

Utrecht Data School onderzoekt vluchtelingendebat op Twitter

Onderzoek naar tweets met #gelukszoeker

Hoe haal je betekenis uit tweets?

De Utrecht Data School analyseert de vluchtelingencrisis op Twitter door netwerkanalyse te koppelen aan semantische indicatoren. *Paul Verhaar en Mirko Tobias Schäfer*

Schäfer

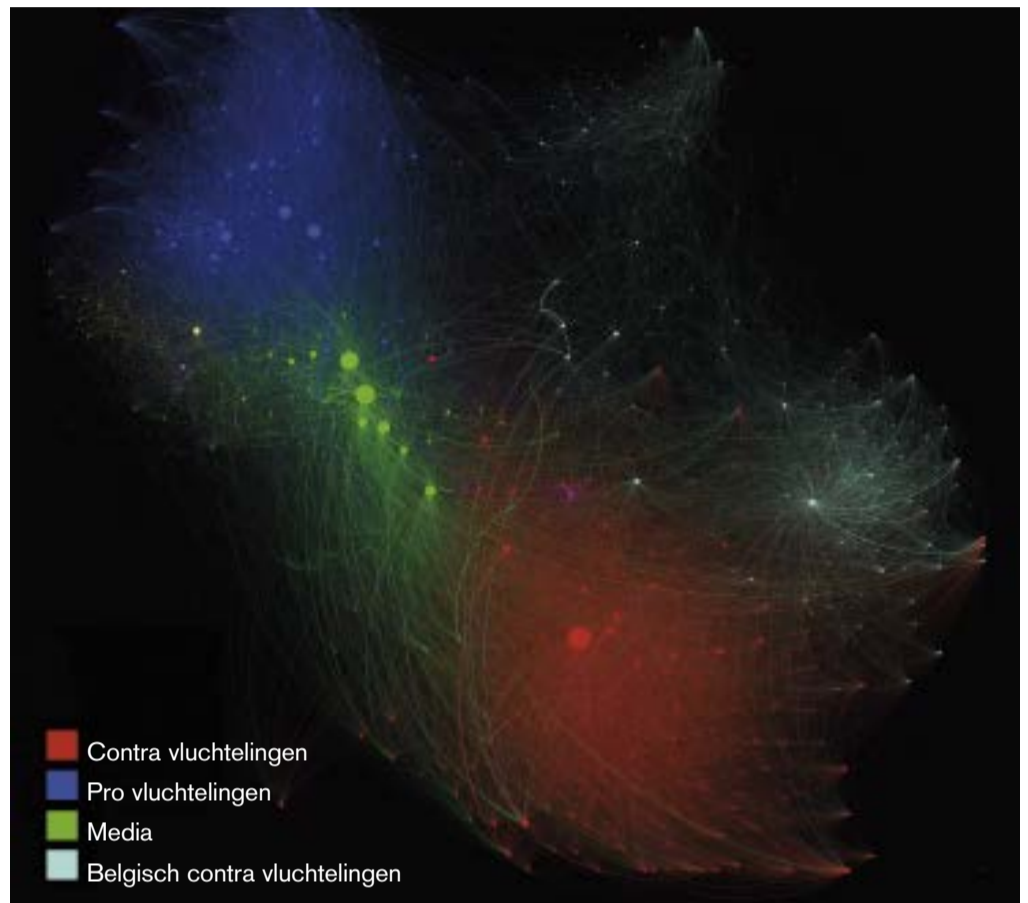
Door het toenemende aantal vluchtelingen nam ook het aantal gerelateerde berichten op social media toe. Talloze berichten werden verspreid over het internet, waarin gebruikers hun mening uitten. Een van de grootste uitdagingen is om betekenis uit deze berichten te halen.

“Ons onderzoek focust zich op het in kaart brengen van het vluchtelingendebat met behulp van netwerkanalyses en linguïstische analyses. Onze totale dataset bestaat uit 561.179 tweets, waarvan 198.100 retweets, met het woord ‘gelukszoeker’. Dit woord brengt een dubbele connotatie met zich mee en is een interessante invalshoek voor onze analyse.”

561.179 tweets

Dit Retweet-netwerk laat (re)tweets van politiek-Nederland zien. Rood staat voor politiek rechts (38% van het totaal aantal retweets), blauw staat voor politiek links (24% van het totaal aantal retweets). De visualisatie laat zien dat het debat duidelijk verdeeld is. De media staan in het midden: hun berichten worden vanuit de twee clusters geretweet. Het Twitter-account van Geert Wilders is het account met de meeste retweets binnen het rechtse cluster. Maar ook PVV-fractievoorzitter René Dercksen is een hoogactieve opinielider in dit gebruikerscluster. In het linkse cluster is het debat veel diverser, een grotere groep aan opinieliders is zichtbaar.

“De netwerkanalyse biedt veel potentie voor lin-



Onze totale dataset bestaat uit 561.179 tweets, waarvan 198.100 retweets, met #gelukszoeker. De blauwe en rode nodes geven 38% en 24% weer van het totaal aantal retweets. Het debat is verdeeld; het netwerk laat een duidelijke polarisatie zien. Ook de positie van de media in de discussie is duidelijk. Bovendien valt op dat het rechtse cluster relatief klein is. Het politiek linkse kamp laat veel verschillende opinieleiders zien, in het rechtste kamp domineert de leider van de PVV. credits UDS

guïstische analyse. Tot op heden zijn er nog geen onderzoeken geweest die netwerkanalyses en linguïstische analyses samenvoegen. Een netwerk kan dienen als basis voor taalanalyse waarin taalpatronen onderzocht en gekoppeld kunnen worden aan de uitkomsten van het netwerk. *Natural language processing* maakt het mogelijk om dieper naar taalrelaties binnen netwerken te kijken met behulp van computationele technieken. De

kracht ligt in de combinatie: nieuwe mogelijkheden om taal online beter te begrijpen en beter te kunnen *minen*. Door deze en andere onderzoekservaringen met grote corpora, zoals twitterberichten, hopen we op termijn specifieke linguïstieke kenmerken van extremistische posities te kunnen traceren, onafhankelijk van netwerkanalyse.”

dataschool.nl

Wint u de Dataprijs '16?

Ook dit jaar wordt de Nederlandse Dataprijs uitgereikt. Een prijs voor een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor aanvullend of nieuw onderzoek.

Heidi Berkhout

Bent u of kent u zo'n onderzoeker met een goed voorbeeld van gedeelde data? U kunt uzelf of een ander via researchdata.nl nomineren. De winnaars van de voorgaande edities zijn in ieder geval enthousiast. Martine de Bruin, winnaar Dataprijs 2014: “De Dataprijs is een bevestiging dat we op het goede spoor zitten” en Marijke Dekker, winnaar Dataprijs 2014: “Door het winnen van de Dataprijs kunnen we nu ook een paar grotere, al langer gewenste verbeterlagen maken.” De prijsuitreiking is op 9 november bij NWO in Den Haag. Ook aanmelden kan online.

researchdata.nl/diensten/dataprijs

Kijk voor meer informatie over de prijs, zoals de drie categorieën, de prijzenpot en de criteria op de website van RDNL. credits RDNL

248 levensverhalen van vluchtelingen

De Stichting Bevordering Maatschappelijke Participatie (BMP) legt verhalen met vluchtelingen van toen en nu vast in het Oral History project ‘Ongekend Bijzonder’. Deze persoonlijke getuigenissen verdienen het om voor de toekomst bewaard te worden. De focus van de interviews ligt op het opbouwen van een nieuw leven in de vier grote steden. Enkele tientallen interviews zijn *Interview Ongekend Bijzonder foto St. BMP* EASY, het online archiveringssysteem van DANS. Een overzicht hiervan met links naar enkele interviews staat in de thematische collectie: Ongekend Bijzonder - Oral History project. (HB)

ongekendbijzonder.nl
DOI: 10.17026/dans-zbd-b88x

E-DATA & RESEARCH

Jaargang 10 | nummer 3

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.

10 JAAR E-DATA

INHOUD

2 Verslagen van events in Gehoord en bijgewoond

3 Metadata van Clusius' brieven in eLaborate

3 RIS Radboud helpt bij duurzaam archiveren

4 & 5 Mijlpaal! Alweer 10 jaar E-data & Research



4 & 5 'Zorgen om de toekomst van data en software'

6 Alliantie strijd voor Text & Data Mining

7 Deze datasets zijn ook sinds kort beschikbaar

8 Patiëntenvolgsysteem: data online beschikbaar

8 Ewoud Sanders: een overdaad aan bronnen



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

eLaborate bevat metadata van honderden brieven

Correspondentie Clusius nu ook online beschikbaar

Dankzij de Clusius-correspondentie zijn honderden brieven van deze Vlaamse geleerde volledig digitaal doorzoekbaar.

Ilja Nieuwland

Carolus Clusius (1526-1609) hield zich bezig met wat destijds 'natuurlijke historie' heette. Hij is vooral bekend geworden dankzij honderden beschrijvingen van nieuwe plantensoorten en het oprichten van de Leidse hortus botanicus. Hoewel veel van Clusius' publicaties al digitaal konden worden bestudeerd via de website van de Missouri Botanical Garden Library, was zijn omvangrijke correspondentie nog niet toegankelijk en doorzoekbaar. Namens Huygens ING (KNAW) en het Scaliger Instituut (Universiteitsbibliotheek Leiden) ontwikkelde Esther van Gelder daarom een online editie van deze rijke correspondentie.



Portret van Carolus Clusius, prentmaker: Cornelis Galle (I) credits Rijksmuseum, hdl.handle.net/10934/RM0001.COLLECT.345840

Netwerk van correspondenten

"Aandacht voor het netwerk van correspondenten rondom een bekende geleerde als Clusius kan een ander perspectief bieden op de ontwikkeling van wetenschappelijke kennis. Clusius werkte samen met honderden mensen die allemaal hun eigen sociale en intellectuele bagage meebrachten bij het bestuderen van de natuurlijke wereld in Europa en daarbuiten. Hun brieven, illustraties, objecten en beschrijvingen laten ons zien hoe natuurlijke historie zich ontwikkelde tot een belangrijk veld in de late zestiende eeuw."

Het Clusius-project, dat van 2004 tot 2011 werd uitgevoerd door de Universiteit Leiden, stortte zich op deze uitwisselingen. Tijdens dit onderzoek kwam Van Gelder op het spoor van honderden ongepubliceerde transcripties van de brieven van Clusius in twintigste-eeuwse onderzoeksarchieven van de UB Leiden.

Huygens ING en Scaliger maakten een grote stap in de richting van een complete digitale editie van

Clusius' correspondentie door het verzamelen en digitaliseren van alle bestaande brieven en transcripties in eLaborate, een collaboratieve transcriptietool ontwikkeld door Huygens ING. De eerste versie van deze *edition in progress* bevat metadata van alle 1.600 bekende brieven, scans van 1.170 brieven en transcripties van bijna duizend brieven. Voor de eerste keer zijn deze brieven nu volledig digitaal doorzoekbaar, waardoor onderzoekers in één klap de introductie en circulatie van nieuwe plantensoorten kunnen volgen. Bovendien is er een inleiding geschreven waarin het collaboratieve editieproces en de herkomst en kwaliteit van de teksten toegelicht worden.

Samen verder

"Vanwege de grote verscheidenheid aan bronnen van gepubliceerde en ongepubliceerde transcripties is er (nog) geen uniforme kwaliteit. Bovendien is er nauwelijks enige annotatie toegevoegd. We presenteren deze editie dan ook als *work in progress* en nodigen onderzoekers en studenten uit om de online editie te doorzoeken en vervolgens nieuwe scans, transcripties en annotaties toe te voegen, bijvoorbeeld via de Clusius Community 2.0. Er ontbreken bijvoorbeeld nog zo'n 600 brieven."

De Clusiusbrieven zullen uiteindelijk opgenomen worden in het Epistolarium, een online databank van zeventiende-eeuwse geleerden-correspondentie ontwikkeld door Huygens ING.

clusiuscorrespondence.huygens.knaw.nl

3 miljoen extra pagina's in online Delpher

De landelijke website voor gedigitaliseerde boeken, kranten en tijdschriften Delpher heeft er drie miljoen krantenpagina's bij. Ruim twintig krantentitels zijn voor het eerst opgenomen in Delpher. Het grootste deel van de nieuwe aanvulling bestaat uit aanvullende jaargangen van titels die al in Delpher waren te vinden. Helemaal nieuw voor Delpher zijn 450 zogenaamde 'overgangsbladen'. Dat zijn kranten uit de veelbewogen jaren 1945-1947, de periode waarin de Nederlandse bevolking overeind krabbelde na de Bevrijding. De weerslag is te vinden in tientallen nieuwe, heropgerichte en 'bovengronds' gekomen verzetskranten. (HB)

delpher.nl



Het Amsterdamsch Dagblad staat nu ook online credits KB

Essentials 4 Data Support verdiept en verbreedt

Frank de Jong, medewerker afdeling Collectievorming, Internationaal Instituut voor Sociale Geschiedenis

Getipt door mijn KNAW-collega van het NIOD heb ik in het najaar van 2015 de cursus Essentials 4 Data Support gevolgd. Mijn werkzaamheden bij de afdeling Collectievorming van het IISG staan voor een groot deel in het teken van het aangaan van nieuwe en onderhouden van bestaande relaties met archiefvormers. Denk daarbij aan de archieven die gevormd worden door vakbonden (FNV en CNV) en politieke partijen (PvdA en Groen Links). Bij recente aanvullingen gaat het om digitaal geboren archief. In onze aanpak hanteren we een aantal basisbegrippen van goed beheer: vertrouwelijkheid (open als het kan, gesloten als het moet), duurzaamheid en terugvindbaarheid. De RDNL-cursus viel samen met interne ontwikkelingen op het gebied van opslaan, beschikbaar stellen en delen van gestructureerde datasets.

GEHOORD & BIJGEWOOND

Ruim 100 datasets uit afgesloten en nog lopende onderzoeksprojecten worden sinds kort aangeboden op de Dataverse-repository. Dankzij de cursus kon ik me goed verdiepen in een aantal technische, juridische en organisatorische zaken rondom datamanagement. Uiterst praktisch vond ik de geboden handreikingen voor de *event*-gerichte acquisitie van datasets. Zo zag ik uitstekende nieuwe kansen voor het verzamelen van datasets die ik op de European Social Sciences History Conference (ESSHC) in Valencia van 30 maart tot en met 2 april 2016 met succes in de praktijk heb kunnen brengen. Een groot deel van de onderzoekers die de ESSHC bezoeken of hun papers presenteren, vindt aansluiting bij het onderzoeksprogramma van het IISG: de geschiedenis van werk, werkenden en arbeidsverhoudingen wereldwijd. Een toenemend aantal onderzoekers legt hun onderzoeksgegevens vast in gestructureerde

datasets. Nieuw verworven datasets zouden idealiter ingezet kunnen worden in de bestaande wereldwijde *collaboratories* bijvoorbeeld rondom stakingen, beroepsstructuur of arbeidsverhoudingen.

Ik probeer onderzoekers met datasets naar de best geschikte repository te verwijzen. Niet alles hoeft immers naar het IISG, doorverwijzen naar DANS, of equivalente buitenlandse instellingen, kan een betere optie zijn.

datasupport.researchdata.nl

NPSO-themamiddag over relatie surveys en beleid

Frans Louwen, senior consultant, marktonderzoeksbureau GfK

Surveydata worden door overheden gebruikt om beleid te ontwikkelen of te evalueren, maar ook als thermometer van de maatschappij. Tijdens de NPSO-bijeenkomst van 22 maart schetsten vijf onderzoekers

een beeld van de relatie tussen surveys en beleid van concrete onderzoeksprojecten.

Rogier van Kalmthout (TNS-NIPO) trapt af en sprak over het meten van de effectiviteit van overheids-campagnes (zoals Veilig vrijen, NL-alert, NIX18). Het doel van deze effectmetingen is vooral gelegen in evaluatie en verantwoording. De vraag bleef wel of deze campagnes tot feitelijke gedragsverandering leiden.

Peteke Feijten (SCP) behandelde onderzoek naar de hervorming van langdurige zorg. Conclusie was dat de combinatie van een kaderwet en gedecentraliseerd beleid zich moeizaam laat sturen door onderzoek. Ook is er sprake van verminderde aandacht voor de kwaliteit van dataverzameling. Positief was de betere samenwerking tussen gemeenten. Na de koffiepauze ging Ann Carton (Studiedienst van de Vlaamse Regering) in op de implicaties van de

survey Sociaal-Culturele Verschuivingen in Vlaanderen. Hiermee wordt trendmatig de relatie van de Vlaamse burgers met regionale indicatoren in kaart gebracht. Aandacht voor de resultaten vanuit pers en politiek is er zeker, maar het gebruik is vrij selectief.

Gerben Huijgen (I&O) en Susan van den Berge (gemeente Rotterdam) belichtten als laatste sprekers van de bijeenkomst de Veiligheidsmonitor Rotterdam en hoe deze wordt ingezet bij lokaal beleid. Surveydata worden gecombineerd met politiekegegevens om te komen tot wijkprofielen en verschillende indexen. Op wijkniveau ontstaat er op deze wijze een signaleringsinstrument.

Voorzitter Vera Toepoel trok samen met de aanwezigen de conclusie dat er een duidelijke, maar niet erg directe relatie is tussen surveys en beleid. Deze relatie wordt versterkt door meer (harde) gegevens uit andere bronnen te combineren met surveydata en meer te focussen op feitelijk gedrag.

npso.net

OVERNEMEN ARTIKELEN

Wilt u een artikel uit dit blad overnemen? Dat mag altijd, maar vermeld wel de bron (E-data & Research) en de naam van de auteur van het artikel. Neem ook contact op met de hoofdredacteur (zie colofon) om door te geven waar artikelen geplaatst worden.

Eerste *researcher-in-residence* van 2016 aan de slag in de KB

‘Samen een vraagstuk induiken’

Op zoek naar literaire personages en hun onderlinge relaties verruimt literatuurwetenschapper Puck Wildschut een paar maanden de universiteit voor de bibliotheek.

Steven Claeyssens

Bij de afdeling Algemene Cultuurwetenschappen van de Radboud Universiteit Nijmegen werkt

Wildschut aan een proefschrift over *Lolita* en *Pale Fire*, twee romans van de Russisch-Amerikaanse auteur Vladimir Nabokov. “In mijn onderzoek probeer ik narratologische en taalwetenschappelijke analytische invalshoeken met elkaar te

verbinden. Daar komen ook kwantitatieve en computationele methoden bij kijken.” Daar houdt ze het echter niet bij. “Ook werk ik samen met enkele taalwetenschappers aan de Radboud Universiteit aan een onderzoeksproject over sprookjes. We proberen om semantische rollen – dat zijn simpel gezegd personages en andere actanten – en de onderlinge relaties tussen die rollen automatisch te laten bepalen, dus met behulp van een computersoftware.” Wildschut is de eerste literatuurwetenschapper die als *researcher-in-residence* bij de KB aan de slag

tieve teksten te herkennen? En werkt een dergelijke analyse ook adequaat op grotere tekstcorpora, zoals de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL)?” De vraag die voorafgaat aan deze vragen luidt: is het mogelijk bestaande analysesoftware, die veelal voor het Engels ontwikkeld is, werkbaarder te maken voor de Nederlandse taal?

Onmiddellijke toegang

Het zijn taaie vragen die zeker niet in een paar maanden volledig van tafel zullen zijn, maar dat tempert

Wildschuts enthousiasme niet: “De KB biedt onmiddellijke toe-

gang tot veel interessant materiaal om deze kwesties aan te pakken. Bovendien geeft het een grote meerwaarde om samen met een onderzoeksprogrammeur en een collectiespecialist het vraagstuk in te duiken.”



“Ik wil narratologie en taalwetenschap verbinden” foto KB Beeldstudio

De resultaten van het onderzoek van Wildschut komen beschikbaar op het KB Research Lab en blog. Na Wildschut is Frank Harbers (RUG) te gast.

kb.nl/organisatie/onderzoek-expertise/researcher-in-residence

Researcher-in-residence

Sinds 2014 nodigt de Koninklijke Bibliotheek (KB) jonge, beloftevolle onderzoekers uit om in de bibliotheek met behulp van digitale technieken een vraagstuk naar keuze te lijf te gaan. Van de samenwerking profiteert zowel de onderzoeker als de bibliotheek: de *researcher-in-residence* krijgt hulp van programmeurs en collectiespecialisten, de KB leert van de wensen en methoden van de onderzoeker om zo collecties en dienstverlening te kunnen verbeteren.

Bestaande of speciaal ontwikkelde tools die daarbij hun diensten bewijzen, worden opgenomen in het KB Research Lab zodat ook andere onderzoekers ze kunnen gebruiken en eventueel uitbreiden.

Een nieuwe Call for Proposals voor *researchers-in-residence* van 2017 wordt begin juni door de KB gepubliceerd.

JONG TALENT

gaat. “Het onderzoek dat ik samen met het Digital Humanities-team van de KB uitvoer, ligt in het verlengde van het sprookjesproject. We hebben een paar centrale vragen. Is het mogelijk om semantische rollen automatisch in langere narra-

FO-BO-model van datacuratie in de praktijk

RIS Radboud ondersteunt bij duurzaam archiveren

Een team van *data librarians* ondersteunt Nijmeegse onderzoekers bij het duurzaam archiveren en toegankelijk maken van hun onderzoeksdata via het CRIS van de instelling. Marion Wittenberg en Mijke Jetten



del (FO-BO) van datacuratie. DANS vervult de rol van backoffice en draagt zorg voor de duurzame archivering en beschikbaarstelling van onderzoeksdata. Het RIS van de Radboud Universiteit, operationeel sinds begin 2016, staat als frontoffice in direct contact met de onderzoeker en zorgt dat onderzoeksdata goed gedocumenteerd worden en voorbereid zijn op hergebruik.

Integriteit en kwaliteit van onderzoek

Bij de Radboud Universiteit is het bestaande CRIS-systeem, Metis, geschikt gemaakt voor de registratie van de metadata van datasets alsmede de upload van de databestanden zelf. Onderzoekers hoeven hierdoor maar van één interface gebruik te maken om informatie over hun onderzoek, publicaties en data te registreren. Data en publicaties kunnen op deze manier ook direct (in het CRIS) aan elkaar gekoppeld worden. Deze oplossing faciliteert niet alleen de onderzoeker, maar zorgt ook dat op organisatieniveau inzicht en controle mogelijk is met betrekking tot de integriteit en de kwaliteit van het onderzoek. Immers: alle gegevens over het onderzoek worden in één primair CRIS-systeem geregistreerd, dat ook als bron functioneert voor de instellingsrepository.

ris.ru.nl

Het Front Office-Back Office-model van Research Data Netherlands: de datasupporters in de onderzoeksinstelling (het frontoffice) vormen een essentiële schakel tussen onderzoekers en het duurzame archief (het backoffice) bron RDNL

FOCUS

komen metadata en onderzoeksdata automatisch in EASY terecht. Een DANS-datamanager doet een laatste controle en publiceert de dataset. Bovendien krijgt de dataset automatisch een DOI. De data zijn nu duurzaam gearchiveerd én toegankelijk voor hergebruik.

Samenwerking Front Office en Back Office

Met deze werkwijze implementeren Radboud Universiteit en DANS het Front Office-Back Office-mo-

Elke onderzoeker van de Radboud Universiteit heeft een persoonlijke RIS-pagina, waarmee data kunnen worden geregistreerd en geüpload via het Current Research Information System (CRIS) van de instelling. Ook de *full text* van bijbehorende publicaties kan geüpload worden. De link tussen dataset en publicatie, inclusief *full text*, wordt zichtbaar in de Radboud Repository.

Ondersteuning door data-librarian

Wanneer de dataset door de onderzoeker geüpload is, controleert de *data librarian* van de Research Information Services (RIS) de metadata op duidelijkheid en begrijpelijkheid. Daarnaast wordt nagegaan of de data voldoende gedocumenteerd zijn, geen privacygevoelige informatie bevatten en of de bestanden het juiste formaat hebben. Als alles in orde is, stuurt de *data librarian* de dataset naar het online archiveringssysteem EASY van DANS. Via een SWORD-interface

Open access interview-bundel

De discussie over open access wordt vooral gevoerd door bestuurders, bibliothecarissen, overheidsdienaren en uitgevers. Maar wat vinden onderzoekers van open access?

De KNAW liet een bont gezelschap onderzoekers aan het woord: jong en ouder; alfa, bèta en gamma. Het palet dat uit de 21 interviews oprijst is inderdaad bont. Hoewel niemand

het belang van kwalitatief hoogstaand wetenschappelijk onderzoek betwist, blijkt open access geen middel om die kwaliteit te bevorderen, en soms zelfs belemmerend te werken. Lees het resultaat in de interviewbundel *Open boek over open access*, kosteloos te downloaden via de KNAW-website. (HB)

knaaw.nl

Nieuwe privacywet is een feit

Het Europese Parlement heeft, na jaren van intensieve onderhandelingen met de Europese Commissie en de EU-lidstaten, de nieuwe privacywet General Data Protection Regulation (GDPR) aangenomen. Deze nieuwe wet heeft ook voor wetenschappelijk onderzoek gevolgen: het beheer van

persoonsdata van proefpersonen en patiënten wordt strenger. De nieuwe regels gaan over twee jaar in. Op nationaal niveau moeten nog aanvullende regelingen worden ontwikkeld, bijvoorbeeld met betrekking tot *informed consent*. (HB) europarl.europa.eu/news/en

Golden Agents verbindt bronnen

De Nederlandse Gouden Eeuw is wereldberoemd, maar hoe verklaren we de creatieve explosie? We hebben verbrokkelde kennis over de productie en weten weinig over de consumptie in de creatieve industrie. Dit verandert door het project *Golden Agents: creatieve industrie en de creatie van de Nederlandse Gouden Eeuw*. Met hulp van een

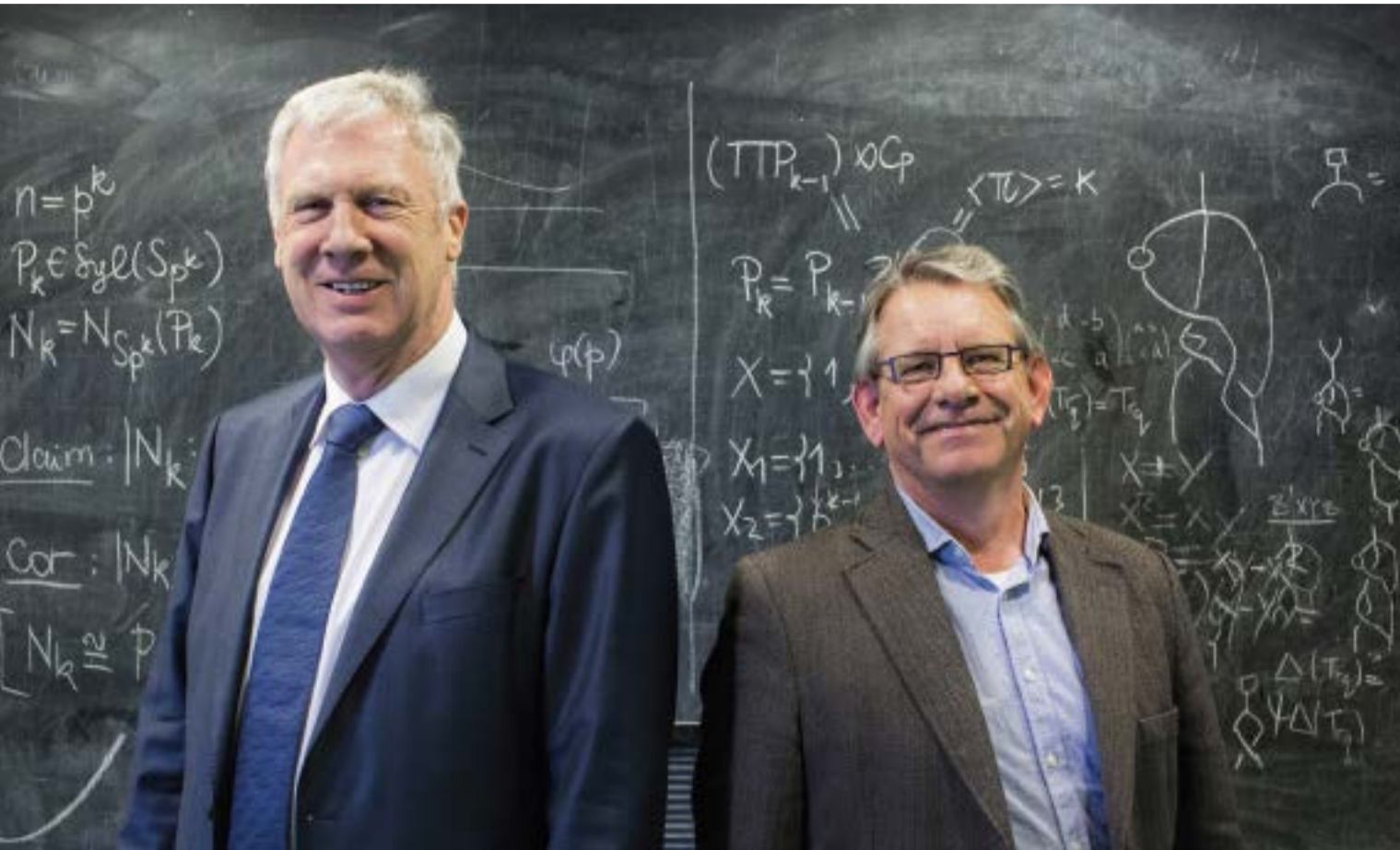
NWO-groot investering verbindt het project bestaande datasets met gegevens over de productie met nieuwe archiefbronnen over de consumptie in de Gouden Eeuw, zoals de notariële akten uit de Gouden Eeuw van het stadsarchief Amsterdam. Interdisciplinair onderzoek naar de creatieve dynamiek van de Gouden Eeuw is zo mogelijk. (MW)



Allegorische voorstelling van Amsterdam als centrum van de wereldhandel credits Pieter Isaacsz., Rijksmuseum, <http://hdl.handle.net/10934/RM0001.COLLECT.358158>

ARIN-NL)

'Ringsakkoord waardig'



Jaap van den Herik (links) en Jan Odijk zien veel positieve ontwikkelingen: "We kunnen terugkijken op mooie projecten met innovatieve resultaten", maar ze delen ook hun zorgen: "We moeten de omslag maken van incidentele naar structurele ondersteuning"

foto Bart van Vliet

Met financiële afspraken over hoe de data die voortkomen uit nieuw onderzoek terecht gaan komen op die centra en beheerd gaan worden."

"Naast een datamanagementplan willen we eigenlijk ook een softwaremanagementplan," vervolgt Odijk. "Het laten draaien en onderhouden van applicaties op servers kost geld. Niet alleen de fysieke opslag, maar meer nog de mensen die dat draaiende moeten houden. Daar moet je geld voor reserveren, want er komt alleen maar meer bij. Het vraagt ook om een life cycle-plan voor data en software: hoe

lang houd je iets direct beschikbaar voor de onderzoeksgemeenschap?"

Wie moet voor zo'n plan zorgen? "De BV Nederland," stelt Van den Herik. "We moeten de omslag maken van incidentele ondersteuning naar structurele ondersteuning. Wij moeten als onderzoekers de regering overtuigen van het belang van onze wetenschappelijke positie in de wereld. We moeten de volksvertegenwoordigers uitleggen dat Nederland anders achter komt te liggen en dat zij daar niet verantwoordelijk voor willen zijn. De enige partij die onze opmerkingen in hun pro-

gramma meeneemt, is D66. Ik wil die andere partijen uitdagen om zich hier ook voor in te zetten. In 2018 zou hierover iets in het regeeringsakkoord moeten staan."

Odijk: "Dat is een ambitieuze doelstelling, maar het zou wel heel goed zijn."

Trots op wat is bereikt

Voorlopig zijn de resultaten van CLARIN-NL en CATCH dus nog volop bruikbaar. Zijn de programmadirecteuren tevreden over het gebruik van de tools en data? "Sinds kort hebben we in Europees verband een monitor die dat bijhoudt, maar daar heb ik geen goed zicht op," vertelt Odijk. "Wel zie ik hoeveel mensen werkelijk onderzoek doen met de CLARIN-infrastructuur. En hoeveel mensen met vragen bij ons komen als er weer een nieuwe aanvraagronde is. Daarnaast zie je ook veel projecten die weer voortkomen uit CLARIN-projecten. En ik zie dat het meer en meer in het onderwijs gebruikt gaat worden. Zo raken ook nieuwe onderzoekers ermee vertrouwd." Van den Herik: "Vergeet inderdaad de mensen niet. Bij CATCH hebben we natuurlijk mooie producten opgeleverd, maar bij elk project waren ook een hoogleraar en mensen van de erfgoedinstelling betrokken. Zo hebben we ruim honderd mensen intensief bij het proces betrokken en die komen ook allemaal weer ergens terecht, met hun kennis. We hebben op die manier veel bereikt en daar kunnen we trots op zijn. En ik ben ook trots op NWO dat ze zo receptief zijn geweest voor onze argumenten voor de noodzaak van deze ontwikkelingen, dat ze die hebben opgenomen en uitgedragen."

catchplus.nl
clarin.nl

Mooie cijfers lezersonderzoek E-data & Research

Lezers van dit blad waarderen de inhoud gemiddeld met een 7,18. De inhoud voldoet voor 96% van de ondervraagden. Mooie cijfers die blijken uit het onlangs gehouden lezersonderzoek.

Marika de Bruijne

E-data & Research verschijnt in het Nederlands en een ruime meerderheid van 85% heeft ook de voorkeur voor deze taal boven het Engels. De meerderheid (57%) van de respondenten zegt het blad per post te ontvangen. 39% ontvangt (ook) de digitale variant of gaat naar de website naar aanleiding van deze e-mail (8%). Voor tweederde was dit al de gewenste manier, 19% gaf tijdens het onderzoek aan het blad liever digitaal te willen ontvangen, waar ze meteen voor konden kiezen. Ongeveer 38% zegt weleens de website edata.nl te bezoeken.

Hoe wordt het gelezen?

Al lezend kiest men de meest interessante onderdelen eruit. Zo'n 41% leest een enkel artikel uit het blad, 17% bladert het blad door en 19% leest veel artikelen. De rubriek 'Sinds kort beschikbaar' wordt het meest gelezen (63%). Ook leest de helft het paginagrote interview.

De overgrote meerderheid van de ondervraagde abonnees vindt dat de thema's die aan bod komen goed verdeeld zijn (84%). Degenen die vinden dat bepaalde thema's te weinig aan bod komen, zouden meer willen lezen over bijvoorbeeld archeologie, sociale wetenschappen, en historische wetenschappen / geschiedenis.

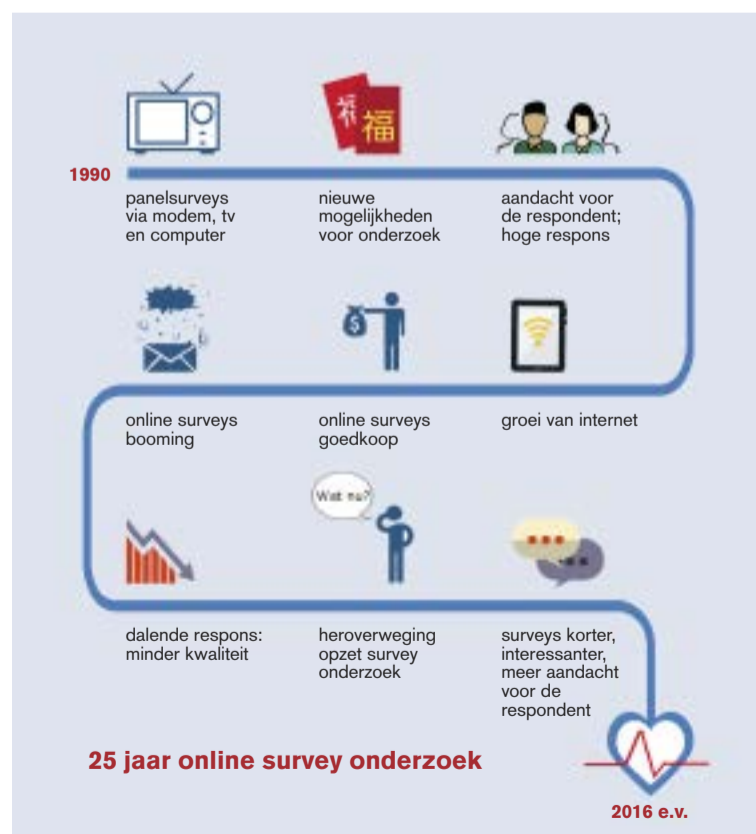
Uw mede-lezers

Het blad bedient meerdere disciplines. Ongeveer een derde (32%) van de lezers heeft een achtergrond in de geesteswetenschappen, 18% in de sociale wetenschappen. De rest van de lezers komt uit andere disciplines. Bijna een kwart (24%) is onderzoeker, 18% werkt in een ondersteunende functie, 17% werkt als hoogleraar of universitair (hoofd)docent en 11% in management-, bestuurlijke of beleidsfuncties.

Meerdere doeleinden

Ruim de helft (52%) zegt informatie uit een artikel weleens te hebben gebruikt. Vaak heeft de informatie de lezer op het idee gebracht voor het (her)gebruiken van data (24%) of heeft de lezer de informatie gebruikt voor onderzoek (23%). Verder wordt de informatie uit het blad gebruikt voor bijvoorbeeld presentaties (13%) of onderwijs (8%).

Alle resultaten uit dit onderzoek zijn indicatief; 525 mensen (7%) hebben de vragenlijst volledig ingevuld. Mocht u niet hebben meegedaan aan het onderzoek maar toch nog een suggestie hebben, mail dan naar edata@dans.knaw.nl edata.nl



AGENDA

7 - 9 juni • Göttingen
Internationale conferentie over elektronisch publiceren Wetenschappers, uitgevers, docenten, librarians, ontwikkelaars en andere belanghebbenden komen samen tijdens dit event en delen ervaringen.
meetings.copernicus.org/elpub2016

8 - 11 juni • St Andrews
CRIS2016
Een door EuroCRIS georganiseerde bijeenkomst met als thema 'Communicating and Measuring Research Responsibility: Profiling, Metrics, Impact, Interoperability'.
eurocris.org

9 - 10 juni • Belval
Digital Humanities Benelux Conference
De jaarlijkse grote DH-conferentie van de lage landen.
dhbenelux.org

13 - 16 juni • Dublin
International Conference on Open Repositories
Een conferentie georganiseerd voor en door de Open Repositories gemeenschap.
or2016.net

6 - 7 juli • Valencia
CARMA
Tijdens deze eerste International Conference on Advanced Research Methods and Analytics (CARMA) kunnen onderzoekers ervaringen delen over het gebruiken van onderzoeksmethoden in de sociale wetenschappen.
carmaconf.org

11 - 16 juli • Kraków
Digital Humanities
Het thema van deze Digital Humanities-conferentie is 'Digital Identities: the Past and the Future'.
dh2016.adho.org

3 - 6 oktober • Bern
iPRES 2016
Deze editie van iPRES wordt georganiseerd door de Zwitserse nationale bibliotheek.
ipr16.organizers-congress.org

13 oktober • Amsterdam
Nationaal eScience Symposium
Het Netherlands eScience Center organiseert voor de vierde keer een nationaal eScience symposium. Het thema is 'Science in a Digital World'.
esciencecenter.nl

9 november • Den Haag
Prijsuitreiking Dataprijs
Wie wint de Nederlandse Dataprijs 2016? Kom naar de prijsuitreiking op 9 november.
researchdata.nl/diensten/dataprijs

Pearl Dykstra, trekker NDSW

'Wat kun jij bijdragen aan andermans werk?'

Pearl Dykstra, hoogleraar Sociologie aan de Erasmus Universiteit Rotterdam, legt hart en ziel in de toekomstige Nationale Data-infrastructuur voor de Sociale Wetenschappen (NDSW). *Jan-Joost de Man*



Pearl Dykstra
foto Angelique van Woerkom

'Geen hokjes maar samenwerken'

Verbinden, samenwerken, over grenzen heen kijken, efficiëntie en vooral ongekende mogelijkheden, dat zijn de sleutelbegrippen voor Pearl Dykstra. Samen met onderzoekers van alle 14 Nederlandse universiteiten zet zij haar schouders onder de investeringsaanvraag in de Nationale Roadmap Grootchalige Infrastructuur voor discipline-overstijgende dataverzamelingen in de sociale wetenschappen. Dykstra: "De focus ligt op de uitdaging: wat kun jij bijdragen aan het werk van een ander? De tijd van afzonderlijke structuren per discipline ligt achter ons. De toekomst is er een van een dracht, niet van aparte hokjes."

Virtueel netwerk

De komende maanden zijn spannend, rond de zomer wordt de uitslag verwacht. Dan komt er antwoord op de vraag of het de initiatiefnemers, met de EUR als indiener van de nationale aanvraag en Pearl Dykstra als wetenschappelijk trekker, gelukt is de handen op elkaar te krijgen voor deze eerste stap op weg

naar een miljoeneninvestering. Dykstra: "Nederlandse survey-onderzoekers behoren sinds de jaren tachtig al tot de wereldtop, zowel methodologisch als inhoudelijk. Maar het zijn afzonderlijke dataverzamelingen: denk aan de grote drie surveys ESS, GGP en SHARE. Doel is om deze dataverzamelingen, én vele andere, te koppelen. Je creëert hiermee een kolossaal virtueel netwerk, met een groeiemodel, dat ons in staat stelt nieuwe vragen te beant-

woorden of oude vragen op een nieuwe manier te onderzoeken." De NDSW biedt zo ongekende mogelijkheden om vraagstukken als migratie, werkgelegenheid, veiligheid en gezondheid aan te pakken, langs empirische weg, vanuit verschillende disciplines, met behulp van actuele bevolkingsgegevens, die in veel grotere mate dan nu voor cross-referenties openstaan.

Met een reuzenstap

Dykstra: "De urgentie én de wenselijkheid van de NDSW is voelbaar. De wetenschap staat in steeds grotere mate in het teken van samenwerken, tussen universiteiten én disciplines. Inmiddels zijn er flinke stappen gezet. Dat geeft energie, bijvoorbeeld om jonge onderzoekers op te leiden." Het breed gedeelde initiële commitment wordt de komende tijd verder vormgegeven. Het gaat daarbij om een nauwe samenwerking tussen de universiteiten (eerste geldstroom), NWO (tweede geldstroom), de ministeries, CBS, DANS en andere kennisinstellingen zoals bijvoorbeeld de planbureaus. Met de financiële armslag die opname in de Nationale Roadmap Grootchalige Infrastructuur met zich mee zal brengen, kunnen de sociale wetenschappen met een reuzenstap een nieuw tijdperk binnentreden.

Text & Data Mining: Europa loopt achter op de VS en Azië

FutureTDM goed op weg

Een omvangrijke internationale alliantie van universiteiten, bibliotheken en bedrijven ijvert voor meer toepassingsmogelijkheden van Text & Data Mining (TDM) in Europa. *Steven Claeysens*

Digitale data zijn essentiële grondstoffen van industrie en wetenschap geworden. Data zijn de virtuele half-fabricaten waarop volwaardige bedrijfstakken en wetenschappelijke disciplines drijven. Data zijn alomtegenwoordig en de technologieën om data te verwerken, worden steeds krachtiger en verfijnder.

Europa kampt echter met een probleem: we benutten dat data-potentieel veel minder effectief dan de VS of Azië. Het FutureTDM-project probeert hier oplossingen voor te vinden.

Drempels slechten

FutureTDM wordt gefinancierd uit het Europese programma voor onderzoek en innovatie, Horizon 2020. Uit Nederland nemen onder meer

de Radboud Universiteit, de Universiteit van Amsterdam en LIBER (de koepel van Europese onderzoeksbibliotheken) deel. FutureTDM beijvert zich voor het slechten van zoveel mogelijk (juridische) drempels die het gebruik van Text & Data Mining-technieken in Europa in de weg staan.

Europese wetgeving

Melanie Imming (LIBER): "De Europese wetgeving zou aangepast moeten worden om het potentieel van TDM in Europa ten volle te kunnen benutten. We analyseren daarom bestaande regelgevingen en beleid in Europa en gaan na hoe deze zich verhouden tot de bredere internationale context."

"We werken ook samen met andere Horizon 2020-TDM-projecten zoals Openminted.eu, dat meer gericht is op de technische aspecten van TDM. En we proberen zoveel mogelijk belanghebbenden te betrekken in de discussie: de overheid en de wetenschap, maar ook bijvoorbeeld uitgevers en KMO's, gezamenlijk probe-

ren we tot oplossingen te komen."

TDM-landschap

"In het eerste jaar van het project brengen we het TDM-landschap in kaart door feedback van alle betrokkenen te verzamelen. Daar zitten we nu middenin. Dat doen we bijvoorbeeld door middel van Knowledge Cafés, zoals eerder dit jaar in het Leidse academiegebouw samen met het Leiden Centre of Data Science, maar ook met enquêtes en interviews. In het tweede jaar brengen we de vergaarde kennis en praktijkvoorbeelden bijeen op een Collaborative Knowledge Base en vertalen we dit naar beleidsadviezen en praktische richtlijnen. Uiteraard organiseren we ook een workshop in Brussel om ons werk daar onder de aandacht te brengen."

Imming: "Alle feedback is welkom. Iedereen kan bijdragen aan FutureTDM: blog, tweet, neem deel aan het Knowledge Café, vul onze vragenlijsten in. Het helpt allemaal."

FutureTDM.eu

KORT

Nieuwe boost voor LISS panel tot 2017

Door een gezamenlijke investering van NWO, Tilburg University en CentERdata kan de dataverzameling in het LISS panel worden voortgezet. De financiering dekt de vaste lasten van de infrastructuur in 2016 en 2017. In het LISS panel kunnen academische onderzoekers in Nederland en daarbuiten hun vragenlijsten of experimenten afnemen. De data verkregen uit deze studies worden gepubliceerd opdat ook andere onderzoekers er gebruik van kunnen maken. Momenteel kent het LISS Data Archief meer dan 1.600 gebruikers. Voor de lange termijn wil CentERdata het LISS panel onderbrengen in een nationale data-infrastructuur voor de sociale wetenschappen. Hierbij zal CentERdata samenwerken met universiteiten, het CBS en NWO Maatschappij- en Gedragswetenschappen. (MdB) lissdata.nl

DANS-subsidie voor kleine dataprojecten

Tot 1 oktober is het mogelijk om een subsidie bij DANS aan te vragen voor kleine dataprojecten (KDP's). Deze zogenoemde KDP-subsidie wordt verstrekt aan haalbaarheidsstudies, verkenningen op het gebied van data-archivering en projecten waarbinnen belangrijke datasets worden beschreven en toegankelijk worden gemaakt. De subsidies zijn vooral bedoeld voor deelgebieden waar nog niet tot nauwelijks aan data-archivering wordt gedaan, en waarbij de subsidie van DANS een stimulans kan zijn om dit juist wel te doen. (HB) dans.knaw.nl

450.000 open access publicaties in NARCIS

In de nationale wetenschapsportal NARCIS zijn nu meer dan 450.000 open access-publicaties toegankelijk. Deze publicaties zijn afkomstig uit de repositories van de universiteiten, KNAW, NWO, HBO (lectoren) en andere wetenschappelijke instellingen, zoals Naturalis Biodiversity Centre, RIVM en TNO. De open access-publicaties zijn voor een groot deel wetenschappelijke artikelen (187.000), rapporten (67.000) en proefschriften (65.000). Naast de open access-publicaties wordt in NARCIS een overzicht gegeven van 785.000 publicaties die voor een groot deel van de academische wereld wel *full text* toegankelijk zijn via de universitaire bibliotheken. (HB) narcis.nl

Eep Talstra Centre for Bible and Computer haalt zichzelf weer in

Databank voor bijbelwetenschappen

Van een enkele computerondersteunde analyse tot het online kunnen raadplegen van rijk geannoteerde Hebreeuwse teksten: het Eep Talstra Centre for Bible and Computer timmert al bijna 50 jaar aan de digitale weg. *Wido van Peursen*

In de 70'er jaren van de vorige eeuw experimenteerden enkele pioniers aan de Vrije Universiteit met de computerondersteunde analyse van de Hebreeuwse Bijbel onder de paraplu van de 'Werkgroep Informatica Vrije Universiteit (WIVU)'. De

WIVU stond onder leiding van Eep Talstra, bijbelwetenschapper met een bijzondere interesse in 'alfainformatica', zoals dat toen genoemd werd. Sindsdien is er een prachtig product opgebouwd: een rijk geannoteerde databank met taalkundige informatie over morfemen, woorden, woordgroepen, zinnen en zinsverbindingen. Talstra ging in 2012 met emeritaat, de naam veranderde in het Eep Talstra Centre for Bible and Computer (ETCBC), om daarmee het pionierswerk te eren.

Tijd ver vooruit

De WIVU was zijn tijd ver vooruit. Op den duur kreeg zij echter te kampen met de wet van de remmende voorsprong. Zo was de databank ontstaan lang voordat Unicode of TEI het licht zagen. Voor de weergave van het Hebreeuws en de taal-

kundige codering werden alleen ASCII-teken gebruikt. De balans: een geweldig waardevol product, maar wel enigszins idiosyncratisch en ontoegankelijk. Vernieuwingslagen hebben geleid tot een grotere toegankelijkheid en gebruikersvriendelijkheid van de databank. Zo heeft een Deense collega op basis van de databank de Bible Online Learner gebouwd, die op zeer uiteenlopende plaatsen in de wereld wordt ingezet voor het onderwijs in het Bijbels Hebreeuws. En bijvoorbeeld in het CLARIN-NL-project SHEBANQ (System for HEBrew Text: ANnotations for Queries and Markup) heeft het ETCBC samen met DANS een webapplicatie en een *query saver* rondom de databank gebouwd. Het is nu voor iedereen mogelijk om de Hebreeuwse tekst online te raadplegen, de taalkundige informatie te

Eep Talstra Centre for Bible and Computer

Het Eep Talstra Centre for Bible and Computer (ETCBC) beheert de WIVU-databank. Deze databank bevat de complete tekst van de Hebreeuwse Bijbel coderingen op woord-, woordgroep-, zins- en tekstniveau. De databank wordt voor diverse doeleinden gebruikt, onder andere voor onderzoeksdoeleinden bij DANS.

zien, zoekopdrachten uit te voeren en op te slaan en hiernaar te verwijzen met een URI en de zoekopdrachten van anderen te raadplegen. In het SHEBANQ-project is de databank geconverteerd naar het LAF-format (Linguistic Annotation Framework), een ISO standaard voor taalkundige annotaties. De uitvoerder

van het SHEBANQ-project, Dirk Roorda, heeft LAF fabric gebouwd: een Python-tool om Python-notebooks te draaien en zeer complexe analyses van de Hebreeuwse databank uit te voeren.

Honderden gebruikers

Immiddels heeft de SHEBANQ-website honderden geregistreerde gebruikers, wordt de dataset gebruikt in projecten van gerenommeerde onderzoekers (bijvoorbeeld het Tiberias-project van de Bar-Ilan University), heeft zich een community gevormd van promovendi en andere onderzoekers die LAF fabric gebruiken en wordt SHEBANQ in binnen- en buitenland ingezet in het onderwijs. Vanaf september 2016 verzorgt het ETCBC een MA-track Biblical Studies and Digital Humanities. bit.ly/ETCBC

SINDS KORT BESCHIKBAAR

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

CentERdata

• Inschattingen van risico's en het effect op investeringsgedrag in aandelen

In 2013 en 2014 is een onderzoek over persoonlijke inschattingen van risico's en het effect op investeringsgedrag in aandelen af-



foto ANP

genomen in het LISS panel door onderzoekers B. Enke en H.M. Gaudecker, University of Bonn. De respondenten werden onder andere uitgenodigd om een inschatting te doen van de waarde van bepaalde aandelen over een half jaar. Na een jaar werden de respondenten op de hoogte gebracht van de actuele AEX-stand en kregen zij de kans om hun antwoord van het voorgaande jaar aan te passen. De datasets 'Subjective Beliefs and Individual Risk Preferences in a Structural Model of Portfolio Choice - 4 parts' zijn beschikbaar via het LISS Data Archive van CentERdata.

lissdata.nl/dataarchive

Ook sinds kort beschikbaar: Studies LISS panel

- CentERdata, november 2015 - december 2015, Personality - Wave 8
- Tooren, M. van den; Rutte, C.G., juli 2012, Type D personality and job resources - Wave 2
- CentERdata, december 2015 - januari 2016, Politics and Values - Wave 8
- Rosema, M.; Huddy, L., augustus 2012, Attachment to political parties
- Choi, S.; Müller, W.; Kariv, S.; Silverman, D., mei 2014 - juni 2014, Consistency and heterogeneity of individual behavior under uncertainty with a special attention to the elderly



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

DANS

• Data landelijk artrose-onderzoek in EASY

Tot eind 2015 coördineerde het Universitair Medisch Centrum Utrecht een landelijk artrose-onderzoek, Cohort Heup en Cohort Knie (CHECK). Tien jaar lang werden 1.000 individuen met vroege symptomen van artrose onderzocht in ziekenhuizen door het hele land. De uitval van deelnemers tijdens de studie was uitermate laag. Zowel CHECK als projectfinancier Reumafonds besloten de resultaten van deze cohortstudie via DANS beschikbaar te stellen. De vier datasets betreffen de baseline-collectie (T0) en follow up-materiaal uit de 10 jaar in aanvullende T-series. De datasets bevatten zowel longitudinale enquêtegegevens over de gezondheidssituatie en -beleving van de deelnemers, alsook klinische data en duizenden röntgenopnames. CHECK en DANS hebben bijzonder veel aandacht besteed aan het anonimiseren van gegevens en beeldmateriaal. Een dataset is open access (for registered users), de anderen zijn beschikbaar op aanvraag.

DOI: [10.17026/dans-252-qw2n](https://doi.org/10.17026/dans-252-qw2n)

Ook sinds kort beschikbaar:

- Allema, Dr. A.B. (Wageningen University) (2010): Dispersal and interface behaviour of carabid beetles in a landscape mosaic: Field study plus model selection. DOI: [10.17026/dans-zd5-uufj](https://doi.org/10.17026/dans-zd5-uufj)
- Centraal Bureau voor de Statistiek (CBS) (2014): Gezondheidsenquête 2014. DOI: [10.17026/dans-xcm-u69z](https://doi.org/10.17026/dans-xcm-u69z)
- Faun, drs. H. (Panteia) (2016): Archivering en digitalisering door cultuurproducerende instellingen. DOI: [10.17026/dans-xtp-djgt](https://doi.org/10.17026/dans-xtp-djgt)
- Hendriks, Dr. M. (NIVEL) (2013): Healthcare professionals' views on feedback of a patient safety culture assessment. DOI: [10.17026/dans-2ae-h7kc](https://doi.org/10.17026/dans-2ae-h7kc)
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK); Centraal Bureau



Röntgenfoto uit het onderzoek credits CHECK

voor de Statistiek (CBS) (2016): WoON2015: release 1.0 - WoonOnderzoek Nederland 2015. DOI: [10.17026/dans-xbt-qc5c](https://doi.org/10.17026/dans-xbt-qc5c)

- Putmans, drs. J.L.C. (2016): Hendrik van Veldeke / Heinrich von Veldeke - Eneïde - Eneasroman - Transkriptionen der Handschriften. DOI: [10.17026/dans-xnu-ukak](https://doi.org/10.17026/dans-xnu-ukak)
- Rieffe, E.C. (Gemeente Den Haag, Dienst Stadsbeheer, Afdeling Archeologie en Natuur- en Milieueducatie) (2016): Laan van Poot - HALO-terrein gemeente Den Haag. DOI: [10.17026/dans-xf9-nybw](https://doi.org/10.17026/dans-xf9-nybw)
- Stoepker, Drs. H. (ARCHEOCOACH) (2016): De middeleeuwse parochiekerk van Brunssum (Limburg). DOI: [10.17026/dans-xqt-uzzc](https://doi.org/10.17026/dans-xqt-uzzc)
- Stokkel, P.J.A. (Gemeente Den Haag, Dienst Stadsbeheer, Afdeling Archeologie) (2012): Steynhof gemeente Den Haag. DOI: [10.17026/dans-xf8-u8u8](https://doi.org/10.17026/dans-xf8-u8u8)
- Wolters, Mr. Dr. P.T.J. (Onderzoekcentrum Onderneming & Recht, Radboud Universiteit) (2016): De invloed van de verhouding tussen schade en boete op de matiging van boetebedingen. Een empirische analyse. DOI: [10.17026/dans-xxz-q9ke](https://doi.org/10.17026/dans-xxz-q9ke)



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

Huygens ING

• BoschDoc: bijna 1.000 documenten over Jheronimus Bosch

BoschDoc ontsluit bijna 1.000 schriftelijke documenten over Jheronimus Bosch, zijn naaste familie, zijn leven en zijn werk. De databank bevat facsimiles, transcripties en vertalingen van handgeschreven en enkele gedrukte teksten tot 1.800. Elk document is voorzien van een informerende tekst over de context, de bron en de relatie tot andere documenten in de databank. De databank is gebouwd binnen de digitale infrastructuur eLaborate van Huygens ING en mede mogelijk gemaakt door de Radboud Universiteit en het Stadsarchief 's-Hertogenbosch.

boschdoc.huygens.knaw.nl



Detail uit 'Wapenboek van Illustre lieve vrouwe broederschap in Den Bosch', met rechtsonder het wapen van 'Hieronymus Aquensis alias Bosch' credits Brabants Historisch Informatie Centrum

Ook sinds kort beschikbaar:

- Clusius: <http://clusiuscorrespondence.huygens.knaw.nl/>
- Ystroom: <http://deystroom.huygens.knaw.nl/edition/>



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

Surveydata kosteloos online beschikbaar

PROFILES: onderzoeksdata over leven met kanker

Hoe gaat het leven verder na een kankerdiagnose? Patiëntenvolg-systeem PROFILES stelt data van een longitudinaal onderzoek onder (ex-)kankerpatiënten beschikbaar.

Marika de Bruijne

PROFILES (PROFIELstudie in het Nederlands) is een patiëntenvolgsysteem waarbij (ex-)kankerpatiënten vragen beantwoorden over de kwaliteit van hun leven. Nicole Horevoorts, onderzoekscoördinator PROFILES, vertelt: “We volgen personen vanaf vlak na de kankerdiagnose tot soms wel 20 jaar erna. Via de PROFILES-registratie kunnen we de antwoorden die patiënten geven, koppelen aan klinische gegevens en zo meer inzicht krijgen in wat er met een patiënt gebeurt na de diagnose en tijdens de behandeling van kanker.”

Waardevol data delen

De surveydata zijn kosteloos online beschikbaar voor wetenschappelijke onderzoekers en bevatten momenteel 23 datasets van 2004 tot en met 2014. Om de datapublicatie te bewerkstelligen, moest eerst een aan-



Het PROFILES-onderzoeksteam foto Merlin Daleman

tal zaken worden opgelost. “Het belangrijkste is het beschermen van de privacy van de patiënten en hoe je de data anoniem maakt. Tevens moest er een keuze gemaakt worden over de manier waarop je data vrijgeeft. Voor veel van onze dilemma’s zijn we in overleg gegaan met CentERdata. Zij hebben het systeem gebouwd en hebben veel ervaring op het gebied van het delen van data,” vertelt Horevoorts.

Een belangrijke keuze was ook hoe de data worden overgedragen. “Je wilt dat de data op een juiste wijze geïnterpreteerd worden, maar ook dat onderzoekers zich vrij voelen om de data te gebruiken,” legt Horevoorts uit. “We hebben gekozen voor een tussenoplossing. We vragen wetenschappers om een formulier in te vullen met hun gegevens

en de onderzoeksvraag. Wanneer iemand anders al bezig blijkt te zijn met een soortgelijke onderzoeksvraag, brengen we de personen in contact met elkaar. Tevens vragen we bij publicaties om een PROFILES-onderzoeker te betrekken, zodat die kan helpen bij juiste interpretatie en keuze voor analysetechnieken.”

Compleet profiel

“Met de huidige data weten we dat globaal een op de drie mensen (lang) na kanker en de behandeling nog allerlei klachten heeft. Deze bevindingen hebben zelfs geleid tot aanpassingen in bepaalde kankerbehandelingen voor mensen met lymfeklier- of baarmoederkanker. We weten echter niet goed waarom de ene patiënt bepaalde klachten wel krijgt en de andere patiënt met

PROFILES
PROFILES (PROFIELstudie) is wetenschappelijk onderzoek naar kwaliteit van leven bij mensen die kanker hebben of hebben gehad. Hiervoor werken onderzoekers van het Integraal Kankercentrum Nederland (IKNL) en Tilburg University (TiU) samen met medisch specialisten van ziekenhuizen in heel Nederland. Kijk voor meer informatie op profilesregistry.nl of neem contact op via info@profielstudie.nl.

exact dezelfde kenmerken nauwelijks. Om dat te kunnen verklaren, gaan we nog meer zaken meten. Hiervoor heeft de PROFILES-onderzoeksgroep (hoofdonderzoeker, Lonneke van de Poll-Franse) recentelijk een nieuwe NWO-investeringsubsidie ontvangen. We gaan bloed afnemen, de leefstijl van mensen bekijken door middel van voedings-apps en internetweegschalen, lichaamsfuncties meten door middel van zogeheten *health patches* (slimme pleisters) en cognitieve testen doen,” vertelt Horevoorts. “Met deze aanvullende data kunnen we een completer profiel van een patiënt maken. Dat zal ons helpen om te begrijpen waarom de ene patiënt wel klachten ontwikkelt en de andere niet.”

profilesregistry.nl

GELEZEN

KWALON special issue: Qualitative research in the digital humanities
Reinoud Bosch

KWALON is het tijdschrift van het gelijknamige platform voor kwalitatief onderzoek. Het tijdschrift biedt ruimte voor de presentatie van en discussie over de ontwikkeling van kwalitatieve sociaalwetenschappelijke onderzoeksmethoden en de toepassing daarvan in alle mogelijke varianten en op alle relevante onderzoeksterreinen. In deze *special* omschrijven José van Dijck, Piek Vossen, Reinoud Bosch, Ruben Verborgh, Susanne Friese, René van Horik en Julia Noordegraaf mogelijkheden voor kwalitatief onderzoek in de digitale geesteswetenschappen. Voorbeelden zijn: breder en complexer onderzoek, nabootsen van menselijke interpretaties, coderen van stappen van kwalitatief onderzoek en ondersteuning vanuit *trusted digital repositories*. De auteurs zien ook uitdagingen. Methoden, annotaties en algoritmes moeten worden verbeterd, verhelderd en geëxpliciteerd, en er is behoefte aan een respectvolle multidisciplinaire samenwerking en experimenten met grote onderzoeksvragen. Het interpretatieproces moet expliciet en controleerbaar worden gemaakt. De auteurs stellen: kwalitatief onderzoek in de digitale geesteswetenschappen gaat een mooie toekomst tegemoet als op de juiste manier wordt gereageerd op bestaande uitdagingen.

tijdschriftkwalon.nl

COLUMN

Een overdaad aan bronnen

Onlangs gaf ik een korte lezing voor historisch letterkundigen. Op een gegeven moment kwamen mijn bronnen ter sprake. Ik noemde enkele websites waar ik die bronnen had gevonden, waaronder Delpher. Zonder twijfel zullen alle lezers van E-data bekend zijn met Delpher, een website die toegang geeft tot onder meer 11 miljoen krantenpagina’s, maar bij die lezing vroeg ik voor de zekerheid: ‘Wie is er niet met Delpher bekend?’

Tot mijn opluchting stak slechts één aanwezige zijn hand op. Toen ik vervolgens wees op de mogelijkheid om in Delpher woorden in elkaars nabijheid te zoeken (met het trucje zoekterm PROX), zag ik enkele toehoorders snel een aantekening maken. Kennelijk waren ze niet op de hoogte van deze geavanceerde zoekmogelijkheid, waar ik vrijwel dagelijks enorm veel baat bij heb.

Alles bij elkaar was het een goede score: slechts één toehoorder die Delpher niet kende en

slechts enkele toehoorders, zo leek het althans, die niet op de hoogte waren van een van de geavanceerde zoekmogelijkheden in deze belangrijke bronnenverzameling.

Ik heb slechtere scores gezien. Sterker nog, onlangs zei iemand van het NIOD (Instituut voor Oorlogs-, Holocaust- en Genocidestudies) tegen mij: wij krijgen hier soms professionele historici over de vloer die niet weten dat vrijwel alle oorlogskranten online te doorzoeken zijn. Tot zo’n vijftien jaar geleden diende je als onderzoeker bepaalde standaardwerken te kennen. Zo kon je niks over de Tweede Wereldoorlog schrijven zonder ‘Lou de Jong’ te raadplegen.

Standaardwerken zijn nog steeds belangrijk, maar het belang van grote datacollecties op internet is zo sterk toegenomen dat je deskundigheid als onderzoeker in hoge mate wordt bepaald door je kennis van dergelijke bronnencollecties.

Ik krijg het daar weleens benauwd van. Ja, ik durf te beweren dat ik sommige digitale collecties, zoals



foto Leo van Velzen

Delpher, goed ken en dat ik optimaal gebruikmaak van de zoekmogelijkheden. Maar welke sites mis ik? Net als de meeste mensen heb ik de neiging om steeds naar dezelfde sites terug te keren. Maar het zijn er zo verschrikkelijk veel, er komt nog steeds zoveel bij, dat ook ik zonder twijfel echt belangrijke sites over het hoofd zie.

Zo is het bijvoorbeeld helemaal niet verstandig om voor oude kranten alleen in Delpher te kijken. Dat is weliswaar de grootste en best ontsloten collectie van oude Nederlandse kranten, maar lang niet alle krantentitels staan erin. Zelf maak ik ook geregeld gebruik van de ‘Krantenbank Zeeland’. En van de krantencollecties die je kunt doorzoeken via

archieven.nl, via digibron.nl (echt een prachtige collectie!), via het Utrechts Archief en via het Archief Eemland. Maar dat blijft slechts een kleine greep, want er is veel en veel meer (zie voor een goed overzicht: kranten-historisch.startpagina.nl).

Mijn ervaring is dat je zelfs goede sites, met enorme bronnencollecties, helemaal kunt vergeten als je ze niet geregeld gebruikt. Ik vind dat onuitstaanbaar, maar het is wel de praktijk. Natuurlijk maak ook ik gebruik van lijstjes met ‘favorieten’ of ‘bladwijzers’, maar die worden al snel zo lang dat ik alleen de aller-aller-belangrijkste gebruik.

Het is een onvoorstelbare luxe om zoveel nuttige bronnen online te kunnen raadplegen, maar het vergt veel tijd en discipline om ze volop en op een gestructureerde manier te benutten.

Ewoud Sanders

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