

Delpher.nl zal komende jaren nog volop groeien

Alle bronnen bij elkaar onder één knop

Full-text kunnen zoeken in één database met alle gedigitaliseerde kranten, boeken en tijdschriften van Nederland. Het lijkt de droom van iedere wetenschapper. Delpher.nl is een flinke stap in de goede richting.

Inge Angevaere

Sinds eind vorig jaar is een nieuwe informatiedienst online: Delpher.nl, een platform waar de universiteitsbibliotheken van Groningen, Utrecht, Leiden en de Universiteit van Amsterdam plus de Koninklijke Bibliotheek (KB) al hun gedigitaliseerde boeken, tijdschriften en kranten full-text beschikbaar stellen aan de wetenschap. Het materiaal komt vooral uit bestaande projecten en websites, zoals *Early Dutch Books Online*, *Historische Kranten* en *Google Books*. "Maar er komt nog veel meer in", legt Andrea Langendoen uit, coördinator van Delpher, "want we oogsten ook al het gedigitaliseerde materiaal uit *Meta-morfoze*, het nationale fonds voor conservering waaraan ook veel kleinere, lokale instellingen deelnemen. We zijn begonnen met 30 miljoen pagina's; eind 2014 moeten dat er al 60 miljoen zijn."

Wat is er zo speciaal aan Delpher?

"Het landelijk karakter," zegt Langendoen direct. "De genoemde initiatiefnemers van Delpher zijn pas het begin van het verhaal. We hopen dat andere instellingen zich zullen aansluiten. Hoe meer hoe beter. Voor onderzoekers is het nu al prettig dat ze niet meer in al die aparte bestanden hoeven te zoeken. Wanneer meer partners zich aansluiten, wordt dat voordeel alleen maar groter."

En wat kan Delpher dat Google niet kan?



Staatssecretaris Sander Dekker van het Ministerie van Onderwijs, Cultuur en Wetenschap (rechts) is enthousiast over de mogelijkheden van Delpher foto Jacqueline van der Kort / KB

Langendoen: "Google brengt je alleen naar webpagina's, Delpher brengt je de hele tekst en mogelijkheden om daarin te zoeken." Maar zelfs full-text zoeken is voor veel wetenschappers niet genoeg, die willen datasets hebben. "Ook daar is aan gedacht", vertelt Langendoen. "De afdeling Dataservices van de KB levert een *Application Programming Interface* (API) en datasets uit Delpher op maat. Tenminste, waar dat mogelijk is. Soms wordt gebruik beperkt door auteursrechten."

Wat ontbreekt er aan Delpher?

Langendoen lacht. "Daar beginnen de klanten vaak het eerst over. Bijvoorbeeld waarom er geen

materiaal in zit dat van oorsprong digitaal is. Wij kijken liever naar de andere kant. Bij het opzetten van een project als Delpher kun je niet alles tegelijk doen. De startsubsidie gold voor gedigitaliseerde bronnen. Die pakken we dus eerst aan."

Hoe handig is Delpher in de praktijk?

Langendoen: "We hebben Delpher uitgebreid getest met een gebruikerspanel, maar wij hopen dat ook de onderzoekers ons zullen laten weten hoe zij deze bèta-versie van Delpher ervaren, wat eraan verbeterd kan worden. We zullen blijven werken aan verbeteringen. En materiaal blijven toevoegen, véél materiaal."

delpher.nl

'Geen data meer te verliezen'

Als je subsidie wilt ontvangen van ZonMw, dan gelden sinds 1 juli 2013 nieuwe subsidiebepalingen. Programmanager Margreet Bloemers legt uit. Rutger Nugteren

Eén van de bepalingen schrijft bijvoorbeeld voor dat de subsidieontvanger moet nagaan of hij gebruik kan maken van bestaande databestanden. En als hij toch een nieuwe dataverzameling gaat aanleggen, moet dat zo gebeuren dat de data voor verder onderzoek herbruikbaar zijn. "ZonMw financiert niet alleen gezondheidsonderzoek, we stimuleren ook het gebruik van de ontwikkelde kennis – om daarmee de zorg en gezondheid te verbeteren. Doordat er echter veelal geen (heldere) voorwaarden werden

gesteld aan het gebruik van onderzoeksdata, bleven deze data vaker dan wenselijk onvindbaar of onbruikbaar voor onderzoekers die niet bij het onderzoek betrokken waren." Aan het woord is Margreet Bloemers. "Het door ZonMw opgerichte programma 'Toegang tot Data' scherpt de regels aan. Een groot deel van het onderzoeksbudget gaat immers op aan het verzamelen van data. Het is onverantwoord om data verloren te laten gaan."

Vanaf 2014 meer eisen

"Onderzoeksdata moeten eenvoudig online te vinden zijn, met voldoende en begrijpelijke beschrijvingen van de data (i.e. goede metadata). Data moeten duurzaam opgeslagen worden. Er zijn uitzon-

deringen mogelijk, maar in principe geldt een minimale bewaartermijn van 5 jaar voor alle data. Als subsidieaanvrager moet je laten zien waar de data te vinden zijn en hoe de data zijn opgeslagen. We hantieren een groeimodel, de eisen die we aan subsidieaanvragers stellen, zullen in de loop van 2014 worden aangescherpt. Uiteindelijk worden projecten beoordeeld en afgerekend op de planning en uitvoering van datamanagement."

ZonMw heeft aan een consortium van vijf partijen (NIVEL, RIVM, BBMRI, CentERdata en DANS), die betrokken zijn bij het verzamelen, verwerken en/of het duurzaam bewaren van data, gevraagd om hulpmiddelen te ontwikkelen voor subsidieaanvragers. Zo ontwikkelt dit consortium momenteel checklists

voor onderzoekers voor het opstellen van een dataparagraaf (voordat de subsidie is toegekend), een datamanagementplan en datamanagementcriteria (als de aanvraag is goedgekeurd). ZonMw richt zich bovendien op bestuurders. "We kweken bewustzijn en begrip. Met het inrichten van voorzieningen zoals een dataloket zijn immers kosten gemoeid. Ook moet worden geïnvesteerd in de kennis over datamanagement. Overigens stuiten we tot nu toe op weinig verzet uit het veld, vrijwel iedereen heeft begrip voor de ingeslagen weg."

Belonen

Margreet: "Het moet volkomen helder zijn dat data niet meer achtergehouden kunnen worden als het

VERVOLG OP PAGINA 2



E-DATA & RESEARCH

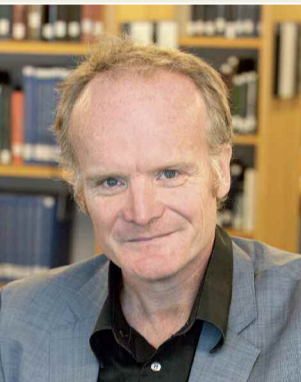
Jaargang 8 | nummer 2

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 De verleiding van het afsnijden van bochten
- 3 Eén loket voor alle surveydata uit Nederland
- 4 Deze datasets zijn sinds kort beschikbaar
- 4 Wie gaat jouw collectie boeken digitaliseren?
- 5 Van Oostendorp aan de slag met KB-data



- 5 Agenda: interessante bijeenkomsten op een rij
- 6 Nederlandse literatuur op de kaart gezet
- 7 It's all in the name: leren van romans
- 8 Online tool PoliMedia maakt analyse mogelijk



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

NCDD zet zichzelf weer terug op de kaart

Heiko Tjalsma

Het was enige tijd stil rondom de Nationale Coalitie Digitale Duurzaamheid (NCDD), maar op 11 november 2013 zette de coalitie zich in het Eye Filminstituut met haar jaarcongres weer op de Nederlandse kaart. Tijdens dit congres presenteerde de NCDD haar werkplan 2013-2018, waren er presentaties uit de verschillende aandachtsgebieden van de NCDD (archieven, cultuur, audiovisuele en wetenschappelijke collecties) en werd een samenwerkingsovereenkomst met de Britse Digital Preservation Coalition ondertekend. In het werkplan initieert de NCDD een aantal samenwerkingsprojecten op het gebied van duurzame toegankelijkheid. Deze projecten moeten oplossingen bieden voor knelpunten van de NCDD-partners.

ncdd.nl

Onderzoekers geven verlanglijst aan Nederlab

Peter Boot

Nederlab is een omgeving die het mogelijk maakt om alle gedigitaliseerde Nederlandstalige teksten van 800 tot heden wetenschappelijk te analyseren. Maar wat zijn de vragen die onderzoekers aan dat corpus gaan stellen, en welke hulpmiddelen hebben ze nodig om die vragen te beantwoorden? Daarover vroeg Nederlab advies aan onderzoekers tijdens een workshop die 4 december jl. werd georganiseerd door Nicoline van der Sijs (Meertens en Nederlab), Karina van Dalen-Oskam (Huygens ING en UvA) en Els Stronks (UU).

Van der Sijs introduceerde Nederlab. In de huidige demoversie zijn de teksten uit de Digitale Bibliotheek voor de Nederlandse letteren

GEHOORD & BIJGEWOOND



NCDD presenteert plannen tot 2018 foto Jacqueline van der Kort

(DBNL) en een jaargang van de krantencollectie van de Koninklijke Bibliotheek (KB) aanwezig. Binnenkort worden de tienduizend boeken uit *Early Dutch Books Online* toegevoegd. Het gaat dus om heel diverse collecties, waarbij de kwaliteit van de teksten en metadata zeer verschillend zijn. Erik Tjong Kim Sang toonde de tools die in eerste instantie worden opgeleverd. Het betreft daarbij onder andere (taalkundige) preprocessing (zoals opwerken van de kwaliteit van de tekst, lemmatiseren, grammaticale analyse) en hulpmiddelen voor transformeren, tellen, analyseren en visualiseren. Er komt een toegang waarmee de onderzoekers via eigen scripts de teksten kunnen analyseren (waarbij vanwege rechtenproblemen de volledige tekst meestal niet toegankelijk zal zijn). Martin Reynaert toonde als inspiratie de OpenSonar zoekinterface dat de inhoud van het Sonar-corpus met hedendaags Nederlands toegankelijk maakt voor leken.

Sommige van de vragen van de historische en letterkundige onderzoekers zouden met de nu beschikbare tools al kunnen worden beantwoord (vragen naar frequenties van

woordgebruik, van termen en de context waarin ze voorkomen, naar verschillen en overeenkomsten tussen tekstversies), maar in veel gevallen bleek er nog een behoorlijke afstand te bestaan tussen de wensen van de onderzoekers en de nu beschikbare tools. Vragen naar bijvoorbeeld de waarschijnlijke auteur van een anoniem overgeleverde tekst of naar de verspreiding van literaire teksten binnen en buiten het Nederlands taalgebied vereisen andere hulpmiddelen en in sommige gevallen ook een koppeling met nog niet in Nederlab aanwezige databronnen. Voorzien aan analytische tools zijn onder meer een *n-gram viewer* en een filter voor passages in vreemde talen, maar daarnaast hebben onderzoekers ook behoefte aan meer experimentele hulpmiddelen voor bijvoorbeeld het herkennen van intertekstualiteit (plaatsen waar de ene tekst de andere citeert), van metaforen, standpunten en onderwerpen. Bij visualisatietools werd vooral gedacht aan geografische visualisaties en visualisaties in de tijd. En omdat analyse staat of valt met kwalitatief goede tekst is ook het opwaarderen van optische tekenherkenning (OCR) en het uniforme-

ren van historische spellingen uiterst belangrijk. Ten slotte, maar dat gaat de mogelijkheden van een enkel project te boven, is het essentieel dat ons verouderde auteursrecht zo wordt aangepast dat in elk geval voor onderzoeksdoeleinden het creëren en openstellen van digitale collecties van recente tekst mogelijk wordt.

nederlab.nl

CLIN 2014: lexicologie en Big Data

Dirk Roorda

Op 17 januari jongstleden organiseerde het Instituut voor Nederlandse Lexicologie (INL) samen met de TST-Centrale de 24e editie van 'Computational Linguistics in the Netherlands' (CLIN) in Leiden. De bijeenkomst was bedoeld voor iedereen die zich bezighoudt met taal- en spraaktechnologie en haar toe-



Titelblad Hebreeuwse bijbel bron Rijksdienst Cultureel Erfgoed

passingen. CLIN telde maar liefst 15 sessies met 60 papers. Zo sprak collega Menzo Windhouwer (DANS/TLA) met Ineke Schuurman (KU-Leuven) over registers van (taal)wetenschappelijke termen waarmee onderzoeksmateriaal beschreven kan worden (ISocat en Relcat) en

presenteerden Martijn Naaijer (VU-Amsterdam) en ik een nieuw instrument om onderzoek te doen naar taalvariatie in het bijbels Hebreeuws. Andere sessies liepen uit een van historische data via lexicologie en semantiek naar de industrie. Gezien de locatie was het bijna voor de hand liggend dat het onderwerp lexicologie extra werd belicht. Dit uitte zich onder andere in de keynote sprekers: lexicologen Patrick Hanks (University of Wolverhampton) en Dirk Geeraerts (Leuven University). De spannende vraag was 'hoe de lexicologie het beste in kan spelen op de mogelijkheden die Big Data bieden'. De corpora worden groter, raken beter geannoteerd en zijn beter doorzoekbaar geworden. Veel hedendaagse artikelen rapporteren hierover. Aan de andere kant is de lexicologie steeds beter gaan beseffen dat een oud ideaal, namelijk het beschrijven van discrete woordbetekenissen, aan richtinggevende kracht verloren heeft. Keynote speaker Hanks verwoordde het als volgt: "een woord heeft geen betekenis maar een betekenispotentieel, en dat potentieel wordt pas gerealiseerd door de patronen waarin het voorkomt". Geeraerts maakte duidelijk dat de informatie over woordgebruik, zoals we die uit de diverse corpora krijgen, meteen het scherpe onderscheid uitwist tussen vaste grond en trends in woordbetekenissen. CLIN24 werd georganiseerd door het INL en de TST-Centrale. Het INL is de plek voor iedereen die iets wil weten over woorden, hun spelling, vorm, betekenis of gebruik door de eeuwen heen. De TST-Centrale is het kennis- en distributiecentrum voor Nederlandstalige tekstverzamelingen, woordenlijsten, wetenschappelijke woordenboeken, spraakcorpora en taal- en spraaktechnologische software.

clin24.inl.nl

VERVOLG VAN PAGINA 1

Subsidie ZonMw

aan ZonMw ligt. We worden hierbij geholpen door de tijdgeest, bijvoorbeeld door de brief van Edith Schippers, minister van Volksgezondheid, Welzijn en Sport (VWS) dd. 23 oktober 2013 over een 'duurzaam informatiestelsel'.

In deze brief geeft ze aan dat standaardisatie en het beschikbaar stellen van zorggegevens essentieel is voor

een zogenaamd 'duurzaam informatiestelsel' met actuele en betrouwbare gegevens over de volksgezondheid en zorg. ZonMw gaat de naleving van de gestelde eisen controleren. We willen geen controle-instituut worden, maar het is bijvoorbeeld mogelijk om een deel van het onderzoeksbudget achter te houden totdat aan alle datamanagement-eisen is voldaan. Of we belonen onderzoekers voor goed datamanagement. Hier wordt nog volop over nagedacht. We leren graag van die wetenschapsgebieden waar het delen van data al langer gemeengoed is."

zonmw.nl/ttd



Margreet Bloemers foto Jeltje Waagenaar

Hoe lang moeten ruwe data bewaard?

Bij onderzoekers leven veel vragen over de bewaartermijn van ruwe onderzoeksdata. DANS zette de regels op een rij. Volgens de Nederlandse Gedragscode Wetenschapsbeoefening, onderdeel III, Controleerbaarheid, moet de bewaartermijn van ruwe onderzoeksdata minimaal 5 jaar zijn. Van een maximumbewaartermijn spreekt deze code niet. Voor medische data met patiëntgegevens geldt wel een maximumbewaartermijn. Deze data moeten na 15 jaar vernietigd worden. Als medische data geanonimiseerd zijn, is er geen sprake van een vernietigingstermijn. Meer daarover staat in het rapport van ZonMw, Inventory Patients Registries in the Netherlands. DANS houdt zich aan de

KORT

hierboven beschreven bewaartermijnen. Binnen andere instituten of organisaties kunnen andere regelingen gelden. (CvZ) dans.knaw.nl

Special Issue E-data & Research

Onlangs heeft E-data & Research een speciale editie uitgebracht, geheel in het teken van onderzoeksinfrastructuren in de sociale en geesteswetenschappen. In deze special staat onder andere een interview met Neelie Kroes, Eurocommissaris voor de digitale agenda. Abonnees van E-data & Research hebben deze special automatisch ontvangen. Wilt u deze special ook ontvan-



gen? Meld u dan aan als abonnee van E-data & Research door een e-mail te sturen naar de redactie (edata@dans.knaw.nl) of kijk op de website van E-data. (ER) edata.nl

‘Datamanagement is niet ‘extra’, maar hoort erbij’

Het College van Bestuur (CvB) van de Erasmus Universiteit Rotterdam (EUR) neemt alle aanbevelingen van de Taskforce wetenschappelijk integriteit over. Een flink deel ervan betreft de omgang met data. *Marjan Grootveld*



Olifantenpaadje Op het eindrapport van de Taskforce staat een olifantenpad. Olifanten bekommen zich niet om geplaveide paden, maar kiezen de kortste route. Als ze die route maar vaak genoeg volgen, ontstaat vanzelf een nieuw pad. Ook mensen creëren ‘olifantenpaden’ als die hen sneller naar hun bestemming brengen. Het paadje op het rapport “symboliseert de alledaagse verleiding van het afsnijden van bochten in het onderzoeksproces”, aldus voorzitter prof. dr. Finn Wynstra. *bron Fietsbult.be*

Het College van Bestuur had de Taskforce in 2012 ingesteld om de bewustwording over wetenschappelijke zorgvuldigheid en integriteit in alle onderzoeksscholen en faculteiten te bevorderen. Het eindrapport van de Taskforce behandelt de onderwerpen: datamanagement en -opslag, training, dilemmaspel, seminarcultuur, integriteitsverklaring, promotiereglement, contractonderzoek, monitoring en mediarelaties. Inmiddels heeft het CvB alle aanbevelingen van de Taskforce overgenomen, waaronder diverse die betrekking hebben op datamanagement en -opslag. Onderzoek dient immers controleerbaar en herhaalbaar te zijn en dat kan alleen als data met voldoende documentatie beschikbaar blijven.

Minimumprotocol

De omvangrijkste werkgroep binnen de Taskforce, de werkgroep Datamanagement, stond onder leiding van DANS. Geïnspireerd door het rapport van de commissie-Schuyt was de kernvraag in interviews met onderzoeksdirecteuren en senior onderzoekers van alle faculteiten: “Welke gegevens zou je moeten opslaan tussen het begin van een onderzoek en het moment van de *peer review*?”. Vanzelfsprekend varieerden de meningen: absolute replicatie is niet in alle soorten onderzoek mogelijk, ‘alles bewaren’ zou in sommige disciplines te duur worden, en uitgever voeren verschillend beleid inzake het beschikbaar stellen van data bij publicaties. Niettemin kon de werkgroep een minimumprotocol opstellen dat aansluit bij het onderzoeksproces van data ver-

zamelen, bewerken, analyseren en erover publiceren. Zo moeten majeure afwijkingen van het oorspronkelijke onderzoeksplan gedocumenteerd en bewaard worden. Dit generieke protocol zal per faculteit worden uitgewerkt in een werkproces dat recht doet aan verschillende onderzoeksmethoden, maar ook duidelijk is over verantwoordelijkheden: “datamanagement hoort bij een professionele werkwijze”, aldus rector magnificus professor Huibert Pols.

Behoeft aan training

Veel onderzoekers bewaren hun gegevens op media die niet per se veilig en duurzaam zijn, zoals Dropbox, zo blijkt verder uit de survey van de werkgroep Datamanagement onder alle EUR-onderzoekers. Onderzoeken aan andere universiteiten geven een vergelijkbaar beeld. Naast vol-

doende en goede mogelijkheden voor dataopslag vragen veel onderzoekers om training en advies op het gebied van datamanagement. Dit is koren op de molen van de universiteitsbibliotheek van de EUR, die evenals andere universiteitsbibliotheek de vraagbaak wil zijn voor datamanagement. Het dilemmaspel *Professionalism and integrity in research*, een ander resultaat van de Taskforce, is inmiddels al opgenomen in trainingen.

eur.nl/integriteit

NOW AVAILABLE!
at your dean/director of research:
The Dilemma Game



Dilemmaspel

Eerste Digital Humanities Benelux conferentie

Heidi Berkhout

In juni wordt de eerste Digital Humanities Benelux conferentie gehouden, een jaarlijks evenement dat geesteswetenschappelijk onderzoek wil stimuleren dat gebruik maakt van computationele hulpmiddelen. Het gaat om een internationaal initiatief, waarvoor onderzoekers uit België, Luxemburg en Nederland de handen ineen slaan. Program chair Mike Kestemont (Universiteit Antwerpen) licht toe: “Er loopt, zo blijkt nu, eigenlijk al bijzonder veel digitaal onderzoek binnen de geesteswetenschappen in de Benelux. Tot voor kort bestond er echter geen platform waarop Benelux-onderzoekers elkaar konden treffen. Met DHBenelux willen we daar verandering in brengen”. Marijn Koolen (Universiteit van Amsterdam) is enthousiast: “We hebben het gevoel dat DHBenelux een belangrijke lacune opvult en willen graag onderzoek bij elkaar brengen vanuit alle geesteswetenschappelijke disciplines, niet alleen vanuit de Benelux, maar ook daarbuiten. En hopelijk ontstaan er vanuit dit evenement spannende nieuwe projecten en samenwerkingen.” Geïnteresseerde onderzoekers kunnen nog tot 1 maart 2014 een abstract indienen voor een presentatie. DHBenelux-2014 wordt gehost door Huygens ING en de Koninklijke Bibliotheek en vindt plaats op 12 en 13 juni in Den Haag. Keynote-spreker is de gerenommeerde Melissa Terras (UCL). dhubenelux.org

WASHP-tool peilt sentiment in twee talen

Erica Renckens

De digitale tool WASHP is in het CLARIN-project Biland uitgebreid zodat tweetalig onderzoek mogelijk wordt. WASHP, ontwikkeld door de Universiteit Utrecht, Universiteit van Amsterdam en de Koninklijke Bibliotheek, is software voor historische sentiment mining van artikelen uit de media. Met behulp van een automatische analyse van tekstbestanden kan een onderzoeker eenvoudig een beeld vormen van de opinie verkondigd in de artikelen. De Biland-onderzoekers verbeterden de WASHP-tool en maakten tweetalige analyse (Nederlands en Duits) mogelijk, zodat historici vergelijkend onderzoek kunnen doen met grote corpora. Het Biland-team gebruikt de tool zelf om te onderzoeken hoe tussen 1860 en 1945 in de Nederlandse en Duitse media werd gesproken over erfelijkheid en eugenetica. biland.nl

Eén loket voor alle surveydata

Vanaf februari is er één portal voor onderzoekers die gebruik willen maken van surveydata. Marion Wittenberg en Eric Balster, werkzaam bij de initiatiefnemers DANS en CentERdata, leggen uit waarom. *Marika de Bruijne*

Survey Data Nederland (SDN) is een portal waarop men kan zoeken naar surveyvragen en -data van grote Nederlandse onderzoeksprojecten. Bij de lancering van de portal deze maand staan de data van Culturele Veranderingen van het SCP, Profielstudie van Tilburg University, het Integraal Kankercentrum Zuid en het LISS panel van CentERdata online. Op termijn is het de bedoeling om meer NWO-

Veranderingsstudies en andere Nederlandse longitudinale surveys via de portal te ontsluiten.

Een stap verder

Waarom een nieuwe website nodig is, terwijl er al overkoepelende portals als NARCIS bestaan, vertelt Marion Wittenberg, projectleider bij DANS. “Survey Data Nederland sluit aan bij diensten zoals NARCIS, maar gaat een stap verder. Met SDN kunnen onderzoekers veel nauwkeuriger zoeken naar vragen en data. Je kunt in één oogopslag bijvoorbeeld zien hoe bepaalde vragen in de verschillende studies gesteld zijn.” Voorheen moest men eerst afzonderlijke projecten opzoeken en daarna in de documentatie duiken om te zien of het onderzoek

eigenlijk wel interessante gegevens bevat. Voor gebruikers is SDN dus één loket met geavanceerde zoekfuncties op basis van rijke metadata. Tevens geeft het toegang tot de achterliggende surveydata.

Eisen duurzame opslag

SDN biedt toegang tot surveys maar ook worden de data duurzaam opgeslagen. “We willen ervoor zorgen dat onderzoekers zich kunnen concentreren op wat ze het liefst willen: onderzoek doen”, zegt Eric Balster, projectleider van SDN bij CentERdata. “Daarom kunnen grotere onderzoeksprojecten bij SDN een ‘all in one’-pakket afnemen waarbij hun data worden ontsloten en duurzaam worden gearhiveerd.” De onderzoeker geeft de data aan SDN

die vervolgens alle taken rondom archivering en disseminatie uitvoert, conform de nieuwe eisen van de Nederlandse subsidieverstrekters zoals NWO of ZonMw. Het is ook mogelijk om slechts een deel van de taken van datapublicatie en -opslag aan SDN uit te besteden, wanneer de onderzoeksgroep bijvoorbeeld de documentatie of de servers in eigen beheer wil houden. Of de portal een succes wordt, hangt af van hoeveel onderzoeksprojecten zich bij de portal aansluiten en hoeveel onderzoekers er gebruik van zullen maken. De initiatiefnemers hebben er in ieder geval het volste vertrouwen in dat SDN aan de onderzoeker iets biedt dat in Nederland nog niet bestond. surveydata.nl

SINDS KORT BESCHIKBAAR

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

CentERdata

• Interpersoonlijke effecten van huilen

Van oktober 2010 tot en met augustus 2012 is het onderzoek 'Inter-personal effects of crying' afgenomen in het LISS panel (Vingerhoets, A.J.J.M.; Millings, A.; Theodoridou, A.; Cornelius, R.R.; Schaafsma, J.; Klei, M. van der). De vragenlijst bestaat uit vragen met betrekking tot huilgedrag, agressie, empathie en 'social bonding'. Tevens is een experiment opgenomen over het effect van het zien van emotionele gezichten op antwoordgedrag. In dit experiment kregen de respondenten foto's te zien van personen met een neutrale, blije, boze, angstige of verdrietige gezichtsuitdrukking. Vervolgens werd hen gevraagd om deze personen te beoordelen op een aantal emoties en karaktertrekken. De data zijn beschikbaar via LISS Data Archive.



Enkele gezichtsuitdrukkingen uit het onderzoek bron Project Inter-personal effects of crying

Ook sinds kort beschikbaar:
Studies LISS panel

- Doorn, J. van; Breugelmans, S.M.; Zeelenberg, M., februari 2013, Pro-Social and Anti-Social Consequences of Guilt
- Simons, A.M.W.; Groffen, D.A.I.; Bosma, H., februari 2013, Does stigmatisation 'explain' why low socioeconomic status is related to poor health?
- Kalmijn, M., juli 2012/augustus 2012, Life History Questionnaire - Wave 1
- Ven, N. van de; Lindenberg, S.M., juni 2013, Status Concerns and Societal Dissatisfaction
- Verhoeven, A.A.C.; Adriaanse, M.A.; Ridder, D.T.D. de; Vet, E. de; Fennis, B.M., maart/april 2013, Reasons to Snack
- Conrad, F.; Zang, C., april 2012/mei 2012, Satisficing
- Tijdens, K., oktober 2009, WageIndicator

- Elshout, M., januari/maart 2013, The Psychology of Vengeance
 - Hagenaars, M.A.; Hagenaars, J.A., juli 2011/september 2012, Tonic immobility in response to trauma: Prevalence and consequences
 - Kroese, F. M.; Adriaanse, M.A.; Evers, C.; Ridder, D.T.D. de, maart/april 2013, Bedtime Procrastination: An Exploration of the 'Who', 'What' and 'Why'
 - CentERdata, juni 2013/juli 2013, Economic Situation: Housing - Wave 6
- Studies Immigrant panel**
- Adams, B.; Vijver, F. van de; Bruin, G. de, februari 2013, The Association between Relational Orientation, Personality, Culture and Life Satisfaction in the Netherlands



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

DANS

• Odyssee brengt historisch perspectief

De Odyssee-collectie in EASY is vier aanvinsten rijker. Oude veldtekeningen, documentatie en vondsten van de projecten Tiel rond 1000, Vechten-Fectio 1946-1947, Mienakker en Keinsmerbrug (Unlocking North Holland's Late Neolithic Treasure Chest) en Wrak Aanloop Molengat zijn nu bestudeerd, uitgewerkt en gepubliceerd en beschikbaar. Dankzij Odyssee, een initiatief van NWO, worden vele archeologische vondsten van de periode 1900 tot 2000 in kaart gebracht en in historisch perspectief geplaatst. Dit leidt tot nieuwe wetenschappelijke inzichten binnen de archeologie.

Ook sinds kort beschikbaar:
Archeologie

- Bazelmans, J.G.A. (2012-01-14), Wrak Aanloop Molengat. Analyse en presentatie van de eerste onderwateropgraving in de Noordzee.
 - Polak, Dr M (Auxilia, archeologisch projectbureau van de Radboud Universiteit Nijmegen) (2013), De Romeinse versterkingen in Vechten-Fectio: Het archeologisch onderzoek in 1946-1947.
 - De Botersteegsvaart en de Vissteegsvaart. R.S. van der Molen, 2013.
 - Wijk bij Duurstede De Geer II Opgraving. ADC, 2013.
- Ruimtelijke wetenschappen**
- Centraal Bureau voor de Statistiek; Kadas-



Odyssee brengt nieuwe wetenschappelijke inzichten foto Evert van Ginkel

ter (2010), Bestand Bodemgebruik - BBG 2010

- Kadaster (2013), TOP10NL, digitaal topografisch bestand - versie april 2013.

Geesteswetenschappen

- Doornik, Dr. ir. J. van (University of Amsterdam, Informatics Institute); Boer, Dr. V. de (VU Amsterdam, Network Institute); Buitinck, drs. L. (University of Amsterdam, Informatics Institute); Ribbens, Dr. K. (NIOD); Veken, drs. T. (NIOD); Grootveld, Dr. M. (DANS); Marx, Dr. M. (Informatics Institute, University of Amsterdam) (2012-12-20), Verrijkt Koninkrijk: Het Koninkrijk der Nederlanden in de Tweede Wereldoorlog in XML en gesemantiseerd; Enriched Kingdom: The Kingdom of the Netherlands during World War II in XML and semanticized.
- Centraal Bureau voor de Statistiek (CBS) (2013), Woning- en Gezinstelling 1947; Woningtelling 1956 en 13e Algemene Volkstelling, 31 mei 1960.

Sociale Wetenschappen

- Centraal Bureau voor de Statistiek (1987-01-01), Enquête Beroepsbevolking - EBB - jaargangen 1987 t/m 2012 - Jaargang 2012.
- Centraal Bureau voor de Statistiek - CBS; Sociaal en Cultureel Planbureau - SCP (september 2010 t/m maart 2011, 2010), Culturele Veranderingen in Nederland 2010 - CV'10.
- Marx, Dr. M. (Universiteit van Amsterdam, Informatica Instituut) (2013-10-01), Verkiezingsprogramma's Tweede Kamer 1977-1998 bijeengebracht door Isaac Lipschits, en verkiezingsprogramma's 2006, 2010 en 2012.

Taal en Literatuur

- Stichting Vrijwilligersnetwerk Nederlandse Taal (©), SVNT; Sijs, Prof. Dr. N. van der (2013), Souter Liedekens ghemaect ter eeren Gods, op alle die Psalmen van Daud (1540).



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

Huygens ING

• Bronnenuitgave Nederlandse ontwikkelingssamenwerking 1949-1989 online

In 1949 begon Nederland met een bescheiden programma voor hulp aan wat destijds 'minderontwikkelde gebieden' heette. 'Ontwikkelingssamenwerking', zoals het terrein vanaf 1967 werd genoemd, mocht zich verheugen in een opgaande maatschappelijke belangstelling en betrokkenheid. De 1.672 documenten verdeeld over 6.000 pagina's die het Huygens ING tussen 1998 en 2009 in boekvorm liet verschijnen, zijn nu ook digitaal te raadplegen en full-text doorzoekbaar. Speciaal voor deze digitale uitgave is een overkoepelende index samengesteld op persoonsnamen en zaken. Van ongeveer 4.800 personen van over de gehele wereld die in de documenten voorkomen, zijn ook biografische gegevens opgenomen.



Landbouwvoorlichtingsproject in Tunesië, begin jaren 60 bron Ministerie van Buitenlandse Zaken

Ook sinds kort beschikbaar:

- Oorkondenboeken Holland en Zeeland: vijf delen met cumulatieve index.
- Persoonlijkheden in het Koninkrijk der Nederlanden: één deel met personenindex.
- Staten en steden van Holland vóór 1544: acht delen.
- Transcripties Dagboekjaar 1823 Willem de Clercq.



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

Kan de KB je privébibliotheek digitaliseren?

In het vorige nummer brak E-datacolumnist Ewoud Sanders een lans voor verzamelaars en geleerden die van hun fysieke bibliotheek af willen, bijvoorbeeld omdat ze kleiner gaan wonen. Liggt daar een taak voor de Koninklijke Bibliotheek (KB)? Om die mooie boekencollecties op te nemen en de schenkers een digitale kopie als dank te geven? Inge Angevaere

E-data vroeg het Jan Bos, hoofd Collecties bij de KB. "Laat ik

voorop stellen dat het een sympathiek idee is van Ewoud Sanders, en ik zou die dienst graag willen leveren, maar we kunnen het niet. De KB heeft een kerntaak, het digitaal beschikbaar stellen van alles wat ooit in Nederland is gemaakt. Daar moeten onze beperkte middelen allereerst voor worden ingezet. We hebben nu 10% digitaal beschikbaar, er wacht dus nog 90%."

Maar zo veel kost het toch niet, de selectie en het maken van digitale scans?

"Was het maar zo simpel", zegt Bos.

"Met name de selectie is veel duurder dan je zou denken. Je moet eerst nagaan wat we al gedigitaliseerd hebben, en van wat er overblijft moet je alle rechten napluizen. Auteurs, samenstellers, vertalers en illustratoren hebben exclusieve rechten op hun werk tot 70 jaar na hun dood, en uitgevers tot 70 jaar na publicatie. Ga dat maar eens uitzoeken voor een partij van 13.000 boeken die vanwege de inhoud bij elkaar is gebracht. En bovendien zal een groot deel vaak niet vallen binnen onze kerntaak, omdat het buitenlandse publicaties zijn."

Kan de KB de scans dan niet in een kluis bewaren totdat de rechten vrijkomen?

Bos: "De verzamelaar wil zo lang niet wachten. En als we hem een digitale kopie geven, kan die op alle mogelijke manieren verder verspreid worden. Volgens de wet mag iedere privépersoon, ook Ewoud Sanders, een kopie maken van zijn bibliotheek voor eigen gebruik. De KB is een instelling die daarvoor toestemming moet hebben van de rechthebbenden. Dus, hoe sympathiek het idee ook is, de praktijk is helaas weerbarstiger."



Bij het digitaliseren van kwetsbaar materiaal komt nog veel handwerk kijken bron KB

Marc van Oostendorp ging aan de slag met de data van de KB:

‘Niet álles op zijn kop door opkomst Big data’

Fonoloog Marc van Oostendorp was de eerste ‘digitale’ fellow van de Koninklijke Bibliotheek (KB) en het Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS). Zijn voorgangers werkten met het analoge materiaal. “Ik wilde vooral eens heel goed nadenken over de relatie tussen big data en de geesteswetenschappen. Daar wordt vaak zo in extremen over gesproken, dat vind ik jammer. En ik had ook een aantal concrete onderzoeksvragen.”

Goed in saai werk

“Nederlanders hebben een voorkeur voor het jambische ritme (tadám tadám tadám tadám tadám), dat vinden we het prettigst. Ik wilde zoeken naar de oudste bronnen met dat ritme in de databestanden van de KB. De Italiaanse dichter Francesco Petrarca zou het ontwikkeld hebben en het zou zich vanuit Italië verspreid hebben over Europa.

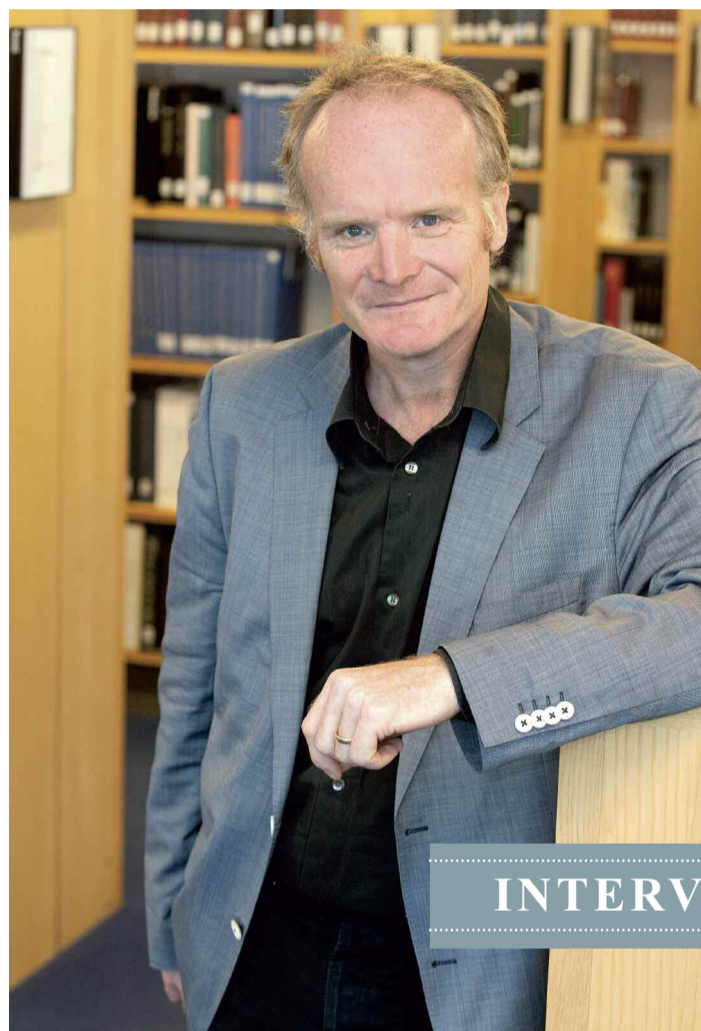
In het Nederlands is er slechts één Middelnederlandse bron, Het leven van Sinte Lutgart, en dan minstens honderd jaar niks. Dat is vreemd. Ik wilde in de buurt van die tekst gaan zoeken naar andere voorbeelden. Ritme kun je in die oude teksten alleen aantonen als ze op rijm zijn, en dan moet je nog aannemen dat de uitspraak ongeveer hetzelfde was als nu. Je moet er vele, vele bronnen voor doorspitten. Het is niet helemaal dom werk, het vereist wel enige kennis van zaken, maar verder is het vooral heel veel en heel saai. En dat is precies werk dat een computer goed kan.”

Mensen zijn te slim

“Wat ik heb onderschat, is iets dat op zich heel erg voor de hand ligt. Het Middelnederlands kent geen vaste spelling, er is eigenlijk geen standaard-Nederlands, er zijn alleen maar dialecten. Ik dacht dat op te lossen door de woorden te vervangen door dezelfde woorden uit een (modern) fonetisch woordenboek waar de klemtonen staan aangegeven. Om die vervolgens te gaan mappen. Maar dat bleek een brug te ver voor een logaritme.

Dat is het probleem van het werken met computers. Je denkt als mens in een aantal stapjes die voor jou volmaakt logisch zijn. Maar er is er altijd één bij die een grote gedachtesprong blijkt te zijn, te groot voor programmeertaal. Eigenlijk is dat ook wel het

Hij was trots op zijn KB-toegangspas en zijn kamer te midden van miljoenen boeken. Maar hij kwam voor de data. Of, liever nog, voor de mensen die hun vak maken van het werken met data. *Inge Angevaere*



INTERVIEW

“Werken met computers dwingt je om heel precies na te denken”

foto Jos Uljee / KB

leuke van met computers werken, dat ze je dwingen om heel precies na te denken. Ze moeten letterlijk alles uitgelegd krijgen. Mensen zijn te slim.”

“Big data zijn voor de geesteswetenschappen wat de telescoop voor de astronomie was”

De onderzoeksvraag

“Die specifieke onderzoeksvraag is dus niet beantwoord in deze periode. Maar we hebben wel heel veel geleerd over het probleem van die spelling. En ik denk dat we het nog wel gaan oplossen in de toekomst. Een volgende onderzoeksvraag is

of je die ritmes ook kunt herkennen in onze moderne zinsbouw. Zulk onderzoek zou je kunnen doen in de databases met kranten en ANP-bulletins van de KB. Samen met een van de programmeurs van de KB ben ik een heel eind gevorderd om dat onderzoek mogelijk te maken. Die samenwerking was heel erg plezierig.”

Ritme van taal zit diep

“Je kunt je trouwens afvragen waarom het ritme van de taal me zo fascineert. In mijn overtuiging is ritme een van de diepste dingen van de taal, letterlijk. We leren het ritme van onze moedertaal al voor onze geboorte, dat is aangevoeld. De geluiden die in de moederschoot doordringen zijn te vaag om klinkers en medeklinkers te kunnen onderscheiden, maar het ritme pikken we al op. Pasgeboren baby’s blijven langer wakker als ze hun moedertaal horen spreken dan wanneer het een andere taal is. Dat ritme zit dus heel diep, je hoort het ook in mu-

ziek, zelfs als het instrumentaal is. Engelse muziek is anders dan Franse muziek. Daar valt nog zoveel te ontdekken.”

Data geen wetenschap

Terug naar de relatie tussen big data en geesteswetenschappen. Van Oostendorp: “Dat was de andere kant van dit fellowship. Er is een soort polemieek gaande. Tussen mensen die zeggen dat we nu eindelijk écht wetenschappelijk werk kunnen gaan doen, omdat we nu pas objectieve gegevens kunnen verzamelen. Dat het vroeger allemaal subjectief was. Ook is er een tendens om te denken dat verzamelen op zich min of meer vanzelf wetenschap oplevert. Aan de andere kant zijn er mensen die zeggen dat dat allemaal onzin is. Dat alles bij het oude moet blijven, dat er weinig verandert door big data. Het lijkt mij evident dat je tussen die extremen een midden moet vinden.”

Nieuw instrument

“Big data zijn een ontzettend belangrijke ontwikkeling in de geesteswetenschappen, vergelijkbaar met de uitvinding van de telescoop. Maar die laatste leidde niet tot een soort ‘telescoopwetenschap’ met als onderzoeksvraag: wat kun je zien door een telescoop? Het debat bleef gaan over hoe het heelal in elkaar zit. Met het blote oog hadden we al veel gezien. Die kennis werd verdiept en verbeterd door de telescoop. Zo zie ik ook de nieuwe ontwikkeling van de digital humanities. Big data vormen een prachtig nieuw instrument, en er gaat veel veranderen, maar je moet veranderen vanuit wat we al weten. We denken al honderden jaren na over taal, en dat is niet zomaar onzin. Dat is niet alleen maar subjectief, dat is ook getoetst op veel verschillende manieren.”

vanoostendorp.nl

Marc van Oostendorp

Marc van Oostendorp is senior-onderzoeker aan het Meertens Instituut en hoogleraar Fonologische microvariatie aan de Universiteit Leiden. Hij publiceert regelmatig op allerhande (online) fora, waaronder *Neder-L*, elektronisch tijdschrift voor neerlandistiek. Hij was KB-NIAS fellow van september 2013 tot en met januari 2014.

AGENDA

4 maart • Hilversum

Preservation Metadata in praktijk

Welke gegevens zijn van belang voor de langetermijntoegang tot digitale bestanden? Deze vraag staat centraal in de workshop georganiseerd door het Nederlands Instituut voor Beeld en Geluid en de Nationale Coalitie voor Digitale Duurzaamheid. De workshop is bedoeld voor iedereen die te maken heeft met het duurzaam bewaren van digitale objecten zowel op uitvoerend als op managementniveau: collectiebeheerders, bibliothecarissen, archivariissen, onderzoekers, ICT-beheerders en ontwikkelaars van software. *ncdd.nl*

4 - 7 maart • Berlijn (Duitsland)

iSchools conferentie Breaking down walls

Op het forum iSchools kunnen onderzoekers en professionals op het gebied van informatiewetenschap elkaar ontmoeten. In maart organiseert iSchools een conferentie met als thema: Breaking down walls | culture, context, computing. *ischools.org/the-iconeference*

13 - 15 mei • Rome (Italië)

CRIS2014: Current Research Information Systems

De conferentie richt zich op recente ontwikkelingen in beheer, beschikbaarheid, kwaliteit en gebruik van onderzoekinformatie. De doelgroep wordt gevormd door onderzoekers, managers, financiers, ICT-experts en beleidsmakers. *cris2014.org*

13 - 16 mei • Berlijn (Duitsland)

IS&T Archiving Conference

De Society for Imaging Science and Technology organiseert sinds 2004 jaarlijks een conferentie op het gebied van digitale conservering van met name cultureel erfgoedmateriaal. *imaging.org/ist/conferences/archiving*

26 - 31 mei • Reykjavik (IJsland)

International Conference on Language Resources and Evaluation

Elke twee jaar wordt de LREC-conferentie georganiseerd. Deze conferentie brengt onderzoekers op het gebied van taaltechnologie bij elkaar. *lrec-conf.org*

27 - 30 mei • Istanbul (Turkije)

Conference on Qualitative and Quantitative Methods in Libraries

Deze zesde editie van de QQML-conferentie is bedoeld voor betrokkenen bij het ontwikkelen, uitvoeren en analyseren van kwalitatieve en kwantitatieve methoden om het functioneren van bibliotheken te verbeteren. *isast.org*

9 - 13 juni • Helsinki (Finland)

Open Repositories 2014

Er zijn verschillende datamanagementsystemen, met verschillende functies en rollen. Netwerken verbinden de repositories. Vandaar het thema van de conferentie: Towards Repository Ecosystems. *or2014.helsinki.fi*

12 - 13 juni • Den Haag

DHBenelux - Conference for Digital Humanities Research

Deze conferentie op het gebied van digitale geesteswetenschappen wordt voor het eerst georganiseerd. De call for proposals meldt dat ook onderzoekers van buiten de Benelux voorstellen kunnen indienen. *dhbenelux.org/dhbenelux-2014-conference*

Nederlandse literatuur blijft niet onopgemerkt in het buitenland

Onderzoekers uit meer dan vijftien landen werken samen om de verspreidingsgeschiedenis en ontvangst van Nederlandse literatuur in kaart te brengen. Het onderzoek wordt ondersteund door een database met gegevens over bewerkingen en vertalingen. *Katalin Márton (Eötvös Loránd Universiteit, Boedapest)*

Het NWO-project *Het is niet onopgemerkt gebleven. An International Network Studying The Circulation of Dutch Literature (CODL)* houdt zich bezig met de overdracht en verwerking van Nederlandse literatuur binnen en buiten het Nederlandse taalgebied. In het project doen ongeveer honderd onderzoekers en neerlandici onderzoek naar elf voorbeeldteksten uit de Nederlandstalige literaire canon, van de middeleeuwse Liederen van Hadewijch tot *De helaasheid der dingen* van Dimitri Verhulst.

Deelnemers aan de elf CODL-werkgroepen inventariseren vertalingen en andere bewerkingen van de ca-



Uit de film 'De helaasheid der dingen' naar het boek van Dimitri Verhulst bron Menuet/idtv film

susteksten in een gemeenschappelijke database. De vrij toegankelijke database wordt door de projectleden in de loop van hun onderzoek voortdurend uitgebreid. De bedoeling is om aan het eind van het project, in 2015, een volledige database te hebben waar alle bewerkingen van de casusteksten in te vinden

zijn. Hiermee wordt het mogelijk de verspreiding van Nederlandse literatuur in kaart te brengen.

De digitale database is letterlijk grensoverschrijdend. Ze maakt het mogelijk om op internationaal niveau onderzoeksmateriaal te delen en inzicht te krijgen in anderms onderzoek. Zo bevordert de data-

base de samenwerking van neerlandici uit verschillende landen en helpt ze bij de uitwisseling van expertise. Dat is vooral belangrijk voor de Neerlandici die buiten Nederland en Vlaanderen in een zeker isolement hun werk doen.

codl.nl

Eerste Nederlandse THATCamp blijkt leerzame unconference

Een conferentie zonder vooropgezet programma, waar technologen en geesteswetenschappers van alle kennisniveaus elkaar ontmoeten en samen de bijeenkomst maken, kan dat? *Peter Boot*

Het bleek te kunnen, getuige *The Humanities and Technology Camp*, het eerste Nederlandse THATCamp op 14 en 15 januari jongstleden bij het Huygens Instituut voor Nederlandse Geschiedenis, mede georganiseerd door het Koninklijk Nederlands Historisch Genootschap. De aanleiding was de lancering van het platform *historici.nl* (zie de vorige E-data & Research). E-data vroeg drie bezoekers om een impressie.

Open en vriendelijk

Lex Heerma van Voss, directeur Huygens ING: "Zo'n unconference brengt mooie anarchistische praktijken boven. Datum, plaats en het onderwerp worden bepaald, maar verder is de invulling *bottom-up* en bijna vrijwillingvrij. De onderwerpen van de sessies werden aangedragen door de deelnemers. Niet door een programmacommissie geaccepteerd, maar in een plenaire vergadering van de 80 deelnemers aan het begin van de dag besproken en aangevuld. Er waren inleiders,

maar het debat bepaalde of zij hun geplande betoog afstaken of dat de aandacht al doende verlegd werd. In plaats van presentaties waren het werkelijk inleidingen op een opmerkelijk open en vriendelijk debat. Dat bleek heel goed te kunnen leiden tot nieuw inzicht in de bruikbaarheid van software voor kwalitatief onderzoek of tot een suggestie hoe het aanbod historische kranten van de KB bruikbaar gemaakt kan worden voor historisch onderzoek."

Leerrijke ervaring

Erik Tjong Kim Sang, computertaalkundige bij het Meertens Insti-

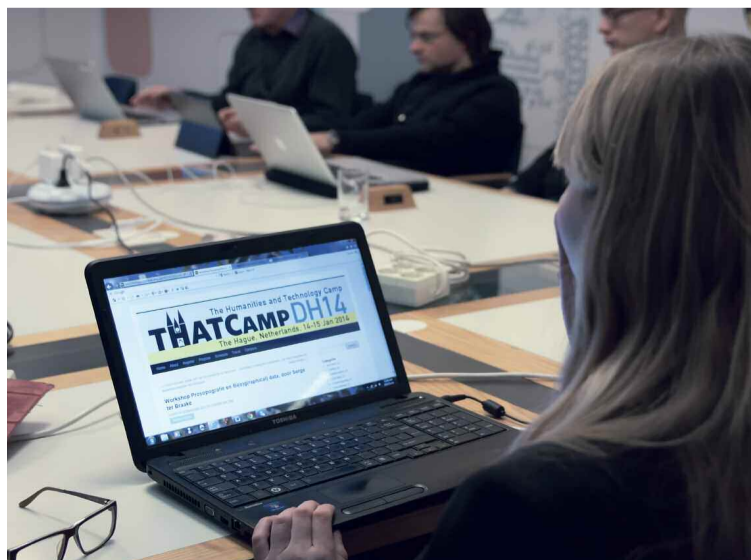
tuut: "Meteen na de inleiding werd aan de deelnemers gevraagd welke onderwerpen ze wilden bespreken. Deze onderwerpen werden verdeeld in 16 workshops van anderhalf uur. Zelf bezocht ik een workshop over digitale tools voor historici en een workshop over Wikipedia. Het formaat van de bijeenkomst is mij goed bevallen. Het was geen probleem om in de eerste sessie genoeg onderwerpen voor de workshops te verzamelen. Ik moet wel zeggen dat ik alleen naar workshops ben geweest die vooraf waren aangemeld. Deze hadden daarom een voor de workshop verantwoordelijke persoon die een ruime inleiding gaf

over het onderwerp. THATCamp The Hague was een leerrijke ervaring, voor mij een eerste kennismaking met dit formaat bijeenkomst. De behandelde onderwerpen waren interessant en de sfeer op de conferentie was goed. Dit is voor herhaling vatbaar."

Open sfeer

Mari Smits, historicus bij Huygens ING: "Het afgelopen jaar nam ik bij diverse gelegenheden al kennis van onderdelen van de digitale geschiedenis. Tijdens het THATCamp heb ik dan ook vooral de gelegenheid benut om te 'ruiken' aan praktische toepassingen. Zo nam ik deel aan een sessie over programmeren met JavaScript, maakte ik kennis met Prezi als alternatief voor Powerpoint-presentaties en initieerde ik zelf een sessie over websites bouwen. Verder raakte ik onder de indruk van de spectaculaire ontwikkelingen rond de herkenning in niet-schriftelijke bronnen. Wat me het meest is opgevallen was de open sfeer. Zonder verplichte nummers – afgezien van de plenaire sessies als start en einde van het THATCamp – lag de regie in handen van de deelnemers zelf. Dit bood tevens de gelegenheid om kennis te maken met vakgenoten in het land."

thehague2014.thatcamp.org



Deelnemers tijdens een workshop aan de slag foto Milo van de Pol

KORT

Standaardwerk WOII verrijkt doorzoekbaar

Het Koninkrijk der Nederlanden in de Tweede Wereldoorlog is sinds kort online doorzoekbaar. Eerder was het standaardwerk van Loe de Jong al integraal te downloaden via de website van het NIOD. Met behulp van een bijdrage van CLARIN-NL verbeterde het NIOD samen met de Universiteit van Amsterdam, de Vrije Universiteit, het Meertens Instituut en DANS de doorzoekbaarheid van De Jongs werk. De data zijn nu bovendien beschikbaar in de vorm van Linked Open Data, zodat de dataset gekoppeld kan worden aan externe datasets, zoals Wikipedia. In hackatons worden programmeurs uitgedaagd om praktische toepassingen voor de datasets te ontwikkelen. (ER) *niod.nl*

DiXiT: netwerk voor digitale edities

In samenwerking met een aantal universiteiten en editie-instituten uit Europa heeft de universiteit van Keulen van de EU subsidie gekregen voor een opleidingsnetwerk voor digitale wetenschappelijke edities. Het doel van het netwerk is om onderzoekers op te leiden die in een tijd van veranderende media en technologieën de betrouwbare bronnenuitgaven kunnen maken waar het historisch en letterkundig onderzoek behoefte aan heeft. In het kader van dit programma kunnen 12 jonge onderzoekers en 5 ervaren onderzoekers worden aangesteld. Bij Huygens ING gaan in april twee onderzoekers aan de slag. Ze richten zich op respectievelijk een business model voor de digitale editie en de mogelijkheden om online invoer voor de digitale editie te combineren met technisch complexere infrastructuur. (PB) *dixit.uni-koeln.de*

Croatian Memories: omgaan met verleden

Croatian Memories is een door het Ministerie van Buitenlandse Zaken gefinancierd project waarin burgers uit Kroatië zijn geïnterviewd over gebeurtenissen tijdens de Tweede Wereldoorlog, de Joegoslavische periode en de oorlog in de jaren negentig. De beschikbaarheid van de getuigenissen beoogt bij te dragen aan de ontwikkeling van een pluriforme en democratische samenleving in Kroatië. Het project is uitgevoerd door 'Documenta - Center for dealing with the past' in Zagreb in samenwerking met de Erasmus Universiteit, Universiteit Twente, DANS en het bedrijf Noterik. Bij veel interviews is een Engelse ondertiteling beschikbaar. Het project is afgerond, de meeste interviews (door DANS duurzaam gearcheveerd) zijn vrij toegankelijk. (RvH) *croatianmemories.org*

Namescape past named entity recognition toe

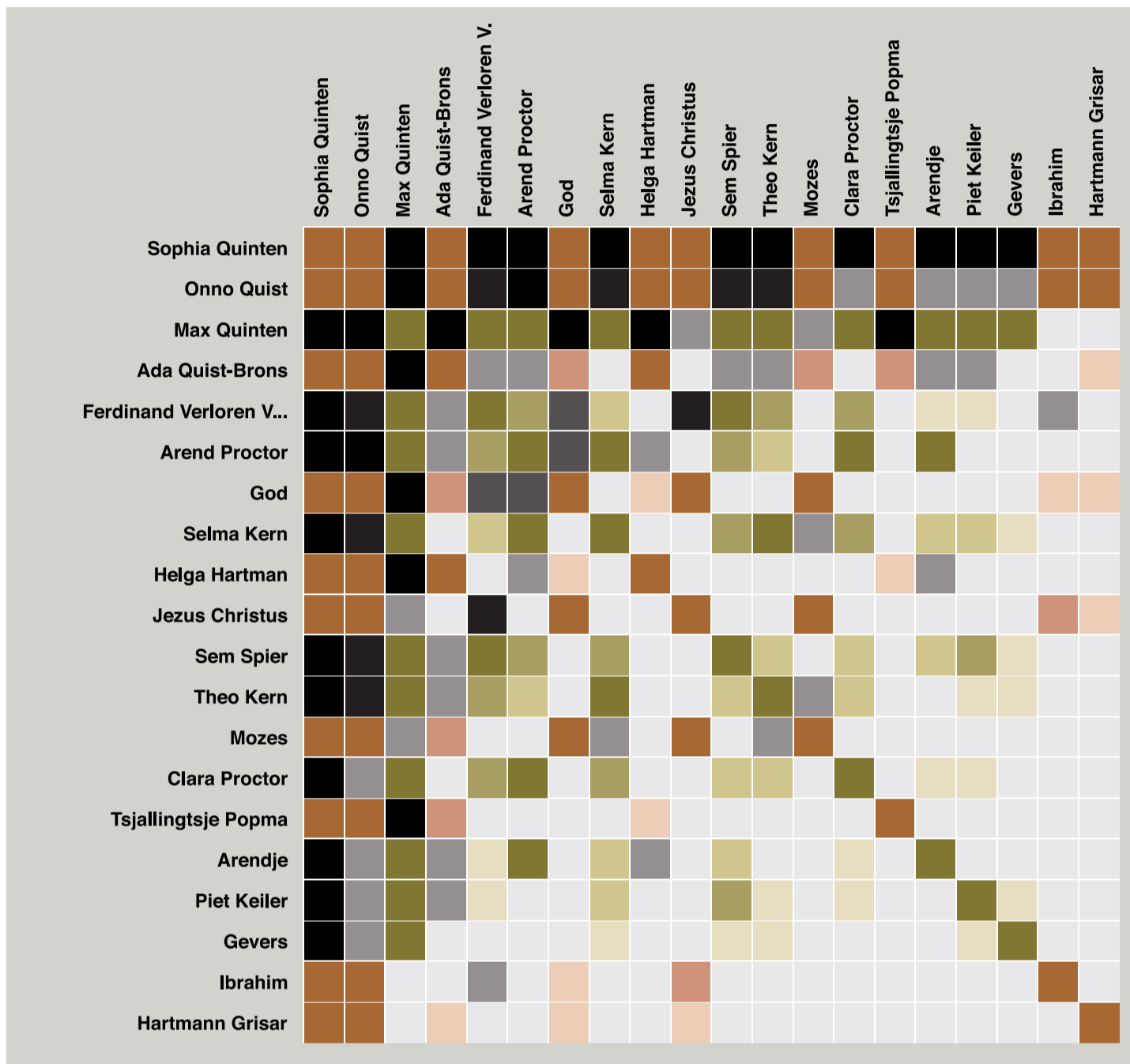
Letterkundig namenlandschap

Wat kunnen we leren over romans door te kijken naar het gebruik van namen? En zijn er voldoende gegevens beschikbaar voor betrouwbare uitspraken? Het Namescape-project brengt het in kaart. Peter Boot

Namen hangen direct samen met een belangrijk aspect van de inhoud (personages en plaatsen) en belichamen ook stilistische keuzes van de auteur. Gebruik van voornamen versus familienamen zegt bijvoorbeeld iets over afstand en intimiteit. Maar voor betrouwbare uitspraken daarover moeten gegevens over een groot aantal romans beschikbaar zijn. In het onlangs afgesloten Namescape-project (een CLARIN-demonstrator project) is van 1.129 literaire werken in kaart gebracht welke namen er voorkomen en hoe die met elkaar samenhangen.

Namen in context

De *Named Entity Recognition* (NER) werd uitgevoerd door een team bij het Instituut voor Nederlandse Lexicologie (INL). Het bleek dat verhalende teksten toch wel anders zijn dan het materiaal waar NER-programmatuur meestal voor wordt ontwikkeld. Jesse de Does (INL): “De namen in literaire teksten zijn vaak nieuw, nog niet uit andere bronnen bekend. De namendichtheid is er ook anders. Voor Namescape werd gebruik gemaakt van een bestaande NER-tagger (de Stanford NER-tagger, nlp.stanford.edu/ner/), getraind op een speciaal ontwikkeld trainingcorpus van ongeveer 1 miljoen tokens. Maar er werd ook een eigen tagger ontwikkeld, die op dit materiaal nog iets beter presteert.” De kwaliteit van de uitvoer van een statistische tagger hangt af van de hoeveelheid trainingmateriaal, waaruit de tagger namen in context leert herkennen. De Does: “Eigenlijk hebben we een samenhangende collectie nodig van al het Nederlandstalig trainingmateriaal voor naamsherkenning”. Het INL ontwikkelde ook een web service waarmee onderzoekers hun eigen teksten kunnen laten analyseren op het gebruik van namen. Onderzoekers kunnen er



Namen in Mulisch' *De ontdekking van de hemel* Personages die in het boek samen voorkomen, hebben dezelfde kleur. De intensiteit van de kleur is gebaseerd op de frequentie waarmee ze samen voorkomen. Zichtbaar is ook dat de naamherkenner soms personages combineert (Quinten is een afzonderlijk personage, niet de achternaam van Max of Sophia).

bestanden uploaden, verwijzen naar een webpagina, of de service aanroepen via eigen scripts.

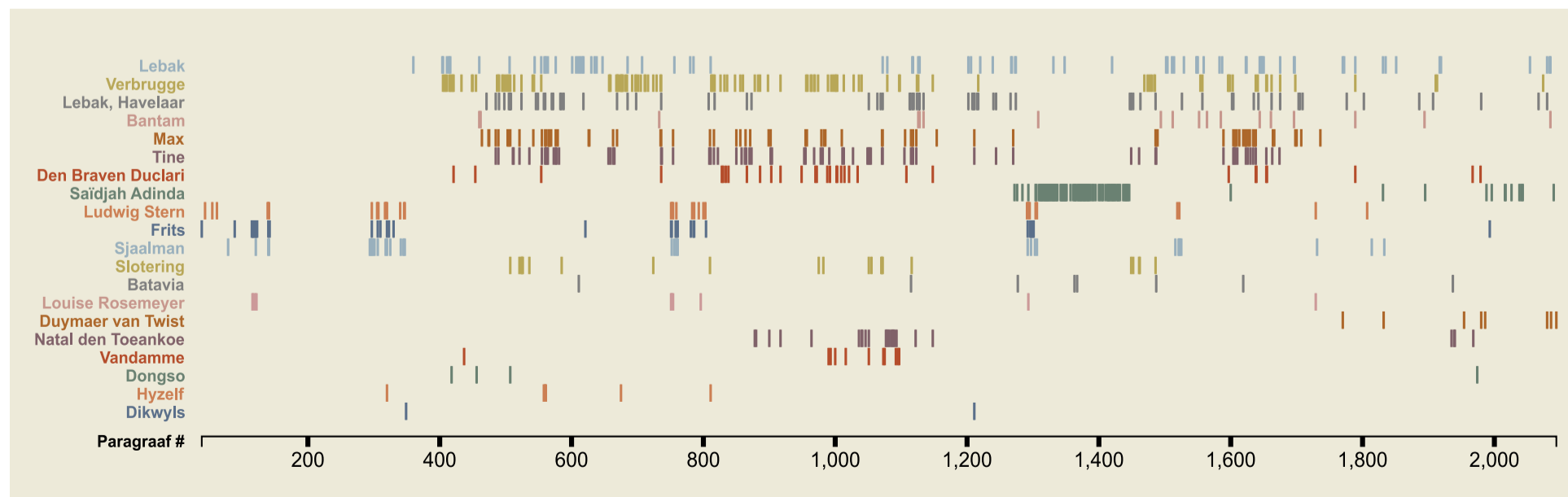
Visualisatie van de resultaten

Onder leiding van Maarten Marx werkten studenten aan de UvA aan de visualisatie van de resultaten. Ze ontwikkelden onder andere een ‘karakterbundel’ die clusters van personages

en de samenhang daartussen toont, een ‘matrixweergave’ van welke personages voorkomen in dezelfde passage, en een ‘barcodeweergave’ van de personages door het boek. “Het is belangrijk dat we nu deze gegevens ter beschikking hebben”, aldus Karina van Dalen-Oskam (Huygens ING), de projectleider van Namescape. Maar het project genereert ook een methodologisch resultaat. Van Dalen:

“Hoe goed de taggers ook zijn, er gaat naar mijn gevoel nog erg veel mis. Letterkundig onderzoekers zijn daar niet aan gewend. Een belangrijke vervolgvraag is hoe we in geesteswetenschappelijk onderzoek moeten omgaan met *noise*.”

namescape.nl
visualizer.namescape.nl
ner.namescape.nl



Barcodeweergave van de namen in Multatuli's Max Havelaar Elk streepje in deze grafiek staat voor een alinea waarin de betreffende persoon verschijnt. De structuur van het boek is goed zichtbaar, met afwisselend hoofdstukken in Amsterdam (Stern, Frits en Sjaalman) en hoofdstukken in Indië. Ook de ingelaste vertelling over Saïdjah en Adinda is goed zichtbaar. Helaas zien we ook nogal wat ruis: de personages Lebak Havelaar, Dikwyls en Hyzelf treffen we in de roman niet aan.

Kranten, radio, tv: allemaal doen ze verslag van politieke debatten. De online tool PoliMedia maakt gecombineerde analyse van deze berichtgeving mogelijk.
Erica Renckens

FOCUS

‘Een gebruiksvriendelijke en innovatieve manier om de media-berichtgeving van de debatten in de Tweede Kamer te onderzoeken.’ Zo beoordeelde de jury van de Linked-Up Challenge afgelopen najaar PoliMedia, een zoekmachine waarin de notulen van de Tweede Kamer zijn gekoppeld aan de radiobulletins van het ANP en het krantenarchief van de Koninklijke Bibliotheek (KB). Het samenwerkingsproject van de Erasmus Universiteit Rotterdam (EUR), TU Delft, VU Amsterdam en het Nederlands Instituut voor Beeld en Geluid won daarmee de eerste prijs in de Europese wedstrijd voor onderwijstools die gebruikmaken van gekoppelde datasets.

Een web van data

PoliMedia maakt gebruik van semantische webtechnologieën om de links tussen de notulen, de kranten en de radiofragmenten te leggen. “Dit betekent dat politici, debatten, krantenartikelen en radiofragmenten allemaal als object worden opgeslagen, en met elkaar worden verbonden via links die hun onderlinge relatie beschrijven”, legt onderzoeker Max Kemman uit. Door collecties op deze manier met elkaar te verbinden, zoekt de gebruiker niet in afzonderlijke collecties met documenten, maar in een web van data. Deze manier van verbinden maakt



Fractievoorzitter Emile Roemer van de SP tijdens het debat over het begrotingsakkoord 2014 in de Tweede Kamer foto Peter Hilt / Hollandse Hoogte

Eerste prijs voor online tool met linked data

zowel kwalitatieve als kwantitatieve analyses mogelijk. “Je kunt bijvoorbeeld vragen welk debat het meest in de krant besproken werd of welke politicus het vaakst aan het woord is tijdens debatten”, aldus Kemman. “Daarnaast is het mogelijk om de informatie uit te breiden, bijvoorbeeld door de politici verder te koppelen aan biogra-

PoliMedia maakt media-analyse politiek debat mogelijk

fische gegevens van Wikipedia of Parlement.com. Hierdoor kunnen vragen worden gesteld aan informatie buiten PoliMedia. Is er bijvoorbeeld een verschil tussen de hoeveelheid verslaglegging over toespraken van Friese en Limburgse politici? Of is media-aandacht gerelateerd aan leeftijd van politici?”

Nieuwe collecties

“PoliMedia is in principe afgerond, maar we zien dat er veel aandacht en interesse voor is”, vertelt onderzoeker Martijn Kleppe. De notulen en de krantenartikelen in PoliMedia, dat is gefinancierd door CLARIN-NL, lopen tot 1995; de

ANP-bulletins slechts tot 1984. “Het zou mooi zijn als we in de toekomst kunnen uitbreiden. Niet alleen met nieuwe, vrij toegankelijke data, maar ook met nieuwe tools, zoals visualisaties.” Omdat de links die PoliMedia maakt tussen de debatten van de Tweede Kamer en de mediaberichten zijn vrijgegeven als Linked Open Data, kunnen ook andere onderzoekers voortbouwen op de relaties tussen politiek en media en zo nieuwe onderzoeksvragen beantwoorden. Via <http://data.polimedia.nl> kunnen ontwikkelaars nu al nieuwe applicaties bouwen op basis van de in PoliMedia verzamelde data. polimedia.nl

GELEZEN

Bijdragen en Mededelingen betreffende de Geschiedenis der Nederlanden (BMGN). Speciaal Engelstalig themanummer 'Digital History'. Vol. 128, nr. 4 (2013).
René van Horik

December 2013 verscheen het BMGN-themanummer 'Digital History' waarin het gebruik van digitale technieken bij historisch onderzoek centraal staat. Het overzichtartikel is van de hand van gastredacteur Gerben Zaagsma. Hij beschrijft de ontwikkeling van 'digital history' sinds de jaren zestig met de nadruk op de situatie in Nederland. Er is aandacht voor de rol van 'Historia en Informatica', de voorganger van E-data & Research. Zaagsma pleit voor een hybride visie op de Geschiedwetenschap waarbij traditionele onderzoekstechnieken gecombineerd worden met de mogelijkheden van informatietechnologie en digitale en gedigitaliseerde bronnen.

Afzonderlijke artikelen in het themanummer zijn gewijd aan drie onderwerpen: de digitalisering van archieven en de gevolgen hiervan voor historisch onderzoek, digitale historische analyse en (re)presentatie van historische kennis met behulp van informatietechnologie. Het themanummer eindigt met een debat tussen Rens Bod, hoogleraar Digital Humanities, en een aantal kritische collega's. Dit debat richt zich op de betekenis en waarde van het gebruik van digitale technieken in geesteswetenschappelijk onderzoek. Zoals gesteld in het voorwoord illustreren de bijdragen in het themanummer zowel de beloften als valkuilen van digitale geschiedenis en vormen zij een uitstekende beschrijving van de huidige stand van zaken.

bmgn-lchr.nl

COLUMN

Afhankelijk

Vroeger nam ik stapels boeken mee als ik op reis ging. Die zocht ik met veel zorg uit. Ze moesten niet alleen passen bij de reis, maar ook niet te veel wegen. Je wordt kritisch als je boeken moet dragen.

Tegenwoordig pak ik het anders aan. Ik neem veel meer boeken mee, dun en dik door elkaar, maar voor vertrek digitaliseer ik ze. Ik lees ze op mijn iPad.

Onlangs reisde ik vijf weken alleen door Birma, Laos en Cambodja. Als je alleen op reis bent is zo'n tablet extra belangrijk, merkte ik. Ik las er mijn boeken en kranten op, deed mijn mail, ik schreef een aantal stukken en af en toe nam ik er een foto mee. Daarmee behoorde mijn iPad tot de belangrijkste dingen die ik bij me had.

Ik baalde dan ook verschrikkelijk toen hij na twee weken, dus halverwege de reis, blokkeerde. 'iPad is geblokkeerd, verbind met iTunes' stond er op het begin-

scherm, dat verder zwart bleef. En wat ik ook probeerde, er was geen beweging meer in mijn iPad te krijgen.

Nou zijn er in Birma, Laos en Cambodja bij mijn weten geen officiële Apple-dealers, maar wel voldoende mensen die veel verstand hebben van Apple-producten. Van hen kreeg ik bij her-



foto Leo van Velzen

haling te horen wat ik ook al op internet had gelezen, namelijk dat er twee manieren waren om mijn iPad ter plekke weer aan de praat te krijgen. Met een bepaalde toetscombinatie kon ik het apparaat resetten. Of ik moest mijn iPad aan een pc hangen nadat ik daar iTunes op had gedownload – en dan resetten.

In beide gevallen zou ik alles wat er op mijn iPad stond, kwijtraken. Dat wil zeggen, alles sinds mijn laatste kopie op iCloud. Dus: alles wat ik in twee weken had geschreven plus alle aantekeningen die ik de afgelopen weken in boeken en artikelen had gemaakt.

Derde oplossing: terug naar huis vliegen en mijn iPad daar aan mijn pc hangen, dan was het pro-

bleem zonder verlies aan data onmiddellijk verholpen. Thuis was veertien uur vliegen verder weg. Ik ben natuurlijk niet naar huis gevlogen en heb mijn reis afge maakt. Maar het was wel een enorme domper om niet meer bij mijn boeken en kranten te kunnen. Mail kan je nog wel bekijken in een internetcafé, hoewel ik dat niet ideaal vind.

Hoe dit probleem is ontstaan, heeft tot nu toe niemand mij goed kunnen uitleggen. Ik hoor dat dit maar zelden voorkomt.

Voor mij was het echter de tweede keer. Eerder dit jaar, bij een andere reis, raakte mijn iPad ook geblokkeerd, toen gelukkig slechts een paar dagen voor terugkomst.

Ik vond het confronterend om te merken hoe afhankelijk ik ben geworden van dergelijke mobiele apparaten. Even voelde ik me werkelijk ontredder. Ik ben er een paar dagen flink sacherijnig van geweest.

Inmiddels heb ik besloten om nooit meer met een iPad op reis te gaan. Althans, niet zolang die mijn persoonlijke iTunes – op mijn pc thuis – nodig heeft om eventuele problemen zonder data-verlies te verhelpen. Ik heb altijd al een hekel gehad aan iTunes, maar dit was voor mij de druppel. Tablets behoren in mijn ogen tot de beste uitvindingen van de afgelopen jaren, maar dan moeten ze wel écht mobiel zijn en niet afhankelijk van één bepaalde pc ergens op de wereld.

Ewoud Sanders

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