaar?’ Of: ‘Werkt dit ook voor andere talen dan het Engels?’ Ik zag iemand fronsen toen een

onderzoeker vertelde dat er nauwelijks tijd was geweest om een tool (die in twee jaar tijd 700.000 euro had gekost), met data te vullen. Het bekende verhaal: de tool was nu voor iedereen beschikbaar, maar werd niet onderhouden. Gelukkig was er wel subsidie voor een nieuwe tool.

Het beste nieuws vond ik dat de Universiteitsbibliotheek Leiden, in samenwerking met studenten, een nuttig lijstje op internet heeft gezet met tools waarvan is vastgesteld dat ze bepaalde dingen goed doen. Dat andere universiteiten in de wereld al tientallen vergelijkbare lijsten op internet hebben gezet, slechts deels met dezelfde tools, bewijst mijns inziens dat niemand in de e-humanities door de bomen het bos meer ziet.

Ewoud Sanders

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