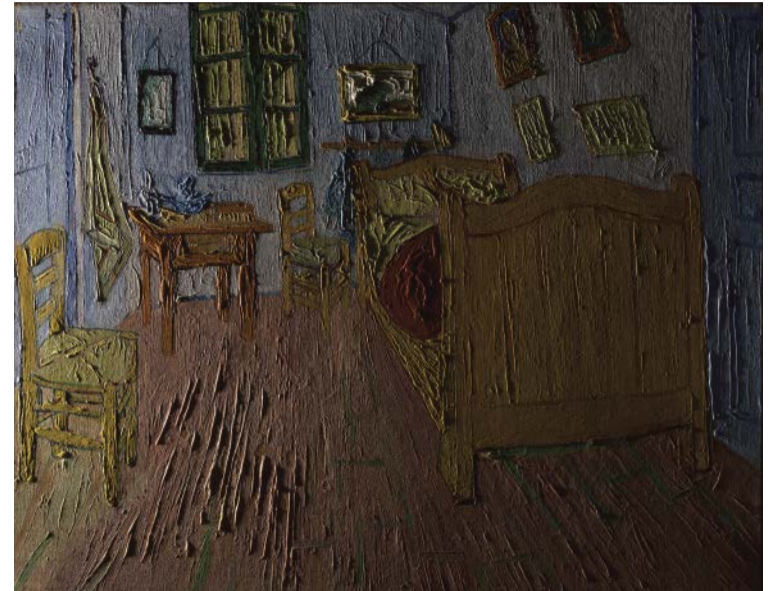


Wetenschappelijke kennis centraal ontsloten

vangoghworldwide.org faciliteert onderzoek



Van De Slaapkamer (links) zijn via vangoghworldwide.org naast de catalogusgegevens van het schilderij ook een uitgebreid tentoonstellingsoverzicht, verwijzingen naar correspondentie en technische data en foto's te vinden, zoals het werk in strijklucht (rechts).

Credits: Van Gogh Museum, Amsterdam (Vincent van Gogh Stichting)

Dankzij een intensieve samenwerking heeft iedereen online toegang tot onderzoeks-informatie over kunstwerken van de bekendste negentiende-eeuwse Nederlandse schilder Vincent Van Gogh. Maarten Heerlien

Het oeuvre van Vincent van Gogh telt zo'n 2000 werken, verspreid over museale en particuliere collecties wereldwijd. Om het onderzoek naar zijn schilderijen en werken op papier beter te faciliteren, lanceerden het RKD (Nederlands Instituut voor Kunstgeschiedenis), het Van Gogh Museum en het Kröller-Müller Museum afgelopen november vangoghworldwide.org, een online platform met wetenschappelijke kennis en informatie over de werken van de schilder.

De behoefte aan één centrale plaats voor deze kennis is groot, aldus Marjon van Schendel, pro-

jectleider van Van Gogh Worldwide. "Er wordt nog steeds veel onderzoek gedaan naar Vincent van Gogh en er zijn regelmatig nieuwe ontdekkingen met betrekking tot diens werk, denk aan de recente vondst van de locatie waar hij zijn laatste werk *Boomwortels* schilderde. Dit nieuwe platform is een inspiratiebron voor verder kunsthistorisch en materiaaltechnisch onderzoek voor kunsthistorici, handelaren, verzamelaars en andere geïnteresseerden."

Centrale rol Linked Open Data

Qua doelstelling is het platform vergelijkbaar met de al sinds 2008 bestaande Rembrandt Database, eveneens beheerd door het RKD. Technisch zit Van Gogh Worldwide anders in elkaar. Van Schendel: "Samen met Spinqe en het Netwerk Digitaal Erfgoed is een infrastructuur ontwikkeld die voldoet aan de Digitaal Erfgoed Referentie Architectuur (DERA), met een centrale rol voor Linked Open Data. Partners stellen hun data beschikbaar conform het Linked Art-model. Binnen het platform worden de collecties verbonden

en verrijkt, bijvoorbeeld met koppelingen naar de Van Gogh brieven (vangoghletters.org) en naar data die het Rijks erfgoedlaboratorium van de Rijksdienst voor het Cultureel Erfgoed beschikbaar stelt. De bronsystemen van de deelnemende instellingen blijven in deze architectuur verantwoordelijk voor de inhoud van de data."

Online raadplegen

De eerste versie van Van Gogh Worldwide werd gerealiseerd met steun van het Mondriaan Fonds en de Vincent van Gogh Stichting en bevat werken van de kunstenaar in Nederlandse collecties, zo'n 1100 in totaal. Van deze werken zijn de objectgegevens, briefverwijzingen, herkomst, tentoonstellings- en literatuurgegevens en materiaaltechnische informatie te raadplegen. Afbeeldingen worden met behulp van de iiif-standaard beschikbaar gesteld. De partners hebben de ambitie om vanaf dit jaar ook werken uit buitenlandse publieke en particuliere verzamelingen toe te voegen.

vangoghworldwide.org

Op weg naar één European Open Science Cloud (EOSC)

FAIR in de praktijk: FAIRsFAIR

FAIRsFAIR is een driejarig Horizon 2020-project voor de implementatie van de European Open Science Cloud. Het project loopt tot maart 2022, maar er zijn al handvatten en tools beschikbaar.

Gerard Coen

FAIRsFAIR staat voor *Fostering Fair Data Practices in Europe*. Het is een samenwerking tussen 22 partners, waaronder 5 Nederlandse: SURF, Sparc Europe, Dutch Techcentre for Life Sciences, Universiteit van Amsterdam en projectcoördinator DANS. Samen richten zij zich op praktische oplossingen om onderzoeksgegevens te laten voldoen aan

de FAIR-principes. Het project loopt nog ruim een jaar, maar er zijn al resultaten beschikbaar.

Beschikbare handvatten

Eén van de resultaten is *F-UJI*, een tool voor het automatisch beoordelen van de FAIRness van onderzoeksdata. Een tweede tool is *FAIR-Aware*, een zelfevaluatie om onderzoekers de FAIR-principes beter te leren begrijpen om daarmee de waarde en impact van hun onderzoeksdata te verhogen. Daarnaast heeft het project een *Synchronisation Force* om de toepassing van de FAIR-principes te vergroten. Zo worden bijvoorbeeld workshops georganiseerd met andere Europese

projecten, EOSC Working Groups en onderzoeksinfrastructuren. Ook begeleidt FAIRsFAIR digitale archieven bij het behalen van de Core-TrustSeal-certificering.

Groeiende community

Naast deze concrete tools groeit ook de Nederlandse gemeenschap van organisaties en professionals die de FAIR-principes in de praktijk toepassen. Zo wordt bijvoorbeeld het implementeren van de principes in het Nationaal Open Science Plan genoemd, hebben onderzoeksfinanciers NWO en ZonMw richtlijnen voor FAIR-datamanagement en werkte de LCRDM-taakgroep *Fair enabling organisations* aan het defi-

niëren van stelregels waarmee een organisatie kan worden beoordeeld op de mate van het faciliteren van de FAIR-principes.

Ook de komende maanden werken de partners verder aan het realiseren van de brede acceptatie van de FAIR-methoden binnen en buiten de *European Open Science Cloud*. fairsfair.eu

**Volg
E-data & Research
ook via Twitter:
@eDataResearch**



**E-DATA &
RESEARCH**

Jaargang 15 | 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, KNAW Humanities Cluster, de Koninklijke Bibliotheek, ODISSEI en het Rijksmuseum.

INHOUD

2
KansenKaart laat zien of afkomst uitmaakt

3
Henk Wals, directeur DANS: "Samen onderzoekers faciliteren"

4
Ook deze datasets zijn sinds kort beschikbaar

5

Van Ossenbruggen pleit voor meer samenwerking

6
Winnaars RDNL Dataprijs 2020 op een rij

7
Tijd besparen met online CLARIAH-tool SASTA

8
Gastcolumnist Klous: "Laat mens én machine aansluiten op behoeften"

E-data & Research liever (tijdelijk) op een ander adres ontvangen? Geef wijzigingen door via edata@dans.knaw.nl.



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

GEHOORD & BIJGEWOOND

CHR: Computatieve methode in de geesteswetenschappen**Folger Karsdorp**

In de afgelopen jaren is het gebruik van computationele, statistische en mathematische methoden sterk toegenomen in de geesteswetenschappen en aanverwante sociale wetenschappen.

Ondanks de onmiskenbare groei van dit onderzoeksgebied, bestond er geen geschikte plaats waar onderzoekers hun computationele werk konden publiceren en presenteren. Om dit gat te vullen, vond van 18 tot en met 20 november 2020 de eerste editie van de workshop *Computational Humanities Research* (CHR) plaats. De workshop, georganiseerd door het KNAW Humanities Cluster in samenwerking met het Alan Turing Instituut, zou in eerste instantie plaatsvinden in Amsterdam, maar door de pandemie werd besloten om er een virtueel evenement van te maken. De virtuele vorm had als voordeel dat de workshop toegankelijk was voor iedereen met een internetverbinding, en zodoende door meer dan 250 deelnemers uit heel de wereld bijgewoond kon worden.

Er waren twee keynote-lezingen. De

eerste werd gegeven door Dong Nguyen, universitair docent aan de Universiteit Utrecht, die sprak over de problemen en moeilijkheden van het toepassen van taaltechnologie op geesteswetenschappelijke data. De tweede keynote werd gegeven door Anne Kandler, senior onderzoeker aan het Max Planck Instituut voor Evolutionaire Antropologie in Leipzig. Zij sprak over statistische en mathematische methoden om de mechanismen die ten grondslag liggen aan culturele verandering beter te kunnen begrijpen.

Een groot verschil met andere (digitale) geesteswetenschappelijke conferenties was dat CHR alleen volledig uitgewerkte artikelen (10 tot 20 pagina's) accepteerde. Uit de 50 inzendingen werden 29 artikelen geselecteerd om gepresenteerd te worden. Deze artikelen passen computationele methoden toe op onderwerpen die voortkomen uit een breed scala aan geesteswetenschappelijke disciplines, zoals geschiedenis, musicologie, literatuurwetenschap, taalkunde en culturele evolutie. De geaccepteerde artikelen zijn open access beschikbaar via de online repository van de CEUR 'workshop proceedings':

ceur-ws.org/Vol-2723/

Community Conference ODISSEI 2020**Suze Zijlstra**

Sinds ODISSEI in het voorjaar van 2020 NWO Roadmap financiering heeft ontvangen, is het project goed op stoom gekomen. Op 24 november hield het consortium een *Community Conference*, die in het teken stond van de mogelijkheden en uitdagingen voor de infrastructuur de komende jaren. De keynote werd verzorgd door Chantal Kemner (UU), zij vertelde over drie centrale uitdagingen van het YOUTH (Youth Of Utrecht) cohort: hoe het omgaat met bescherming van privacy, hoe het datakwaliteit en reproduceerbaarheid waarborgt en hoe het data combineert en verrijkt. Dit nodigde de vraag uit welke overkoepelende rol ODISSEI in de sociale wetenschappen op deze gebieden kan gaan spelen.

Hierna presenteerde Daniel Oberski (UU) het *ODISSEI Social Data Science Team* (SoDa), opgericht om onderzoekers te helpen met computationele en statistische expertise bij het uitvoeren van onderzoek met vernieuwende methoden en data. Ook deelde Jan van der Laan (CBS) de nieuwe *CBS Network File* en de manier waarop deze gebruikt kan

worden in onderzoek met CBS microdata.

De volgende sessie draaide om vernieuwend onderzoek dat ondersteund werd door ODISSEI. Zo presenteerde Mara A. Yerkes (UU) haar onderzoek naar genderongelijkheid tijdens de coronacrisis, waarvoor zij een *LISS Corona Grant* heeft ontvangen. Trond Husby (PBL) deelde voorlopige resultaten van zijn onderzoek naar de lange-termijn gevolgen van het Vinex-beleid op woon-werk afstand en woonkwaliteit. Voor dit onderzoek maakte hij gebruik van de ODISSEI Secure Supercomputer, waardoor hij zijn analyses met een veel grotere snelheid in een streng beveiligde omgeving kon uitvoeren.

Een paneldiscussie over de uitdagingen waarvoor sociale wetenschappers door de coronacrisis staan, sloot de middag af. Uit de discussie werd duidelijk dat het niet alleen belangrijk was dat relevant onderzoek werd gedaan, maar dat deze ook goed toegankelijk moet zijn voor beleidsmakers. Ook hier kan ODISSEI in de toekomst een rol spelen.

odissei-data.nl

Winnaars LISS Call bekend

Onderzoekers werkzaam bij ODISSEI-deelnemers konden tot eind september 2020 via de LISS Call hun onderzoeksvoorstel indienen om data te verzamelen via het LISS panel. Inmiddels zijn de winnaars bekend.

In 2021 zal door deze winnaars onderzoek gedaan worden naar onderwerpen zoals: tijdsbesteding en uitgaven van huishoudens voor en na Covid-19, eetgedrag in afleidende situaties, gebruik van kinderopvang en wereldburgerschap. Het volledige overzicht van alle winnaars en hun onderzoeken staat op odissei-data.nl. Alle verzamelde data worden op termijn beschikbaar gesteld. (EdC)

dataarchive.lissdata.nl



Credits: Bruno/Germany, Pixabay

Maakt afkomst uit?**KansenKaart.nl schetst rijk beeld van kansengelijkheid in Nederland**

Waar je in Nederland wordt grootgebracht, kan sterk samenhangen met je latere inkomen, blijkt uit de onlangs gelanceerde interactieve website KansenKaart.nl.

Suze Zijlstra

Econoom Bastian Ravesteijn (Erasmus School of Economics) en zijn team brachten met gepseudonimiseerde CBS-microdata het gemiddelde jaarinkomen in 2017 en 2018 in kaart van Nederlanders die tussen 1982 en 1987 werden geboren. KansenKaart.nl toont dit gemiddelde jaarinkomen per gemeente of wijk waar deze dertigers in 1995 woonden.

Inkomensverschillen

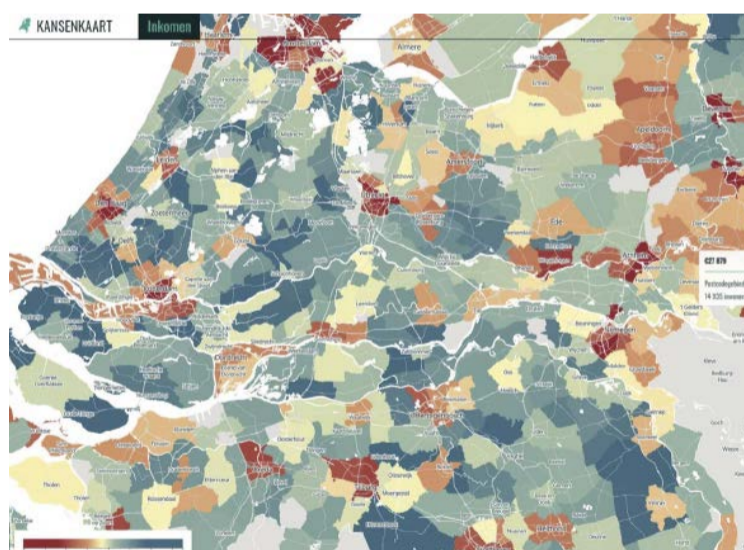
Ravesteijn en zijn team documenteerden flinke inkomensverschillen tussen mensen met ouders met lage en hoge inkomens en tussen regio's

in Nederland. Maar ook als het ouderlijk inkomen gelijk werd gehouden, bleek dat er substantiële verschillen waren tussen en zelfs binnen gemeentes.

Sterke infrastructuur

Het onderzoek profiteerde van de in Nederland aanwezige onderzoeksmogelijkheden. "In Nederland was het dankzij de data-infrastructuur bij het CBS en de financiële ondersteuning vanuit ODISSEI mogelijk om dit onderzoek uit te voeren. Dit zou in de meeste andere Europese landen enorm kostbaar of zelfs onmogelijk zijn geweest. Zo loopt Nederland internationaal voorop als het gaat om het vergaren en delen van kennis over belangrijke maatschappelijke vraagstukken, zoals het ontstaan en tegengaan van kansengelijkheid", licht Bastian Ravesteijn toe.

Om de onderzoeksresultaten voor een breed publiek toegankelijk te



KansenKaart.nl toont het gemiddelde jaarinkomen per gemeente of wijk waar dertigers in 1995 woonden.

KansenKaart.nl ©MapTiler ©OpenStreetMap contributors

maken, ontwikkelden de onderzoekers de website KansenKaart.nl, geïnspireerd door de Amerikaanse Opportunity Atlas (opportunityatlas.org). Ravesteijn: "Burgers, beleids-

makers en onderzoekers kunnen eenvoudig online zien hoe omstandigheden in de jeugd samenhangen met uitkomsten later in het leven." Afgelopen januari zijn extra data en

resultaten toegevoegd aan de Nederlandse website.

Verdere ontwikkeling

Ravesteijn en zijn team zullen de komende tijd het project nog verder gaan ontwikkelen. Hiervoor zullen ze gebruikmaken van de ODISSEI Secure Supercomputer (OSSC). Met de OSSC kunnen ze met enorme rekenkracht de grote hoeveelheid data analyseren in een beveiligde omgeving die voldoet aan de strenge regelgeving rondom het beschermen van privacy. "Voor het doen van analyses voor het onderzoek rondom de website KansenKaart.nl zijn we in de reguliere beveiligde CBS Remote Access Environment enorm veel tijd kwijt. Het is fantastisch dat we middels *parallel computing* op de OSSC nu een rijk beeld kunnen schetsen van kansengelijkheid in Nederland", sluit Ravesteijn af.

KansenKaart.nl

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.

Interdisciplinariteit drijfveer voor afstemming

“Samen onderzoekers optimaal faciliteren”

In december lanceerde DANS een nieuw vijfjarenprogramma, waarin samenwerking centraal staat. E-data sprak met Henk Wals, sinds april 2020 directeur van DANS. *Heidi Berkhout*



Directeur DANS Henk Wals: “Laten we allemaal doen waar we goed in zijn, maar intussen onze diensten voor onderzoekers en dataprofessionals op elkaar afstemmen.”
Credits: Bart van Vliet

“Ontwikkelingen in het onderzoek moeten leidend zijn voor wat wij als datadienstverleners doen. En één van de belangrijkste ontwikkelingen is inter- en multidisciplinair onderzoek.” Aan het woord is Henk Wals, directeur DANS. “Er wordt in toenemende mate aan ‘de wetenschap’ gevraagd om oplossingen voor urgente maatschappelijke problemen, zoals het klimaat of pandemieën. Dat vereist een natuurwetenschappelijke en biomedische insteek, maar voor oorzaken en oplossingen kom je ook bij menselijk gedrag uit, dus bij de geestes- en sociale wetenschappen. Om antwoorden te kunnen geven, moeten onderzoekers datasets uit verschillende disciplines kunnen correleren. De wetenschap is echter grotendeels disciplinair georganiseerd, het datalandschap heeft die indeling overgenomen. Binnen domeinen en disciplines kunnen data totaal verschillende vormen aannemen en lopen methodes en standaarden uiteen. Dat maakt combineren van datasets lastig.”

Daarbij komen steeds grotere hoeveelheden data beschikbaar. Soms zijn de datasets extreem groot, soms speelt privacywetgeving een rol. Wals: “Was schaarste aan data vroeger het probleem van de onderzoeker, tegenwoordig is dat de overvloedige beschikbaarheid. Het probleem voor interdisciplinair onderzoek is: hoe vinden onderzoekers data buiten hun eigen domein en hoe combineren ze deze met het juiste analytische gereedschap?”

Data én tools

Het antwoord ligt volgens Wals in een gedistribueerd mondiaal ecosysteem waarin data en analytisch gereedschap gevonden, bijeengebracht en gekoppeld kunnen worden. “Die weg zijn we aan het inslaan, maar het is een

“Laten we er voor zorgen, dat onderzoekers ons als bos gaan zien, niet als de bomen die het zicht daarop verhinderen”

enorme uitdaging. Het begint ermee dat het gros van onderzoeksdata nog niet eens beschikbaar is. Onderzoekers komen niet toe aan het delen van hun data en worden hier onvoldoende in gestimuleerd en gefaciliteerd. Dan zijn er nog allerlei kwaliteitsaspecten die aandacht vereisen; de output is zo goed als de input, wat onderzoekers vooral merken als ze onderzoek willen herhalen. Vervolgens zouden data en andere onderzoeksartefacten als softwarecomponenten, broncodes en workflows vindbaar, toegankelijk, interoperabel en herbruikbaar moeten zijn. En niet alleen voor mensen, ook computers moeten begrijpen wat het object is en wat ermee kan en mag. En dat dus allemaal over de grenzen van domeinen, disciplines en andere silo's heen.” Wals vervolgt: “Zo'n datalandschap met elkaar creë-

ren, kan niet zonder afstemming, afspraken en samenwerking. Daaraan wordt gewerkt, zowel internationaal als nationaal, bijvoorbeeld in de European Open Science Cloud (EOSC) en het Nationaal Programma Open Science (NPOS). Het zou goed zijn als we deze inspanningen nog wat intensiever en doelmatiger kunnen maken. DANS wil daar graag aan bijdragen.”

Samenwerking als strategie

In het onlangs gepubliceerde DANS-programma 2021-2025 is samenwerking dan ook een van de speerpunten. Wals: “De paragraaf over samenwerking is zowel een statement als een strategie. DANS is op mondiaal niveau een van de grootste dienstverleners voor onderzoeksdata. We hebben maar één doel: onderzoek optimaal faciliteren. Maar ook als grote speler kunnen we dat niet zonder samenwerking met andere serviceproviders. We moeten er als dienstverleners samen voor zorgen dat onderzoekers ons als bos gaan zien, en niet als de bomen die het zicht daarop verhinderen. Laten we dus allemaal doen waar we goed in zijn, maar tegelijkertijd onze diensten op elkaar afstemmen, onze systemen en workflows koppelen en in elkaars verlengde zetten, kortom doen wat nodig is om interdisciplinaire data-analyses makkelijk te maken voor onderzoekers. Die zouden bij wijze van spreken niet eens moeten merken bij welke dienstverleners data en tools vandaan komen.”
dans.knaw.nl

AGENDA

Door het coronavirus worden veel events geannuleerd of online georganiseerd. Raadpleeg de websites van organisatoren voor actuele informatie.

17 - 19 februari • online

Conferentie Open Science

Een forum voor onderzoekers, bibliothecarissen, beleidsmakers en andere belanghebbenden over Open Science-ontwikkelingen.

open-science-conference.eu

16 - 19 maart • online

ENDORSE conferentie

De eerste editie van de Europese conferentie over referentiedata en semantiek.

op.europa.eu/en/web/endorse

13 - 15 april • online

SURF Research Week

Nationale conferentie over onderzoek en ICT.

surf.nl/agenda/surf-research-week

20 - 22 april • online

RDA-bijeenkomst

Het thema van deze 17e RDA-bijeenkomst is: 'Opening data for global challenges'.

rd-alliance.org/plenaries

18 - 20 mei • online

IASSIST 2021

Jaarlijkse conferentie voor data-professionals over datamanagement, data delen, digitale archieven en meer.

iassistdata.org/conferences/iassist-virtual-2021

KORT

OPROEP: Data gevraagd voor leidingnet 2070

Dagelijkse watervraagpatronen worden bepaald door verschillende factoren, zoals de locatie van mensen (thuis, werk), het klimaat (droogte, hitte), overheids-subsidies en technische ontwikkelingen (waterloze wc). Naast deze kennis van waterverbruik zijn onderzoekers van KWR Water Research Institute op zoek naar data en modellen met betrekking tot demografische en stedelijke ontwikkelingen, verkeersstromen en gedragsbeïnvloeding. KWR wil namelijk de invloed van deze ontwikkelingen op de watervraagpatronen tot 2070 kunnen voorspellen met een zogenoemd *agent-gebaseerd model*, om een toekomstbestendig leidingnet te kunnen ontwerpen. Heeft u data en/of modellen die voor KWR interessant kunnen zijn? Neem contact met ze op. (MB)
kwrwater.nl

Nieuwe Call ODISSEI Microdata Access Grant

Eind februari opent de nieuwe call voor de Microdata Access Grant (MAG) van ODISSEI. De deadline voor indienen van voorstellen is 30 april 2021. Medewerkers van deelnemers van ODISSEI kunnen de MAG aanvragen. Een MAG geeft ontvangers gratis toegang tot CBS microdata voor hun project in de CBS Remote Access Environment. Op de website van ODISSEI is meer informatie te vinden over de projecten die eerder een MAG hebben ontvangen. Zodra de call is geopend, staan ook de voorwaarden voor de aanvraag van 2021 online. Schrijf je in voor de ODISSEI nieuwsbrief om op de hoogte te blijven. (SZ)
odissei-data.nl

Henk Wals

Henk Wals is sociaaleconomisch historicus. Voor hij naar DANS kwam, was hij directeur van het Huygens Instituut voor Nederlandse Geschiedenis en daarna van het Internationaal Instituut voor Sociale Geschiedenis. Deze twee instituten bundelden in 2016 hun krachten met het Meertens Instituut in het KNAW Humanities Cluster. Wals was een van de initiatiefnemers en de eerste directievoorzitter van dit 250 FTE-tellende onderzoekcluster. Vooruitlopend daarop had Henk Wals in 2011 al een fusie van het Huygens Instituut (KNAW) en het Instituut voor Nederlandse Geschiedenis (NWO) bewerkstelligd. Vanaf april 2020 is hij directeur van DANS.

SINDS KORT BESCHIKBAAR

Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij CentERdata en DANS.

CentERdata

• Data LISS Core Study 'Politics and values'

De nieuwste data van de LISS Core Study 'Politics and values' is beschikbaar. Jaarlijks worden aan de panelleden, die een representatieve afspiegeling vormen van de Nederlandse bevolking, vragen met betrekking tot politiek en hun normen en waarden voorgelegd. Bij deze 12e meting zijn gegevens van ongeveer 5.500 respondenten verzameld van december 2019 t/m maart 2020. De gegevens gaan onder andere over politieke interesse, stemgedrag, vertrouwen in de overheid, houding ten opzichte van geslachtrollen en het huwelijk, familiesolidariteit en (economisch) conservatisme. De data van alle metingen zijn beschikbaar via LISS Data Archive.

dataarchive.lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Smeekes, A., Wildschut, T., Sedikides, C., maart 2019, Nostalgia 2019 - Wave 1
- Smeekes, A., Wildschut, T., Sedikides, C., juli 2019, Nostalgia 2019 - Wave 2
- Yerkes, M., Besamusca, J., van der Zwan, R., april 2020, Gender inequalities in times of the COVID-19 pandemic - Wave 1
- Jacobs, K., september 2019, Evaluation Democracy Festival
- Schwarz, N., maart 2019, Your opinion on sexual harassment
- Leopold, L., maart 2019, Health, wellbeing and bodyweight
- Müller, R., Aysolmaz, B., Meacham, D., juni-juli 2019, Algorithmic Decision Making Systems
- Jacobs, D., Kaufmann, W., juni 2019, Public participation

- CentERdata, november-december 2019, Health - Wave 12
- CentERdata, oktober-november 2019, Social Integration and Leisure - Wave 12
- CentERdata, juli-augustus 2020, Economic Situation: Housing - Wave 13
- CentERdata, juni-juli 2020, Economic Situation: Income - Wave 13
- CentERdata, september-oktober 2019, Family and Household - Wave 12



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

DANS



Credits Stadsarchief Amsterdam

• Kweekschool voor de Zeevaart 1792-1943

Via EASY is het databestand *Pupils of the Amsterdam Maritime Institute 1792-1943* beschikbaar. Het bestand bevat gegevens van rekruten van de Kweekschool voor de Zeevaart gedurende de periode 1792-1943. Deze dataset is voor een groot deel gebaseerd op het Archief van het Vaderlandsch Fonds ter Aanmoediging van 's Lands Zee-

dienst. Het bestaat uit persoonlijke pagina's die voor elke student werden bijgehouden. De data zijn gekoppeld aan de dienstplichtregisters en sollicitatiebrieven. Daarnaast zijn over iedere student gegevens beschikbaar over lengte, ouderlijke sterfte, religie, pokkenvaccinatie, geboorteplaats en sociaaleconomische achtergrond.

Quanjer, M.A.B. (Radboud University Nijmegen); Kok, Prof. Dr. J. (Radboud University Nijmegen) (2020): *Pupils of the Amsterdam Maritime Institute 1792-1943*.

DOI: 10.17026/dans-zhg-gmwg

Ook sinds kort beschikbaar:

De volgende datasets zijn beschikbaar via DANS:

- Ekamper, Dr P (Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW)) (2020): World War II municipal liberation dates in the Netherlands, 1944-1945. DOI: 10.17026/dans-zcq-7grx
- Both, dr. S. Stephanie (LUMC) (2020): The Effect of Differential Disgust Conditioning and Subsequent Extinction versus Counter-conditioning Procedures on Women's Sexual Responses to Erotic Stimuli. DOI: 10.17026/dans-z4k-6n4w
- Elzinga, Prof. dr. B.M. (Leiden University) (2020): RE-PAIR: Relations and Emotions in Parent-Adolescent Interaction Research & follow-up RE-PAIR during COVID-19 pandemic. DOI: 10.17026/dans-x7n-p9n9
- Haenen, A.P.J. (Radboud University); Teerenstra, dr. S. (Radboud University); Liefers, drs. J. (Radboud University); Huis, dr. A.M.P. (Radboud University); Hulscher, prof. dr. M.E.J.L. (Radboud University) (2020): Hand hygiene compliance in 14 Dutch long-term care facilities.

DOI: 10.17026/dans-z7e-kh44

- de Jong, Prof.mr.dr. E (Utrecht University) (2019): Rechterlijke motiveringsstijlen en maatschappelijke acceptatie van uitspraken.

DOI: 10.17026/dans-xsu-gct8

- Snijder, dr R (OAPEN Foundation) (2020): Open access book usage data - how close is COUNTER to the other kind?

DOI: 10.17026/dans-x9z-4q2w

- Schuwer, Dr R (Fontys University of Applied Sciences, The Netherlands); Baas, M (Saxion University of Applied Sciences) (2020): What about reuse? A study on the use of open educational resources in Dutch higher education.

DOI: 10.17026/dans-zz9-bang

- Sociaal en Cultureel Planbureau - SCP (2019): Continu Onderzoek Burgerperspectieven 2018 - COB2018.

DOI: 10.17026/dans-xwb-v93p

- Peters, dr. R.J.B. (Wageningen Food Safety Research, Wageningen University & Research); Bouwmeester, dr. ir. H. (Toxicology, Wageningen University & Research) (2020): Exposure to microplastic associated chemicals upon oral consumption of microplastics.

DOI: 10.17026/dans-x3h-cbpb

- Peat, Dr DC (Leiden Law School) (2019): Dataset for 'The tyranny of choice and the interpretation of standards: why the European court of human rights uses consensus.'

DOI: 10.17026/dans-zd8-v3j6

- Oosten, Dr RMR van (Leiden University) (2017): Reconsidering ceramics and trade using big data: the significance of stoneware distribution in the Low Countries, 1200-1600.

DOI: 10.17026/dans-zrz-vq88



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

Cultureel erfgoed als proeftuin om data te linken

LD Wizard is Steen van Rosetta

De Steen van Rosetta bleek een sleutel om de ene taal in een andere taal om te zetten. Linked Data is ook zo'n sleutel: deze techniek vertaalt data om ze te laten communiceren met andere data.

Mathilde Jansen

Stel, een kunstmuseum bezit een schilderij waarop een bepaald huis is afgebeeld. Om dit schilderij vast te leggen in een collectie, moet bepaalde informatie aan het schilderij worden gekoppeld. Bijvoorbeeld het adres van het huis. Maar misschien had de straat vroeger een andere naam, of heet de straat over 50 jaar anders. Zo'n informatielabel is dus kwetsbaar. Beter is om zo'n schilderij een identificatienummer van het Kadaster te geven. De Linked Data Wizard biedt de oplossing. "Sinds drie maanden is er de Linked Data Wizard", vertelt projectleider Ivo Zandhuis trots. "Als erfgoed-

stelling heb je alleen een Excel-sheet nodig waarin wordt beschreven welke schilderijen of foto's je hebt. De LD Wizard legt precies vast wat elke kolom in de tabel betekent, en zorgt er zo voor dat de data gelinkt kunnen worden aan andere data."

Erfgoed en meer

Of het nu gaat om actuele panden bij het Kadaster, of films die volgende week in de bioscoop draaien, de LD Wizard is voor alle soorten data te gebruiken. Zandhuis: "Maar voor erfgoed heb je andere velden nodig dan voor films en plaatsen op een kaart. Daarom hebben we ons in eerste instantie gericht op één domein, namelijk erfgoed. We hopen dat van daaruit andere domeinen aanhaken. Het Kadaster heeft het bijvoorbeeld al opgepakt, en een eigen variant gemaakt die ook geografische coördinaten om kan zetten. Zij hebben dus een functie toegevoegd die voor hun domein van belang is. Zo moeten



Linked Data kan worden vergeleken met de Steen van Rosetta; beide zijn een sleutel om data te laten communiceren met andere data.

Credits: Kris Asard, Pixabay

we samenwerken als community. Als het goed is, is dit project nooit af."

Digitaal erfgoedcoaches

Zelf werkt Zandhuis voor het Netwerk Digitaal Erfgoed (NDE). Daar zorgt hij dat data uit het collectieregistratiesysteem van een instelling als Linked Open Data gepubliceerd wordt. Daarnaast heeft het NDE de opdracht om andere erfgoedinstellingen op weg te helpen. Zo organiseerde Zandhuis onlangs een workshop voor digitaal erfgoedcoaches. Elke provincie in Nederland heeft zo'n coach aangesteld om lokale musea en historische verenigingen te helpen bij hun digitalisering. Zandhuis: "Na instructie konden zij een Excel-sheet in korte tijd omzetten in Linked Data. Dat is echt een stap voorwaarts."

ldwizard.netwerkdigitaal erfgoed.nl

Informaticus Jacco van Ossenbruggen:

“Multidisciplinair onderzoek is heel belangrijk én heel leuk”

Als leider van de onderzoeksgroep User Centric Data, co-director van Civic AI Lab en lid van ODISSEI's Management Board is Jacco van Ossenbruggen een veelzijdig man. E-data sprak met hem. *Suze Zijlstra*

Zijn er overeenkomsten in de manier waarop techniek wordt gebruikt in de geestes- en sociale wetenschappen? En wat is uw rol?

“Ik zie mezelf als echte informaticus, één die geïnteresseerd is in de context waarin technologie wordt toegepast. Allebei de disciplines maken grote stappen in het gebruik van digitale tools. Hun onderzoeksmethodes zelf moeten ook veranderen. Dat vind ik een heel interessant proces om bij betrokken te zijn. Bij de geesteswetenschappen is *tool criticism* gelanceerd als onderzoekspraktijk. Zelfs bij geesteswetenschappers heerst soms nog het idee dat wat uit een computer komt, neutraal is. Dat is natuurlijk niet zo. De computer doet wat geprogrammeerd is, met data die mensen geselecteerd hebben. Mensen realiseren zich dat steeds beter. Dan is de volgende vraag: kunnen de bij bronnenkritiek gebruikelijke vragen ook gesteld worden over digitale bronnen, zijn onze tools en systemen daar geschikt voor? Dat is de uitdaging voor mij en mijn vakgenoten.”

Universiteiten passen hun curriculum aan op digitale toepassingen. Van Ossenbruggen signaleert dit ook, maar: “Soms gaat een cursus nog te veel over hoe je een bepaalde tool gebruikt. Het mag van mij wel kritischer gaan over wat de machine wel en niet kan, en of dat wel of niet past in de gebruikte onderzoeksmethode.”

U heeft met zowel CLARIAH als met ODISSEI gewerkt. Wat kunnen deze infrastructuren van elkaar leren?

“ODISSEI draait vaak om gevoelige data, terwijl bij CLARIAH de aanname vaak is dat de data open zijn of uiteindelijk gaan worden. De aandacht voor veiligheid is daarom groter bij ODISSEI. Andersom heeft CLARIAH verschillende KNAW-instituten die een natuurlijke partner zijn binnen de geesteswetenschappen, om CLARIAH op lange termijn te ondersteunen. Dit zouden de sociale wetenschappen meer moeten hebben, want dat maakt het makkelijker om een menselijke infrastructuur te bouwen die langer duurt dan een project.”

Zou er meer samenwerking moeten zijn?

“Maatschappelijke vragen vereisen multidisciplinair onderzoek, dus ik ben voorstander van samenwerking, maar dit komt met extra overlegstructuren. Het is zaak dat je een hel-



Van Ossenbruggen: “De geestes- én sociale wetenschappen maken grote stappen in het gebruik van digitale tools. Hun onderzoeksmethodes zelf moeten ook veranderen. Daar wil ik graag bij betrokken zijn.” Credits Bart van Vliet

INTERVIEW

‘Laten we elkaar op een goede manier sterker maken’

der gemeenschappelijk doel hebt. Als ODISSEI erin zou slagen een gemeenschappelijk portal te maken waarin *alle* in Nederland door sociale wetenschappers gebruikte en gemaakte onderzoeksdata vindbaar en herbruikbaar is, dan zou dat al een grote bijdrage zijn aan het onderzoekslandschap. Maar ook samenwerking tussen ODISSEI en een aantal geesteswetenschappelijke gebieden zou voor de hand liggen. Als je een platform voor contentanalyse uit CLARIAH zou combineren met sociale netwerkanalyse uit ODISSEI, dan is het wel van belang dat verschillende systemen met elkaar kunnen praten.

ter van de onderzoekers gaat. En dat natuurlijk op zo’n veilige manier dat de onderzoeker de data zelf nooit te zien krijgt, maar alleen de niet-gevoelige uitkomst van de analyse ziet. Maar: is onze wetgeving hier al op aangepast? Als de wetenschapper de data niet ziet, hoeft er dan ook geen toestemming voor gebruik te zijn? Dat uitpluizen is baanbrekend onderzoek. De combinatie van techniek en juridische kennis is cruciaal.”

‘Kennisoverdracht over digitale toepassingen mag wel wat kritischer’

U bent voorstander van multidisciplinaire teams, maar hoe ziet een multidisciplinair project eruit?

“Het is een beetje de vloek van multidisciplinair onderzoek: iedereen vindt het nodig, maar onderzoeksvoorstellen moeten toch vaak door één discipline worden goedgekeurd, of publicaties moeten passen in een monodisciplinair tijdschrift. Ik probeer altijd een ‘computer science hoekje’ af te bakken. Een infrastructuurproject zoals ODISSEI wil wetenschappelijke vragen beantwoorden en onderzoeksprototypes maken om hypothesen te testen. Dan probeer je de prototypes ook geschikt te maken voor een latere productieversie binnen de infrastructuur. Je belooft een innovatieve infrastructuur te bouwen, er moeten nieuwe dingen onderzocht worden waarvan we nu nog niet weten wat het antwoord is. Multidisciplinair onderzoek is heel belangrijk, maar vooral ook heel erg leuk. Dat zit in de verschillende mensen die je ontmoet, wat betreft culturen, wetenschappelijke richting en afkomst, maar ook in het verschil in onderzoeksvragen: een sociaalwetenschappelijke of geesteswetenschappelijke onderzoeksvraag doet mij ook weer anders denken over mijn eigen technologische achtergrond, ik moet het ook wat anders doen als zij mijn tools willen gebruiken. Het is leuk om uit je comfort zone te gaan. Op een goede manier kan je elkaar sterker maken.”

cwi.nl/people/jacco-van-ossenbruggen

Jacco van Ossenbruggen

Jacco van Ossenbruggen is informaticus. Hij houdt zich bezig met de toepassing van technologie in de sociale en geesteswetenschappen. Hij is leider van de onderzoeksgroep User Centric Data aan de VU, co-director van het onlangs gelanceerde Civic AI Lab en lid van ODISSEI's Management Board.

Ik heb veel gedaan op het gebied van linked data van erfgoedinstellingen. Hoe stel je bijvoorbeeld culturele data beschikbaar voor AI op een manier dat data gelijkvormig genoeg zijn om er gemeenschappelijk mee te werken, terwijl ze ook specifiek genoeg blijven om recht te doen aan de unieke kant van een culturele instelling? Die trade-off zie je bij ODISSEI ook. Data en metadataschema's moeten gelijkvormig genoeg zijn om in een gemeenschappelijk systeem nuttige dingen te doen, maar ook flexibel en uniek genoeg om unieke eigenschappen van een dataset naar voren te laten komen.”

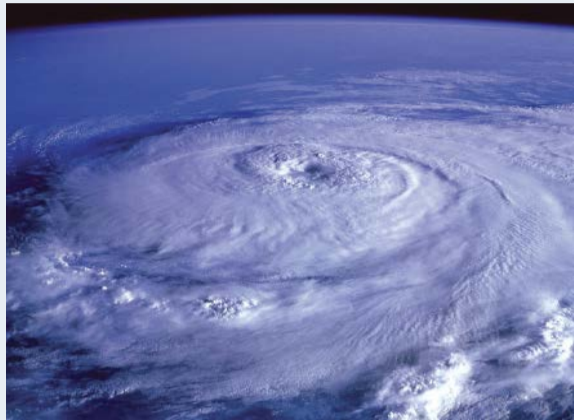
SSH heeft de informatiewetenschap nodig, maar wat heeft SSH te bieden?

“Wat ik bij ODISSEI heel inspirerend vind, is de grote hoeveelheid gevoelige onderzoeksdata. Om daar goed mee om te gaan, met de idealen van open en reproduceerbare wetenschap toch de privacy van de mensen waarover de data gaat te waarborgen, dat vind ik heel interessant. Zoals gebeurt binnen ODISSEI bij de groep van Michel Dumontier in Maastricht: zij kijken naar manieren om AI en data-analyse-algoritmes naar de databron te brengen, in plaats dat de data naar de compu-

**Winnaar - exacte en technische wetenschappen:
STORM**

De STORM-dataset bevat 10.000 jaar aan synthetische tropische cycloonsporen. Dit zijn tropische cyclonen die theoretisch mogelijk, maar (nog) niet voorgekomen zijn. De gegevens kunnen worden gebruikt om het tropische cycloonrisico te berekenen in alle (kust) regio's die vatbaar zijn voor tropische cyclonen. Het is de eerste publiek toegankelijke dataset die op wereldwijde schaal tropische cyclonen representeert, vooral interessant voor risico-analisten, onderzoekers en beleidsmakers.

doi.org/10.4121/uuid:82c1dc0d-5485-43d8-901a-ce7f26cda35d



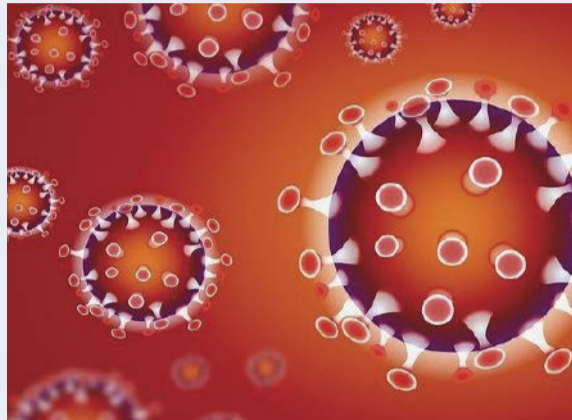
Orkaan Elena, Golf van Mexico, 1 september 1985.

Credits: Image Science and Analysis Laboratory, NASA-Johnson Space Center

**Winnaar - medische en levenswetenschappen:
CoronaWatchNL**

De CoronaWatchNL-dataset bevat openbare en accurate gegevens over COVID-19 in Nederland. De dataset bestaat uit data uit nieuwsrapporten en betrouwbare statistische gegevens, bijvoorbeeld afkomstig van het RIVM, LCPS, NICE en Nationale Corona Dashboard. Dit project standaardiseert en publiceert de gegevens en maakt ze vindbaar, toegankelijk, interoperabel en herbruikbaar. Onderzoekers, epidemiologen en beleidsmakers hoeven niet meer zelf deze data te verzamelen en samen te voegen.

github.com/J535D165/CoronaWatchNL



Credits: CoronaWatchNL

**Winnaar - sociale en geesteswetenschappen:
17-eeuwse kranten**

Deze dataset bevat een transcriptie van de oudste Nederlandstalige kranten, gepubliceerd tussen 1618 en 1700; het gaat om 14 titels, 6.184 kranten met circa 120.000 artikelen en circa 20 miljoen woorden. Ruim 200 vrijwilligers hebben in vijf jaar tijd alle 20 miljoen woorden handmatig overgetikt en gecorrigeerd, waardoor een betrouwbaar full-text-doorzoekbaar onderzoekscorpus is gecreëerd. De dataset is relevant voor onderzoekers van alle historische disciplines binnen de geesteswetenschappen.

delpher.nl/kranten



Eén van de online beschikbare kranten is *Avec privilège de nos-seigneurs les Etats de Hollande et de West-Frise*, uitgegeven door J.T. du Breuil te Amsterdam 1698/01/09. resolver.kb.nl/resolve?urn=ddd:010551041:mpeg21:p001

Online toegankelijk maken en delen van data beloond

STORM, CoronaWatchNL en 17-eeuwse kranten winnaars Dataprijs

Eind 2020 werd aan drie inspirerende datasets de RDNL Dataprijs 2020 uitgereikt, erkenning gevend aan de inspanning om de data te willen delen. E-data zet de winnaars op een rij. *Heidi Berkhout en Deirdre Casella*

Ingebod in het partnerevent *Tgether We Share* reikte RDNL voor de 6e keer de Dataprijs uit. Naast een keynote over de Europese Open Science Cloud door Dataprijs-juryvoorzitter prof. dr. Karel Luyben (Rector Magnificus Emeritus en Nationaal Coördinator Open Science),

zeven interactieve parallelsessies georganiseerd door RDNL-partners en dataexperts uit het Nederlandse datalandschap en aanmoedigingsprijzen, vormden de uitreiking van de Dataprijs 2020 de finale van het event.

Nieuwe mogelijkheden

Binnen de categorie exacte en technische wetenschappen prees de jury de winnaar mede omdat de data zich bevinden in een stabiele omgeving, open toegankelijk zijn en nieuwe onderzoeksmogelijkheden met zich meebrengen. Onderzoeker Nadia Bloemendaal (Instituut voor Milieu-

kunde, Vrije Universiteit van Amsterdam), wiens dataset STORM de Nederlandse Dataprijs won, gaf aan “het prijzengeld te besteden aan een interactieve website waarop het risico van een tropische cycloon wordt gevisualiseerd.”

Vakwerk

Jonathan de Bruin (Universiteit Utrecht) nam namens het team de prijs van de categorie medisch en levenswetenschappen in ontvangst voor CoronaWatchNL. De jury gaf aan: tegenwoordig is veel informatie en data publiek beschikbaar. Data op dusdanige wijze aanbieden dat

anderen er echt verder mee kunnen, is vakwerk. De Bruin noemde in zijn dankwoord “aanbevelingen te doen voor het creëren van FAIR-datasets in crisissituaties, gecombineerd met inzichten uit RDM, Open en Citizen Science, zodat data op snel en transparant gepubliceerd kunnen worden als de tijd dringt.”

Unanieme winnaar

Winnaar in de categorie geestes- en sociale wetenschappen werd Nicoline van der Sijs (Instituut voor de Nederlandse Taal, Delpher/KB). Namens de jury een unanieme winnaar in deze categorie. Want, aldus de

jury: de data zijn publiek toegankelijk en bruikbaar voor onderzoekers uit allerlei disciplines. De toepassingen zijn legio en ook het feit dat wordt gewerkt met vrijwilligers geeft aan dat deze collectie iedereen raakt. Winnares Van der Sijs benadrukte de waarde en potentie van het combineren van FAIR-datapraktijken met Citizen Science binnen de crowdsourced research database van de winnende dataset getranscribeerde 17e-eeuwse kranten. Ze sprak haar dank uit aan de vrijwilligers die het delen van de gegevens mogelijk maakten.

researchdata.nl

NPOS-project in beeld

Wat kenmerkt een data steward?

Waar halen data stewards hun kennis vandaan en hoe kan Nederland dit professioneler aanpakken? Een NPOS-project onderzocht het. *Marjan Grootveld*

“Willen we Open Science tot dagelijkse praktijk maken, dan hebben we veel data stewards nodig. Hun verantwoordelijkheden en expertise moeten passen bij het onderzoek dat zij ondersteunen en bij hun positie in de organisatie.” Het zijn uitspraken van Margreet Bloemers, projectleider FAIR-data bij onderzoeksfinancier ZonMw. Tot op heden ontbreken echter duidelijke opleidingstrajecten en functiebeschrijvingen voor data stewards. Tijd voor een verkennend onderzoek. Vanuit het Nationaal Platform Open Science (NPOS) werd

opdracht gegeven voor het project Professionaliseren van data stewardship: competenties, training en educatie.

Vele betrokkenen

Een LCRDM-taakgroep nam de opdracht aan, samen met het Dutch Techcentre for Life Sciences. Ook andere partijen verbonden zich aan het project: de Vereniging van Universiteiten, de Nederlandse Vereniging van Hogescholen, de Nederlandse Federatie van Universitair Medische Centra, het Promovendi Netwerk Nederland en SURF.

De taakgroep verzamelde case studies van data stewards in Nederlandse instellingen. Daaruit bleek onder andere een tekort aan formele opleiding voor data stewards (als uitzondering werd de RDNL-training Essenti-

als 4 Data Support opgemerkt). Datacompetenties uit een eerdere studie werden gebruikt om bestaande trainingen te annoteren. Het project leverde bovendien componenten op van een functieprofiel voor data stewards, met taken zoals “adviseert over compliance met regelgeving” en “identificeert de behoefte aan data-infrastructuur”, met hun uitwerking op verschillende functieniveaus.

Data steward skills tool

De voornaamste aanbeveling aan de projectpartners is om hiermee concrete functieprofielen op te stellen. Dit helpt de onderzoeksinstituten bij het uitvoeren van hun databeleid en biedt data stewards een carrièreperspectief. Het ontwikkelen van een data steward skills tool is een andere aanbeveling.



Hierin komen taken, benodigde competenties en passend trainingsaanbod samen. Met verende – Nederlandse maar ook internationale – inspanningen is op termijn bovendien een gecertificeerd curriculum wenselijk.

openscience.nl

DOI: 10.5281/zenodo.4320505

Linguïsten besparen tijd met CLARIAH-tool

Online tool SASTA analyseert taal

Automatische grammaticale analyse is niet alleen nuttig voor de wetenschap. Ook de klinische praktijk kan deze tool goed gebruiken, zo laat het CLARIAH-project SASTA zien.

Erica Renckens

“Voor een goed behandelplan is een grammaticale analyse van de afwijkende taal essentieel”, vertelt Margo Zwitserlood, die als klinisch linguïst en onderzoeker in het Pento Audiologisch Centrum Amersfoort vooral jonge kinderen met taal- en spraakproblemen ziet. “We nemen dan een gesprek met het kind op, dat we transcriberen en vervolgens analyseren. Daar ben je in totaal vier tot vijf uur mee bezig. Logopedisten hebben daar onvoldoende tijd voor of het ontbreekt hen aan kennis.”

Maatschappelijke impact

In het project SASTA werkt zij daarom met collega's van de Vereniging voor Klinische Linguïstiek en CLARIAH-onderzoekers aan

een programma voor de automatische analyse van spontane taal. Dit programma is niet alleen bedoeld voor kindertaal, maar ook voor afwijkende taal bij volwassenen met afasie. “De basis van SASTA bestaat uit bestaande onderdelen uit de CLARIAH-infrastructuur, zoals de Alpino-parser en de GrEtel-toepassing voor treebanks”, vertelt Jan Odijk, hoogleraar taal- en spraaktechnologie aan de Universiteit Utrecht. De parser ontleedt de zinnen, waarna met GrEtel in de grammaticale structuren gezocht kan worden. “Het is mooi om te zien hoe deze technieken ook maatschappelijke impact kunnen genereren.”

Eén druk op de knop

SASTA maakt met één druk op de knop een complexiteitsanalyse van de transcriptie. Daarin staat onder andere hoe vaak verschillende woordsoorten en zinsconstructies voorkomen, en hoe dit zich verhoudt tot normaal taalgebruik. De uitkomst van de analyse wordt verwerkt in de scoreformulieren van drie veelgebruikte diagnostische methodes. “Dat scheelt al heel veel werk, geweldig”, aldus Zwitserlood.

In een vergelijking tussen automatisch en



Logopediste Henrieke Tissingh en de zevenjarige Noud Thus tijdens een afspraak. Credits: logopediepraktijk Puur Logopedie, Harderwijk

manueel uitgevoerde analyses scoort SASTA erg hoog, tussen de 75 en 90%. Odijk: “Het systeem vindt soms woordsoorten en zinsconstructies die de linguïsten zelf niet uit de data haalden; het werkt dan dus beter dan de mens.” Wel heeft SASTA nog moeite met afwijkende taal: het probeert er correct Nederlands in te zien. Zo analyseert het programma een zin met een congruentiefout als twee zinnen en ziet hij ‘stukkies’ als een samenstelling van ‘stuk’ en ‘kies’. Odijk: “De foutenanalyse doen we daarom nu nog handmatig,

maar de betrokken onderzoekers werken aan de automatisering hiervan.” De transcriptie zal voorlopig ook nog handmatig moeten gebeuren, omdat kinderspraak moeilijk automatisch te herkennen is, zeker als deze afwijkend is. Zwitserlood: “Maar ook daar zijn wetenschappers volop mee bezig, onder andere in Nijmegen. Alle tijdswinst is welkom!”

Een voorlopige versie van SASTA is momenteel online beschikbaar voor linguïsten via: sasta.hum.uu.nl

Data Nationaal Vrijheidsonderzoek beschikbaar

LISS panel meet draagvlak Dodenherdenking en Bevrijdingsdag

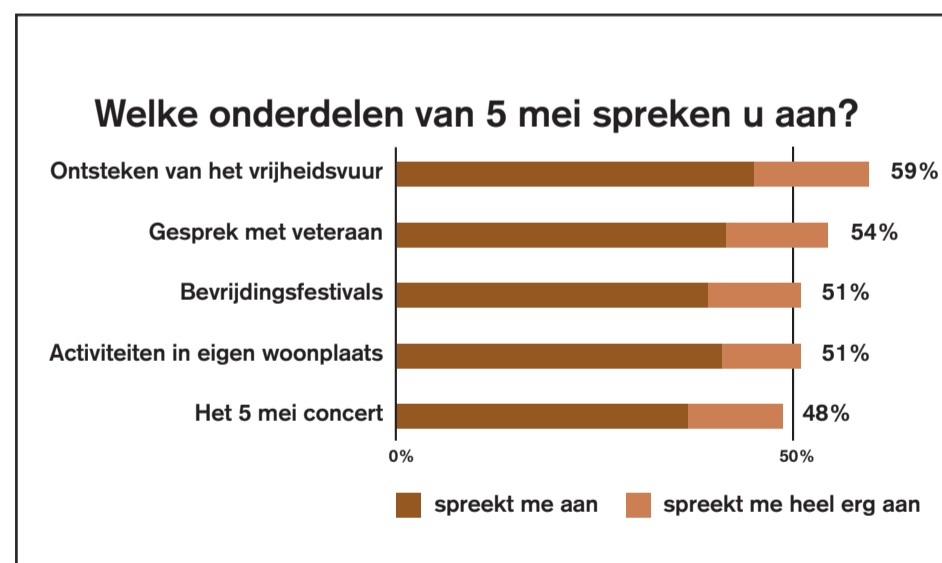
Uit het Nationaal Vrijheidsonderzoek blijkt dat een ruime meerderheid van de Nederlandse bevolking de Dodenherdenking op 4 mei en Bevrijdingsdag op 5 mei belangrijk vindt.

Marije Oudejans

Het Nationaal Vrijheidsonderzoek van Nationaal Comité 4 en 5 mei bestaat al sinds 2001. Voor de jaarlijkse meting wordt vanaf 2020 gebruik gemaakt van het LISS panel van CentERdata. Aan dit onderzoek, waarin wordt gevraagd naar de ervaringen van een jaar eerder, hebben ruim 1.700 personen meegedaan. Dezelfde personen worden de komende jaren gevolgd.

Beleving van vrijheid

In het onderzoek komt het thema ‘draagvlak voor 4 en 5 mei’ ieder jaar aan bod. Daarnaast worden om het jaar andere thema's belicht. Zo werd in de meting van maart 2020 ingezoomd op ‘beleving van vrijheid’. Waar denken men-



sen aan bij het begrip vrijheid? Welke gebeurtenissen uit het verleden zijn bepalend voor de ideeën die iemand heeft over vrijheid? De Tweede Wereldoorlog blijkt de gebeurtenis die met afstand het meest bepalend is geweest voor de ideeën die men heeft over

vrijheid en onvrijheid.

Samenhang in data

88% van de Nederlandse bevolking vindt de Nationale Herdenking op 4 mei belangrijk of heel belangrijk. De groepen die hierin wat af-

De afbeelding toont welke vijf onderdelen van 5 mei de respondenten (heel erg) aanspreken. Alle resultaten staan in het onderzoeksrapport van het Nationaal Comité 4 en 5 mei, beschikbaar via 4en5mei.nl. Credits: 4en5mei.nl/onderzoek/nationaal_vrijheidsonderzoek/nationaal-vrijheidsonderzoek-2020

wijken, maar waarvan ook de meerderheid de herdenking (heel) belangrijk vindt, zijn jongeren (75%) en mensen met een niet-westerse migratieachtergrond (70%). Maar liefst 89% hield op 4 mei 2019 om 20:00 uur twee minuten stilte tijdens de Dodenherdenking. Verder blijkt uit het onderzoek dat 4 en 5 mei voor veel mensen met elkaar samenhangen, waarbij de Dodenherdenking voor de meesten net iets belangrijker is dan Bevrijdingsdag.

De door het LISS panel verzamelde data zijn beschikbaar via: dataarchive.lissdata.nl

LISS panel

Het LISS panel is met 5000 huishoudens en zo'n 7500 panelleden een representatieve afspiegeling van de Nederlandse bevolking en sinds 2007 hét panel voor wetenschappelijk, maatschappelijk en beleidsrelevant onderzoek. Joris Mulder, sinds 2020 LISS coördinator: “Representativiteit, datakwaliteit, innovatie en een optimale respons staan bij ons centraal. Hiervoor werken we samen met onder andere ODISSEI, CBS, DANS en SURF. Met regelmaat worden innovatieve experimenten in het LISS panel afgenomen en steeds vaker wordt survey onderzoek gecombineerd met data science. Ook nemen we deel aan het ODISSEI Portal en worden nieuwe projecten voor studies met sensors en wearables gepland.” (EdC) lissdata.nl

SWORD voor LISS Data Archive en DANS

Het LISS Data Archive bevat gegevens over het leven van Nederlanders en hun mening over verschillende onderwerpen, verzameld in een groot aantal onderzoeken onder de leden van het LISS panel. Voor duurzame opslag voor langere termijn worden de data ook gearchiveerd bij DANS. Sinds kort is er een geautomatiseerde interface tussen het LISS Data Archive en DANS. Met deze zogenaamde SWORD-interface (Simple Web-service Offering Repository Deposit) worden de data grotendeels automatisch doorgestuurd naar DANS. Zo is dubbel invoerwerk niet meer nodig en worden de data sneller doorgegeven. (MdB) dans.knaw.nl

Na elke zoekronde een stapje slimmer

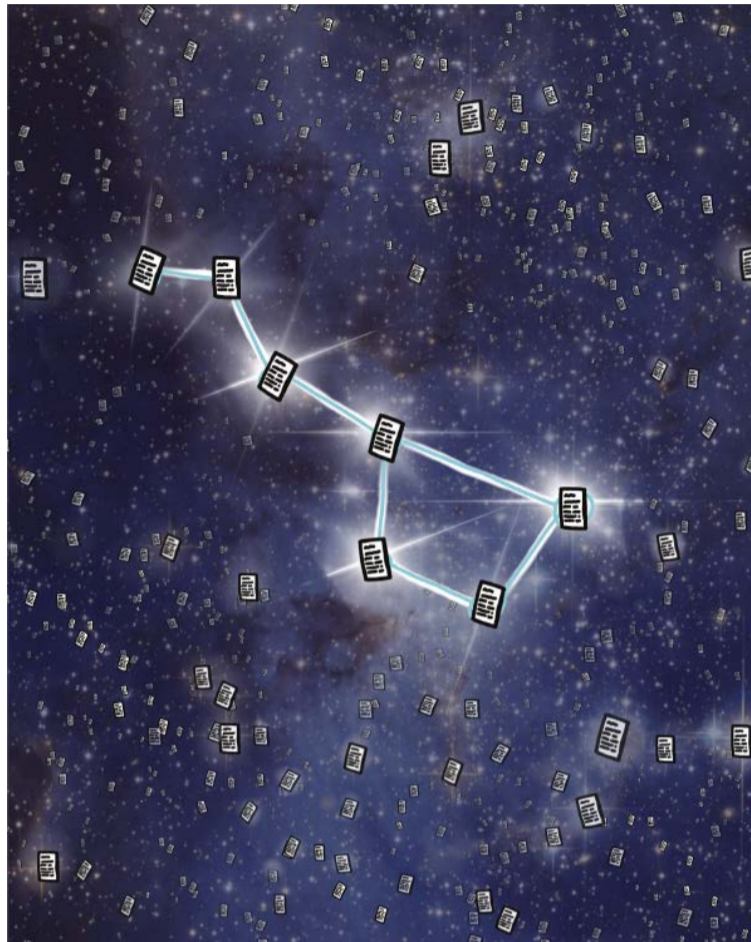
Software ASReview doorzoekt sneller digitale data

Een team aan de Universiteit Utrecht ontwikkelde onder leiding van Rens van de Schoot software die onderzoekers kan ondersteunen bij het snel vinden van relevante artikelen. *Suze Zijlstra*

In veel wetenschappelijke vakgebieden komt er met een enorme snelheid een grote hoeveelheid nieuwe artikelen bij. Dit maakt het schier onmogelijk om artikelen binnen een redelijke tijd te doorzoeken en te beoordelen. Rens van de Schoot, hoogleraar statistiek aan de Universiteit Utrecht, vroeg zich af of dit handiger kon. Samen met Daniel Oberski, UHD aan dezelfde universiteit, ontwikkelden ze ASReview: Automated Systematic Review.

Elke ronde slimmer

Rens van de Schoot: “Als je zoekt in een bestaande database, dan gebruik je één keer de zoektermen en krijg je in één keer de resultaten. ASReview kent een iteratief proces: als mens laat je de machine weten of een artikel wel of niet relevant is. De machine leert hiervan, ordent de artikelen opnieuw en geeft nieuwe resultaten ter beoordeling; het proces begint opnieuw. Deze active le-



Met ASReview vind je de juiste papers in een oneindig heelal aan publicaties. Credits: ESA/Hubble (spacetelescope.org/images/heic0607a) CC BY 4.0, recolouring and illustrations by statistics-illustrated.com

arning cycle maakt de machine elke zoekronde een stapje slimmer, de mens blijft het ‘orakel’. Het programma is sinds een half

jaar door iedereen vrij te gebruiken. Het vakgebied van de gebruiker is irrelevant: het kan een database als PubMed zijn, LexisNexis, maar het

programma is ook in te zetten voor oud-Griekse teksten of krantenartikelen. Van de Schoot: “De data moeten wel digitaal beschikbaar zijn. Het moet ook een grote database zijn, zo groot dat je het als mens niet meer kan lezen. De taak van de machine is om relevante teksten er tussenuit te vissen.”

Open source software

ASReview is open source software en op de computer van de gebruiker te installeren. Daarmee blijft het data-eigenaarschap, waaronder gegevens over iemands zoektocht, in handen van de gebruiker zelf. Samen met het ODISEI Social Data Science Team wordt ondersteuning gegeven aan onderzoekers die gebruik willen maken van de software.

Verder ontwikkelen

Van de Schoot sluit af: “Nu moeten we de wetenschap ermee gaan bedrijven. Bovendien willen we het verder ontwikkelen. Recente simulaties hebben laten zien dat ASReview wat betreft selectie nu al dezelfde kwaliteit levert als mensen, maar dan met een grotere snelheid. We gaan verder onderzoeken welke machine learning modellen je het best kan gebruiken per tekstsoort.” asreview.nl

GELEZEN

Linking Knowledge – Linked Open Data for Knowledge Organisation and Visualisation. Ergon, Würzburg. Richard Smiraglia, Andrea Scharnhorst. 2021.

Hoe kunnen archivariissen, informatiekundigen en computerwetenschappers het beste met onderzoekers uit de geesteswetenschappen samenwerken en hen ondersteunen? Dit boek wil bijdragen aan een diepere dialoog tussen informatieprofessionals en geesteswetenschappers. Bovendien bevat het bijdragen uit het project Digging into the Knowledge Graph en andere Linked Data projecten zoals Golden Agents en Open Knowledge Graph. Het boek benadert overdracht en organisatie van kennis op drie manieren. Ten eerste worden er fundamentele vragen gesteld over de organisatie van kennis, bijvoorbeeld: wat is een concept en hoe kan dit met behulp van termen beschreven worden? Ten tweede wordt ingegaan op de rol van thesauri, classificaties en ontologiën. Hoe moeten deze kennisorganisatiesystemen als Linked Data gepubliceerd worden en welke effect heeft dit op de kennisproductie in de wetenschap? Als laatste wordt ingezoomd op de vraag welke taal en terminologie ontwikkeld moeten worden. Een must read voor informatieprofessionals en onderzoekers!

COLUMN

Mens en machine

En hoe bepalen we die nauwkeurigheid dan?

Antwoorden op die vragen hadden mooi meegenomen kunnen worden in de richtlijn van EBA. Een gemiste kans, maar wel een goede zaak dat regulering en toezicht geüniformeerd wordt. Net als bij Uber en AirBnB is het belangrijk dat we er op kunnen vertrouwen dat men zorgvuldig omgaat met taxi's, woningen en



Credits foto: Don Wijns

waardebepalingen. Daarbij moeten we ons wel realiseren dat de tijd waarin we simpelweg konden stellen dat de mens het beter zal doen, inmiddels ver achter ons ligt. Toch is het nog steeds een ingesloten reflex bij hoe we naar nieuwe technologie kijken: we nemen deze nieuwe technologie nadrukkelijk de maat maar vragen ons nauwelijks af of de klassieke werkwijze wel zo goed was. Want heel eerlijk: hoe objectief is die menselijke taxateur eigenlijk?

Een mooi voorbeeld was dit voorjaar te zien in Engeland. Net als in Nederland konden leerlingen in Corona tijd geen landelijk eindexamen doen en werden eindlijsten op school dan maar ingevuld door de docenten zelf. De overheid besloot deze eindlijsten met een computermodel te ‘normaliseren’ op basis van historische resultaten van de onderwijsinstelling. Het gevolg was dat leerlingen op scholen die in het verleden relatief zwak hadden ge-

presteerd naar beneden werden bijgesteld, ook als een leerling zelf juist goed presteerde op die school. Niet verrassend kwamen scholen uit armere gebieden, gezien de historie, er vaker slecht vanaf.

Britse kranten stonden vol verhalen van leerlingen die fors naar beneden werden gecorrigeerd en daardoor konden fluiten naar toelating op de universiteit van hun keuze. De overheid moest bakzeil halen en trok het algoritme terug. En dus was opeens het oordeel van de docent gewoon weer doorslaggevend. Dat gaf een unieke kans om te beoordelen of die docenten dan wel objectief waren. Het resultaat laat zich raden: de docentenbeoordelingen lagen gemiddeld 12% hoger dan de eindcijfers van vorig jaar. Niets menselijks is docenten blijkbaar vreemd. En dat was ook precies de oorspronkelijke aanleiding voor een correctie door een algoritme.

Terug naar de waardebeoordeling van huizen. Daar gloort wel wat hoop. Want de toezichthouder staat wel een tussenvorm toe: de desktoptaxatie, een mengvorm tussen een computergestuurde waardebeoordeling en het oordeel van een taxateur. En dat is eigenlijk precies datgene waar het echt om gaat in een wereld vol data en algoritmen: nieuwe samenwerkingsvormen tussen mens en machine ontdekken, die uiteindelijk beter aansluiten op onze behoeften.

Sander Klous

Prof. dr. Sander Klous, gepromoveerd in hoge energie fysica, is hoogleraar Big Data Ecosystems for Business and Society (Universiteit van Amsterdam) en verantwoordelijk voor data & analytics (KPMG). Hij werkte aan verschillende projecten bij de deeltjesversneller op CERN, waar hij ook deel uitmaakte van het Nobelprijs-winnende Atlas-experiment dat het Higgs-Boson ontdekte. Zijn boeken *Wij zijn Big Data* en *Vertrouwen in de Slimme Samenleving* werden genomineerd voor managementboekprijzen. Klous geeft de column door aan *Nart Wielgaard*.

En een computer ziet niet of jij een state-of-the-art keuken in je huis hebt. Dat is kortweg de redenering van criticasters over modelmatige waardebeoordelingen van huizen die meer en meer de rol van de taxateur overnemen. Huizenbezitters kunnen tegen een fractie van de kosten van een gewone taxatie een rapportje krijgen waarmee ze de bank tevreden kunnen stellen. Maar nu krijgen ze de wind juist tegen. Een richtlijn van de Europese Bankenautoriteit (EBA) staat deze alternatieve versie van de taxatie niet meer toe vanaf medio 2021.

Er is van alles te zeggen over dit thema – en dat is in de media ook vrij uitgebreid gedaan. Bijvoorbeeld: hoe nauwkeurig moet die waardebeoordeling eigenlijk zijn? Wie voor een huis van 8 ton een waardebeoordeling nodig heeft voor een hypotheek van 6 ton heeft een minder nauwkeurige waardebeoordeling nodig dan een echtpaar dat deze waarde nodig heeft voor het verdelen van de boedel in een echtscheiding. Dus moet de gewenste nauwkeurigheid geen rol spelen bij de prijs van die waardebeoordeling? Hoe duurder hoe nauwkeuriger.