

Commissie Schuyt pleit voor toegang tot data en ambtseed

Onderzoekers moeten elkaar vaker controleren

Om wetenschappelijke fraude tegen te gaan, zijn geen nieuwe regels nodig. Maar wat dan wel?

Marika de Bruijne

Meer onderlinge waakzaamheid en sociale controle, een ambtseed en het breed toegankelijk maken van onderzoeksgegevens. Dat zijn enkele aanbevelingen van de Commissie Schuyt in haar adviesrapport *Zorgvuldig en integer omgaan met wetenschappelijke onderzoeksgegevens*, dat op 21 september op een symposium in Amsterdam werd gepresenteerd.

Voor het rapport werden wetenschappers in verschillende disciplines bevraagd over de onderzoekspraktijk in hun vakgebied. Aanleiding was onder andere de fraude van hoogleraar Diederik Stapel in 2011. "Wat tot nu toe ontbreekt, is bewustwording van de voordelen van zorgvuldig opgezette gegevensverzameling. En een gericht beleid van verantwoordelijke instanties om dit bewustzijn op de werkvloer aan te wakkeren en te (blijven) bewaken", aldus de Commissie Schuyt. In hoofdlijnen pleit de Commissie voor een bredere toegankelijkheid van onderzoeksgegevens en voor activiteiten om bestaande integriteitsregels tot een vitaal element van wetenschapsbeoefening te maken. Met dit laatste moet al in het onderwijs aan toekomstige wetenschappers worden gestart. Ondertekening van een ambtseed behoort wellicht tot de mogelijkheden.

Een voorrecht

Een actievere houding en betere begeleiding van wetenschappers is ook nodig. Naast het intellectuele debat moet er meer dialoog komen over de praktische kanten van onderzoek. Een promotor of onderzoeksleider die regelmatig op de werkvloer een praatje maakt met de onderzoekers, is



illustratie iStockphoto

eerder op de hoogte als iets dreigt mis te gaan. Werken in teamverband draagt bij aan sociale controle van elkaars werk. "Het is een voorrecht wetenschapper te mogen zijn", zegt Hans Clevers, president van de KNAW. Hij daagt onderzoekers uit om zelf in actie te komen.

Labjournaals

Er bestaan binnen disciplines *good practices* en controlemechanismen, waar andere disciplines uit zouden kunnen putten, zoals het bijhouden van logboeken en labjournaals. De verschillen tussen wetenschapsgebieden zijn te groot voor één strategie. "Wie de schoen past, trekke hem aan", schrijft het adviesrapport. De Commissie doet ook een beroep op instanties. Een kritischere cultuur bij onderzoeksinstituten, aandacht voor datamanagement tijdens opleidingen en externe

controles op instellingsniveau, suggereert de Commissie. Ook bijvoorbeeld uitgever van *peer reviewed journals* kunnen verplicht stellen dat onderliggende onderzoeksdata standaard toegankelijk zijn voor andere onderzoekers.

De vraag die op de casus van Stapel volgde - wat de aard en de omvang van het wetenschappelijke wangedrag in Nederland eigenlijk is - blijft ook na dit rapport onbeantwoord. Door de uiteenlopende beweringen en grote verschillen tussen de wetenschapsgebieden kon de Commissie geen algemeen oordeel geven over de praktijk van omgang met onderzoeksgegevens. Ze pleit daarom voor een grondig en representatief wetenschappelijk onderzoek door de organisaties TNO, NWO, VSNU en KNAW.

www.knaw.nl

Manuscripten koning Arthur op tafel

Online database over koning Arthur en de ridders van de ronde tafel overstijgt cultuur- en taalgrenzen. Peter Boot

Eén van de belangrijkste genres in de westerse literatuur wordt gevormd door de verhalen over koning Arthur en de ridders van de ronde tafel. De verhalen waren al in de middeleeuwen populair in heel Europa, en zijn dus ook overgeleverd in manuscripten uit verschillende talen en cultuurgebieden.

In het door CLARIN ondersteunde project *Arthurian Fiction* is door ontwikkelaars van Huygens ING een online database gemaakt met gegevens over de verhalen, de manuscripten en hun bewaarplaatsen. Zoals projectleider Bart Besamusca in de begeleidende video op de website toelicht, weerspiegelen de verhalen de thema's waarover de culturele elite van Europa in de mid-

deleeuwen nadacht. Voor de (vergelijkende) cultuurgeschiedenis zijn de verhalen dus van groot belang, en wereldwijd zijn er zo'n veertienhonderd onderzoekers betrokken bij het onderzoek daarnaar.

Online database

In de database kan op een groot aantal ingangen worden gezocht, zoals op de voornaamste held van het verhaal, de bewaarplaats van het manuscript, periode, geografische oorsprong. Zelfs aan magische objecten (zoals ringen, drankjes, en natuurlijk de graal) is gedacht. Zoekacties kunnen steeds worden aangescherpt via het principe van de *faceted search*: naast de lijst met treffers worden de zoekcriteria getoond waarmee aanvullende eisen aan het zoekresultaat kunnen worden gesteld. Zoekacties kunnen worden opgeslagen en ook via een URL in publicaties worden opgenomen.



De Arthur-verhalen leven voort, zoals op deze postzegel
foto Museum voor Communicatie

Hoe volledig is de database? Desgevraagd licht Besamusca toe dat in principe alle bekende handschriften zijn opgenomen, maar dat de mate van detail nog erg wisselt. De gegevens zijn al verzameld in een eerder door NWO ondersteund

project. Het is de bedoeling dat een beperkte groep redacteurs de gegevens gaat aanvullen. Er worden ook nog steeds nieuwe handschriften, vaak fragmenten, ontdekt.

Norris Lacy, erevoorzitter van de International Arthurian Society, is enthousiast over de database. "De database brengt gegevens bijeen die voorheen alleen fragmentarisch toegankelijk waren, en maakt ze bovendien doorzoekbaar. Vooral het feit dat de database cultuur- en taalgrenzen overstijgt, is van belang." Gevraagd naar resterende wensen verklaart Besamusca een tevreden mens te zijn. Ook de bouw van de database verliep heel prettig, mede dankzij de meedenkende ontwikkelaars. "Ik zie uit naar de mogelijkheid om in de toekomst verwijzingen toe te voegen naar gedigitaliseerde manuscripten."

www.arthurianfiction.org



E-DATA & RESEARCH

Jaargang 7 | 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: CBS, CentERdata, CLARIN-NL, DANS, Huygens ING, Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica.

INHOUD

- 2 Interessante bijeenkomsten in de Agenda
- 3 Digitale pelgrimverslagen bron voor onderzoek
- 3 Text mining in onderzoek naar referentieculturen
- 4 Deze datasets zijn sinds kort beschikbaar
- 5 Peter Wittenburg over de digitale wetenschap



- 6 Valorisatie centraal in CATCHPlus
- 7 HISGIS neemt je mee naar het verleden
- 8 Nominaties Nederlandse Dataprijs 2012 bekend
- 8 Ewoud Sanders blijft er van schrikken



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

ICPSR-zomerschool kwantitatieve methoden

Thijs van den Broek
en Sedef Turper

Elk jaar organiseert het International Consortium of Political and Social Research (ICPSR) voor studenten en onderzoekers in de sociale wetenschappen een zomerschool kwantitatieve methoden. Het programma biedt een breed scala aan cursussen op het gebied van statistiek, onderzoeksontwerp en sociale methodologie. De cursussen worden gegeven door vooraanstaande onderzoekers en aangevuld met lab-sessies. Behalve de reguliere sessies is er ook een grote keuze uit korte seminars die de *state of the art* in de sociale wetenschappen toelichten. Een bijzondere verrijking vormen de *Blalock Memorial Lecture Series*, veelzijdige en prikkelende colleges waar nieuwe ontwikkelingen op het gebied van kwantitatieve onderzoeksmethoden worden gepresenteerd. Douglas Steinley behandelde *Latent Class Analysis*, een reeks technieken voor het schatten van niet-geobserveerde, categorische variabelen uit de structurele relaties tussen wel geobserveerde variabelen. In de familie-sociologie wordt deze datagedreven methode bijvoorbeeld gebruikt om typen familierelaties op te sporen. Chris Zorn hield een inspirerende bespiegeling over de implicaties voor sociale wetenschappers van gigantische databestanden. Zulke bestanden, bijvoorbeeld afkomstig van sociale media, klantenkaarten of logs van telecommunicatiegedrag, kunnen leiden tot een meer inductieve benadering van onderzoek. Computervaardigheden worden voor onderzoekers misschien wel belangrijker dan methodologi-

sche vaardigheden. Evan Ringquist presenteerde de mogelijkheden van meta-analyse. Het gaat daarbij om technieken om op een systematische wijze tot generaliseringsen te komen uit een groot aantal



Studenten aan de slag bij cursus sociale methodologie foto ICPSR

losse studies. Vooral bij omstrede, elkaar tegensprekende theorieën kan meta-analyse van belang zijn. Dit type analyse is sterk in opkomst, maar kent ook methodologische knelpunten. Overigens is het ICPSR-zomerprogramma ook een uitgelezen gelegenheid om te netwerken. Het programma voorziet ook in picknicks en andere informele gelegenheden om professionele en persoonlijke contacten te leggen. Mede dankzij een bijdrage van DANS konden Thijs van den Broek en Sedef Turper deze ICPSR-zomerschool bijwonen.

www.icpsr.com

Digitale diversiteit bij Digital Humanities '12

Peter Boot

Onder het motto 'Digitale diversiteit

GEHOORD & BIJGEWOOND

in cultuur, taal en methode' werd deze zomer in Hamburg de conferentie Digital Humanities 2012 gehouden. Digital Humanities is de belangrijkste conferentie binnen dit vakgebied, en wordt ook steeds groter: inmiddels zijn er vijf parallelle tracks. Daaronder is een track met *short papers*: met achttien minuten inclusief discussie nog niet zo kort als in sommige andere vakgebieden, maar toch misschien eerder gericht op tempo dan op diepgang. De diversiteit neemt inderdaad toe, hoewel in veel opzichten de Engelse taal en cultuur blijven domineren, niet alleen als *lingua franca* (logisch) maar ook als object van onderzoek. De situatie verbetert langzaam, met een toenemend aantal gedelegeerden uit Japan en China, en eigen organisaties voor Digital Humanities onderzoek in Canada, Japan en Duitsland.

Een opvallende trend, vond ik, was de beperkte aandacht voor digitale edities. Bij vorige Digital Humanities-bijeenkomsten waren de digitale editie en de rol van TEI (*Text Encoding Initiative*) vaak een be-



Twee deelnemers Digital Humanities 2012 foto Lou Burnard

langrijk onderwerp. Misschien wordt de editie steeds meer een onderdeel van grotere collecties en TEI meer een vanzelfsprekend stuk van de infrastructuur.

De meest inspirerende papers waren in elk geval heel divers en hadden betrekking op het opsporen van intertextualiteit, toegepast op de invloed van Voltaire in de Encyclopédie; op beïnvloeding in een corpus van 3.000 negentiende-eeuwse romans, met intrigerende visualisaties; op vaststellen van stemma's (afstammingsrelaties) tussen middeleeuwse handschriften; op visualisatie van de strategieën van vroegmoderne drukkers op basis van gedetailleerde codering van eigenschappen van hun titelpagina's; op het gebruik van bepaalde grammaticale constructies bij de vaststelling van auteurschap; en op hulpmiddelen voor het exploreren van e-mailcollecties, speciaal voor musea en archieven die de digitale nalatenschap van schrijvers en kunstenaars beheren.

www.dh2012.uni-hamburg.de

38^e IASSIST-conferentie behandelt authenticiteit

Rob Grim

Tegenwoordig ziet elke *digital* curator zich geconfronteerd met *digital born* archiefmateriaal. Maar hoe kan het publiek er zeker van zijn dat een digitale informatiebron authentiek is en niet bewerkt? Dergelijke vragen stonden centraal in de *keynote* van Mike Wash op de 38^e IASSIST-conferentie, die begin juni in Washington werd gehouden met als motto *Data Science for a Con-*

nected World: Unlocking and Harnessing the Power of Information. Wash' werkgever, het Nationaal archief (NARA), is verantwoordelijk voor de archivering van alle documenten van regering en congres. Informatie- en datamanagement en infrastructuur van NARA staan voor grote uitdagingen rondom *Big Data* en het verwerken van enorme hoeveelheden data. Wash besprak een aantal spraakmakende voorbeelden uit de praktijk. De data van de 2010 Census – 330 TB groot – werden door een vrachtwagen afgeleverd. Bij de publicatie van de 1940 Census werd de eerste twee dagen voor 200 TB aan foto's en documenten gedownload. De kosten om dergelijke pieken in de vraag op te vangen, zijn enorm. En kan men nog zoals voorheen gebruikelijk was elk document waarop een president te zien is of genoemd wordt in het archief opnemen? De archivering van de Obama-periode loopt wellicht al in de petabytes. Een indrukwekkende presentatie was ook die van Julia Lane over wetenschapsbeleid en een methodologie voor het meten van wetenschappelijke opbrengsten, het resultaat van een groot NSF-project (veertig jaren). Haar opdracht was een samenhangend stelsel van *science metrics* te ontwikkelen dat een beter inzicht geeft in de wetenschappelijke productie, vanaf de subsidieaanvraag tot aan de opbrengsten, zoals publicaties, patentaanvragen of databestanden. Met behulp van *data mining* brengt het project het volledige netwerk in kaart van personen en onderwerpen die bij het onderzoek betrokken zijn. Deze aanpak levert veel objectieve en uiterst waardevolle informatie op.

www.iassist2012.org

AGENDA

11 oktober • Amsterdam

Talentendag vrouwelijke wetenschappers

NWO en het Landelijk Netwerk Vrouwelijke Hoogleraren organiseren een talentendag om vrouwelijke wetenschappers te helpen met hun carrière in de universitaire wereld en om de doorstroom van vrouwelijke wetenschappers naar hogere posities te verbeteren.

www.nwo.nl/pumpyourcareer

11 - 12 oktober • Amsterdam

eCommons 'Sustainable Futures for Digital Archives'

De conferentie, gericht op audiovisuele archieven, behandelt twee thema's: 1. digitalisering, archivering en presentatie van film, audio, video en fotografie en 2. de toekomst van archieven: toegang en modellen voor beheer en behoud.

<http://ecommons.eu/>

18 oktober • Den Haag

Prijzuitreiking Nederlandse Dataprijs 2012

Op 18 oktober wordt de winnaar van de Nederlandse Dataprijs voor humaniora en sociale

wetenschappen en voor exacte en technische wetenschappen bekend gemaakt. De prijsuitreiking bijwonen kan, aanmelden is wel verplicht (via de website van DANS).

www.dans.knaw.nl

22 - 24 oktober • Barcelona (Spanje)

1st EUDAT Conference

EUDAT (EUropean DATA infrastructure) stimuleert samenwerking tussen Europese onderzoeksinfrastructuren. Speerpunten tijdens de conferentie zijn datamanagement, toegang tot onderzoeksdata en digitale archivering.

www.eudat.eu/1st-conference

26 oktober • Amsterdam

Science for Society: How?

Een symposium georganiseerd door de KNAW over wetenschaps- en innovatiebeleid in internationaal perspectief.

www.knaw.nl/Pages/DEF/33/588.bGFuZz:1OTA.html

6 - 7 november • Frascati (Italië)

Alliance Permanent Access (APA) Conference

Een groot aantal Europese organisaties op het gebied van digitale duurzaamheid maken deel uit van 'Alliance for Permanent Access'. Het

thema van deze conferentie is 'Value from data now and into the future'.

www.alliancepermanentaccess.org/index.php/2012/08/06/apa-2012-conference-6-7-nov/

22 - 24 november • Amsterdam

9^e conferentie van de European Society for Textual Scholarship

De conferentie bestudeert verschillen in benadering in het uitgeven van historische en letterkundige bronnen. Thema's zijn onder andere het omgaan met massale hoeveelheden data, 'mash-ups' van verschillende soorten bronnen en de mogelijkheden van crowdsourcing.

www.textualscholarship.nl/?p=10313

3 - 4 december • Rotterdam

Digital Erfgoed Conferentie 2012

Tijdens conferentie met als thema 'Beter Weten, interactie erfgoed en wetenschap', staat de vraag centraal hoe erfgoedinnovatie en wetenschap elkaar beïnvloeden. Met andere woorden: 'What can science do for you and what can you do for science?'.

www.den.nl/DEconferentie2012

10 december • Florence (Italië)

Data Seal of Approval Conference 2012

Op deze conferentie komt onder andere informatie aan bod over de wijze waarop repositories het keurmerk aan kunnen vragen. De conferentie maakt deel uit van het symposium 'Cultural Heritage online'.

www.dans.knaw.nl

14 - 16 januari 2013 • Amsterdam

8th International Digital Curation Conference (IDCC)

De IDCC conferentie brengt mensen samen die datasets maken, beheren en gebruiken. Het thema van de conferentie is 'Innovation: driving the Data Science agenda'.

www.dcc.ac.uk/events/idcc13

12 - 15 februari 2013 • Forth Worth (VS)

iConference Scholarship in Action

De iConference is een jaarlijkse bijeenkomst van wetenschappers, afkomstig uit verschillende wetenschapsgebieden, die zich bezig houden met informatie aspecten in de hedendaagse samenleving. iConference deelnemers onderzoeken de grenzen van informatiestudies, verkennen nieuwe concepten en ideeën en nieuwe technologische en conceptuele configuraties.

www.ischools.org/iConference13/2013index/

Pelgrimsverslagen: 'Dichterbij kun je niet komen'

Wat bezielt de pelgrim?

Kwalitatief onderzoek naar de motieven van moderne pelgrims is tijdrovend werk. Een database met geannoteerde verslagen van pelgrims moet daar verandering in brengen.

Erica Renckens

Zelf zullen de onderzoekers er niet zo snel aan beginnen. Veel te ver. En veel te massaal. Toch leggen elk jaar meer en meer mensen de pelgrimstocht naar Santiago de Compostela in Spanje af. Met name de laatste honderd kilometer – minimaal benodigd voor het verkrijgen van de 'compostela', het certificaat – is zó druk, dat inmiddels gezocht wordt naar alternatieve routes. Wat bezielt deze mensen om deze lange tocht te voet, te fiets of te paard af te leggen?

Motieven van pelgrims

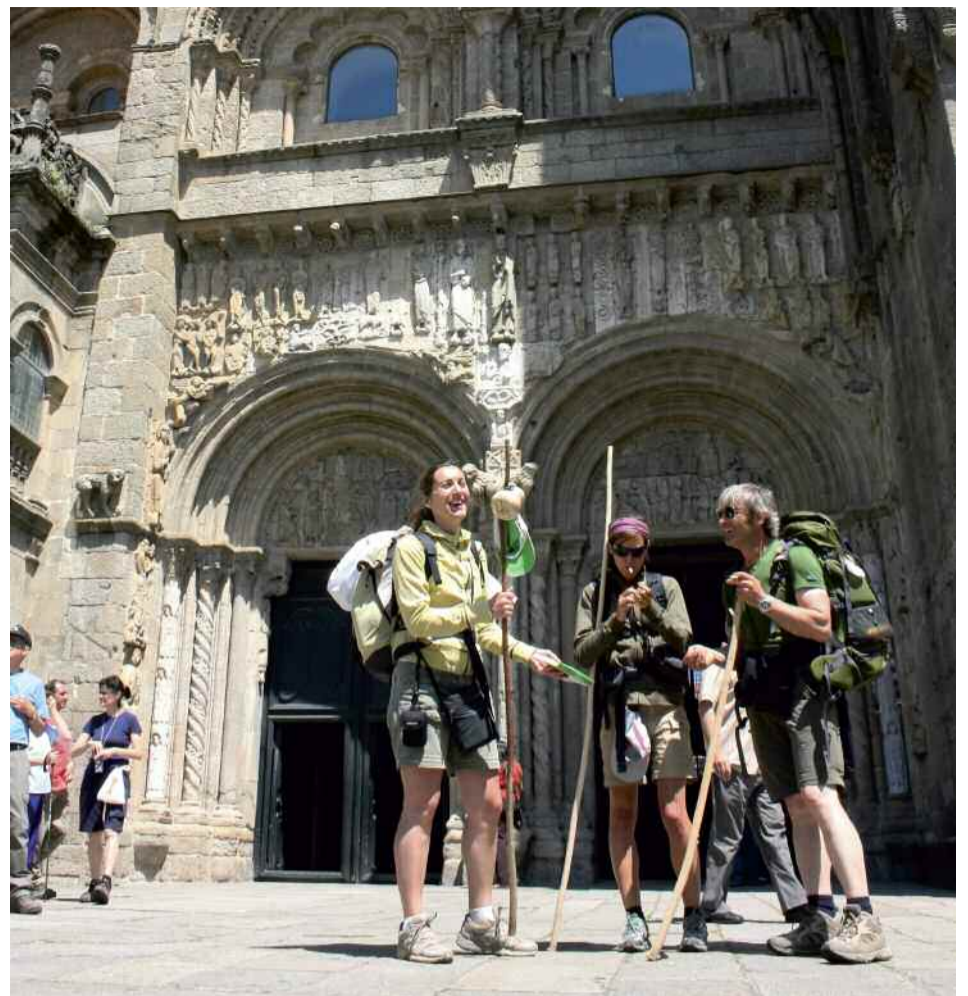
Paul Post en Laurie Faro, beiden verbonden aan Tilburg University (School of Humanities, onderzoeksgroep Religie en Ritueel), doen al jaren onderzoek naar de motieven van pelgrims. "We ontdekten al vrij snel dat de verslagen die de pelgrims zelf van hun tocht schreven een interessante bron voor ons onderzoek waren. Dichterbij kun je niet komen",

vertelt Post, terwijl hij bladert door het tijdschrift van het Nederlands Genootschap van Sint Jacob, waar meer dan tienduizend Nederlandse pelgrims lid van zijn. "Het kost alleen heel veel tijd om zulke verslagen handmatig te analyseren. Meestal kun je niet meer dan tien verslagen in je onderzoek meenemen."

Daar moet het *Corpus of Pilgrimage Narratives* verandering in brengen. In deze database, gefinancierd door CLARIN-NL, zullen een paar duizend Santiago-pelgrimsverslagen vanaf de jaren '80/'90 opgenomen worden. De Tilburgse onderzoekers werken in dit project samen met het *Meertens Instituut*, dat veel ervaring heeft in het ontwikkelen van digitale corpora. "Alle verslagen zullen uitvoerig worden voorzien van metadata, zodat de gebruiker beschikt over informatie over de auteur, de bron en trefwoorden", aldus Faro. In de database zullen niet alleen verslagen uit genootschapstijdschriften worden opgenomen, maar ook verslagen als blogs en andere sociale media. Het corpus zal nooit 'af' zijn: nieuwe verslagen kunnen doorlopend toegevoegd worden.

Persoonlijke beleving

Met geavanceerde zoektechnieken kunnen de gebruikers de pelgrimsverslagen inhoudelijk analyseren. "We hebben eerder een model ontwikkeld waarin hedendaagse rituelen grof-



Pelgrims aangekomen in Santiago de Compostela, Spanje foto Toni Sanchez Poy

weg worden verbonden met een vijftal zogenoemde 'fields of the sacred': religie & spiritualiteit, kunst & cultuur, healing, gedenken en recreatie", vertelt Post. "Met de database willen we onderzoeken welke velden een rol spelen voor Santiago-pelgrims. Ze voelen zich méér dan alleen wandelaars. Aan welke velden ontleen ze identiteit? Zijn er meerdere velden samen in het spel? En veranderen deze velden tijdens de tocht?"

"We zijn ook benieuwd naar de invloed van het medium op de inhoud van het verslag", vertelt Faro. "Op weblogs en sociale media gaat het er vooral om jezelf te presenteren. Je

schrijft niet voor jezelf, maar om te delen met anderen. We verwachten dat daardoor de traditionele focus op cultuurhistorische aspecten van de tocht zal verschuiven naar de persoonlijke beleving van de pelgrim."

Voor Faro en Post zal het project geslaagd zijn als de database er is, blijft groeien en onderzoekers er ook daadwerkelijk gebruik van maken. Voor dat laatste lonkt er al een concreet perspectief: vanaf 2013 begint een Tilburgse e-humanities-promovendus een onderzoek naar het profiel van de 'moderne pelgrim' op basis van het corpus.

Schatgraven naar sentimenten

E-onderzoek naar referentieculturen

Is 'text mining' bruikbaar bij onderzoek naar referentieculturen? Een samenwerking tussen UU, UvA en Huygens

ING. Peter Boot

Is het Amerikaanse bedrijfsleven een positief of een afschrikwekkend voorbeeld? Hoe zagen Nederlanders de Amerikaanse consumptiecultuur of media? En kunnen we *text mining* gebruiken om antwoorden op deze en soortgelijke vragen te vinden in de steeds grotere digitale corpora die beschikbaar komen?

Samenwerking

NWO heeft in het *Horizon-programma* een subsidie van 2 miljoen toegekend voor een onderzoek naar deze vragen. Het project *E-onderzoek naar Referentieculturen: De opkomst van de Verenigde Staten in het publieke discours in Nederland, 1890-1990* is aangevraagd door Joris van Eijnatten (UU) in samenwerking met Toine Pieters en Jaap Verheul (UU), Maarten de Rijke (UvA) en Charles van de Heuvel (Huygens ING). Het vijfjarige programma onderzoekt hoe Nederlanders aan-

keken tegen de producten, praktijken en ideeën die met de Verenigde Staten werden geassocieerd. De VS zijn daarbij een voorbeeld van een 'referentiecultuur': een cultureel model dat al dan niet geïmiteerd wordt



Uitingen zoals dit affiche uit 1963 beïnvloeden de beeldvorming over Amerika foto Stichting NAGO

maar uit de discussie in geen geval is weg te denken.

Uitgangspunt vormt de collectie historische kranten en tijdschriften van de Koninklijke Bibliotheek. Daarin worden in eerste instantie entiteiten (personen, organisaties) onderkend en de relaties daartussen geïnventariseerd. Ook andere digitale collecties zullen in het onderzoek worden betrokken, zoals die van het Instituut voor Beeld en Geluid, of meer specialistische collecties van bijvoorbeeld het verzekeringswezen. Aan de hand daarvan worden de gevonden entiteiten voorzien van contextinformatie, zodat het systeem de tekst tot op zekere hoogte 'begrijpt'. Met *sentiment mining* wordt vervolgens onderzocht hoe deze entiteiten worden gewaardeerd. De onderzoeker krijgt zo veel sneller dan voorheen inzicht in de wisselende mening over, bijvoorbeeld, Amerikaanse consumptieartikelen.

Zo makkelijk als Google

Gevraagd naar de doelstelling van het project verklaart Jaap Verheul, één van de projectleiders: "Er bestaat al veel technologie voor het doorzoeken van grote bestanden. Voor geesteswetenschappers is die technologie echter slecht toegankelijk. We zijn op zoek naar een

manier om deze techniek net zo makkelijk inzetbaar te maken als Google. Daarvoor zullen de informatici zich ook verdiepen in de processen van het historisch onderzoek. In het project werken twee mediators speciaal aan de vertaling van onderzoekerswensen naar techniek en de inpassing van de techniek in de onderzoekspraktijk. Daar gaan we ook over publiceren."

Zijn zulke intermediairs nog steeds nodig? José de Kruif, ICT-consulent in Utrecht en nauw betrokken bij deze aanvraag: "Zeker. Het gaat niet zomaar om softwaregebruik, of om het op maat laten bouwen van tools. Het gaat ook om een andere manier van denken en een andere methodologie. Als wij een voldoende 'geesteswetenschapvriendelijk' instrumentarium afleveren, wordt het gebruik van digitale technieken in de toekomst misschien makkelijker." Jaap Verheul: "En alleen de praktijk kan dat leren. We willen toe naar een iteratieve methode van onderzoek, waarin fasen van interpretatie (kwalitatief) en kwantitatieve toetsing elkaar afwisselen. Wat het beste werkt, kun je alleen samen uitvinden. Wat dat betreft zien we onszelf wel als vroedvrouw bij de geboorte van deze nieuwe techniek."

SINDS KORT BESCHIKBAAR

Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij het Centraal Bureau voor de Statistiek, CentERdata, Data Archiving and Networked Services en Huygens ING.

CBS

- **Geregistreerde (weg)voertuigen van personen, 2010-2011**

Dit bestand bevat technische kenmerken en eigendomsinformatie over geregistreerde (weg)voertuigen in Nederland. De voertuigen zijn onderdeel van het actieve park en behoren aan natuurlijke personen. Voorbeelden zijn personenauto's, bestelauto's, bromfietsen, motorfietsen, vrachtwagens, autobussen en trekkers. De data zijn afkomstig uit het jaarbestand van de Rijksdienst voor Wegverkeer (RDW).



Eerste blauwe kentekenplaat (2 witte letters en 4 cijfers) in Nederland bron RDW

- **Ook sinds kort beschikbaar:**

- Arbeidsmarktpanel, 1999-2009
- Populatie Actieve Bedrijven, 2007-2012
- Regionale Eenheden Base, 2006-2011
- Koppelbestand nieuwe naar oude baan-definitie, 2006-2009
- Community Innovation Survey, 2010
- Hoogste opleiding, 1999-2009
- Burgerlijke staat van personen in de GBA, 2011
- Migratiebewegingen van personen in de GBA, 2011
- Nationaliteitsinformatie van personen in de GBA, 2011
- Recent arbeidsverleden van personen in Nederland, 2004-2008
- Adresnummer met gemeente- wijk- en buurtcode, 2012
- Adresnummers met soort woonruimten longitudinaal vanaf 1992, 2009
- Energieverbruik, 2010



Deze bestanden zijn beschikbaar via www.cbs.nl/micro-data. Bezoek deze site of scan de QR code.

CentERdata

- **Studies Immigrant panel, A. Aydinli; F. van de Vijver, Volunteering, 2011**

Heeft het denken aan vrijwilligerswerk effect op de bereidheid om te helpen? Dit hebben Arzu Aydinli en Fons van de Vijver in december 2011 onderzocht onder immigranten. In hun studie *Volunteering* werden respondenten willekeurig verdeeld in drie groepen. Ze werden gevraagd om te denken aan een gebeurtenis waarin ze iemand anders geholpen hadden, of aan een situatie waarin ze dachten aan of spraken over vrijwilligers. Een derde groep moest denken aan gewone dagelijkse routines. Vervolgens beschreven



Kinderen helpen bij het leren lezen foto Shutterstock

de respondenten deze gebeurtenissen. Daarna gaven ze aan in hoeverre ze bereid zouden zijn om verschillende taken, zoals een groep kinderen leren lezen, te verrichten. De data uit het onderzoek zijn beschikbaar via LISS Data Archive.

- **Ook sinds kort beschikbaar:**

- **Studies LISS panel**

- J. Yabar; R. Pieters; J. Leenheer, Tilburg Consumer Outlook Monitor, wave 2011 June
- E. Wisse, Decreasing Trust in Democratic Governments, wave 5, 2011
- R. Winkelmann; R. Studer, Visual Analogue Scale and Item Response Time Measurement in Happiness Research Part 1 + Part 2, 2011
- M. van Overveld; CentERdata, Financial Literacy, 2011
- R. Meuleman; M. Lubbers, Nationalism and the national dimension of cultural consumption, wave 1, 2011
- CentERdata, Religion and Ethnicity, wave 5, 2012
- O. Paccagnella; F. Bassi; R. Variata, Vignette evaluation of customer satisfaction - VECS project - wave 1, 2011
- CentERdata, Health, wave 5, 2011
- CentERdata, Social Integration and Leisure, wave 5, 2012

- **Studies Immigrant panel**

- O. Celenk; F. van de Vijver, Relationship satisfaction and attitudes, 2011



Deze bestanden zijn beschikbaar via www.lissdata.nl/data archive. Bezoek deze site of scan de QR code.

DANS

- **SKON, CBS, Henk van der Kolk (Universiteit Twente), Kees Aarts (Universiteit Twente), Jean Tillie (Universiteit van Amsterdam), Nationaal Kiesonderzoek, 2010 - NKO 2010; Dutch Parliamentary Election Study 2010 - DPES 2010**

Het Nationaal Kiesonderzoek (NKO) is een uniek databestand met betrekking tot het Nederlandse electoraat. De kern van het NKO bestaat uit een survey-onderzoek rond de verkiezingen voor de Tweede Kamer onder een representatieve steekproef uit de kiesgerechtigde bevolking van Nederland. In de vraaggesprekken wordt een groot aantal politieke opvattingen en prioriteiten, percepties van politieke verschijnselen, politieke gedragingen (waaronder uiteraard stemgedrag) en andere kenmerken van de respondenten vastgesteld. Sinds kort zijn de data van het

NKO 2010 beschikbaar. Het bestand met alle variabelen is op aanvraag en onder speciale voorwaarden beschikbaar bij het Centrum voor Beleidsstatistiek van het CBS. In het online archiveringssysteem EASY is een geanonimiseerd bestand beschikbaar. <https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:50534>

- **Uitbreiding data historische tellingen**

Onlangs zijn de datasets van historische tellingen in EASY aangevuld met 117.000 images van de Beroeps- en Bedrijfstelling 1930 en de Bedrijfstellingen 1950 en 1963. Binnenkort volgen 109.000 images van de Woonstellingen 1947 en 1956 en de Volkstelling 1960. Betrokken partijen zijn het CBS en DANS, samen werken zij aan de digitale ontsluiting van historische tellingen via EASY en www.volkstellingen.nl. Het bestuderen van deze gegevens zal onderzoekers nieuwe inzichten opleveren over de Nederlandse bevolking in een periode van snelle industrialisering en verstedelijking, gedurende de tweede helft van de negentiende en de eerste helft van de twintigste eeuw. [urn:nbn.nl:ui:13-tkk-5kw](http://nbn.nl/ui:13-tkk-5kw)

VOLKSTELLING.	
IN	
'LANDS VAN UTRECHT.	
UTRECHT.	
	Zielen. Grond- verg.
Utrecht	33294
NIEUWERSLUIS.	
Abcoude	1059
Nichtevegt	293
Nederhorst den Bergh	520
Vreeland	605
Kortenhoef	527
Ankeveen	174
Baanbrugge	441
Hoenderloot	693
Oudekoop	178
Loenen, Stigs en Nieuwerluis	174
Thamen	582
Uithoorn	640
Denemark	658
Oudbuysen	384
	270
	39399
	80

Volkstelling in 's Lands van Utrecht 1795 bron www.volkstellingen.nl

- **Ook sinds kort beschikbaar:**

- **Archeologie**

- Groothedde, M. (2011-09-01), De vorstelijke palts van Zutphen, 2011
- Alma, X.J.F.; ADC ArcheoProjecten, 's-Hertogenbosch, Dode Nieuwstraat Minderbroedersklooster, 2011

- **Historische Wetenschappen**

- Museon, getuigenverhalen Atlantikwall Den Haag: Interviews 01 t/m 15
- M. Marx - Universiteit van Amsterdam - Instituut voor Informatica, Political Mashup Dataset 1814-2012

- **Taalkunde**

- CLARIN-NL, TDS Curator - Typological Database TDS Curator Documentation 2012
- **Sociale Wetenschappen**

- Inspectie van het Onderwijs, Toezichtkaart (voorheen Kwaliteitskaart) Voortgezet Onderwijs 2012
- Gjalte-Jorn Peters, Threatening communication: a critical re-analysis and a revised meta-analytic test of fear appeal theory 2012
- Centraal Bureau voor de Statistiek en Rijkswaterstaat, Onderzoek Verplaatsingen in Nederland 2011
- R. Knegt - Universiteit van Amsterdam Hugo Sinzheimer Instituut, Werkgeverskosten van

arbeidsgerelateerde schade 2011

- Jean Tillie, Universiteit van Amsterdam, opkomst en stemgedrag migranten gemeenteraadsverkiezingen 1994-2010



Via <https://easy.dans.knaw.nl/> zijn deze bestanden beschikbaar. Bezoek deze site of scan de QR code.

Huygens ING

- **Een vriendschap in het teken van 's Lands financiën. Briefwisseling tussen Elias Canneman en Isaac Jan Alexander Gogel 1799-1813, bewerkt door Mieke van Leeuwen-Canneman**

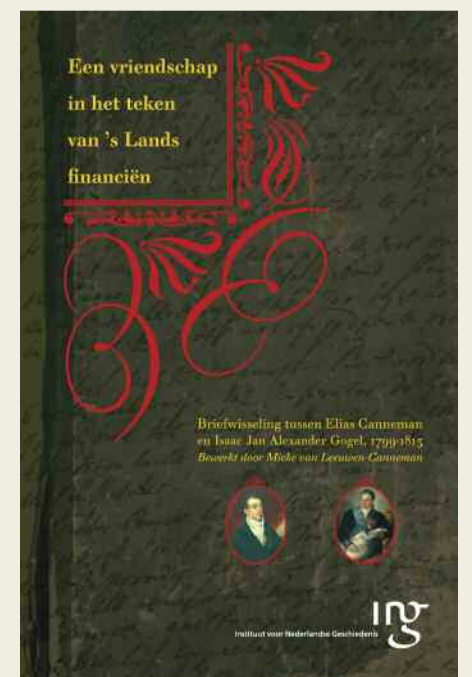
Elias Canneman en Isaac Gogel hebben een belangrijke rol gespeeld in het nieuwe, nationale, financiële beleid in de jaren die volgden op de Bataafse Omwenteling van 1795. De briefwisseling, bewerkt door Mieke van Leeuwen-Canneman, laat niet alleen een uniek kijkje achter de schermen van het politieke bedrijf zien, maar vult het beeld aan dat bestudering van alleen officiële bronnen oproept. In het bijzonder vormt zij een bron van kennis over de financieel-economische en politieke verhoudingen, over de samenwerking tussen de financiën van de staat en het particuliere koopmansbedrijf en over de operationele kant van transacties tussen deze partijen. Maar ook biedt deze privé-correspondentie een beeld van de twee invloedrijke hoofdpersonen en de tijd waarin zij leefden. www.historici.nl/retroboeken/gogel_canneman

- **Ook sinds kort beschikbaar:**

- Sicke Benninge, Croniken der Vrescher Landen mijten Zieven Seelanden ende der stad Groningen
- Landsheerlijke archieven van Gelre, Holland, Zeeland en het Sticht
- Bronnen tot de bouwgeschiedenis van den Dom te Utrecht, 3 delen
- Dutch Asiatic Shipping 1595-1795, het eerste, inleidende deel



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



Omslag boek 'Een vriendschap in het teken van 's Lands financiën' bron www.historici.nl

Peter Wittenburg, grondlegger van The Language Archive:

‘De vraag van morgen moet je vandaag kunnen beantwoorden’

Eén jaar na lancering van ‘The Language Archive’ spreekt E-data & Research met Peter Wittenburg. Grondlegger van TLA, werkzaam bij het Max Planck Instituut voor Psycholinguïstiek vanaf de start en (op papier) met pensioen. Hij vertelt ons hoe hij de digitale geesteswetenschappen heeft zien opgroeien.

Heidi Berkhout en Erica Renckens

De methode

“Moet je je voorstellen. 35 jaar geleden wilden we handgebaren tijdens het spreken meten. We installeerden twee infraroodcamera’s en plakten infraroodlampjes op een hand. Samen maakten ze de driedimensionale observatie mogelijk van een bewegende hand. De methode was erg storingsgevoelig. Zo kon je de hand niet draaien, daarmee verbrak je de verbinding tussen licht en camera. Vervolgens had een computer een nacht nodig om de data te verwerken. Tegenwoordig is minder inspanning nodig, zijn er laboratoria gespecialiseerd in deze onderzoeksmethode. Zij gebruiken bijvoorbeeld zestien camera’s in plaats van twee. Uiteindelijk lijkt het wel of alles makkelijker wordt, waardoor meer en betere experimenten mogelijk zijn.”

De analyse

“In 1976, bij de start van het MPI, hadden we maar één PDP-11/55, toentertijd een supersnelle computer. Er werkte hier een Engelse onderzoeker die we ‘het spook’ noemden. Hij deed de zware berekeningen voor experimenten met de infraroodtechnologie ‘s nachts, zodat de computer overdag beschikbaar was voor andere werkzaamheden. Als wij ‘s ochtends binnenkwamen, ging hij weg, slapen. Dit heeft hij zeker een jaar volgehouden. Wat een pionierswerk.”

De inspiratie

“Deze zomer ontmoette ik een invloedrijke onderzoekster tijdens een workshop. Vol trots vertelde ze dat ze sneller en beter analyses kon uitvoeren op haar onderzoeksdata dankzij de mogelijkheid om in ELAN (een professionele tool voor het maken van annotaties bij video- en audio-opnames, red.) te zoeken met reguliere expressies. Er was een wereld voor haar opengestaan. Ik vind dat schattig en grappig. ELAN maakt dit type onderzoek al jaren mogelijk. Maar het duurt altijd vrij lang voor nieuwe technieken zijn doorgedrongen tot kleinere afdelingen op universiteiten. Onderzoekers moeten de tijd en rust vinden om zich erin te verdiepen, zodat ze zich het systeem eigen kunnen maken. De wijze waarop deze dame over haar recente ontdekking vertelde, enthousiasmeert anderen om deze applicatie ook te gebruiken. Zo verspreidt het gebruik



Peter Wittenburg: “Uiteindelijk lijkt het wel of alles makkelijker wordt” foto Heidi Berkhout

Peter Wittenburg

Peter Wittenburg werkt sinds de oprichting in 1976 bij het MPI. Peter was ruim een jaar hoofd van The Language Archive, dit najaar zal een opvolger gekozen worden en zal hij als adviseur bij TLA betrokken blijven. Het doel van TLA is om digitale taalbronnen op te slaan en te behouden, om onderzoekers en andere geïnteresseerde gebruikers daartoe toegang te verlenen en om nieuwe technologieën die taalonderzoek bevorderen te ontwikkelen en te integreren. Naast TLA is Peter actief binnen digitale infrastructuur zoals CLARIN (Common Language Resources and Tools Infrastructure) en EUDAT (European Data Infrastructure).

zich verder. Zij verovert op dat moment de wereld. En dat doet mij goed.”

De wereld anno 2012

“Tegenwoordig beschikken we over betere technologieën waarmee we veel meer data kunnen analyseren. Dit zorgt ervoor dat, met name binnen observationeel onderzoek, een heel ander type onderzoek mogelijk is. Hoe hebben talen zich in de loop der tijd ontwikkeld? Zo’n onderzoeksvraag kunnen we sinds een jaar of vier dankzij de technologie beantwoorden door gebruik te maken van algoritmes uit de erfelijkheidsleer. We kunnen nu veel meer taalkundige kenmerken meenemen in onze kwantitatieve analyse.”

De uitdaging

“Eén van de grootste uitdagingen op dit moment is dat veel video- en audiomateriaal beschikbaar is, maar dat lang niet alles geannoteerd is. Begrijpelijk, er is sprake van een factor 1:35; het kost 35 uur om 1 uur audio handmatig te transcriberen. We hebben 80 terabyte in ons archief, dus het is praktisch onmogelijk om al het materiaal te annoteren. Maar zonder

annotaties is geen taalkundig onderzoek mogelijk. De vraag is: hoe is dit materiaal toch te ontsluiten? We werken aan een technologie om in audio- en video-signalen semi-automatisch patronen te vinden. In een bepaalde taal worden bijvoorbeeld meerdere werkwoorden geclusterd. We segmenteren dan een typisch voorbeeld van zo’n clustering om kenmerkende *audiofeatures* eruit te halen. Vervolgens kijkt de software of dit patroon vaker voorkomt in de opname. Het zijn en blijven kansberekeningen die alsnog handmatig gecontroleerd moeten worden, maar onderzoekers kunnen zo veel makkelijker het complete materiaal doorzoeken. Wij ontwikkelen dit in samenwerking met andere gespecialiseerde

instituten. Zo combineren we onze kennis over automatische spraakverwerking met kennis over audioverwerking (IAIS Institute Bonn) en kennis over videoverwerking (Heinrich Hertz Institute Berlijn).”

De toekomst

“Wat blijft, is dat het MPI alles zo open mogelijk wil maken. Zodat wetenschappers de kans krijgen van de nieuwste methoden en technologieën gebruik te maken. We zouden graag zien dat nog meer partijen hun technologieën beschikbaar stellen. Maar het is lastig, want het ‘open stellen van gebruik voor anderen’ betekent dat je moet investeren in iets waar je zelf al bij kan. Dat is niet eenvoudig, dat is missiewerk en kost geld.

Wij stellen op termijn onze patroonherkenningsalgoritmes als *web-services* beschikbaar. Om intensief rekenwerk op data door vele anderen mogelijk te maken, kopiëren we onze gegevens en de benodigde *services* naar een machine met grote computerkracht, denk bijvoorbeeld aan het computercentrum SARA. Andere onderzoekers kunnen hun data op dezelfde wijze ter beschikking stellen via het

project EUDAT. Aan EUDAT doen alle ‘SARA’s’ van Europa mee, zodat de gebruiker uit een groot data-aanbod kan kiezen. Onderzoekers spreken dan vanuit hun eigen computer bijvoorbeeld in ELAN een *web-service* aan, die de opgevraagde resultaten berekent en de annotaties terugstuurt. Het gebruik van deze ‘supercomputers’ past ook goed in de CLARIN-gedachte: maak de *data* en *tools* beschikbaar, zodat ze optimaal door iedereen gebruikt kunnen worden.

Het data- en toollandschap is momenteel nog sterk gefragmenteerd, dat is historisch zo gegroeid. Er bestaan verschillen in formaten,

in technologieën, in vragen. Er is dus ook niet één oplossing, één standaard. Verschillen-

de organisatie, zoals MPI, DANS en het Meertens Instituut, werken daarom samen aan gemeenschappelijke antwoorden op syntactische en semantische vraagstukken. Op het gebied van tools en diensten mag er best wel competitie zijn, maar als je dezelfde gebruikers bedient, moet je tenminste wat standaarden en *best practices* betreft samenwerken om het leven voor gebruikers makkelijk te maken.”

De Wet van Wittenburg

“Wees pragmatisch, kijk naar oplossingen die op de vloer gebruikt kunnen worden. En wees proactief. Als je denkt dat de wetenschapper morgen met een vraag kan komen, moet je vandaag werken aan de oplossingen. Het kost vaak ruim vijf jaar om een nieuwe technologie of methode – het gaat hier dus niet om een snel te maken script – breed in te kunnen zetten. En dus moet je vijf jaar eerder zijn dan de wetenschapper. Dat betekent dat je risico’s moet nemen. Dat doen wij, wij kijken altijd naar wat in de toekomst zou kunnen gebeuren. Zo kunnen we de wetenschap competitief houden. En maken we de grote vloed aan data voor de wetenschap toegankelijk.”

INTERVIEW

Al beïnvloed door één blik op advertentie

Marika de Bruijne

Naar advertenties op het internet, in bladen of op reclameborden wordt meestal niet of nauwelijks gekeken. Toch maakt het uit hoe lang men aan advertenties wordt blootgesteld. Zelfs als het verschil enkele seconden is, kan al een effect in attitudevorming en mate van herinnering worden aangetoond. Dit heeft Millie Elsen, die werkzaam is bij CentERdata, samen met Rik Pieters van Tilburg University onderzocht. De effectiviteit van lange of juist extreem korte, subliminale blootstelling was al bekend, maar de meeste praktijksituaties liggen juist daartussenin. Daarom heeft Elsen mensen gedurende 100 milliseconden tot enkele seconden naar verschillende advertenties laten kijken. Conclusie: die ene blik heeft al effect. Elsens proefschrift *Thin and Thicker Slices: How Advertising Effectiveness Depends on Exposure Duration* is vanaf oktober beschikbaar via Tilburg University Repository. <https://repository.uvt.nl/>

Voorganger E-data via EASY toegankelijk

Heidi Berkhout

Onlangs is de complete reeks van het blad *Historia & Informatica*, één van de voorlopers van E-data & Research, gearchiveerd. Het blad verscheen op initiatief van het Nederlands Historisch Data Archief (NHDA, later NHDA/NIWI), de Vereniging voor Geschiedenis en Informatica (VGI) en het Internationaal Instituut voor Sociale Geschiedenis (IISG) tussen 1994 en 2005. De reeks geeft als collectie een duidelijk beeld van de ontwikkelingen op het gebied van geschiedenis en informatica.

<https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:50714>



De voorpagina van de eerste H&I foto Dagmar Striebal

DANS gebruikerspanel

Wilt u meedenken over de diensten van DANS? Ga dan naar www.dans.knaw.nl en word lid van het gebruikerspanel.



Patricia Alkhoven foto Bart van Vliet

Erfgoedwereld digitaal steeds comfortabeler

CATCHPlus liep officieel per 1 juli af, maar aan de agenda van project-leider Patricia Alkhoven is dat nog niet te merken.

Erica Renckens

“Alle veertien deelprojecten zijn klaar of bevinden zich in de afrondende fase. Ze worden nog getest, gedocumenteerd en geëvalueerd”, vertelt Patricia Alkhoven. “Dat betekent dat we druk zijn alle relevante informatie te verzamelen, zodat ze voor iedereen via onze website in te zien zijn.”

CATCHPlus is een vervolgproject op CATCH, dat staat voor *Continuous Access to Cultural Heritage*. In CATCH werken onderzoekers uit de geesteswetenschappen en informatica samen met erfgoedinstellingen aan technieken om de digitale collecties te beheren. CATCHPlus gaat verder waar CATCH ophoudt. Waar is dat? “CATCH resulteerde in verschillende demonstratieversies, halfproducten en dissertaties”, legt Alkhoven uit. “In CATCHPlus zijn enkele van deze demo’s en halfproducten doorontwikkeld tot volwaardige producten die daadwerkelijk door erfgoedinstellingen worden gebruikt. Valoriseren staat dus centraal in CATCHPlus.”

Van de veertien CATCH-deelprojecten zijn er zeven geselecteerd voor doorontwikkeling tot eindproducten. Daarnaast coördineerde het CATCHPlus-projectbureau zelf de ontwikkeling van enkele centrale diensten. De zo ontwikkelde software varieert van een automatische herkenner voor oude handschriften tot een online module die aanbevelingen kan doen op basis van het

persoonlijke profiel van bezoekers. De centrale diensten bevatten onder andere een vocabulairebank en digitale workspaces waar de collectiebeheerder in kan werken.

Open data

Is de Nederlandse erfgoedwereld wel klaar voor deze toepassingen? “Digitaliseren van analoge bronnen en het opnemen van *born digital*-bestanden maken inmiddels deel uit van de dagelijkse praktijk van erfgoedinstellingen”, verzekert Alkhoven. “De afgelopen jaren heb ik het

‘digitale klimaat’ binnen de sector zien veranderen naar een veel opener attitude, zeker bij de grotere organisaties. De kleinere haken dan maar wat graag aan.”

“Ik heb de indruk dat de erfgoedwereld en de wetenschap langzaam naar elkaar toe aan het groeien zijn en dat ook bijvoorbeeld het gebruik van open data daaraan bijdraagt”, aldus Alkhoven. “In de *Topsector Creatieve Industrie* speelt hetzelfde: kennisinstellingen werken samen met de private sector. Dat doet CATCHPlus al vanaf het begin. Nu is het zaak die investeringen zo goed mogelijk te verzilveren en duurzaam verder te ontwikkelen.”

Behoud van resultaten

“We willen natuurlijk dat de resultaten uit CATCHPlus behouden blijven en onderdeel gaan uitmaken van het bedrijfsproces van de erfgoedinstellingen. Daarom ontwikkelen we samen een businessmodel dat aansluit bij de missie en de doelen van de organisatie”, aldus Alkhoven. “Daarnaast zoeken we een plek waar alle ontwikkelde software

ondergebracht kan worden, zodat ze maximaal geëxploiteerd en doorontwikkeld kunnen worden. We zoeken een ‘erfgoedmakelaar’ die deze taken voor zijn rekening neemt. Deze makelaar kan werken vanuit een instelling of vanuit een samenwerkingsverband van grotere erfgoedinstellingen.”

DE2012

Daarnaast zullen de resultaten uit CATCHPlus gepresenteerd worden tijdens de Digitaal Erfgoedconferentie (DE), 3 en 4 december in De Doelen in Rotterdam. “We organiseren DE2012 samen met DEN (Digitaal Erfgoed Nederland, red.), die ook de voorgaande edities van DE organiseerde”, aldus Alkhoven. “Ik vind het belangrijk dat CATCHPlus niet als project ten grave wordt gedragen, maar dat de resultaten naadloos overgaan in een nieuwe organisatie of bijdragen aan de ontwikkeling van nieuwe producten. DE2012 focust daarom niet op de eindproducten, maar op de nieuwe mogelijkheden voor onderzoek en digitaal erfgoedbeheer.”

Het thema van DE, dat jaarlijks wordt afgewisseld met de internationale variant DISH, is dit jaar *Beter weten – Interactie Erfgoed en Wetenschap*. “Hoewel er veel aandacht zal zijn voor de resultaten van CATCHPlus, is het programma veel breder. Alle mogelijke onderwerpen op het snijvlak van wetenschap en digitaal erfgoed komen aan bod. De samenwerking tussen deze twee vakgebieden kan nog zoveel meer mooie resultaten opleveren.”

Een goed voorbeeld

Zelf is Patricia Alkhoven een goed voorbeeld van het feit dat de ontwikkelingen niet stil zullen komen te liggen nu CATCHPlus in zijn afrondende fase zit. Sinds februari werkt ze twee dagen per week aan de Universiteit van Amsterdam als coördinator van het Center for Digital Humanities. “We laten jonge onderzoekers kortlopende projecten uitvoeren met een *private partner*.

De onderzoekers kunnen zo ervaring opdoen bij bedrijven in de creatieve industrie en de bedrijven kunnen nieuwe producten creëren vanuit de vraag van de wetenschap en erfgoed.”

Daarnaast is Alkhoven ook coördinator van innovatienetwerk CLICK Cultureel Erfgoed, waar ze partijen bij elkaar probeert te brengen die elkaar verder zouden kunnen helpen. “De ‘gouden driehoek’ van samenwerking tussen erfgoed, wetenschap en bedrijven speelt natuurlijk ook een essentiële rol in de *Topsector Creatieve Industrie*. Door deze drie samen te brengen, kunnen ze op termijn wellicht alle drie profiteren.”

www.catchplus.nl

KORT

Community van experts van start

SURF nodigt medewerkers van hoger onderwijs- en onderzoeksinstituten uit om inzichten, kennis en ervaringen op het gebied van onderzoeksdata te delen in de Special Interest Groep (SIG) Research Data. Voorloper van deze ‘community van experts’ is het Onderzoeksdata Forum. Leden van de kerngroep van de SIG Research Data zijn: NWO, DANS, universiteiten, RIVM, Nationale Coalitie Digitale Duurzaamheid (NCDD), het Nederlands eScience Center en SURF. Samen met andere participanten streeft de SIG Research Data naar samenhang in Nederlandse initiatieven op het gebied van kwaliteit, betrouwbaarheid, management en toegankelijkheid van data. Geïnteresseerd? Aansluiten bij deze en andere SIG’s kan via de website. (AvdK) www.surfspace.nl

Alle teksten onder één virtueel dak

Het project ‘Nederlab - een laboratorium voor onderzoek naar de veranderingspatronen in de Nederlandse taal en cultuur’ gaat begin 2013 van start. In dit project werken het Meertens Instituut, Huygens ING, de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL), het Instituut voor Nederlandse Lexicologie (INL), de Koninklijke Bibliotheek (KB) en verschillende universiteitsbibliotheken samen. Zij zullen alle Nederlandstalige teksten van de achttiende eeuw tot nu verzamelen onder één virtueel dak. Geesteswetenschappers zullen vanaf eind 2014 via de Nederlab-website deze teksten kunnen doorzoeken en analyseren met binnen Nederlab ontwikkelde tekstanalyse software. Nederlab wordt onder andere gefinancierd door NWO, KNAW, CLARIAH en CLARIN. De totale investering is vier miljoen euro. (ER)

Online toegang CBS microbestanden

Sinds kort is het mogelijk beveiligde microbestanden van het Centraal Bureau voor de Statistiek kosteloos uit het online archiveringssysteem EASY van DANS te downloaden. Beveiligde microbestanden zijn bestanden die zodanig bewerkt zijn dat herkenning van individuen en huishoudens uitgesloten is. Voorbeelden van deze bestanden zijn: Enquête Beroepsbevolking (EBB), Consumenten Conjunctuuronderzoek (CCO) en Leefsituatie Allochtone Stedelingen (LAS). Dit is onder andere het resultaat van de nieuwe overeenkomst tussen het CBS en DANS met als doel het beschikbaar stellen van CBS data voor wetenschappelijk onderzoek. (JvdS) <https://easy.dans.knaw.nl>

Wie greep wil krijgen op de historische ruimte van een stad of regio, hoeft al lang niet meer in ouderwetse kaartenbakken te zoeken. *Heidi Berkhout*

GIS staat voor Geografisch Informatiesysteem. Een GIS maakt het mogelijk om ruimtelijke informatie in kaarten op een scherm te tonen. Het bekendste GIS is wel het navigatiesysteem *TomTom*. Ook voor historische kaarten biedt GIS de oplossing. Alhoewel software voor GIS voorhanden is, is aan uniforme historische basisgegevens echter nog steeds een groot gebrek. Het HISGIS-programma van de Fryske Akademy brengt verandering.

Oude data in modern jasje

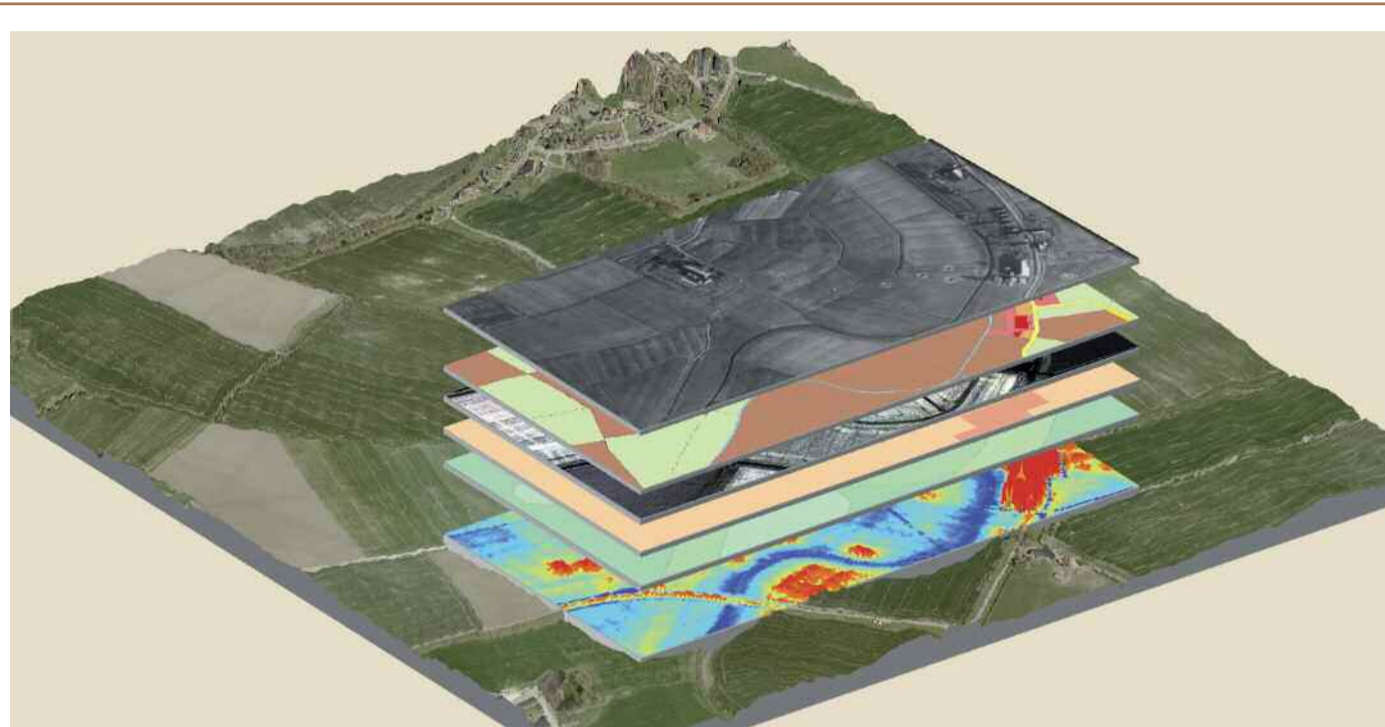
“HIS staat voor historisch. HISGIS is dus historische informatie in een modern jasje. De Fryske Akademy gebruikt hiervoor als uitgangspunt het kadaster uit de Franse tijd (1812-1832).” Aan het woord is Hans Mol, projectleider HISGIS. “Aan deze informatie kan vervolgens andere informatie gekoppeld worden. Denk hierbij aan bodemkaarten, luchtfoto’s, digitale hoogtekarten maar ook vroegmoderne manuscriptkaarten. Zelf vullen we dit kadaster aan met informatie van diverse archieven en gewestelijke stichtingen voor de uitgave van kadastrale atlanten. Bruikbare informatie is informatie die goed in het Rijksdriehoekstelsel is geplaatst. Dit is het Nederlandse stelsel van coördinaten, de grondslag voor geografische aanduidingen en bestanden. Door gebruik te maken van dit stelsel, past alle informatie op elkaar en zijn de kaarten voor analyses bruikbaar.”

HISGIS Amsterdam 1832-1851

Het HISGIS-programma lijkt zich te lenen voor een verdere uitbouw op nationaal niveau. Immers, voor verschillende steden en regio’s zoals Friesland, Groningen, Utrecht, Overijssel en Noord-Drenthe, is de informa-

HISGIS-projectleider Hans Mol aan het woord

Terugreizen in de tijd met Fryske Akademy



Kaartlagen zijn op diverse manieren met elkaar te combineren

Het oudst gedigitaliseerde kadaster stamt uit 1832. Op deze kaart sluiten andere kaartlagen aan, bijvoorbeeld chronologische en thematische kaartlagen. Denk bijvoorbeeld aan de topografisch-militaire kaart van circa 1865 en de hoogtekarte van de Actueel Hoogtebestand Nederland. Deze lagen zijn op diverse manieren met elkaar te combineren. *bron J. Feikens, Fryske Akademy*

tie reeds beschikbaar. “Op termijn willen we inderdaad heel Nederland in kaart brengen”, aldus Mol. “Maar”, zo vervolgt hij: “om het belang van HISGIS te onderstrepen, en om ook technisch meer greep op de digitale kartering (het in kaart brengen van stedelijke gebieden) te krijgen, concentreren we ons ook op het project *HISGIS Amsterdam 1832-1851*.” Dit project, gestart in december 2011 met het

Nederlands Interdisciplinair Demografisch Instituut (NIDI), wordt gesubsidieerd door KNAW en DANS. Het heeft een looptijd van anderhalf jaar. Mol: “HISGIS Amsterdam is opgezet naar het model van HISGIS Fryslân. Het biedt de oudste kadastrale kaart van 1832 met bijbehorende eigenaars- en perceelinformatie in digitale vorm. Deze gegevenslaag, die met achtergrondkaarten en andere data verrijkt

kan worden, is bedoeld als kapstok voor het ophangen van digitale kennis over het landschap, het erfgoed en de sociale en economische ontwikkelingen in het verleden. Voor Amsterdam zal dit worden ontwikkeld naar de huidige gemeentegrens anno 2012. Het project voorziet tevens in een koppeling aan de huisnummerregistratie van 1850-1860, zodat ook de daarop aansluitende bevolkingsgegevens uit de jaren 1851-1853 ruimtelijk precies in beeld kunnen worden gebracht. Het te verwerken materiaal beslaat zo’n 170 minuutplannen (oudste kadastrale detailkaarten) en 60.000 perceelnotities met informatie over bijvoorbeeld eigenaar, grootte, opbrengst, beschreven in circa 2.000 bladzijden registers.”

Nieuwsgierig?

Het project *HISGIS Amsterdam 1832-1851* loopt tot mei 2013. Voor wie nu al nieuwsgierig is: informatie over de aanpak en tussentijdse bevindingen van Hans Mol en Johan Feikens (Fryske Akademy), Peter Ekamper (Nederlands Interdisciplinair Demografisch Instituut) en Jan Hartmann (Universiteit van Amsterdam) staat op de Amsterdam-pagina op de HISGIS-website.

www.hisgis.nl



Amsterdam anno 1832 verrijkt met digitale data

Eerste impressie van HISGIS Amsterdam, de oude binnenstad: de oudste kadastrale kaart van 1832 met bijbehorende eigenaars- en perceelinformatie in digitale vorm. Deze gegevenslaag, die met achtergrondkaarten en andere data verrijkt kan worden, is bedoeld als kapstok voor het ophangen van digitale kennis over het landschap, het erfgoed en de sociale en economische ontwikkelingen in het verleden.

bron J. Feikens, Fryske Akademy

Oproep

Onderzoekers die over seriële data over negentiende eeuwse Amsterdam met huisnummercodering beschikken, worden van harte uitgenodigd om gebruik van de kadastrale kapstok te maken. Neem voor meer informatie contact op met Hans Mol via hmol@fryske-akademy.nl.

Nominaties Nederlandse Dataprijs bekend

Uit zeventien inzendingen zijn drie datasets genomineerd voor de Nederlandse Dataprijs. Juryvoorzitter Wijnand Mijnhardt licht een tipje van de sluier op. Heidi Berkhout

Het beoordelen van de inzendingen was niet eenvoudig. “Het palet inzendingen was buitengewoon breed. Van mooie literaire dataverzamelingen voor geïnteresseerde onderzoekers en sociaalwetenschappelijke databases tot archeologisch-historische datasites voor een breder publiek. De drie genomineerde datasets hebben stuk voor stuk iets vernieuwends, iets aantrekkelijks en ook iets maatschappelijks. Het wordt een spannende prijsuitreiking”, aldus Mijnhardt. Een korte toelichting van de juryvoorzitter per nominatie.

Geoplaza: schitterende belofte

“Geoplaza is een schitterende belofte. Het portaal biedt nu al de mogelijkheid de meest uiteenlopende kaartbestanden met elkaar te combineren. Met de toevoeging van nieuwe bestanden zal de bruikbaarheid ervan, zo verwacht de jury, alleen maar toenemen. Naast onderzoekers zal Geoplaza ook voor de Nederlander die steeds meer geïnteresseerd is in zijn fysieke omgeving een prachtig oriëntatiemiddel worden.”

Geoplaza is een door de universiteitsbibliotheek van de Vrije Universiteit in samenwerking met het SPINlab (Spatial Information labo-



En de winnaar is...

Op 18 oktober wordt de winnaar van de Nederlandse Dataprijs bekend gemaakt. Naast deze drie genomineerden voor de Dataprijs humaniora en sociale wetenschappen zijn ook drie datasets genomineerd voor de Dataprijs exacte en technische wetenschappen. Kijk voor deze genomineerden en andere informatie over de prijsuitreiking op www.dans.knaw.nl.

De bronzen sculptuur voor de uiteindelijke winnaar
foto Tjavidar Iliev

ratory) gecreëerd portaal om toegang tot geo-grafische datasets (GIS) te verlenen. Door middel van dit portaal kunnen uiteenlopende GIS-bestanden worden geraadpleegd, zoals bijvoorbeeld gedigitaliseerde en gegeoreferente historische kaarten en raster- en vectorbestanden van het Kadaster of het CBS.

geoplaza.uvu.vu.nl

Political Mashup: onmisbaar instrument

“Political Mashup kan bijvoorbeeld uitgroeien tot een onmisbaar instrument voor politici en geïnteresseerde burgers die hun volksvertegenwoordigers op de voet willen volgen. Voor historici die geïnteresseerd zijn in de grote lijn, is de website ook een *Fundgrube* (plaats waar men veelvuldig kan vinden wat men zoekt, red.)”

Political Mashup maakt grote hoeveelheden politieke data digitaal doorzoekbaar. Het betreft data vanaf 1946 zoals partijprogramma's en kamerhandelingen. Behalve het digitaal ontsluiten van deze data streeft het project ernaar om deze data semantisch doorzoekbaar te maken op bijvoorbeeld trefwoorden, thema's, personen en gebeurtenissen. Political Mashup is een gezamenlijk start-up project van UvA-informatica, RUG Documentatiecentrum Nederlandse Politieke Partijen en UvT-computer linguïstiek.

<http://politicalmashup.nl>

Het Veteranenproject: prachtig monument

“Het Veteranenproject heeft vooral een grote maatschappelijke betekenis. In een land dat zich steeds maar weer beroemt op zijn antimilitarisme

maar ondertussen wel aan allerlei militaire acties meedoet, is deze site met interviews van veteranen een prachtig monument voor die talloze militairen die overal ter wereld Nederlandse belangen hebben moeten dienen. De site kan helpen de waardering voor de veteraan in de Nederlandse samenleving te vergroten.”

De collectie van het Interviewproject Nederlandse Veteranen bestaat uit meer dan duizend interviews van veteranen van conflicten en militaire missies waar Nederland bij betrokken was. Van soldaat tot generaal, van hospik tot para, het resultaat is een gevarieerde collectie. Door hun verhaal dragen de geïnterviewde veteranen bij aan een schat aan informatie voor beleidsmakers, onderzoekers, media en het geïnteresseerde publiek. Het project is een initiatief van Stichting het Veteraneninstituut.

<http://interview.veteraneninstituut.nl>

De jury en secretaris

De jury voor de humaniora en sociale wetenschappenprijs bestaat uit **prof. dr. W.W. Mijnhardt**, wetenschappelijke directeur van het Descartes Centre en hoogleraar Vergelijkende Wetenschapsgeschiedenis aan de Universiteit Utrecht (voorzitter), **drs. S. Krauwer**, Universiteit Utrecht, coördinator en voorzitter Executive Board CLARIN.EU en **prof. dr. W.B.G. Liebrand**, directeur SURF. Als secretaris van deze jury treedt op **drs. Heiko Tjalsma** (DANS).

GELEZEN

Controversies around the Digital Humanities, in: Historical Social Research, volume 37 (2012), nummer 3, Manfred Thaller (ed.).

Peter Boot



Het gaat niet slecht met de Digital Humanities, integendeel. ‘Funding agencies’ zien het belang van

en grote projecten (CLARIN, Bamboo) bouwen aan een digitale infrastructuur. Enkele jaren geleden riepen de Blackwell Companions (to Digital Humanities en to Digital Literary Studies) het beeld op van een terrein van studie dat een volwaardig vakgebied was geworden. Nu verschijnt, kort na Debates in the Digital Humanities (Minneapolis 2012), een tweede bundel met discussies over principiële elementen van het vak: *Controversies around the Digital Humanities*.

De bundel is het resultaat van een discussiebijeenkomst in Keulen in april 2012 waar Manfred Thaller, hoogleraar historisch-cultuurwetenschappelijke informatiekunde, bij controverses steeds twee opponenten had uitgenodigd. De papers van de opponenten zijn in deze bundel verzameld, voor het merendeel zeer lezenswaardig. Wat jammer genoeg ontbreekt is een synthetiserende discussie.

www.cceh.uni-koeln.de/events/CologneDialogue

COLUMN

Je blijft er van schrikken

Voor mij was het een van de eerste dingen die ik leerde op de universiteit: zoeken in bibliotheken. Er was een vak aan gewijd, je moest er literatuur over lezen en onder begeleiding van een docent werd je op een dag losgelaten in een universiteitsbibliotheek.

Tegenwoordig is internet de belangrijkste bron van informatie. Studenten krijgen wel te horen dat ze ook gebruik moeten blijven maken van boeken en tijdschriften in bibliotheken, maar in de praktijk beginnen ze altijd te zoeken op internet.

Maar leren studenten van meet af aan hoe ze geavanceerd kunnen zoeken op internet? Met behulp van filters en zoekoperatoren, in databanken en op het vrije web? Ruim een jaar geleden was ik aanwezig bij de lancering van de website *Early Dutch Books Online* – een collectie van ruim 11.000 gedigitaliseerde boeken uit de periode 1781-1800. Tijdens een klein symposium presenteerden diverse promovendi en hoog-

leraren hun onderzoek. Bij herhaling zeiden zij: jammer genoeg kunnen we dit soort dingen nog niet onderzoeken. Het ging daarbij om historisch letterkundige kwesties.

Ik schrok daarvan. Hoogleraren en promovendi – dan heb je het toch over de frontlinie van de wetenschap. Veel van wat volgens hen nog niet te onderzoeken was op internet, was dat wel degelijk – en al jaren.

Naar aanleiding van dat symposium heb ik een gidsje geschreven dat aanvankelijk *Eerste Hulp Bij*



foto Leo van Velzen

e-Onderzoek heette en inmiddels *Slimmer zoeken op internet*. Er zijn het afgelopen jaar 80.000 exemplaren van verspreid, vooral bij universiteiten, hogescholen en onder journalisten. Ook ben ik workshops gaan geven over geavanceerd zoeken op internet, voor dezelfde doelgroep. Sinds die workshops weet ik: het is nog veel erger dan ik dacht. Uit een onderzoek van Google bleek enkele jaren geleden dat driekwart van de docenten op middelbare scholen en de helft van de universitaire docenten de toetscombinatie *ctrl/cmd + f* niet kent. Als ze iets hebben gezocht op internet en ze zien het niet meteen op een website staan, dan lezen ze hele pagina's door. Terwijl je met de toetscombinatie *ctrl/cmd + f* snel kunt zien waar iets staat en hoe vaak het op een webpagina voorkomt.

Toen ik over dat onderzoek van Google las, dacht ik: dat lijkt mij zwaar overdreven. Misschien geldt dit voor de VS, maar toch zeker niet voor gidsland Nederland.

Inmiddels weet ik beter. Aanvankelijk vertelde ik in workshops een beetje lacherig over die toetscombinatie ('ik durf er bijna niet over te beginnen'), omdat het tot de absolute basiskennis hoort. Maar tientallen toehoorders – studenten, hoogleraren, promovendi, zelfs journalisten – bleken er nooit van te hebben gehoord. Laat staan van meer geavanceerde zoektrucs.

Je moet er niet aan denken hoeveel tijd mensen verspillen met ongericht, ongestructureerd en kritiekloos zoeken op internet. Volgens een recent rapport van de Universiteit Twente (getiteld *Ctrl Alt Delete*) verdoen Nederlanders

op internet gemiddeld vijf procent van hun werktijd – een schadepost van 19 miljard euro. Ik vermoed dat dit percentage in werkelijkheid nog veel hoger ligt.

Wat mij het meest verbaast is dat het onderwijs zo traag reageert op deze feiten. Op een groep van twintig vierdejaarsstudenten voor wie ik onlangs een workshop gaf, waren er welgeteld drie die meer wisten over slimmer zoeken – vooral omdat ze zelf websites bouwden. De anderen waren bijna klaar met hun studie, maar ze hadden er totaal geen kaas van gegeten.

Jonge, slimme mensen – maar ze hadden geen idee hoe ze belangrijkste hedendaagse bron van informatie optimaal konden gebruiken.

Ik blijf er van schrikken.

Ewoud Sanders

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