



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

E-data & Research 1403

Berkhout, Heidi

published in

E-data & Research
2020

document version

Publisher's PDF, also known as Version of record

document license

CC BY

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Berkhout, H. (editor) (2020). E-data & Research 1403. *E-data & Research*, 14(3), 1-8.
<https://www.edata.nl/1403/pdf/1403.pdf>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

Dendro4Art combineert dendrochronologie en kunsthistorie

Houtonderzoek verrijkt kennis over kunst

Het RKD - Nederlands Instituut voor Kunstgeschiedenis stelt een grote database over houtonderzoek in kunsthistorische context beschikbaar. *Sytske Weidema*

Door het meten van jaarringen in hout stellen dendrochronologen vast vanaf wanneer een kunstwerk kan zijn gemaakt. De gemeten gegevens tonen de breedtes van iedere waargenomen jaarring in een plank die is verwerkt in het kunstobject. De jaarringbreedte verschilt per boomsoort en wordt bovendien beïnvloed door het klimaat tijdens de groei van een boom. Daardoor ontstaan unieke variaties in de jaarringbreedtes per periode en houtsoort. Dit jaarringpatroon wordt vergeleken met jaarringchronologieën, referentiekalenders van duizenden bomen. Zo worden de leeftijd en herkomst van het hout bepaald.

Wood for Goods

Duizenden van deze dendrochronologische gegevens zijn nu bijeengebracht en gecombineerd met achtergrondinformatie en literatuur. Het resultaat is online beschikbaar gemaakt in Dendro4Art, een internationale samenwerking van het RKD - Nederlands Instituut voor Kunstgeschiedenis, het Center for Art Technological Studies and Conservation van de National Gallery of Denmark en verschillende onderzoekers. De basis wordt gevormd door dendrochronologie-rapporten, meetreeksen en -werktekeningen



Dendrochronoloog dr. Domínguez Delmás (UvA) bestudeert het hout van het schilderij *Groentenmarkt* van Joachim Beuckelaer (KMSK Antwerpen). Onderzoek naar het hout in kunstwerken kan extra informatie blootleggen over deze objecten. Credits: Dr. Marta Domínguez Delmás

van em. prof. Peter Klein van de Universität Hamburg vanaf de jaren '70 tot heden.

Dr. Marta Domínguez Delmás leidt aan de UvA het NWO Veni-project *Wood for Goods* en is als wetenschappelijk adviseur verbonden aan Dendro4Art: "De combinatie van dendrochronologische en kunsthistorische data maakt de RKD-database uniek. In andere dendrochronologische databases ontbreekt de relatie met de historische context. Deze combinatie geeft ons een waardevol gereedschap om bredere onderzoeksvragen te beantwoorden,

bijvoorbeeld over historische houthandel in Noord-Europa en de ontwikkeling daarvan door de tijd heen."

Nieuwe data

Het portaal is gekoppeld aan RKD-technical, een database met uiteenlopende technische onderzoeksgegevens over kunstwerken. Ieder 'dendro-record' bevat per kunstwerk gegevens zoals houtsoort, jongste-, en aantal jaarring(en), mogelijke boom-veldatum, vanaf wanneer het kunstwerk kan zijn vervaardigd, enzovoort. Via het

RKD wordt continu nieuwe data toegevoegd aan Dendro4Art.

Bij dendrochronologie is diversiteit en duurzaamheid van de data belangrijk. Domínguez Delmás: "Digitale repositories zoals Dendro4Art zijn nodig om data en metadata over kunsthistorische objecten voor de lange-termijn te kunnen borgen. Dat bevordert de transparantie van onderzoeksresultaten en bevordert het hergebruik van deze data voor verder onderzoek. Dat is belangrijk, want toegang tot kunstobjecten is beperkt."

dendro4art.org

Rijksuniversiteit Groningen brengt studenten de basis van data bij

Succesvolle minor Data Wise

De minor *Data Wise* aan de Rijksuniversiteit Groningen leert studenten om vanuit verschillende disciplines met data te werken.

Lucas van der Meer

"Studenten breken hun hoofd over vragen als: hoe kan data worden gebruikt als bewijs? Is het ethisch verantwoord om algoritmen toe te passen op een maatschappelijk probleem? Hoe kunnen we inzicht krijgen met behulp van visualisaties?" Aan het woord is Gert Stulp, socioloog aan de Rijksuniversiteit Groningen (RUG) en mede-initiator van de minor *Data Wise*. "De minor is

opgezet om studenten in aanraking te brengen met verschillende invalshoeken van verantwoord data-gebruik, zodat ze in hun carrière met relevante experts kunnen communiceren."

Leren van elkaar

In september 2019 startte de minor met het maximum van 60 studenten. Programmadirecteur Anne Beaulieu: "De kern van de minor is een *collaborative data project* waarin een team van vijf studenten werkt aan een data gerelateerde casus van een externe organisatie zoals Philips of het Nederlandse Rode Kruis. Ter voorbereiding op het dataproject

krijgen de studenten de vakken *introduction to data*, waarin ze een basis en 'gezamenlijke taal' worden bijgebracht en *dynamics of multidisciplinary teamwork*, waarin studenten leren in teams te werken en elkaars vaardigheden te herkennen en waarderen. Het is prachtig om te zien dat elke student vanuit diens eigen expertise kijkt – van bedrijfskunde tot psychologie en astronomie tot geschiedenis – maar dat ze alleen gezamenlijk tot een oplossing komen. Hierdoor leren de studenten de taal van elkaars discipline te spreken." Gert vult aan: "Een student sociale wetenschappen heeft meer ervaring met een kri-

tische kijk naar data over mensen en het trekken van gerechtvaardigde conclusies, een student informatica zal vaardiger zijn in het verwerken en analyseren van data." Studenten kiezen ook nog drie uit zes verdiepende keuzevakken.

Voor onderzoekers

Op de vraag of elke sociale- of geesteswetenschapper goed moet kunnen programmeren, antwoorden beide ontkennend. Anne: "Niet iedereen hoeft te kunnen programmeren om iets zinnigs met data te kunnen doen, net zoals niet iedereen weet

Lees verder op pagina 2



E-DATA & RESEARCH

Jaargang 14 | 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, KNAW Humanities Cluster, de Koninklijke Bibliotheek, ODISSEI en het Rijksmuseum.

INHOUD

3 Hoe zoekt de onderzoeker naar secundaire data?

4 Ook deze datasets zijn sinds kort beschikbaar

4 Teamwork verbetert OCR gotische druk



5 Van Eijnatten wil research community versterken

6 Jong Talent Stork maakt data machine-leesbaar

7 ODISSEI meet sociale hartslag Nederlanders

8 Met AI naar algoritme met 90% score

8 Gastcolumnist Marieke over tijd voor nieuwe app

E-data wordt gratis toegezonden aan relaties van de stakeholders. Ook een uitgave ontvangen? Mail de redactie: edata@dans.knaw.nl.



Scan deze QR-code om edata.knaw.nl te bezoeken.

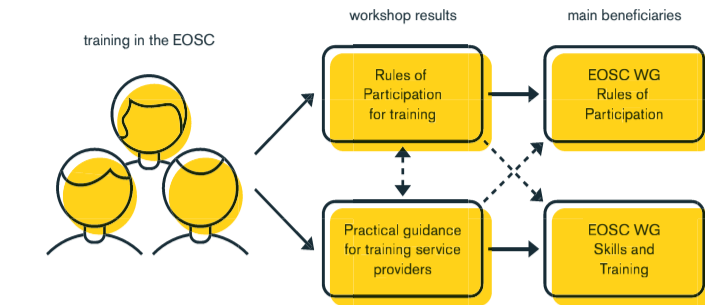
GEHOORD & BIJGEWOOND

SURF VRE
informatiebijeenkomst

Lucas van der Meer

In het huidige onderzoekslandschap wordt interdisciplinair en internationaal samenwerken, het delen van data en het beschikken over *high performance* rekenkracht steeds belangrijker. Een *Virtual Research Environment* (VRE), of virtuele onderzoeksomgeving, biedt hiervoor een uitkomst. Hoger onderwijsinstellingen kunnen eenvoudig hun eigen schaalbare VRE opzetten met de door SURF ontwikkelde Research Cloud. Kennis hierover werd op 7 april gedeeld via Zoom, met een recordaantal van 90 deelnemers.

Rogier de Jong, Ivar Janmaat en Hylke Koers (allen SURF) lichtten enkele componenten van SURF Research Cloud toe: SURF Research Drive, SURF Research Access Management (SRAM) en iRODS. SURF Research Drive stelt onderzoekers in staat om op een veilige manier bestanden met elkaar te delen. SRAM, voorheen Science Collaboration Zone, biedt mogelijkheden om de identiteit van gebruikers vast te stellen, deze in groepen in te delen en toegang te verlenen. SRAM wordt naar verwachting in het derde kwartaal van 2020 beschikbaar gesteld. iRODS biedt oplossingen voor researchdatamanagement, waaronder annotatie en herleidbaarheid van onderzoeksgegevens. De recent aangebrachte koppeling tussen iRODS en Research Drive wordt de komende periode getest en uitgebreid. SURF verwacht dat het ook iRODS binnenkort kan aanbieden. Andere componenten van SURF Research Drive zijn het centraal administreren van gebruik en het virtualiseren van omgevingen. De organisatoren benadrukten dat het VRE-ecosysteem continue in ontwikkeling is en dat SURF-leden hier inspraak in hebben. Ten slotte lichtte Elo Bosma toe hoe het Erasmus MC (EMC) VRE's heeft ingericht, waarbij is gekozen voor een combinatie van Microsoft Azure plus compo-



Tijdens de EOSC Training workshop brainstormden dertig trainingscoördinatoren over de vraag hoe trainingsmaterialen en diensten op het gebied van trainingen gemakkelijker toegankelijk kunnen worden.

Credits: Janno de Jong

nenten van SURF. Hierin werkt het EMC in het consortium anDREA nauw samen met het Radboud UMC. Het EMC had gepland eind 2020 zo'n 100 VRE's te hebben ingericht, maar verwacht door thuiswerken dat punt veel eerder te bereiken.

Meer informatie over de volgende VRE-bijeenkomst staat op de website van SURF.

surf.nl

Trainingsworkshop
over aanbod in EOSC

Ellen Leenarts

In allerlei Europese projecten en onderzoeksinfrastructuren wordt gewerkt aan het ontwikkelen en aanbieden van services in de European Open Science Cloud (EOSC) voor onderzoekers en onderzoeksinstellingen, services op het gebied van dataopslag, open en FAIR data, maar bijvoorbeeld ook voor de analyse van big data. Om deze services goed te kunnen gebruiken, wordt een scala aan trainingen georganiseerd voor verschillende doelgroepen, variërend van service providers tot onderzoekers. Zo kwamen eind februari dertig trainingscoördinatoren uit verschillende Europese projecten en onderzoeksinfrastructuren in Den Haag bijeen om te zorgen dat trainingsmaterialen en diensten op het gebied van trainin-

gen, net als services en data, gemakkelijker toegankelijk worden. In samenwerking met de Europese infrastructuur OpenAIRE, EGI en EUDAT en met subsidie van het EOSC-secretariaat was een tweedaags programma opgezet. Na korte introducties door Rene Belsø van de EOSC Working Group on Rules of Participation en Iryna Kuchma van de EOSC Skills and Training Working Group was het tijd voor de trainingscoördinatoren om gezamenlijk aanbevelingen te formuleren op het gebied van *Rules of Participation* voor het geven van trainingen in de EOSC en om praktische richtlijnen op te stellen voor aankomende trainers en trainingsorganisaties die willen deelnemen aan de EOSC.

Gedurende de talrijke breakout sessies werd gewerkt naar een rapport met aanbevelingen. Dit rapport is aan de EOSC-werkgroepen gepresenteerd en openbaar gemaakt via Zenodo.

DOI: 10.5281/zenodo.3739055

Webinar SSHOC over
dataopslagdienst voor SSH

Marieke Willems

Op 18 maart organiseerde SSHOC een webinar voor serviceproviders van het Consortium van European Social Science Data Archives (CESSDA), om te bespreken wat hun

ideeën zijn over een SSHOC Dataverse service.

SSHOC staat voor Social Sciences & Humanities Open Cloud. SSHOC heeft als doel om initiatieven van de huidige Europese onderzoeksinfrastructuren op het gebied van Sociale- en Geesteswetenschappen (SSH) beter op elkaar en op de Europese Open Science Cloud (EOSC) te laten aansluiten, om zo onderzoekers binnen het SSH-domein beter van dienst te zijn.

Eén van de diensten die SSHOC ontwikkelt, is een dataopslagdienst voor SSH-instellingen. Deze nieuwe dienst, gebaseerd op Dataverse-software, zal worden aangepast aan de behoeften van de Europese onderzoeksinfrastructuren, denk hierbij aan vertalingen van de User Interface, aanpassing van het metadata formaat en de ontwikkeling van data viewers. Dataverse (dataverse.org) is open source software, ontwikkeld door het instituut voor kwantitatieve sociale wetenschappen (IQSS) van de Universiteit van Harvard. SSHOC Dataverse wordt ontwikkeld onder leiding van DANS.

SSHOC's virtuele discussie startte met een presentatie van de huidige functionaliteit, gevolgd door een overzicht van nieuw te ontwikkelen features. Na deze presentaties werd input van de deelnemers verzameld, en richtte de geëngageerde discussie zich op essentiële vereisten voor een dergelijke service, voorkeuren, organisatie en noodzakelijke training. Bijna alle CESSDA serviceproviders waren vertegenwoordigd. In totaal namen 47 mensen deel aan de discussie.

De volgende stap in de SSHOC Dataverse ontwikkeling zijn virtuele discussies met drie andere Europese onderzoeksinfrastructuren binnen het project: DARIAH, CLARIN en E-RIHS. De opnamen en presentatie van de virtuele discussie staan op de website van SSHOC.

sshopencloud.eu/sshoc-webinar-cessda-service-providers-dataverse

KORT

EC geeft 5 miljoen
euro aan ESS

European Social Survey (ESS) heeft van de Europese Commissie zo'n 5 miljoen euro aan financiering ontvangen. Daarmee kan ESS sinds begin dit jaar het project 'SUSTAIN 2' uitvoeren. ESS is een tweemaaljaarlijkse onderzoek in verschillende Europese landen. Duizenden burgers beantwoorden tijdens persoonlijke interviews een gestandaardiseerde vragenlijst, met vragen over bijvoorbeeld menselijke waarden, immigratie en klimaatverandering. Een groot deel van de nieuwe financiering komt ten goede aan centrale ondersteunende tooling voor de dataverzameling in 12 landen. ESS werkt hierbij nauw samen met CentERdata. CentERdata gaat verschillende nationale teams voorzien van software voor *computer-assisted personal interviewing* (CAPI), management van vertalingen, steekproefbeheer en data-oplevering. (MdB) europeansocialsurvey.org

Tool voor verhogen
FAIRness data

In het kader van het FAIRsFAIR-project werkt DANS aan een FAIR self-assessment tool voor data deponeerders. De tool biedt zo'n vijftien assessmentvragen die alle aspecten van FAIR reflecteren. Door deze vragen te beantwoorden, ontvangt de deponeerder een score, met bijbehorende tips voor het verhogen van de FAIRness van de dataset. De deponeerder kan deze tips opvolgen alvorens de data bij DANS te deponeren. Een eerste versie van de tool genaamd SATIFYD bevindt zich momenteel in de testfase. Naast SATIFYD werkt DANS aan een generieke versie van de tool, die door andere repositories eenvoudig aangepast en geïmplementeerd kan worden. (JdV) satifyd.dans.knaw.nl/

Rust zacht

Machteld Maris (1970-2020)

Stil en verdrietig zijn we om het overlijden van Machteld Maris. Machteld was redactielid van dit blad namens het KNAW Humanities Cluster in de periode 2017-2019. We hebben haar leren kennen als een bevlogen mens, met hart voor het blad en aandacht voor iedereen. We wensen haar naasten veel kracht in deze moeilijke tijd.



Machteld Maris.

Credits: Martin van Welzen

Vervolg van pagina 1

Minor RUG
Data Wise

hoe een auto precies werkt om goed te kunnen rijden." Gert voegt toe: "Maar het is wel nuttig om minstens een beetje te weten wat programmeren precies inhoudt om zinvol met computerexperts te kunnen spreken." Toch vindt Gert het jam-

mer dat studenten in de sociale- en geesteswetenschappen niet meer in aanraking komen met programmeren: "Opleidingen kunnen helpen om de drempel van het werken met computercode te verlagen door voor statistiekonderwijs het technische programma R te gebruiken in plaats van een programma met een volledige grafische interface als SPSS. Hierdoor leren ze iets van programmeren, maar krijgen ze als bonus ook meer inzicht in hun data."

Anne: "Onze studenten krijgen een goede basis om in de wetenschap te werken. Onderzoek wordt steeds meer in teams en in gesprek met de maatschappij gedaan. Het zou mooi zijn als de aanpak van dit programma uitgebreid kan worden naar *graduate programmes* voor onderzoekers. Er is behoefte aan innovatie én verantwoord gebruik van data, en dat leer je bij *Data Wise*."

rug.nl/gmw/education/minor-datavise

COLOFON Uitgever: E-data & Research. Redactieadres: Anna van Saksenlaan 51, 2593 HW Den Haag, 070-3494450, edata@dans.knaw.nl, edata.nl.

Hoofd-/eindredacteur: Heidi Berkhout. Redactie: Marika de Bruijne, Steven Claeysens, Maarten Heerlien, Mathilde Jansen, Lucas van der Meer, Erica Renckens, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning. Aan dit nummer werkten mee: Valentijn Gilissen, Kathleen Gregory, Marjan Grootveld, Marieke Houben-Van Herten, Janno de Jong, Rutger van Koert, Ellen Leenarts, Iris Muste, Steamwork Graphics, Thijs Stork Photography, Jerry de Vries, Sytske Weidema, Marieke Willems.

Opmaak: Colette Sloots, Haarlem. Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Oplage: 5200 papier, 5000 digitaal.

ISSN: 1872-0374. We hebben getracht alle belanghebbenden met betrekking tot het gebruikte beeldmateriaal te benaderen. Degenen die menen rechten te kunnen doen gelden, kunnen zich tot ons wenden. Toezending papieren en/of digitale versie is kosteloos aan relaties van de stakeholders en studenten in de alfa- en gamma-richtingen.

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.

De cruciale rol van data-archieven

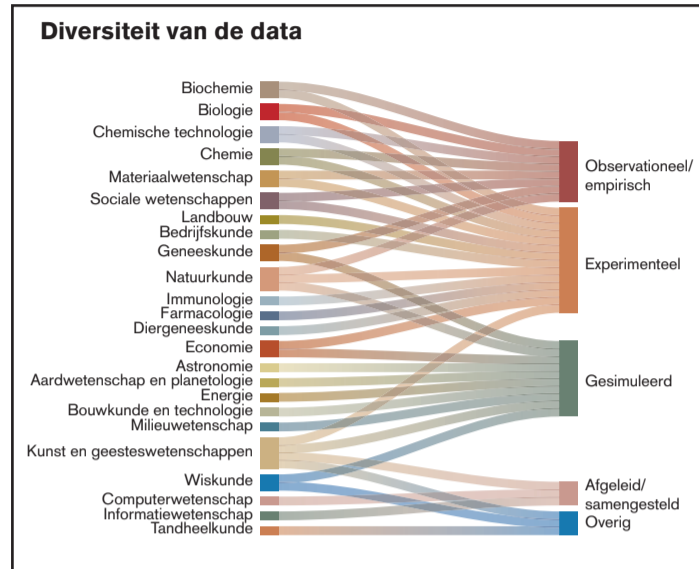
Hoe zoeken gebruikers naar data van andere onderzoekers?

Hergebruik van door anderen gecreëerde data kan een veelbelovende onderzoeksstrategie zijn, maar er is weinig bekend over de manier waarop gebruikers naar dergelijke data zoeken. *Kathleen Gregory*

Data-archieven spelen een belangrijke rol bij het ontsluiten van zogenoemde secundaire data. Toch is er maar weinig bekend over de manier waarop gebruikers naar deze data zoeken. Wat zijn hun motieven? Waarvoor hebben ze de data nodig? Wat voor soort data zoeken ze?

1.677 respondenten

Onlangs hebben onderzoekers van DANS, de Universiteit Maastricht en de Universiteit van Amsterdam de resultaten van het grootste bekende onderzoek naar deze vragen gepubliceerd in het kader van een project gefinancierd door NWO creative Industries Grant in samenwerking met Elsevier. 1.677 respondenten uit 105 landen, uit verschillende disciplines, met verschillende rollen en in verschillende fasen in hun academische loopbaan, vulden de vragenlijst in. De uitkomsten van het onderzoek, gepubliceerd in het artikel



Uit 'Diversiteit van de data' (linker grafiek) blijkt de diversiteit van de data die onderzoekers uit verschillende disciplines nodig hebben. Iets meer dan de helft van de respondenten geeft aan data van buiten de eigen discipline nodig te hebben.

'Gebruik van data' (rechter grafiek) toont het kerngebruik van de data door de respondenten; 71% zei data te gebruiken als basis voor een nieuwe studie. De grafiek laat zien dat het datagebruik wordt bepaald door discipline domeinen en methoden; zo worden domeinen die doorgaans gebruikmaken van computeronderzoeksmethoden, geassocieerd met het gebruik van data voor model- en algoritme-input. Voor beide afbeeldingen geldt: statistisch significante associaties gedetecteerd met Bonferroni-correctie voor gelijktijdige paarsgewijze marginale onafhankelijkheid; $n = 1677$; significantieniveau: $p < 0,05$. Credits: Kathleen Gregory en Steamwork Graphics

Lost or found? Discovering data needed for research (Gregory, K., Groth, P., Scharnhorst, A. & Wyatt, S. (2020). Harvard Data Science Review. DOI: 10.1162/99608f92.e38165eb), geven inzicht in de databehoeften van de respondenten, de bronnen en strategieën die ze gebruiken om data te vinden en de criteria die

ze hanteren om die data te evalueren.

De onderzoeksgegevens zijn vrij toegankelijk via EASY: Gregory, K.M (Data Archive and Networked Services) (2020): Data Discovery and Reuse Practices in Research. Dit artikel is een vertaling van het

Engelstalige artikel van Kathleen Gregory, M.A., MSLIS, PhD-kandidaat, werkzaam bij het project Re-SEARCH bij Research & Innovatie binnen DANS.

DOI: 10.17026/dans-xsw-kkeq

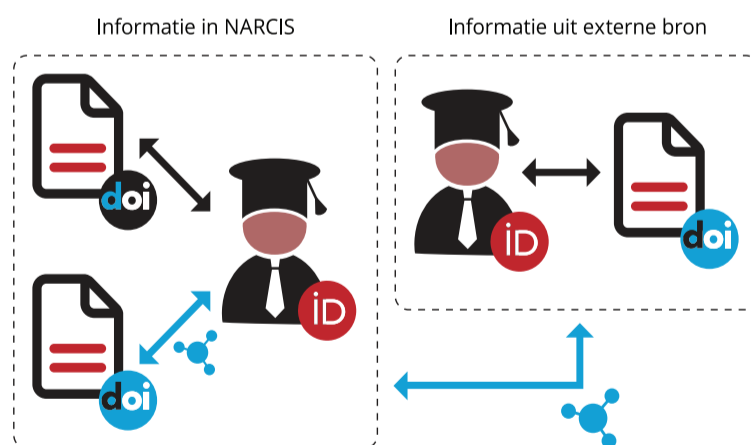
DANS verbindt en verrijkt beschikbare informatie in NARCIS

PID Graph van de Nederlandse wetenschap

Benieuwd naar informatie over de Nederlandse wetenschap? De portal NARCIS.nl biedt steeds meer informatie door de automatische koppeling van persistent identifiers in een PID Graph.

Ricarda Braukmann

Een persistent identifier (PID) is vergelijkbaar met een Burgerservicenummer voor digitale objecten; het is een uniek en blijvend nummer waaraan allerlei informatie is gekoppeld. Het gebruik van PIDs is van groot belang om op een betrouwbare manier te kunnen verwijzen naar het juiste artikel of de juiste persoon of organisatie. Aan veel wetenschappelijke informatie is dan ook zo'n PID toegekend. Door deze PIDs en door de koppeling van een PID aan een andere PID, verbetert de vindbaarheid, toegankelijkheid en navigatie van onderzoeksgegevens. Het koppelen



PID Graph maakt relaties zichtbaar. In dit voorbeeld zijn twee papers van 1 onderzoeker via NARCIS beschikbaar. Van 1 paper is bekend dat het hoort bij deze onderzoeker. Omdat NARCIS eigen informatie linkt met informatie uit externe bronnen (bijvoorbeeld ORCID), wordt ook de link met het tweede paper bekend. De NARCIS PID Graph haalt deze externe informatie op en integreert het in NARCIS, waardoor de link tussen het tweede paper en de onderzoeker ook zichtbaar wordt, gevisualiseerd met het blauwe PID Graph symbool. Credits: Ricarda Braukmann

van PIDs kan automatisch doordat PIDs ook leesbaar zijn voor computers. Het Europese project FREYA onderstreept de kracht van PIDs en bouwt aan een infrastructuur voor PIDs als essentieel onderdeel van de

European Open Science Cloud (EOSC).

De nationale portal NARCIS.nl biedt informatie over (open access) publicaties, datasets, onderzoeksprojecten, wetenschappers en onder-

zoeksinstellingen. Onlangs is de informatie in NARCIS verrijkt met behulp van bestaande en nieuwe PIDs. Zo zijn onderzoeksorganisaties nu eenduidig identificeerbaar in NARCIS doordat ze nu de nieuwe PID research organisation identifier (ROR) bevatten.

Relaties zichtbaar

Door PIDs met elkaar te verbinden in een PID Graph kunnen meer relaties zichtbaar gemaakt worden in NARCIS. Zo vergelijkt NARCIS bijvoorbeeld de informatie van het portaal van ORCID (een PID voor onderzoekers) met de beschikbare informatie in NARCIS en linkt publicaties aan de onderzoeker als deze nog niet als auteur was herkend. Op dezelfde manier is een link gemaakt met unpaywall.org, waardoor NARCIS nu ook informatie geeft over open access versies van bepaalde wetenschappelijke artikelen. Een laatste voorbeeld is de beschikbaarheid via NARCIS van de informatie van altmetric.com, een webservice over de online impact van papers of datasets op bijvoorbeeld twitter of facebook. De NAR-

CIS PID Graph maakt het ook mogelijk om resultaten te verzamelen van bepaalde onderzoeksprojecten. In een pilot project hebben het Donders Instituut, de Radboud Universiteit en DANS publicaties en datasets uit het NWO-gefinancierde project Language in Interaction aan elkaar gekoppeld.

De PID Graph biedt dus veel mogelijkheden om de informatie in NARCIS te verrijken. Hiervoor is het wel van belang dat PIDs door onderzoekers en instellingen worden gebruikt en meegeleverd in de metadata.

narcis.nl

project-freya.eu

AGENDA

Door het coronavirus zijn veel evenementen geannuleerd of is het onduidelijk of evenementen doorgaan. Veel organisaties bieden digitale alternatieven, minstens zo interessant. We adviseren u om online te zoeken naar actuele informatie.

SINDS KORT BESCHIKBAAR

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

CentERdata

• Denkend aan Nederland

Wat zijn de meest typerende kenmerken voor Nederland? En welke factoren dragen bij aan het gevoel van verbondenheid met Nederland? Dit heeft het Sociaal en Cultureel Planbureau (SCP) onderzocht in 2019. Het doel was om de Nederlandse identiteit



Credits: SCP

in beeld te brengen. Daartoe is gebruikgemaakt van het LISS panel. Vanwege de omvang van het onderzoek werden de vragen in twee metingen afgenomen van juli tot en met september 2019. Mede op basis van deze data is het 'Sociaal en Cultureel Rapport 2019 - Denkend aan Nederland' gepubliceerd. De data zijn beschikbaar via LISS Data Archive.

lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Abidi, L.; Nilsen, P., april 2017, Implementation of alcohol prevention in healthcare in the Netherlands
- Kok, L., april 2017, Pension designs and continued working after retirement
- Portegijs, W., juni 2018, Emancipatiemonitor 2018
- CentERdata, oktober-november 2018,

Social Integration and Leisure - Wave 11

- CentERdata, juni-juli 2019, Economic Situation: Income - Wave 12
- CentERdata, juli-augustus 2019, Economic Situation: Housing - Wave 12



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

DANS

• Nieuw in EASY:

Maritieme opgravingsdossiers

Sinds de inpoldering van de Wieringermeer zijn honderden scheepswrakken aangetroffen, vergaan op de voormalige Zuiderzee. De documentatie hiervan varieert van eenvoudige meldingen tot volledig uitgevoerde archeologische opgravingen. Het gaat om waardevolle en unieke brondocumentatie voor (scheeps)archeologisch onderzoek. Stichting Batavialand beheert zowel de ar-



Credits: DOI: 10.17026/dans-2z5-jmy2

cheologische objecten als het bijbehorende archief van de maritieme rijkscollectie namens de Rijksdienst voor het Cultureel Erfgoed. Onlangs heeft de Stichting de papieren (scheeps)archeologische opgravingsdocumentatie gedigitaliseerd. Deze bijzondere collectie wordt nu toegankelijk gemaakt via EASY. DOI: 10.17026/dans-x6z-3dnp.

Ook sinds kort beschikbaar:

De volgende datasets zijn open access beschikbaar via het online archiveringsysteem EASY van DANS:

- Berkel, dr. R. van (Utrecht University) (2020): Versterking methodisch werken via HRM. DANS. DOI: 10.17026/dans-x3w-7q4b.
- Farace, dr. D. (GreyNet International) (2020): Grey Literature Resources generate and drive Awareness to the Circular Economy. DANS. DOI: 10.17026/dans-zh-zkg3z.
- Frankena, dr. K. (Wageningen University) (2020): ROMAN, Few-Foods-Diet and ADHD in Practice. DANS. DOI: 10.17026/dans-xn4-6pjh.
- Gregory, K.M. (Data Archive and Networked Services) (2018): Data Discovery and Reuse Practices in Research. DANS. DOI: 10.17026/dans-xsw-kkeq.
- Heine, F.A. (Tilburg University) (2020): Using Moral Foundations in Government Communication to reduce Vaccine Hesitancy. DANS. DOI: 10.17026/dans-xuv-vyzk.
- Jordanov, drs. M.S. (RAAP) (2020): Kasteelpark IJsselstein, gemeente IJsselstein, een archeologische opgraving. DANS. DOI: 10.17026/dans-z33-gtvv.

- Leemans, L.H. (Radboud University) (2020): A mutualism between unattached coralline algae and seagrasses prevents overgrazing by sea turtles. Ecosystems. DANS. DOI: 10.17026/dans-25p-82rx.
- Lutkie, T. (2019): De pot en de ketel: Nederlandse dagbladen en hun oordeel over communisme en fascisme, 1918 - 1939. DANS. DOI: 10.17026/dans-zeq-tnrx.
- Moretta, dr. T.M. (Department of General Psychology, University of Padova) (2019): Data from problematic and non-problematic Facebook users who performed a Go/Nogo task with Facebook-related, pleasant, unpleasant, and neutral pictures and a self assessment manikin (SAM). DANS. DOI: 10.17026/dans-zqm-d9zh.
- Nollen, drs. J.H. (Gemeente Breda) (2020): Breda Kasteelplein (AO). DANS. DOI: 10.17026/dans-zxr-3xtd.
- Scholtens, J. (Commissariaat voor de Media) (2019): Representatie van mannen en vrouwen in Nederlandse non-fictie televisieprogramma's 2019. DANS. DOI: 10.17026/dans-27s-4q6g.
- Sociaal en Cultureel Planbureau (SCP) (2018): Vrouwen in besluitvorming 2018 - VIB2018. DANS. DOI: 10.17026/dans-26j-7rw8.
- Westen, dr. C.J. van (University of Twente) (2020): Landslide inventory of the 2018 monsoon rainfall in Kerala, India. DANS. DOI: 10.17026/dans-x6c-y7x2.



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

KORT

Nieuwe directeur DANS omarmt open science

Sinds 1 april is Henk Wals directeur van DANS. Zijn visie is helder: "Waar het om draait, is de beweging richting *open science*. Naarmate onderzoeksdata en -resultaten sneller en beter gedeeld worden, versnelt de kenniscirculatie en boekt de wetenschap in een hoger tempo resultaten. In Nederland houden ruim honderd organisaties zich bezig met data, opslag, infrastructuur, etc. Hoe ordenen we dat landschap, welke afspraken maken we over het verbinden van onderzoeksgegevens en hoe voorkomen we duplicatie van services? Gelukkig zijn er initiatieven als het Nationaal Platform Open Science en de European Open Science Cloud. Samen bewegen we richting een netwerkorganisatie met goede taakverdeling, coördinatie en afspraken. DANS heeft alles in zich om een nuttig knooppunt te vormen in een netwerk van Nederlandse en Europese instellingen die bijdragen aan de data-infrastructuur. Wij zijn bereid ons aan zo'n rol te committeren. En KNAW en NWO steunen deze gedachte, is mij verzekerd." (HB)

dans.knaw.nl

Mooie resultaten met Optical Character Recognition

Teamwork verbetert OCR gotische druk

OCR is een interessante tool met vele toepassingen. Of het ook werkt voor Nederlandse gotische druk, werd tijdens de workshop *ICT with Industry* onderzocht. *Rutger van Koert*

Optical Character Recognition (OCR) staat voor optische tekenherkenning: een methode waarbij een computer door middel van patroonherkenning tekens uit een afbeelding haalt. OCR werkt over het algemeen vrij goed op modern materiaal. Helaas gaat de kwaliteit van de herkenning achteruit naarmate het materiaal ouder is. Ook bij 'vreemde' fonts, vlekken en vervuiling verslechtert de kwaliteit. Genoeg motivatie om tijdens de jaarlijkse, door het ICT Research Platform Nederland (IPN) georganiseerde workshop *ICT with Industry* afgelopen februari aan deze wetenschappelijke uitdaging te werken.

Vier subproblemen

Door het team werden vier subproblemen gedefinieerd: preprocessing inclusief voorbereiden van de scans, segmentatie van de scans op woord- of zinsniveau, herkenning (de daadwerkelijke OCR) en postprocessing, het automatisch corrigeren van fouten van de herkenning. Samen met Mirjam Cuper (KB) zorgde ik voor scans, transcripties en rekenkracht voor de machinelearning, Jerry Guo (TU Delft) probeerde diverse algoritmes uit. Visueel was de verbetering goed zichtbaar, maar de resulterende OCR-output verbe-



Titel-scan van het boek 'Geleentheit van s Hertogen-Bosch' door Pieter Bor, geschreven in 1630. De binnenzijde van dit boek bevat teksten in gotisch schrift, OCR-technieken maken het onderzoekers makkelijker om de teksten te gebruiken. Credits: KB

terde nauwelijks. Voor de segmentatie, het tweede subprobleem, gebruikten we ARU-net. Samen met Xue Wang (CS, Leiden University) trainde ik het systeem op het detecteren van spaties en woorden met hulp van data van de ALTO-xml van een commerciële OCR-provider. We controleerden de resultaten weer visueel, op sommige pun-

ten was er zelfs een verbetering ten opzichte van de commerciële provider. Vervolgens werd Monk door Lambert Schomaker (AI/ML RuG) ingezet om data te labelen en ging Mahya Ameryan (AI, RuG) woorden herkennen met machinelearning. 88% van de woorden bleek correct te zijn herkend, een mooie score! Als laatste namen Koen Dercksen (Radboud Universiteit) en Konstantin Todorov (ILLC, UvA) het nabewerken op zich door gebruik te maken van BERT, gefinetuned op het tekstcorpus van de Meertens Kranten (1662-1795) en aansluitend een LSTM encoder-decoder netwerk. Met het softwareplan van Adriëne Mendrik (e-Science Center) kunnen we resultaten kwantificeren en meten wat daadwerkelijk de beste opties zijn voor specifieke onderdelen.

Flinke verbetering

Via *ICT with Industry* hebben we, naast een leuke week met slimme mensen uit de wetenschap en het bedrijfsleven, mooie resultaten bereikt. Samen concluderen we trots dat het mogelijk is om de herkenning van Nederlands gotisch drukwerk flink te verbeteren. Het KNAW Humanities Cluster en de KB gaan kijken hoe deze pijplijn voor vroegmoderne druk verder kan worden ontwikkeld.

ict-research.nl/ict-with-industry

Rutger van Koert is Lead Engineer Team Images bij het KNAW Humanities Cluster.

Joris van Eijnatten, directeur eScience Center:

‘Je ziet dat het eScience Center heel actueel wordt’

Per 1 januari 2020 trad cultuurhistoricus Joris van Eijnatten aan als nieuwe directeur van het Nederlands eScience Center, het nationale research software-instituut. Komend najaar presenteert hij een nieuw strategisch plan voor het instituut.

Steven Claeysens

Het gaat goed met het Nederlands eScience Center. Opgericht in 2011 om ‘multidisciplinair en data-intensief wetenschappelijk onderzoek in Nederland te versterken en het gebruik van innovatieve ICT voor onderzoek te stimuleren’, groeide het uit tot de belangrijkste organisatie op dat terrein voor wetenschappelijk Nederland. Onderzoekers uit alle wetenschapsdisciplines kunnen meedingen naar de calls van het eScience Center. Honorering levert financiering en de bijstand van een ervaren Research Software Engineer (RSE) van het eScience Center op. In enkele jaren groeide het instituut van 25 naar 70 mensen, een internationale evaluatiecommissie publiceerde vorig jaar een lovend rapport.

Kennisoverdracht

Bij de start van Van Eijnatten vielen hem een aantal zaken op. “Er was een evaluatierapport, een zelfevaluatie en input uit interviews met interne en externe betrokkenen en belanghebbenden. Maar toen ik de stukken las, viel me op dat het eScience Center erg missie-gedreven was. De missie *Enabling digitally enhanced research* is helder, maar gaf geen richting. In de nieuwe strategie staat nu een heldere visie, die maakt dat we keuzes kunnen maken. Het hoofddoel is een *robust research community* die in staat is digitale tools en methoden toe te passen. Hiervoor willen we veel meer onze kennis gaan delen met anderen.”

“Vorig jaar hebben we community officers aangesteld, mensen die naar buiten treden om te kijken waar we trainingen kunnen verzorgen, waar welke vragen leven en hoe we daarop in kunnen spelen. Begin dit jaar organiseerden we een machine learning-cursus. We hadden 100 aanmeldingen, uit alle vakgebieden, maar konden er maar 25 accommoderen. In een week tijd hielpen we de cursisten zodanig op weg dat ze zelf verder konden. Ik voorzie dat we dat nog veel vaker zullen doen. Er is een enorme vraag hiernaar en we kunnen op dat vlak goed werk doen.”

Digital Competence Centers

NWO wil dat kennisinstellingen hun expertise op het gebied van researchdatamanagement en onderzoekssoftware op één plek bundelen: in een Digital Competence Center (DCC). Van Eijnatten: “Bij universiteiten en onderzoeksinstituten werken mensen aan de digitale infrastructuur en mensen aan het beheren van data, de datastewards. Daar investeren ze fors in. Maar de derde component zie je nog amper: de software engineers. Zij zitten te vaak verscholen in onderzoeksgroepen. Voor het eScience Center is het van belang



Joris van Eijnatten, directeur eScience Center: “Als je vijf jaar software in het onderwijs kan laten draaien, dan bereik je meerdere dingen tegelijk: verduurzaming én disseminatie. Dat zijn het soort constructies waar we heen willen.” Credits: Bart van Vliet

INTERVIEW

‘Research is voor ons het toverwoord. Wij bedienen het onderzoek’

om onze kennis onder meer via die DCC’s onder onderzoekers te verspreiden. Zo wordt ook de disseminatie geregeld.”

“Het verhaal van het eScience Center is ook een historisch verhaal. Er is een infrastructuur neergelegd vanaf de Tweede Wereldoorlog, met grote IBM-machines, computers, het internet, gevolgd door een enorme datagolf. Nu zitten we in de volgende fase: wat doen we met de software? Zowel vanuit de data-kant als vanuit de ICT-kant zie je een beweging richting de toegepaste kant. Dat moet ergens bij elkaar komen. Het eScience Center wordt zo heel actueel.”

Software sustainability

“Onze engineers hebben ideeën over hoe je software kunt verduurzamen, maar ook ideeën over

hoe je de kwaliteit van software kunt garanderen. Ook dat is kennisoverdracht. Daarmee hangt samen dat we eigenlijk meer zichtbaar willen zijn en meer zicht ook willen hebben op wat er precies gebeurt met software als we die eenmaal gemaakt hebben. Software blijft nu te vaak op de plank liggen. Dat zie je wereldwijd in heel veel projecten. Dat is erg inefficiënt, het gevaar is dat je een herhaling van zetten krijgt. We willen daarom commitment vragen van universiteiten. We blijven onderzoekers rechtstreeks bedienen bij het beantwoorden van onderzoeksvragen – dat is de kern van ons bestaan en dat gaat niet veranderen – maar we gaan onderzoekers die reageren op een call ook vragen om commitment van hun eigen instelling. Denk aan het inzetten van software in het onderwijs, een manier om de software een paar jaar te verduurzamen.”

Software en de FAIR-principes

“*Open science* leeft heel sterk in het Center. Onze wetenschappers staan vierkant achter *open science*, daar zijn ze mee opgeleid. Al onze software is open source. Waar dat nuttig is, willen we de software ook op een kwalitatief hoogwaardig niveau wegzetten: dusdanig gedocumenteerd dat anderen er meteen mee aan de slag kunnen. In dat verband hebben we samen met DANS fair-software.eu opgezet. Op dit moment worden de FAIR-principes vooral toegepast op data, maar een aantal FAIR-principes kun je zeker vertalen naar software.”

“Er werd bij het eScience Center veel gesproken over ‘scientific breakthroughs’. Als geesteswetenschapper let ik ook op de retorica. ‘Scientific’ is gericht op de ‘sciences’. Daar kom je bij de ‘social sciences’ misschien nog mee weg, maar bij de ‘humanities’ wordt deze taal niet gebruikt. En een ‘breakthrough’ binnen de geesteswetenschappen? Misschien op bepaalde gebieden en misschien dat methodologieën doorbraken kunnen zijn, maar kennisdoorbraak? Nee, het is meer verschuiving van perspectief wat wij zien. Dit soort termen probeer ik te vervangen, zodanig dat iedereen zich herkent in onze nieuwe strategie en zich uitgedaagd en welkom voelt bij het eScience Center.”

esciencecenter.nl

KORT

Platform Digitale Infrastructuren SSH gelanceerd

In het najaar van 2019 is het Platform Digitale Infrastructuren SSH (PDI-SSH) gelanceerd. PDI-SSH is een initiatief van het SSH-beraad en de digitale infrastructuren CLARIAH en ODISSEI. Het platform heeft drie taken: het ontwikkelen van een strategie voor digitale infrastructuur in Nederland, middelen toekennen aan digitale infrastructuure voorzieningen en de afstemming van die voorzieningen. De eerste *call for proposals* is recentelijk gesloten. De uitslag wordt in juni verwacht. (LvdM) pdi-ssh.nl

Let's play FAIR: DANS Data Game

Speciaal voor onderzoekend Nederland is de DANS Data Game ontwikkeld. Het kwartet geeft een indruk van het onderzoeksdataveld en is speciaal voor het 15-jarig bestaan van DANS geproduceerd. Het spel wordt op verzoek per post bezorgd, maar omdat Corona het fysiek spelen niet toestaat, is ook een online versie van het spel ontwikkeld. Meer informatie staat op de website van DANS. (HB) dans.knaw.nl



Het kwartet kan fysiek en online worden gespeeld.

Credits: DANS

Inschrijving gebruik ODISSEI Secure Supercomputer open

De ODISSEI Secure Supercomputer (OSSC) biedt onderzoekers de mogelijkheid om hun gekoppelde CBS Microdata te analyseren met behulp van de SURFsara supercomputer Cartesius. Na een succesvolle piloffase wordt de OSSC binnenkort voor ODISSEI-deelnemers opengesteld. Iedereen die goed gebruik kan maken van de OSSC kan een project op de wachtlijst plaatsen. (LvdM) edu.nl/7jh9q



De OSSC maakt gebruik van de rekenfaciliteiten van de Cartesius supercomputer van SURFsara. Credits: SURFsara

Friese teksten door online pijplijn

Tool voor taalkundig onderzoek Fries

De online tool **UDPipe Frysk** kent woordsoorten toe aan teksten in het Fries. Een dergelijke basistool ontbrak nog voor de tweede rijkstaal.

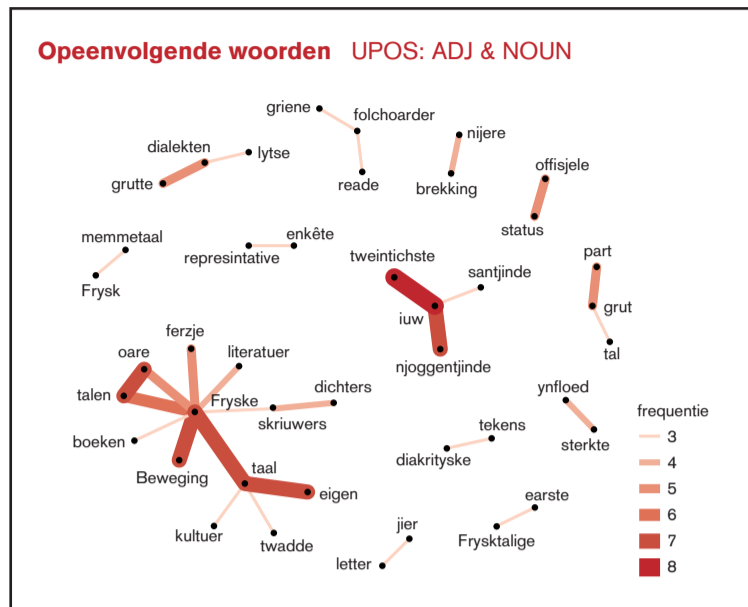
Erica Renckens

Onlangs verscheen de eerste update van de webapp **UDPipe Frysk**, die eind januari werd gelanceerd. Deze tool maakt taalkundige analyse van Friese teksten mogelijk. In de ingevoerde tekst worden de losse tokens (woorden) herkend en voorzien van lemma's en woordsoorten (POS-tags).

Webapp

“Een dergelijke basistool voor taalkundig onderzoek bestond nog niet voor de tweede rijkstaal in Nederland, het Fries”, vertelt Hans Van de Velde, die als projectleider aan de Fryske Akademy verantwoordelijk was voor de ontwikkeling van de tool. “POS-tags zijn belangrijk, omdat woordsoorten soms contextafhankelijk zijn. In de zin ‘De bern krige iisfrij’ (‘De kinderen krijgen ijsvrij’) is iisfrij bijvoorbeeld een zelfstandig naamwoord, maar in de zin ‘De mar is hielendal iisfrij’ (‘Het meer is volledig ijsvrij’) een bijvoeglijk naamwoord.”

Onderzoekers kunnen de webapp gebruiken voor onderzoek naar bijvoorbeeld taalverandering, syntactische verhoudingen, auteursher-



Op basis van het Wikipedia-artikel ‘Frysk’ (nl.wikipedia.org/wiki/Westerlauwers_Fries) kan **UDPipe Frysk** teksten analyseren. Zo laat de rechter afbeelding zien dat zelfstandige naamwoorden (NOUN) het meest frequent zijn gebruikt, gevolgd door voorzetsels (ADP) en lidwoorden (DET). De linker grafiek laat de combinaties zien van bijvoeglijk naamwoord (ADJ) en zelfstandig naamwoord (NOUN). Uit de tekst zijn wel de titels, opschriften, bijschriften, tabellen, links en referenties weggelaten. Credits: *UDPipe Frysk*

kenning, sentiment-analyse of voor de ontwikkeling van automatische vraag-antwoordsystemen. Van de Velde: “De gebruiker typt zelf een Friese tekst in, uploadt deze of voert een Friese website in.” Hierna verschijnt een tabel met voor elk token het lemma en de woordsoort. Deze output kan vervolgens in verschillende formaten (txt, excel, CoNLL-U) gedownload worden voor verdere analyse.”

Wilbert Heeringa, programmeur bij de Fryske Akademy, ontwikkelde

de pijplijn, samen met Gosse Bouma van de Rijksuniversiteit Groningen. Martha Hofman (Fryske Akademy) helpt bij het handmatig annoteren van het trainingscorpus. Heeringa gebruikte het project *Universal Dependencies* (UD). “Dat project ontwikkelt een universeel annotatieschema dat cross-linguïstisch vergelijken mogelijk maakt. Zo kunnen vergelijkbare constructies in verschillende talen op een consistente manier worden geannoteerd, terwijl ook taalspecifieke an-

notaties worden toegestaan als die nodig zijn.”

1.547 zinnen

Heeringa trainde de **UDPipe Frysk** met 1.547 zinnen uit het Oersettercorpus. Dit corpus is in 2012 ontwikkeld voor *Oersetter*, een automatische vertaalservice voor het Fries en het Nederlands. Het bevat onder andere nieuwsberichten, romans, wetenschappelijke teksten en historisch-culturele teksten.

In de eerste update, die half mei ver-

scheen, is dit trainingscorpus verder uitgebreid met meer zinnen. Bovendien zijn er data toegevoegd die *dependency parsing* mogelijk maken, zodat ook de grammaticale structuur van een zin met de onderlinge relaties tussen woorden in kaart gebracht kan worden. Ten slotte wordt ook gewerkt aan een analyse van de kwaliteit van de POS-tagging. **UDPipe Frysk** is mede tot stand gekomen dankzij financiering van CLARIAH-PLUS.

fryske-akademy.nl

Tool voor omzetten in machine-leesbare data

Historische biodiversiteit digitaal ontsluiten

Lise Stork helpt onderzoekers om moeilijk toegankelijke gegevens te verwerken met computationele technieken. Eind 2019 ontving ze hiervoor de **Young eScientist Award**. Maarten Heerlien

Onderzoeksinstellingen en natuur-musea herbergen veel archieven waarin gegevens over historische biodiversiteit zijn vastgelegd. Deze data zijn nog altijd relevant, maar de toepassing ervan wordt bemoeilijkt door het complexe, vaak ondoor-dringbare karakter van deze archieven. Lise Stork, PhD kandidaat aan het Leiden Institute of Advanced Computer Science (LIACS), combineert in haar onderzoek verschillende computationele modellen om dergelijke archieven toegankelijk te maken en zo het wetenschappelijk onderzoek te accelereren.

Digitaal vindbaar

Er zitten verschillende uitdagingen aan het ontsluiten van de informatie in het soort manuscripten dat Stork gebruikt, veelal soortbeschrijvingen en -schetsen van wetenschappers op



Winnaar van de **Young eScientist Award 2019** Lise Stork helpt onderzoekers om moeilijk toegankelijke gegevens te verwerken met computationele technieken. Credits: *Thijs Stork Photography*

onderzoeksexpedities in gebieden met een rijke flora en fauna. Voorbeelden van uitdagingen zijn bijvoorbeeld de kwaliteit van het handschrift, meertaligheid en verouderde terminologie. Stork gebruikt een innovatieve mix van methoden en technieken om belangrijke stukjes

informatie in de manuscripten machine-leesbaar, en daarmee digitaal vindbaar te maken. “Eerst modelleer ik elementen die in de manuscripten voorkomen aan de hand van achtergrondkennis uit het domein, bijvoorbeeld taxonomie, anatomie en geografie. Vervolgens ge-

JONG TALENT

‘Relatief eenvoudig historische manuscripten omzetten naar machine-leesbare data’

bruik ik beeldherkenning om deze specifieke elementen - soortnamen, anatomische kenmerken en locaties - automatisch terug te vinden. Deze elementen maak ik vindbaar aan de hand van de standaarden van het kennisdomein, met behulp van semantische webtechnieken.”

Naar een webomgeving

Haar onderzoek maakt deel uit van het NWO-project *Making Sense of Illustrated Handwritten Archives*. Stork heeft een workflow ontwikkeld waarmee onderzoekers relatief eenvoudig historische manuscrip-

ten kunnen omzetten in machine-leesbare data. De volgende stap is de ontwikkeling van een schaalbare en duurzame webomgeving, waarin wetenschappers historische onderzoeksarchieven betekenisvol kunnen ontsluiten. Zo kan er efficiënt door deze manuscripten worden gezocht en kunnen relaties worden blootgelegd. De ontwikkeling van deze omgeving wordt ondersteund door het eScience Center, dat in november 2019 de Young eScientist Award toekende aan Stork voor dit idee.

Mens centraal

Hoewel de focus in het onderzoek ligt op historische biodiversiteitsdata, zijn de resultaten van het onderzoek van Stork en haar mede-onderzoekers breder toepasbaar: “De essentie van deze technieken en workflow is dat de mens centraal staat: we helpen mensen bij lastige keuzes door ze van de juiste informatie te voorzien en suggesties te doen, daar waar beelddata met gestructureerde, terugkerende informatie een rol speelt.”

liacs.leidenuniv.nl/~storkl/

Twaalf onderzoeken ODISSEI Call 2020 in beeld

ODISSEI meet sociale hartslag van Nederlanders met dataverzameling

Wat zijn de gevolgen van de coronacrisis? Hoe staan Nederlandse patiënten tegenover kunstmatige intelligentie? Een verslag van 12 actuele dataverzamelingen gehonoreerd door ODISSEI.

Marika de Bruijne

In het najaar van 2019 konden ODISSEI-deelnemers een verzoek indienen om kosteloos data te verzamelen in het LISS panel van CentERdata. Dit panel is een representatieve afspiegeling van het bevolkingsregister van het CBS. De tien gehonoreerde onderzoeksprojecten variëren in onderwerpen van de mentale gezondheid van Nederlanders tot politieke onvrede. In totaal waren er 39 aanmeldingen. De vragenlijsten worden in de loop van 2020 afgenomen.

In maart 2020 veranderde het leven door de COVID-19 pandemie. Sociale wetenschappers volgen de maatschappelijke veranderingen op de voet. Om actueel onderzoek naar de sociale impact te ondersteunen, opende ODISSEI een extra call. Hiermee konden ODISSEI-deel-



Het LISS panel van CentERdata is een representatieve afspiegeling van het bevolkingsregister van het CBS. Credits: Eak K. via Pixabay

nemers een korte vragenlijst afnemen in het LISS panel over het coronavirus. Deze call had een verkort tijdschema. Begin april zijn twee van de 43 voorstellen geaccepteerd en de dataverzamelingen zijn gelijk in april gestart.

De 12 gehonoreerde voorstellen staan op deze pagina weergegeven. Op termijn krijgen wetenschappers toegang tot de data van deze 12 voorstellen via het LISS Data Archive. Dit draagt bij aan het doel van ODISSEI om een

LISS Data Archive vernieuwd

Het LISS Data Archive is recent vernieuwd. De website van deze data repository is gebruiksvriendelijker en beter geschikt gemaakt voor mobiele devices. De zoekfunctie is verbeterd en men kan weer een eigen 'winkelmand' met variabelen vullen. Ook is de beveiliging aangescherpt door tweefactor-authenticatie. Het LISS Data Archive biedt zo'n 12 jaar aan data aan, afkomstig uit onderzoeken in het LISS panel. De vragenlijsten, codeboeken en gerelateerde publicaties zijn vrij beschikbaar. Na gratis inschrijven zijn de data vrij toegankelijk voor wetenschappelijk en maatschappelijk relevant onderzoek.

dataarchive.lissdata.nl

nationale data-infrastructuur voor de sociale wetenschappen in Nederland te bieden. Later dit jaar wordt nog een nieuwe reguliere call geopend.

odissei-data.nl

De twaalf onderzoeken op een rij

Understanding attitudes towards refugees as compared to immigrants

Inge Hendriks, Marcel Lubbers, Michael Savelkoul

Dit onderzoek richt zich op de vraag in welke mate de houding van mensen tegenover vluchtelingen enerzijds en migranten anderzijds verschilt, en of dit ook geldt voor de verklaringen voor de houding.

Retrospective informal caregiving careers

Ellen Verbakel

In dit onderzoek staat de vraag centraal hoe de zorg van mantelzorgers zich ontwikkelt gedurende de mantelzorgperiode en welke langetermijneffecten dat heeft op hun arbeidsmarktpositie. Dit onderzoek combineert een survey in het LISS panel met bestaande data uit de LISS kernstudie.

Gender inequalities in times of the COVID-19 pandemic

Mara A. Yerkes et al.

Het thuiswerken in verband met de coronacrisis brengt een nieuwe dynamiek voor ouders met zich mee. Dit onderzoek richt zich op de veranderingen die daarmee gepaard gaan in de taakverdeling van betaald en onbetaald werk tussen mannen en vrouwen. De data worden gecombineerd met de LISS kernvragenlijsten over arbeid en huishouden.

An algorithm a day keeps the doctor away?

Marieke Haan

Kunstmatige intelligentie en algoritmes worden steeds meer toegepast in de geneeskunde en in de klinische praktijk. In dit onderzoek wordt 1) via een survey nader onderzocht hoe de algemene Nederlandse populatie hier tegenover staat en 2) middels een experimenteel design uitgezocht wat de effecten van verschillende schalen zijn op het antwoordgedrag van deze populatie.

Who is politically intolerant?

Linda Bos

Dit onderzoek naar politieke moraliteit, extremisme en intolerantie bouwt voort op een studie die in de vorige ronde van de ODISSEI-call is gehonoreerd. Het onderzoek maakt gebruik van de LISS kernstudie om verklarende factoren voor verschillende vormen van politieke intolerantie te analyseren.

Measuring identification with multiple political parties

Martin Rosema

De mate waarin mensen zich met één of meerdere politieke partijen identificeren, is het onderwerp van deze studie. In een longitudinaal onderzoek met drie metingen wordt gekeken in hoeverre deze identificatie stabiel is en met welke kenmerken dit samenhangt.

The Dutch Mental Health Continuum Short Form

Gerben J. Westerhof, Peter ten Klooster

GGZ-cliënten hebben soms moeite met het invullen van vragenlijsten om hun mentale gezondheid te meten middels het instrument MHC-SF. Om dit te voorkomen, is een herziene versie van het instrument ontwikkeld die in het LISS panel getest wordt.

The Corona outbreak and the school shutdown

Thijs Bol

Een van de maatregelen om het coronavirus te bestrijden, is het sluiten van de scholen waardoor kinderen thuis onderwijs moeten volgen. Aan ouders in het LISS panel wordt gevraagd hoe hun kinderen dit doen. Deze data worden gekoppeld met registerdata om effecten op de lange termijn te volgen.

What affects individual outcomes? Objective neighborhood characteristics or perceptions?

Heleen Janssen, Maarten van Ham

Door een combinatie van LISS panel data en CBS microdata wordt onderzocht in hoeverre objectieve kenmerken van een buurt of juist subjectieve ervaringen van een buurt invloed hebben op de beleving en opvattingen van mensen.

What is the relation between financial scarcity, decision-making and social exclusion?

Leon P. Hilbert, Frank T. Doolaard, Marret K. Noordewier, Wilco W. van Dijk

Dit onderzoek analyseert hoe financiële problemen zich in de tijd ontwikkelen en hoe dit relateert aan het nemen van financiële beslissingen en gevoelens van sociale uitsluiting.

How to reduce the education gap in participation in citizens' initiatives?

Vivian Visser, Willem de Koster, Jeroen van der Waal

In dit onderzoek wordt aan de hand van een vignettenstudie gekeken of de welbekende opleidingsverschillen in bereidheid tot deelname aan burgerinitiatieven afhangen van de manier waarop de uitnodiging wordt geformuleerd.

The Willingness, Opportunity and Ability to Prepare for Automation at the Workplace

Giedo Jansen, Suzanne Janssen, Mark Levels

Dit onderzoek gaat onder meer na welke arbeidsmarktkenmerken samenhangen met de bereidheid van mensen om nieuwe vaardigheden te leren voor het omgaan met toenemende automatisering en robotisering op de werkplek.

Het Nederlandse fotografische geheugen is in toenemende mate online beschikbaar. Hoe zorg je ervoor dat je daarin doeltreffend kunt zoeken? Artificial Intelligence-technieken kunnen helpen.

Mathilde Jansen

De meeste beeldherkenningsalgoritmes zijn getraind op modern fotomateriaal en presteren daarom minder goed op historische beelden. Door de algoritmes te trainen op historisch materiaal kan het fotografisch geheugen aanzienlijk worden verbeterd. Melvin Wevers van het KNAW Humanities Cluster werkt aan zo'n algoritme voor de twee miljoen foto's uit de periode 1945-2004 van Fotopersbureau De Boer. De foto's liggen in de depots van het Noord-Hollands Archief. Een deel is gedigitaliseerd en wordt vanaf deze zomer door vrijwilligers van velehanden.nl van labels voorzien. Het door Wevers ontwikkelde algoritme is daarbij leidend.

Sneeuw of zand?

Bestaande beeldherkenningstechnieken werken soms al heel goed, legt Wevers uit, omdat de algoritmes getraind zijn op specifieke datasets voor specifieke taken. "Zelfrijdende auto's zijn bijvoorbeeld heel goed in het herkennen van verkeersborden. Maar veel objecten verschillen door de tijd heen te veel: een telefoon ziet er anders uit dan tien jaar geleden. Bovendien bevat



Het labelen van afbeeldingen gebeurt deels handmatig. Ook wordt gebruikgemaakt van labels die eerder zijn toegevoegd. In samenspraak met het Noord-Hollands Archief komt Wevers zo tot een lijst met labels.

Credits: Noord-Hollands Archief

2 miljoen foto's online doorzoekbaar

Met AI zoeken in historisch fotomateriaal

ten veel datasets geen historisch materiaal. "En dan hebben we het nog niet over de kwaliteit van de afbeelding of de techniek. Veel oude foto's zijn in grijstinten; een sneeuwlandschap en een strand zijn moeilijk van elkaar te onderscheiden. Mogelijke oplossingen zijn om met AI oude afbeeldingen automatisch te laten inkleuren, of bestaand trainingmateriaal juist zwart-wit te maken."

Toch werkt Wevers met bestaande algoritmes voor een eerste indicatie.

"Een bestaand model is op heel veel afbeeldingen getraind. Sportwedstrijden worden bijvoorbeeld makkelijk herkend. Maar veel van die afbeeldingen komen van websites als Flickr, en hebben een sterke Amerikaanse bias." Het labelen van de afbeeldingen gebeurt daarom deels handmatig. Ook maakt de onderzoeker gebruik van labels die eerder zijn toegevoegd door de fotografen van persbureau De Boer. In samenspraak met het Noord-Hollands Archief komt Wevers zo tot een lijst met labels. Dat wordt de in-

put voor het Vele Handen-project: mensen kunnen straks kiezen uit de labels die het computeralgoritme genereert.

Naar 90 procent

Het algoritme heeft nu een correcte score van 85 procent, Wevers streeft naar 90 procent. "Ik hoop uiteindelijk een algoritme te ontwikkelen dat niet alleen toepasbaar is op deze collectie, maar ook op andere collecties, zoals die van het Nationaal Archief."

velehanden.nl

GELEZEN

Recommendations for Services in a FAIR data ecosystem

D. Bangert, E. Hermans, R. van Horik, M. de Jong, H. Koers, M. Mokrane

FAIR data ontstaan en worden gebruikt in een complex wetenschappelijk landschap (FAIR ecosysteem); het omvat ICT-diensten, data stewardship en samenwerking binnen en buiten instellingen en disciplines. Daarbij is de vraag: hoe FAIR-aligned zijn de diensten? Deze vraag hebben de Europese initiatieven FAIRsFAIR, RDA Europe, OpenAIRE, EOSC-hub en FREYA gesteld in een reeks workshops. In het heldere eindrapport *Recommendations for services in a FAIR data ecosystem* springen vier aanbevelingen eruit: gebruik services voor persistent identifiers (ook voor organisaties en instrumenten); op standaarden gebaseerde metadata worden idealiter automatisch gegenereerd; maak FAIR-alignment en het delen van data onderdeel van onderzoeksevaluaties; werk internationaal samen aan de uitdagingen voor het uitvoeren van de FAIR-principes. Ook zijn een aantal aanbevelingen acties verbonden: bibliotheken willen de waardering voor FAIR-alignment en FAIR data aanwakkeren, onderzoeksinstellingen zetten in op data stewardship en service providers op domeinspecifieke ontologieën en tools voor repositories. "Work in progress", aldus de auteurs, en belangrijk om voort te zetten in beleid, infrastructuur en onderzoekspraktijk.

DOI: 10.5281/zenodo.3585742

23:30 - 07:00

Slapen, met partner

07:00 - 07:10

Snoozen, alleen

07:10 - 07:20

Douchen, alleen; nevenactiviteit: zingen.

Voor het tijdsbestedingsonderzoek (TBO) dat het CBS samen met het SCP eens per 5 jaar uitvoert, vullen respondenten naast een algemene vragenlijst ook 7 dagen lang een papieren dagboek in. Per 10 minuten geven zij aan welke activiteit zij hebben gedaan, of zij daarnaast nog iets deden en met wie ze dat deden. Hoe mensen hun tijd besteden, zegt iets over hun interesses, mogelijkheden en beperkingen. Het TBO geeft daarmee ook een mooi zicht op maatschappelijke verschijnselen en hoe die veranderen in de tijd. Zo komt het, vergeleken met tien jaar eerder, bijvoorbeeld vaker voor dat mensen meerdere activiteiten op hetzelfde moment uitvoeren, het 'multitasken'. SCP heeft een mooi overzicht gemaakt in de vorm van

Tijd voor een nieuw tijdsbestedingsonderzoek

een 'card stack', een verzameling digitale kaarten die ingaan op een aspect van de tijdsbesteding van Nederlanders. We zijn onze respondenten dankbaar dat ze de tijd (!) nemen om een hele week hun activiteiten bij te houden, want dat is heel veel werk.

Maar ook het verwerken van al die papieren boekjes op dusdanig wijze dat de privacy wordt geborgd, kost de nodige uren: inscannen, coderen, opschonen, het maken van een databestand bruikbaar voor onderzoekers. Redenen om samen met het SCP een app te ontwikkelen die digitaal veilig is, en die het de respondent gemakkelijker maakt zijn of haar tijdsbesteding bij te houden. Eerder voerde het SCP ook al experimenten met CentERdata op

dit terrein uit. Het bijkomend voordeel van een app is dat het de kwaliteit van de invoer verbetert (je kunt controleregels inbouwen als tijds slots worden vergeten) en de verwerking vereenvoudigt (je hebt immers alles al digitaal binnen). Het CBS is al aan het experimenteren met een app die verplaatsingen meet voor het onderzoek Onderweg in Nederland, en ook wordt er gewerkt aan een app die het inscannen van bonnetjes



van je boodschappen mogelijk maakt, voor het Budgetonderzoek. Beide zijn onderzoeken die met slimme apps het voor respondenten minder tijdrovend kunnen maken en zo hopelijk de respons op het onderzoek kunnen verhogen.

Natuurlijk zijn er mitsen en maren, zoals: 'Willen mensen wel een app installeren?' of 'Is het niet een hele selectieve groep die meedoet?'. Om dit soort vragen te beantwoorden, wordt de app eerst getest in een veldtest. Voor het TBO hopen we dit eind dit jaar te kunnen doen. Als de test positief is, wordt de app verder ontwikkeld. In eerste instantie voor het TBO, en met wat aanpassingen ook voor andere onderzoeken die tijd meten, zoals onderzoeken naar vrije tijd en naar het

gebruik van social media. Er zijn genoeg mogelijkheden, bijvoorbeeld door slim gebruik te maken van de sensoren die standaard in deze apparaten aanwezig zijn. Een walhalla voor een onderzoeker, uiteraard met een uiterst scherp oog voor privacy, doelmatigheid en informatiebeveiliging.

22:50 - 23:00

Tandenpoetsen, alleen; nevenactiviteit: whatsappen

23:00 - 23:20

Nog even lezen, alleen

23:20 - 07:00

Slapen, met partner

Marieke Houben-Van Hertem

Marieke is projectmanager/statistisch onderzoeker bij het CBS. Als clustercoördinator *Leefsituatie* heeft zij zicht op alle persoons-enquêtes die het CBS doet op het gebied van leefsituatie, sociale samenhang en welzijn. Daarnaast werkt Marieke mee aan de gezondheidsstatistiek van het CBS. Marieke geeft de pen door aan Karin Hagoort.

COLUMN