

Data Huygens ING gebruikt in Augmented Reality

VOC-archieven tot leven met Data Experience

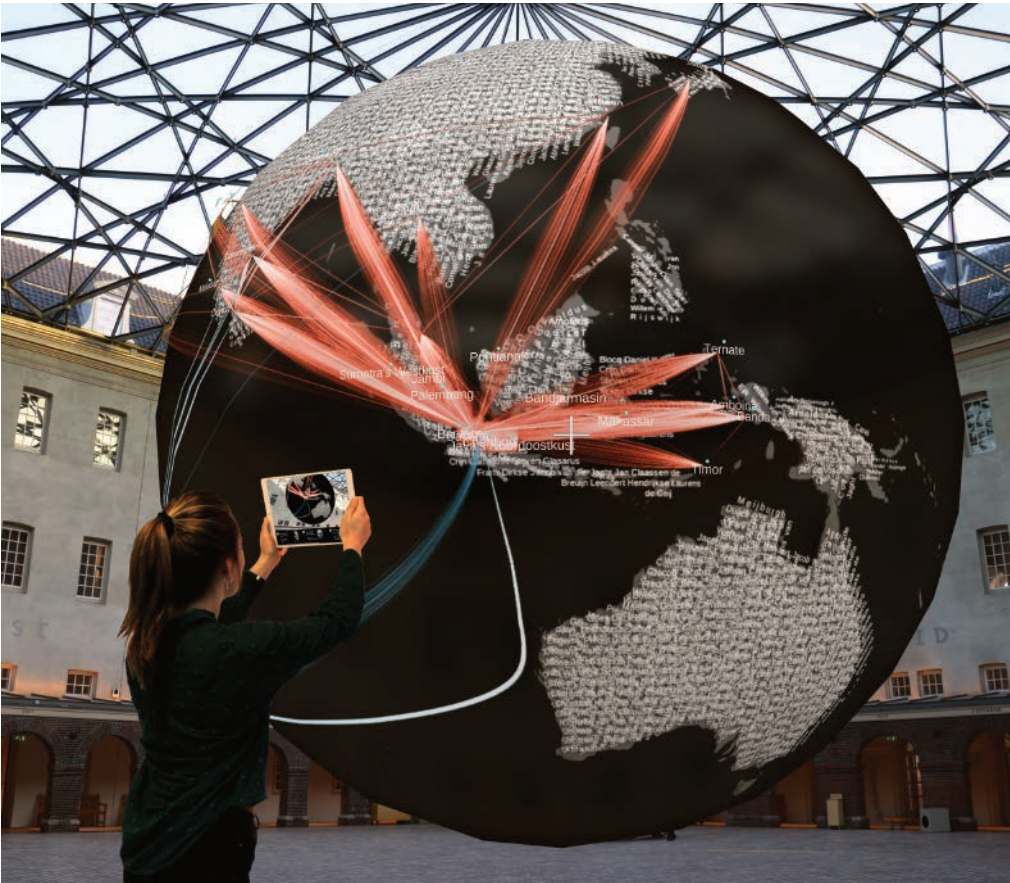
De VOC Data Experience biedt museumbezoekers een unieke kijk in het leven van opvarenden van VOC-schepen. De data zijn deels afkomstig van het Huygens Instituut voor Nederlandse Geschiedenis. *Mathilde Jansen*

Het totale VOC-archief wordt geschat op zo'n 25 miljoen pagina's. Deze liggen voornamelijk in het Nationaal Archief, maar zijn voor een deel ook te vinden op andere plekken op de wereld waar de VOC actief is geweest. Informatie uit die archieven is de afgelopen decennia verwerkt in verschillende databases. Het Huygens ING beheert twee grote, online toegankelijke datasets.

Het leven van een opvarende

Om de grote hoeveelheid data toegankelijk te maken voor een breed publiek, ontwikkelde het Huygens ING de VOC Data Experience. Samen met historicus Jelle van Lottum en andere collega's van het Huygens ING reconstrueerde data-expert Lodewijk Petram de afgelopen jaren de carrières van bemanningsleden aan de hand van gedigitaliseerde data uit soldijboeken. Het resultaat wordt met hulp van Augmented Reality (AR) aan bezoekers gepresenteerd, allereerst in Het Scheepvaartmuseum.

Petram legt uit hoe het werkt: "Bezoekers van Het Scheepvaartmuseum zien via een iPad de binnenplaats van het museum als AR-ruimte, met daarin honderdduizenden namen van opvarenden. Bezoekers kunnen zo bijvoorbeeld kennismaken met Christiaan Keems, die maar liefst zeven keer meevoer met de VOC.



Een bezoeker van Het Scheepvaartmuseum ziet via een iPad de binnenplaats van het museum als AR-ruimte. Credits: foto Studio Bertels

Van abstract naar visueel

Het visualiseren van de data in AR heeft volgens Petram zeker toegevoegde waarde. "Bezoekers moeten echt om zich heen kijken, via de iPad is niet alles in één oogopslag te zien. Het voelt alsof ze zelf op onderzoek kunnen gaan. Bezoekers worden zich ook beter bewust van de enorme omvang van de data. De VOC Data Experience toont bijvoorbeeld ook een wereldkaart op de vloer. Bezoekers zien onder hun voeten schepen uit de Republiek vertrekken naar Azië, waarbij

het aantal bemanningsleden steeds minder wordt. Dat komt veel meer binnen dan een tabel met abstracte cijfers."

De VOC Data Experience is een initiatief van het Huygens ING in samenwerking met Studio Bertels. De financiering komt uit het Stimuleringsfonds voor Creatieve Industrie van de Rijksoverheid. De tentoonstelling zal ook te zien zijn in andere musea. vocdataexperience.nl

ODISSEI Social Data Science Team

Eerste hulp bij Data Science

Kun je tijdens sociaalwetenschappelijk onderzoek ondersteuning gebruiken bij computationele uitdagingen? Neem dan contact op met het ODISSEI Social Data Science Team. *Suze Zijlstra*

Sociale wetenschappers maken in toenemende mate gebruik van grote hoeveelheden data, die ze met steeds complexere technieken analyseren. Omdat niet alle wetenschappers altijd thuis zijn in de nieuwste computationele onderzoeksmethoden, heeft ODISSEI vorig jaar met Daniel Oberski (Universiteit Utrecht) het Social Data Science Team (SoDa) opgericht.

Erik-Jan van Kesteren, teamleider SoDa, licht toe: "Voor het beantwoorden van onderzoeksvragen lopen sommige onderzoekers tegen technische obstakels aan. De leden van SoDa beschikken over vaardig-

heden om dergelijke obstakels weg te nemen." Het team bestaat uit research engineers en statistici met technische kennis om complexe vraagstukken te behandelen. Bovendien hebben ze goede contacten bij verschillende ODISSEI-organisaties, waardoor ze snel relevante aanvullende expertise kunnen inschakelen.

Handvatten

Van Kesteren benadrukt dat het belangrijk is dat sociale wetenschappers handvatten krijgen om nieuwe uitdagende onderzoeksvragen te kunnen beantwoorden. Er zijn immers steeds meer grote datasets beschikbaar, niet alleen bij dataprojecten, maar bijvoorbeeld ook via het CBS. De vraag naar vaardigheden om hoogwaardig onderzoek te verrichten met dergelijke omvangrijke datasets, zal blijven toenemen.

Wetenschappers hoeven geen grote

projecten uit te voeren om bij het team aan te kloppen. SoDa organiseert maandelijks een online Data Drop-in. Iedereen kan aanschuiven om specifieke onderzoekskwesties voor te leggen of algemene vragen te bespreken. Van Kesteren: "Laatst keken we met een onderzoeker of het mogelijk was de data van een bepaalde website te scrapen. Het kunnen kleine obstakels zijn die een onderzoeker tegenhouden om iets wel of niet te doen, even overleggen kan heel goed zijn. Soms is een vraag snel beantwoord, andere keren besteden we wat meer tijd aan een specifieke kwestie."

Grotere projecten zijn natuurlijk ook welkom. Soms leidt een hulpvraag zelfs tot een uitgebreidere samenwerking. In zo'n geval dragen de teamleden van SoDa bij aan de output van een project, of dit nu een we-

Lees verder op pagina 2

Welkom CBS

Vanaf 2022 levert het Centraal Bureau voor de Statistiek (CBS) weer een bijdrage aan dit blad. Annette Scherpenzeel, hoofd van de afdeling Microdata Services: "Via een beveiligde omgeving, met gepseudonimiseerde data en onder strikte voorwaarden kunnen ook anderen onderzoek doen met de databestanden van het CBS. Zo profiteert de samenleving optimaal van de data die het CBS beheert, er is op deze manier immers veel meer onderzoek mogelijk dan het CBS zelf zou kunnen doen. Via dit blad berichten we graag over nieuw beschikbare data en over nieuwe mogelijkheden om met deze data te werken."

cbs.nl



E-DATA & RESEARCH

Jaargang 16 | nummer 1

Nieuwsbrief over data en onderzoek in de alfa- en gamma-wetenschappen.

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: Centerdata, CLARIAH, DANS, KNAW Humanities Cluster, de Koninklijke Bibliotheek, ODISSEI en het Rijksmuseum.

INHOUD



- 3 FAIR Signposting biedt online navigatiehulp
- 3 Benieuwd naar ervaring met Speech to Text?
- 4 Taskforces maken zich op voor vervolgstap EOSC
- 5 "Weer een stap dichterbij één data-infrastructuur"
- 6 Ook deze datasets zijn sinds kort beschikbaar
- 7 Omgaan met data: FAIR én CARE steeds belangrijker
- 8 Columnist Van der Poel: delen als default

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

SHARE Netherlands
User Conference

Suze Zijlstra

Tijdens de SHARE Netherlands User Conference van 25 juni stond het gebruik van data uit de Survey of Health, Ageing and Retirement in Europe centraal. SHARE is de grootste Europese panelstudie in de sociale wetenschappen. Doordat de dataverzameling al bijna twintig jaar loopt, biedt het waardevolle longitudinale data over het leven en de gezondheid van Europeanen ouder dan 50 jaar.

De Nederlandse tak van SHARE organiseerde deze gebruikersconferentie, met het doel gebruikers bij elkaar te brengen en onderzoeksresultaten uit te wisselen. Adriaan Kalwij, teamleider van SHARE Netherlands, zat de online conferentie voor.

De eerste helft van de bijeenkomst richtte zich op recent uitgevoerd onderzoek. Zo presenteerde Yuri Pettinichchi (Max Planck Institute) over de kennis die ouderen hebben over hun toekomstige pensioen. Nikos Theodoropoulos (University of Cyprus) liet zien hoe de arbeidsparticipatie van een persoon beïnvloed



Tijdens de SHARE Netherlands User Conference stond het gebruik van data uit de Survey of Health, Ageing and Retirement in Europe centraal. Credits: foto Laura Thonne via Unsplash

Social Data
Science Team

Vervolg van pagina 1

tenschappelijk artikel is, een onderzoeksaanvraag of een softwarepakket. Van Kesteren sluit af: “Wij zijn erg vrij in de manier waarop we onderzoek vooruit helpen. Wij delen niet alleen onze technische vaardigheden, maar ook ons enthousiasme voor innovatief computationeel onderzoek.”

odissei-data.nl/soda

GEHOORD & BIJGEWOOND

wordt door de afname van de gezondheid van diens partner. Judith Bom (Erasmus Universiteit Rotterdam) liet zien hoe vaak en hoe snel ouderen worden opgenomen in verzorgingstehuizen in Nederland, Denemarken en de Verenigde Staten. Het laatste deel van de conferentie richtte zich op actuele ontwikkelingen en dataverzamelingen. Tijdens de COVID-pandemie heeft SHARE veel data verzameld over de invloed van deze gezondheidscrisis op het welzijn van ouderen. Zijn mensen eenzamer dan voor de pandemie, hebben zij vaker last van depressieve gevoelens? En in hoeverre beïnvloedde de pandemie familierelaties? Juist omdat SHARE al veel data heeft van voor de pandemie, is het goed te zien op welke manieren de coronacrisis het leven van ouderen heeft beïnvloed. Tijdens de conferentie presenteerde Johanna Bristle (Max Planck Institute) de eerste resultaten van de SHARE Corona Survey.

share-project.org

DANS Data Trail 2:
Workshop From Here
to the EOSC

Ricarda Braukmann

Op 28 september vond de tweede editie van de DANS Data Trail plaats – een online reeks waarin verschillende thema’s rondom Research Data Management en Open Science aan bod komen. In deze bijeenkomst stond de weg naar Open Science en de implementatie van de European Open Science Cloud (EOSC) op de agenda.

Tijdens de eerste helft van de workshop kwamen verschillende sprekers aan het woord. Als eerste presenteerde Valentino Cavalli (EGI) een landschapsanalyse uitgevoerd door het EOSC Synergy-project die de status van Open Science in zeven Europese landen vergeleek. Vervolgens zoomde Peter Doorn (DANS)



De DANS Data Trail behandelde de weg naar Open Science en de implementatie van de European Open Science Cloud (EOSC).

Credits: Stephanie Bosschaert

in op de Nederlandse situatie en de verschillende manieren waarmee Open Science en EOSC-readiness gekwantificeerd kunnen worden. Ook al zijn er nog vrij weinig gestandaardiseerde cijfers beschikbaar, wel wordt steeds meer vooruitgang geboekt. Zo stijgt het aantal gecertificeerde repositories en groeit het aantal open access datasets. Maurice Bouwhuis (SURF) vertelde vervolgens over de Nederlandse data-infrastructuur en de ondersteuning hierin door SURF en DANS. Initiatieven zoals het Nationaal Platform Open Science (NPOS), de lokale Digital Competence Centers (DCC’s) en thematische DCC’s die disciplinebrede ondersteuning moeten leveren, spelen in dit landschap een belangrijke rol. Ellen Leenarts (DANS) vertelde als laatste spreker over het belang van training voor Open Science. Zij gaf een inkijkje in de plannen van de nieuwe EOSC Association Task Forces en met name in de plannen van de Task Force Upskilling Countries to engage in EOSC, die zal gaan kijken naar opleiding en training rondom Open Science in de verschillende Europese landen en hoe nationale activiteiten elkaar kunnen versterken.

In het laatste deel van de workshop werd het publiek aan het werk gezet. Door middel van een aantal vragen werd Open Science in Nederland bediscussieerd. Een centraal punt die in de discussie naar voren kwam, was de vertaling naar de onderzoekspraktijk.

De aanwezigen gaven aan, open te staan voor alle stappen op het gebied van Open Science, maar vooral ondersteuning en overleg te kunnen gebruiken in de uitvoering en implementatie. Initiatieven zoals de DANS Data Trail hopen hier een bijdrage aan te kunnen leveren.

Voor wie meer wil weten over data publiceren, archiveren, herbruikbaar

en interoperabel maken, en voor wie benieuwd is naar de ontwikkelingen op het gebied van Open Science, Research Data Management-planning en beschikbare tools, organiseert DANS diverse trainingsactiviteiten.

Leenarts: “Kennis en ervaring uit Europese dataprojecten wordt door DANS vertaald naar de Nederlandse praktijk. Geregeld organiseren we interactieve events voor betrokkenen in het datalandschap. Voorbeelden zijn de RDNL-training Essentials 4 Data RDNL, de CESSDA Data Management Expert Guide en DANS Data Trails. Kijk voor ons aanbod op de website van DANS.”

dans.knaw.nl

CESSDA Roadshow
on COVID-19

Heidi Berkhout

Op 30 september organiseerde het Consortium of European Social Science Data Archives (CESSDA) een online roadshow over COVID-19-data. Het belang van goed gedocumenteerde en snel beschikbare data over deze pandemie werd gedurende twee uur vanuit verschillende oogpunten belicht.

Als eerste vertelde Alle Bloom (UKDA) over het vinden van

COVID-19-data in de CESSDA Data Catalogue (CDC). De CDC aggregiert metadata van verschillende sociaalwetenschappelijke dataarchieven in Europa: meer dan 200 COVID-19-datasets, veelal enquêtes en opiniepeilingen, zijn op die manier centraal vindbaar. Twee voorbeelden van beschikbare datasets werden vervolgens gepresenteerd door onderzoekers Julia Partheymüller (University of Vienna) en Anna Zamberlan (University of Trento).

In de tweede helft van de bijeenkomst werd aandacht besteed aan het belang van Research Data Management (RDM) en het werk van de nationale CESSDA Service Providers.

Ricarda Braukmann (DANS) introduceerde de Data Management Expert Guide (DMEG), een gratis online RDM-training voor jonge onderzoekers. Gecertificeerde data repositories, zoals DANS en andere CESSDA Service Providers, kunnen onderzoekers helpen om COVID-19-data veilig op te slaan en (onder voorwaarden) beschikbaar te stellen voor hergebruik. Zowel DANS en AUSSDA - het Oostenrijkse dataarchief voor de sociale wetenschappen - gebruiken Dataverse als repositorysoftware en bieden op maat gemaakte services voor onderzoekers.

Otto Bodi-Fernandez (AUSSDA) presenteerde vervolgens een fast track publishing route die gedurende de pandemie is opgericht om COVID-19-data extra snel beschikbaar te stellen. Flexibiliteit is volgens Bodi de sleutel als men adequaat wil reageren op een zich ontvouwende globale crisis.

De sprekers waren het eens dat data snel beschikbaar moeten zijn, zo bleek gedurende de paneldiscussie waarmee het online event werd afgesloten. Een goede samenwerking tussen onderzoekers en data-archieven verbonden aan CESSDA speelt hierin een cruciale rol.

cessda.eu



Tijdens de online CESSDA-roadshow over COVID-19-data werd het belang van goed gedocumenteerde en snel beschikbare data vanuit verschillende oogpunten belicht. Credits: CESSDA

Machinevriendelijke wegwijzers op en rond landingspagina's

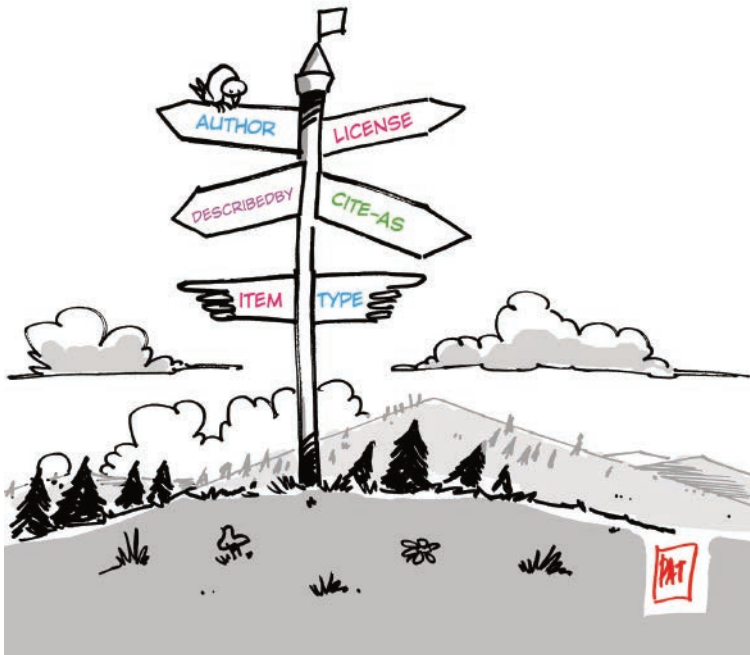
FAIR Signposting biedt online navigatiehulp voor bots

FAIR Signposting biedt een online wegwijzer voor bots op zoek naar wetenschappelijke output. DANS realiseerde onlangs een implementatie die als inspiratie kan dienen voor andere wetenschappelijke repositories.
Herbert Van de Sompel

Wie zich op het wetenschappelijke web begeeft, is vertrouwd met landingspagina's voor onderzoeksoutput. Ze komen in vele geuren en kleuren, het vinden van specifieke informatie op die pagina's is soms een uitdaging. Uiteindelijk wordt de link naar een artikel, dataset of een bibliografische citatie vaak wel gevonden. Met een beetje geluk wordt ook de Persistent Identifier (PID) gelokaliseerd om, in plaats van het webadres van de landingspagina, deze PID duurzaam naar het werk te laten verwijzen.

Tallose links

Voor webbots is de situatie veel lastiger. Een bot ziet op een landingspagina tallose links en kan niet zomaar differentiëren naar welk soort informatie de links leiden en of ze nuttig zijn voor de taak die vol-



De implementatie van DANS kan als inspiratie dienen voor andere repositories. Credits: illustratie Patrick Hochstenbach

bracht moet worden. Diensten die gebruik maken van bots gaan hierdoor noodzakelijkerwijs heuristisch te werk en houden per type landingspagina bij welk soort informatie via welke link te vinden is. Een behoorlijke klus voor bijvoorbeeld referentiemanagers zoals Zotero en archiveringsdiensten zoals LOCKSS. En het is uiteraard een significante

drempel voor het creëren van nieuwe bot-gebaseerde diensten.

Hulp bij navigatie

FAIR Signposting werd in het leven geroepen om dit probleem aan te pakken en bots te helpen bij het navigeren van het wetenschappelijke web. FAIR Signposting plaatst machinevriendelijke wegwijzers op en

rond landingspagina's. Technisch zijn dit links voorzien van een type-ring. Zo leidt de link met het type 'describedby' naar een bibliografische beschrijving, 'cite-as' naar de PID, en 'item' naar het artikel of de dataset.

Dataverse

In het kader van het ODISSEI roadmap project implementeerde DANS 'FAIR Signposting' in Dataverse, een open source repository platform. Een aanpak met garanties voor stabiliteit op de lange termijn: de basale ingrediënten voor deze uniforme machinetoegang zijn links en link types gestandaardiseerd door de Internet Engineering Task Force. Door de implementatie van 'FAIR Signposting' zullen bots makkelijker hun weg vinden om en rond datasets in Dataverse installaties zonder gebruik te moeten maken van een specifieke API. Maar het wetenschappelijke web wordt pas écht machinevriendelijk als ook andere platforms 'FAIR Signposting' ondersteunen. DANS hoopt dan ook anderen te inspireren hetzelfde te doen voor hun wetenschappelijke repositories. Wees lief voor bots en zeg het voort! signposting.org/FAIR

Centerdata deelt ervaring LISS panel

Tips en tricks over werken met Speech to Text

Wat zijn de voordelen en uitdagingen van audio-antwoorden in een online vragenlijst? Centerdata deelt de ervaring met spraaktechnologie Speech to Text.
Joris Mulder

Het LISS panel, dat beheerd wordt door Centerdata, verzamelt sinds 2007 data door middel van online vragenlijstonderzoek. Panelleden beantwoorden de vragen door een antwoord aan te klikken of, in het geval van een open vraag, een antwoord te typen. Dit levert een scala aan waardevolle gegevens op voor onderzoekers en beleidsmakers.

Twee experimenten

In aanvulling op deze klassieke methode van online dataverzameling kan het LISS panel ook *Speech to Text* in online vragenlijsten toepassen. Het afgelopen jaar zijn er twee experimenten uitgevoerd waarbij deze spraaktechnologie succesvol in online vragenlijsten werd ingezet. In deze experimenten konden respondenten hun open antwoorden in-



spreken in plaats van typen. De gegeven antwoorden zijn op twee manieren opgeslagen: als audiobestand en als transcriptie naar tekst. De winst van deze methode van dataverzameling is dat er, naast het primaire antwoord op de vraag, waardevolle aanvullende data worden verzameld. Denk bijvoorbeeld aan de toon van een respondent bij een antwoord, de linguïstische en verbale capaciteiten en de grootte van iemands woordenschat. Dit kan inzicht geven in het cognitief functioneren, de sociaaleconomische status of het verbaal redeneervermogen van een respondent. Tevens bieden

de data de mogelijkheid tot sentiment- of emotieanalyses (al dan niet gerelateerd aan een bepaald onderwerp waar mensen meer of minder binding mee hebben) en *Natural Language Processing* (het analyseren van natuurlijke taal met behulp van kunstmatige intelligentie).

Aandachtspunten

Naast deze voordelen zijn er ook enkele nadelen en aandachtspunten. Zo zijn er soms technische beperkingen, bijvoorbeeld omdat een respondent geen werkende microfoon heeft of een te oude (niet geschikte) browser gebruikt. Ook wil een deel

van de panelleden überhaupt niet deelnemen aan deze vorm van dataverzameling. Hierdoor kan er (selectieve) non-respons ontstaan. Een belangrijk aandachtspunt is privacy. Een stem geldt als persoonlijke data en kan tot persoonsidentificatie leiden. Data die middels *Speech to Text* worden verzameld, dienen dan ook onder voorwaarden opgeslagen te worden. Het LISS panel zorgt er bovendien voor dat alle audio-bestanden en transcripts door een medewerker worden gecontroleerd op eventuele persoonlijke informatie die men in het antwoord geeft (zoals een naam of plaatsnaam).

Beschikbare data

De data van de *Speech to Text*-experimenten komen eind 2021 beschikbaar in het LISS Data Archive en zijn te gebruiken voor eigen onderzoek. Voor meer informatie kan contact worden opgenomen met Joris Mulder, Survey Onderzoeker bij Centerdata, via joris.mulder@centerdata.nl.

centerdata.nl

KORT

Meld je aan voor ASReview Academy

Sneller systematisch literatuuronderzoek uitvoeren kan met ASReview. Deze software leert van gebruikers welke resultaten relevant zijn en verbetert hierdoor het zoekproces in de relevante databases. Om ASReview zo goed mogelijk te gebruiken, kunnen onderzoekers zich aanmelden voor een ASReview Academy, maandelijks georganiseerd in samenwerking met ODISSEI. Instromen kan elke maand, op verschillende niveaus (van beginners tot gevorderden) wordt in anderhalf uur steeds diepgaander kennisgemaakt met de software. (SZ) odissei-data.nl/nl/asreview-nl

Collecties als data bij KBR België

Met het project DATA-KBR-BE wil de KBR, de Koninklijke Bibliotheek van België, een verbeterde toegang op dataniveau van haar digitale collecties realiseren. Samen met de universiteiten van Gent en Antwerpen gaat de KBR onder andere de IT-infrastructuur optimaliseren en wordt een open data platform ontwikkeld. Hierlangs zullen de onderliggende databestanden van gedigitaliseerd erfgoed beschikbaar worden gesteld, wat als voedingsbodem dient voor onderzoek in de Digital Humanities. Met het project volgt de KBR het Amerikaanse Collections as Data Initiative (collectionsasdata.github.io). DATA-KBR-BE wordt gefinancierd door het Belgische instituut voor federaal wetenschapsbeleid BELSPO in het kader van het BRAIN 2.0 programma en loopt tot en met 2022. (MH) kbr.be/nl/projecten/data

Beurs voor OpenHuygens.nl

Het Onderzoeksfonds KNAW-instituten heeft het Huygens ING een beurs toegekend voor het versneld online overzetten van de collectie van het Huygens ING naar OpenHuygens.nl. Het Huygens ING beschikt over bijna tweehonderd resources op het gebied van Nederlandse letterkunde, geschiedenis en wetenschapsgeschiedenis. Samen vormen deze bronnen, teksten en datasets een fundamenteel onderdeel van de digitale infrastructuur voor de geesteswetenschappen. Met OpenHuygens.nl, een nieuwe data-omgeving die de collectie beter toegankelijk maakt voor innovatieve, digitale toepassingen, realiseert het Huygens ING duurzame toegang tot deze belangrijke bronnen voor onderzoek naar de Nederlandse geschiedenis en cultuur. De eerste versie van een werkend OpenHuygens.nl wordt in het voorjaar van 2022 gelanceerd. (MJ)

Task Forces European Open Science Cloud Association presenteren toekomstplannen

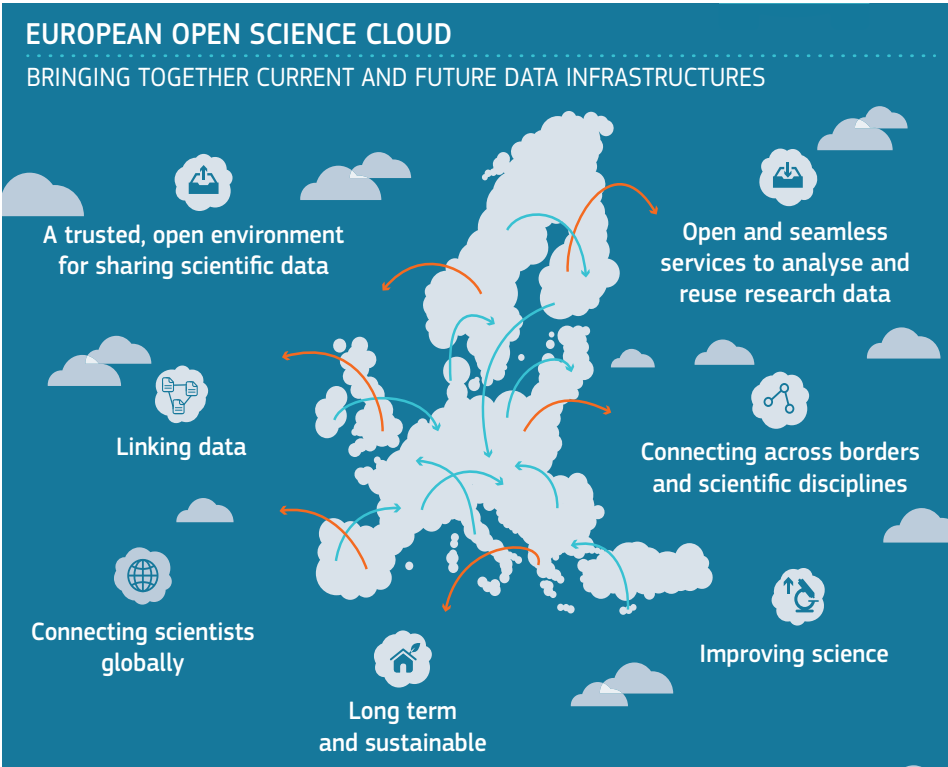
Volgende fase voor één Europese data-infrastructuur

Veertien Task Forces van de European Open Science Cloud Association presenteerden in juli hun plannen voor de Europese wetenschappelijke data-infrastructuur.

Ricarda Braukmann

Eén virtuele omgeving die elke Europese onderzoeker toegang geeft tot de nieuwste services voor het vinden en analyseren van data. Een plek waar data duurzaam opgeslagen en gedeeld kunnen worden voor nieuw innovatief onderzoek, gebaseerd op Open Science en FAIR-dataprincipes. Dat is de visie van de European Open Science Cloud (EOSC).

Om deze ambitie te realiseren, moeten bestaande infrastructuren op elkaar aansluiten en beleid en regelgeving geharmoniseerd worden. Een enorme klus, waar sinds 2018, onder meer gefinancierd door de Europese Commissie, honderden organisaties aan meewerken, waaronder ook Nederlandse organisaties zoals DANS, SURF, DTL en TUDelft. Sinds vorig jaar vertegenwoordigt en coördi-



De afbeelding laat belangrijke uitgangspunten binnen de vorming van de EOSC zien. Credits: EOSC

neert de EOSC-associatie het brede scala aan stakeholders die van EOSC een succes moeten maken. De associatie stelde begin dit jaar veertien Task Forces in rondom vijf centrale thema's:

implementatie van de EOSC, technische uitdagingen, metadata en datakwaliteit, onderzoekscarrières en opleiding, en duurzaamheid van de EOSC. In juli presenteerden alle Task

Forces via een charter de activiteiten voor de komende twee jaar.

Duurzame opslag

“Data duurzaam opslaan zodat het beschikbaar blijft voor nieuw onderzoek, is essentieel voor de wetenschap en de basis van de EOSC.” Aan het woord is Ingrid Dillo, adjunct directeur van DANS. “Als nationaal expertisecentrum en data repository is DANS al jaren betrokken bij de EOSC. Met onze kennis over metrics voor FAIR-data, semantische interoperabiliteit, uitwisseling van training en educatie, en onze visie rondom preservatie van onderzoeksdata, hebben we een fundamentele bijdrage kunnen leveren aan vier charters. Hierdoor zorgen we ervoor dat de behoeftes uit de Nederlandse onderzoeksgemeenschap ook op Europees niveau gehoord worden.” Belangrijke focus is ook het certificeren van archieven. Dillo: “Onderzoekers moeten vertrouwen hebben in de systemen die de EOSC gaat gebruiken. Dat vertrouwen kan worden gegeven met gezamenlijke richtlijnen waar datarepositories aan moeten voldoen.”

Het werk van de Task Forces neemt twee jaar in beslag en dient als basis voor de volgende implementatiefase van de EOSC.

eosc.eu

Onderzoekers maken Anne Franks werkwijze als schrijver inzichtelijk

Volledig oeuvre van Anne Frank online

Onlangs lanceerden het Huygens ING en de Anne Frank Stichting een online editie van de manuscripten van Anne Frank. De website is toegankelijk in zestig landen wereldwijd, helaas (nog) niet in Nederland. Mathilde Jansen

Iedereen kent wel het rood-wit geruite dagboek van Anne Frank. Daarnaast zijn nog twee schriften bewaard waarin ze tot augustus 1944 bijna dagelijks schreef. In mei van dat jaar besloot Anne haar dagboeken te gaan bewerken tot een roman, in briefvorm. Na de oorlog combineerde haar vader, Otto Frank, de dagboeken en de roman in het boek dat de titel Het Achterhuis kreeg. Maar dat is niet alles. Anne schreef ook nog korte verhalen en verzamelde fragmenten uit boeken in haar ‘mooie-zinnen-boek’. Alle originele manuscripten zijn sinds september online raadpleegbaar via annefrankmanuscripts.org.

Geheimen ontrafelen

“In een digitale editie kunnen we veel inzichtelijker maken hoe Anne de teksten heeft geschreven”, zegt Peter de Bruijn. Hij is één van de onderzoekers van het Huygens ING

die de afgelopen jaren de manuscripten bestudeerden. “We laten bijvoorbeeld zien welke passages uit het dagboek terecht zijn gekomen in het boek Het Achterhuis en hoe Anne ze bewerkte. We hebben alles ontcijferd, ook wat ze heeft doorgehaald.” Met speciale technieken konden de onderzoekers zelfs achterhalen wat er achter afgeplakte pagina's stond. Zijn daarmee alle geheimen onthuld? Toch niet. “Uit het jaar 1943 missen een of twee dagboeken. Om die lacune op te vullen, heeft Otto Frank destijds de dagboeken en de roman tot één geheel gecombineerd.” In de online editie worden Annes originele manuscripten nu afzonderlijk uitgegeven, waarbij ook de onderlinge verbanden getoond worden.

Toegankelijk in 2037

Hoe mooi dit ook allemaal klinkt, er



Transcriptie van een brief uit Het Achterhuis. Anne schrijft over het wegvoeren van joden. Ze voegt er naderhand in rode inkt aan toe: ‘Wij nemen aan dat de meesten vermoord worden’.

Credits afbeelding manuscript: Collectie Anne Frank Stichting, Amsterdam

is één grote ‘maar’. De website is toegankelijk in zestig landen wereldwijd. Nederland hoort daar niet bij. In buurlanden Duitsland en België hebben internetgebruikers wel toegang. Dit heeft alles te maken met de auteursrechten, die berusten bij het Anne Frank Fonds in Basel. In Nederland is de website pas

vanaf 2037 voor iedereen toegankelijk. Maar het belangrijkste is misschien toch wel dat dit onderzoek een completer beeld geeft van Anne Frank als persoon en als romanschrijver. De Bruijn: “Door tekstvergelijkend onderzoek weten we nu precies hoe ze te werk is gegaan: we zien dat ze

in haar dagboek al heel erg aan het experimenteren is met literaire vormen. Daardoor is het veel meer dan een gewoon dagboek. In haar roman werkt ze die vorm verder uit, en wordt het meer en meer óók literatuur.” huygens.knaw.nl/projecten/onderzoek-manuscripten-anne-frank

DANS lanceert domeingerichte Data Stations

“Stap voor stap richting een nationale data-infrastructuur”

Onderzoekers die data willen bewerken, delen, analyseren en archiveren, kloppen steeds vaker aan bij DANS, het nationaal expertisecentrum en repository voor onderzoeksdata. *Heidi Berkhout*

Op verzoek van onderzoekers en dataprofessionals bewaart DANS bijna 200.000 datasets gegarandeerd duurzaam. Het KNAW/NWO-instituut is sinds 2006 een belangrijke speler binnen de Nederlandse en wereldwijde onderzoeksgemeenschap. De komende periode gaat DANS domeingerichte Data Stations aanbieden, diensten die zijn toegesneden op de behoeften binnen wetenschappelijke disciplines. Daarmee wordt DANS naast een archief voor datasets uit voltooide onderzoeksprojecten, ook een omgeving waarin onderzoekers datasets kunnen vormen, bewerken, analyseren en delen. Henk Wals, directeur DANS, licht toe: “De omstandigheden en behoeften binnen wetenschappelijke domeinen lopen nogal uiteen. Datasets in de natuurwetenschappen of de levenswetenschappen kunnen bijvoorbeeld gigantisch groot zijn, en in de levenswetenschappen of de sociale wetenschappen bovendien privacy-gevoelig. In de geesteswetenschappen zijn sets meestal kleiner, maar vaak complex, bijvoorbeeld omdat historische data incompleet en tijdreeksen inconsistent zijn. De metadataschema’s en de gebruikte tooling verschillen ook per discipline. De roep om domeinspecifieke werkplaatsen voor data werd de afgelopen jaren steeds groter.”

Cross-disciplinaire aanpak

Wals vervolgt: “Tegelijkertijd zien we dat onderzoek steeds vaker een cross-disciplinaire aanpak vereist. Interoperabiliteit over domeinen heen wordt belangrijker, maar het datalandschap is gelaagd: het begint bij de individuele dataruimte van een onderzoeker, dan het niveau van instituties, dan domeinen, dan het nationale niveau en uiteindelijk internationaal. Op elk niveau bevinden zich organisaties met diensten voor onderzoekers. Voorbeelden naast DANS zijn 4TU.ResearchData, SURF en DTL. Voor het maximaal faciliteren van cross-disciplinair onderzoek is het belangrijk dat de organisaties op de knooppunten van alle lagen samenwerken en één infrastructuur vormen. Uiteindelijk maakt het voor een onderzoeker niet uit waar data en software vandaan komen, als het maar op elkaar aansluit.”

Vier Data Stations

Terug naar de Data Stations. Er komen er vier: Archaeology, Social Sciences & Humanities, Physical & Technical Sciences en Life, Health & Medical Sciences. Voor de laatste twee wordt afgestemd met respectievelijk 4TU.ResearchData en DTL over waar DANS aanvullend kan opereren of op welke data-niches DANS zich het beste kan richten. Ook wordt nauw samengewerkt met SURF, onder meer voor veilige opslag van de data. Via een interne procedure selecteerde DANS Data Station Managers die inhoudelijk goed thuis zijn in de domeinen. Wals: “Het idee achter de Data Stations is om dicht tegen het onderzoek aan te kruipen, aan te sluiten bij onderzoeksinfrastructuren en uit te gaan van de noden en wensen die in verschillende communities leven. Een Data Station moet je zien als een conglomeraat van services, met in de kern een instantie van Dataverse. Elk station wordt bemenst door Data Station Managers die vanuit hun eigen onderzoeksachtergrond het wetenschapsdomein en de mensen daarin goed kennen en dus constructieve verbindingen kunnen leggen.”

Voor onderzoekend Nederland

Door het nieuwe dienstenpakket en samenwerking op verschillende lagen zet DANS stappen richting een samenhangende basisinfrastructuur voor onderzoekend Nederland. Wals sluit af: “Samen met de andere datadienstverleners en de onderzoeksinfrastructuren in de verschillende wetenschapsdomeinen willen we een compleet pakket vormen dat gebruikers een naadloze ervaring biedt. Lokale Digital Competence Centers kunnen dan van dit dienstenpakket gebruik maken om onderzoekers te helpen met hun data te werken en deze FAIR te deponeren. Zo zou uiteindelijk een heldere en efficiënte nationale infrastructuur voor research data kunnen ontstaan.”



Hella Hollander
“Data afkomstig uit archeologisch onderzoek duurzaam opslaan, waarvoor de data ook op termijn vindbaar en herbruikbaar zijn, kan via DANS. Mocht je daar vragen over hebben, laat het mij weten.”



Ricarda Braukmann
“In de sociale wetenschappen werken we vaak met gevoelige data. Ik wil er graag aan bijdragen dat deze data goed bewaard blijven en, waar mogelijk, op een veilige manier hergebruikt kunnen worden om nieuwe vragen te beantwoorden.”



Femmy Admiraal
“In een veld met zo veel verschillende soorten data vind ik het een fantastische uitdaging om de geesteswetenschappen te ondersteunen in duurzame toegankelijkheid, op een manier die precies bij hun data past.”



Cees Hof
“Door wetenschappelijke data gestructureerd te archiveren en te delen, dragen we bij aan betere wetenschap voor een meer gezonde, groene, duurzame, inclusieve en innovatieve toekomst. Via mijn werk bij DANS kan ik een belangrijke bijdrage leveren aan deze missie.”



Marion Wittenberg
“DataverseNL is meer dan een veilige online omgeving waarin deelnemers zelf hun data-management kunnen inrichten, het is een samenwerkingsverband vol ‘good practices’ op het gebied van research data management.”

Data Station Archaeology
Hella Hollander is verantwoordelijk voor het Data Station Archaeology. Ze studeerde archeologie (Universiteit Leiden) en is projectleider van verschillende infrastructuren, projecten en andere co-creaties zoals ARIADNE, 4CH, Iperion-HS, Pure3D, Limes en Exalt. Ook is ze lid van de stuurgroep van E-rihs.NL waarin Nederlandse partners werken aan de Europese infrastructuur voor erfgoedwetenschap.

Data Station Social Sciences & Humanities
Ricarda Braukmann en **Femmy Admiraal** zijn de managers van het Data Station Social Sciences & Humanities.
Ricarda studeerde Psychologie (Universiteit Twente) en Cognitieve Neurowetenschappen (Radboud Universiteit) en promoveerde op een interdisciplinair onderzoeksproject naar de vroege ontwikkeling van autisme (Radboud Universiteit). Ze is actief binnen (inter)nationale infrastructuren zoals ODISSEI en CESSDA. Ook is ze betrokken bij de ontwikkeling van trainingen op het gebied van Research Data Management. **Femmy** is taalkundige en antropoloog (Universiteit van Amsterdam), zij promoveerde op een diepgaand onderzoek naar linguïstische codering van ruimtelijke referentie in de bedreigde taal Baure (Bolivia). Zij is actief in Europese infrastructuren DARIAH en CLARIN, en in verschillende onderzoeksprojecten, zoals PARTHENOS en Polifonia.

Data Station Physical & Technical Sciences en Data Station Life, Health & Medical Sciences
Cees Hof is het aanspreekpunt van deze Data Stations. Cees is zijn carrière begonnen als aquatisch ecoloog (Radboud Universiteit) en onderzocht als promovendus (Universiteit van Amsterdam) en post-doc (University of Bristol) hoe fossielen ontstaan en de rol van fossielen in evolutionaire reconstructies. Ervaring met het delen van data en de werking van data-infrastructuren deed hij op als coördinator van verschillende grote (open) dataprojecten, zoals de bouw van de de Global Biodiversity Information Facility (GBIF). Cees is ook een van de Research Data Management trainers binnen DANS.

Data Vault
DANS biedt naast de Data Stations ook de Data Vault; een beveiligd, betrouwbaar en gecertificeerd repository voor de lange termijn opslag van data. De Data Vault huisvest alle datasets die (ooit) door onderzoekers aan DANS zijn toevertrouwd, en funktioneert ook als een repository voor grote datacollecties afkomstig van universiteiten en onderzoeksinstituten. **Cees Hof** is aanspreekpunt voor de Vault.

DataverseNL
Naast de Data Stations biedt DANS ook DataverseNL. DataverseNL is een platformdienst waarin universiteiten, hogescholen en onderzoeksinstituten hun eigen instellingsrepository kunnen onderbrengen. De aldaar gedeponeerde datasets zijn tevens vindbaar via de domeingerichte Data Stations. Aanspreekpunt voor DataverseNL is **Marion Wittenberg**. Marion heeft Sociologie en sociaal-wetenschappelijke Informatica gestudeerd (Universiteit van Amsterdam) en heeft jarenlange ervaring als datamanager en projectmanager op het gebied van datacuratie.

SINDS KORT BESCHIKBAAR

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



Credits: foto Surface via Unsplash

Centerdata

Eenzaamheid voorkomen in een periode van social distancing (COVID-19)

In mei 2020 hebben Dr. Lodder, Dr. van Scheppingen, Dr. De Caluwé en Dr. Reitz van het departement ontwikkelingspsychologie, Tilburg University, de impact van de coronacrisis op sociale contacten en eenzaamheid onderzocht in het LISS panel. Eén van de conclusies is dat afnames in de hoeveelheid contact (met name met collega's en kennissen) en afnames in de ervaren kwaliteit van dat contact (met name met mensen buiten het huishouden) samenhangen met een toename in eenzaamheid. Ook bleek dat het gebruik van meer copingstrategieën (zoals videobellen of tuinieren) samenhang met minder eenzaamheid. De data van dit onderzoek zijn beschikbaar via het LISS Data Archive.

DOI: 10.17026/dans-zne-2bat

Ook sinds kort beschikbaar:

Studies LISS panel

- Hartevelde, E., augustus 2019, Affective polarization. DOI: 10.17026/dans-zff-t4nn.
- Hilbert, L., februari 2020, Social exclusion, control and financial scarcity - February 2020. DOI: 10.17026/dans-29g-3bjg.
- Rosema, M., Huddy, L., februari 2020, Party attachment - Wave 1. DOI: 10.17026/dans-xw7-sxry.
- Bos, L., maart 2020, Political intolerance. DOI: 10.17026/dans-zzn-x33w.
- Van der Velden, P., maart - april 2020, Victims in Modern Society - Follow-up measurement. DOI: 10.17026/dans-zws-ffz7.
- Verbakel, E., maart 2020, Retrospective informal care career - Main measurement. DOI: 10.17026/dans-xyf-v7vu.
- Von Gaudecker, H., april 2020, Effects of the Outbreak of COVID-19 - Part 2
- Haan, M., Ongena, Y., april 2020, Artificial Intelligence in Medicine. DOI: 10.17026/dans-zfc-smnw.
- Westerhof, G., ten Klooster, P., mei 2020, The Dutch Mental Health Continuum Short Form - Revised. DOI: 10.17026/dans-25d-n6kr.
- Steinmetz, S., mei - juni 2020, Elderly and Loneliness Wave 1 - Wave 2. DOI: 10.17026/dans-xyz-zk56, DOI: 10.17026/dans-zfy-ryrz.
- Bol, N., Antheunis, M., juni - augustus 2019, Video consultation. DOI: 10.17026/dans-zvq-8jzj.
- Hendriks, I., Lubbers, M., Savelkoul, M., juni 2020, Understanding attitudes towards refugees as compared to immigrants
- Jansen, G., Janssen, S., Levels, M., juni 2020, Resilience towards robotization. DOI: 10.17026/dans-232-jp33.

- Rooduijn, M., juni 2020, Corona measures. DOI: 10.17026/dans-2zq-c994.
- Janssen, H., van Ham, M., juli 2020, Neighbourhood perceptions. DOI: 10.17026/dans-22g-rgt5.
- Lo Iacono, S., Juli 2020, Effect of the Pandemic on Social Trust. DOI: 10.17026/dans-zk3-vyph.
- Verbakel, E., juli 2020, Retrospective informal care career - Follow-up measurement. DOI: 10.17026/dans-z6w-rd24.
- Hilbert, L., augustus 2020, Social exclusion, control and financial scarcity - August 2020. DOI: 10.17026/dans-z3a-22m8.
- Visser, V., de Koster, W., van der Waal, J., november 2020, How to reduce the education gap in participation in citizens' initiatives. DOI: 10.17026/dans-xas-y9fv.
- Centerdata, 2020, Initial Questionnaire - 2020. DOI: 10.17026/dans-xvp-87as.
- Centerdata, april - mei 2020, Work and Schooling - Wave 13. DOI: 10.17026/dans-xgb-wdnc
- Centerdata, september - oktober 2020, Family and Household - Wave 13. DOI: 10.17026/dans-xxu-7mzv
- Centerdata, december 2020 - maart 2021, Politics and Values - Wave 13. DOI: 10.17026/dans-xkc-x8sz.
- Centerdata, mei - juni 2021, Personality - Wave 13. DOI: 10.17026/dans-zry-ptb4.



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

DANS

Interviews in Conservation Research

Interviews met kunstenaars, restauratoren en verzamelaars geven informatie over kunstwerken. Deze interviews zijn echter zelden duurzaam gearchiveerd of gemakkelijk toegankelijk voor onderzoekers, omdat hun bestaan als primaire bron relatief onbekend is. De thematische collectie Interviews in Conservation Research doorbreekt dit en geeft een unieke kijk op het leven achter de schermen van kunstwerken en cultureel erfgoed in Nederland. Het is één van de belangrijkste resultaten van het door NWO-gefinancierde gelijknamige KIEM-project door Sanneke Stigter (Universiteit van Amsterdam) in samenwerking met de Rijksdienst voor het Cultureel Erfgoed en het Rijksmuseum, en is een partnerproject van het NICAS (Netherlands Institute for Conservation+Art+Science) en DANS, als onderdeel van het Interviews in Conservation Initiative, Oral History for Conservation Research. De groeiende collectie bestanden zijn online beschikbaar via DANS. DOI: 10.17026/dans-zja-88ta

Ook sinds kort beschikbaar:

- Nivette, Dr AE (Utrecht University) (2021): A global analysis of the impact of COVID-19 stay at home restrictions on crime. DANS. DOI: 10.17026/dans-xuf-a75p.
- Manen, MSc E.L.M. van (Vrije Universiteit Amsterdam) (2020): Experiences of Dutch maternity care professionals during the first wave of COVID-19 in a community based

maternity care system.

DANS. DOI: 10.17026/dans-zjh-9ae3.

- Borensztajn, D (Erasmus MC) (2021): Shortness of breath in children at the emergency department: variability in management in Europe. DANS. DOI: 10.17026/dans-xhp-qbjg.
- Peat, Dr DC (Leiden University) (2021): Perception and Process sentiment analysis search terms. DANS. DOI: 10.17026/dans-xkj-grcp.
- Heine, Dr. F.A. (Tilburg University); Wang, Dr. T. (Tilburg University); Wittebloostuijn, Prof.dr. A. van (Vrije Universiteit Amsterdam) (2018): Data: Self-Sacrifice for the Common Good under Risk and Competition. DANS. DOI: 10.17026/dans-zqw-rj7j.
- Dijk, C.N. van (Centre for Language Studies, Radboud University); Wonderen, E. van (2021): Effect sizes of experimental studies on cross-linguistic influence in bilingual children. DANS. DOI: 10.17026/dans-z6a-yw53.
- Slijper, H.T. (Business Economics, Wageningen University & Research) et al (2019): Survey data on perceived farm resilience, risk management, risk preferences, and risk perceptions from 11 European countries. DANS. DOI: 10.17026/dans-xgp-sr7c.
- Lenoir, MA M.L. (Gemeente Den Haag, Dienst Stadsbeheer, Afdeling Archeologie.) (2019): Dedemsvaartweg 178-392 en 161-181, Gemeente Den Haag. DANS. DOI: 10.17026/dans-z25-pwzc.
- Emhardt, SN (Open Universiteit Nederland) (2021): Study: Effects of displaying the teacher's gaze and mouse cursor cues in video lectures on students' learning. DANS. DOI: 10.17026/dans-zyn-rmy2.
- Konkol, Dr. M. (University of Twente, Faculty of Geo-Information Science and Earth Observation); Jager-Ringoir, Drs. K.A. (University of Twente, Faculty of Geo-Information Science and Earth Observation); Zurita-Milla, Prof. Dr. R. (University of Twente, Faculty of Geo-Information Science and Earth Observation) (2021): Open Educational Resources - Basic concepts, challenges, and business models. DANS. DOI: 10.17026/dans-zcr-e2mw.
- Tiessen-Raaphorst, Drs. Z.H. (Nederlandse Sportraad) (2019): Sportbeleving bevolkingsonderzoek 2019. DANS. DOI: 10.17026/dans-zn5-mw2f.



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



Credits: foto Louise Wijnberg

AGENDA

14, 21 en 28 oktober • Online

ASReview Academy

Deze Academy leert om met de ASReview-software sneller systematisch literatuuronderzoek te doen. Instappen kan op 3 momenten en niveaus: 14-10 (beginners), 21-10 (advanced) en 28-10 (simulation). odissee-data.nl/nl/asreview-nl

19 - 21 oktober • Online

EGI Conference 2021

Het thema van dit jaar is 'Beyond the Horizon: Shaping the Digital Future'. egi.eu/egi-conference/2021-beyond-the-horizon

19 - 22 oktober • Beijing

iPRES 2021

17e internationale conferentie over digitale preservatie. ipres2021.ac.cn

20 oktober • Online

SoDA Data Drop-in

Maandelijks online moment voor het bespreken van uitdagingen op data science gebied. odissee-data.nl/soda

26 - 28 oktober • Dún Laoghaire

i-Society 2021

Internationale conferentie over 'the information society'. i-society.eu

30 oktober - 2 november • Salt Lake / Online

ASIS&T 2021 Annual Meeting

Het thema van dit jaar is 'Information: Equity, Diversity, Inclusion, Justice and Relevance'. asis.org/am21

2 - 4 november • Wenen

Conference on Cultural Heritage and New Technologies

Jaarlijkse conferentie over cultureel werelderfgoed en de laatste technologische ontwikkelingen. chnt.at

4 november • Wereldwijd

World Digital Preservation Day

Het thema dit jaar is 'Digits: for Good'. dpconline.org/events/world-digital-preservation-day

15 - 18 november • Online

SURF Research Bootcamp

SURF organiseert samen met o.a. DANS en het NL eScience Center in 4 middagen een online DCC Research Bootcamp. surf.nl

18 november • Utrecht

ODISSEI Community Conference

Conferentie over de ontwikkeling van de computationele sociale wetenschappen in de afgelopen vijf jaar en de toekomst. odissee-data.nl/nl/2021/07/save-the-date-odissee-community-conference-2021-nl

25 november

Jubileumviering Centerdata

Centerdata bestaat dit jaar 25 jaar, en viert dit op 25 november met lezingen en workshops. centerdata.nl

29 november • Parijs / Online

EDDI 2021

De 12e jaarlijkse Europese DDI-gebruikers conferentie. eddi-conferences.eu/eddi-2021

ODISSEI-directeur Pearl Dykstra over vijfjarig bestaan ODISSEI:

“Data-infrastructuur is een gedeeld goed”

Deze maand bestaat ODISSEI, de nationale onderzoeks-infrastructuur voor de sociale wetenschappen, vijf jaar.

Suze Zijlstra

ODISSEI is in 2016 opgericht om samenwerking tussen dataverzamelaars te bevorderen en grote datasets bij elkaar te brengen. Inmiddels zijn 42 Nederlandse onderzoeksinstituten en sociaalwetenschappelijke faculteiten betalend deelnemer geworden. Wetenschappelijk directeur van ODISSEI, Pearl Dykstra, merkt dat het draagvlak toeneemt: “Er is in de sociale wetenschappen duidelijk een toegenomen erkenning dat een data-infrastructuur een gedeeld goed is.”

Financiering vanuit NWO in 2020 stelde ODISSEI in staat om de infrastructuur verder uit te breiden. Zo wordt een portal ontwikkeld om al bestaande datasets veel beter doorzoekbaar te maken. Daarnaast ondersteunt ODISSEI onderzoekers op verschillende manieren.

Aanbod

Het maakt met beurzen CBS-microdata gratis of tegen een gereduceerd tarief toegankelijk, waardoor de hoeveelheid wetenschappelijke gebruikers van deze data aanzienlijk is toegenomen. Bovendien houdt ODISSEI het LISS panel structureel in stand, zodat deze belangrijke dataverzameling kan worden voortgezet. Ook biedt ODISSEI technische hulp aan wetenschappers die computationele projecten willen gaan ontwikkelen.

Een belangrijke infrastructurele mijlpaal is al behaald: in het najaar van 2020 is de ODISSEI Secure Supercomputer (OSSC) gelanceerd, in samenwerking met SURF en het CBS. De OSSC stelt onderzoekers in staat om in een streng beveiligde omgeving onderzoek te doen aan de hand van CBS-microdata en andere datasets, met veel grotere rekenkracht dan voorheen mogelijk was. Hiermee kunnen analyses veel sneller en efficiënter worden uitgevoerd.

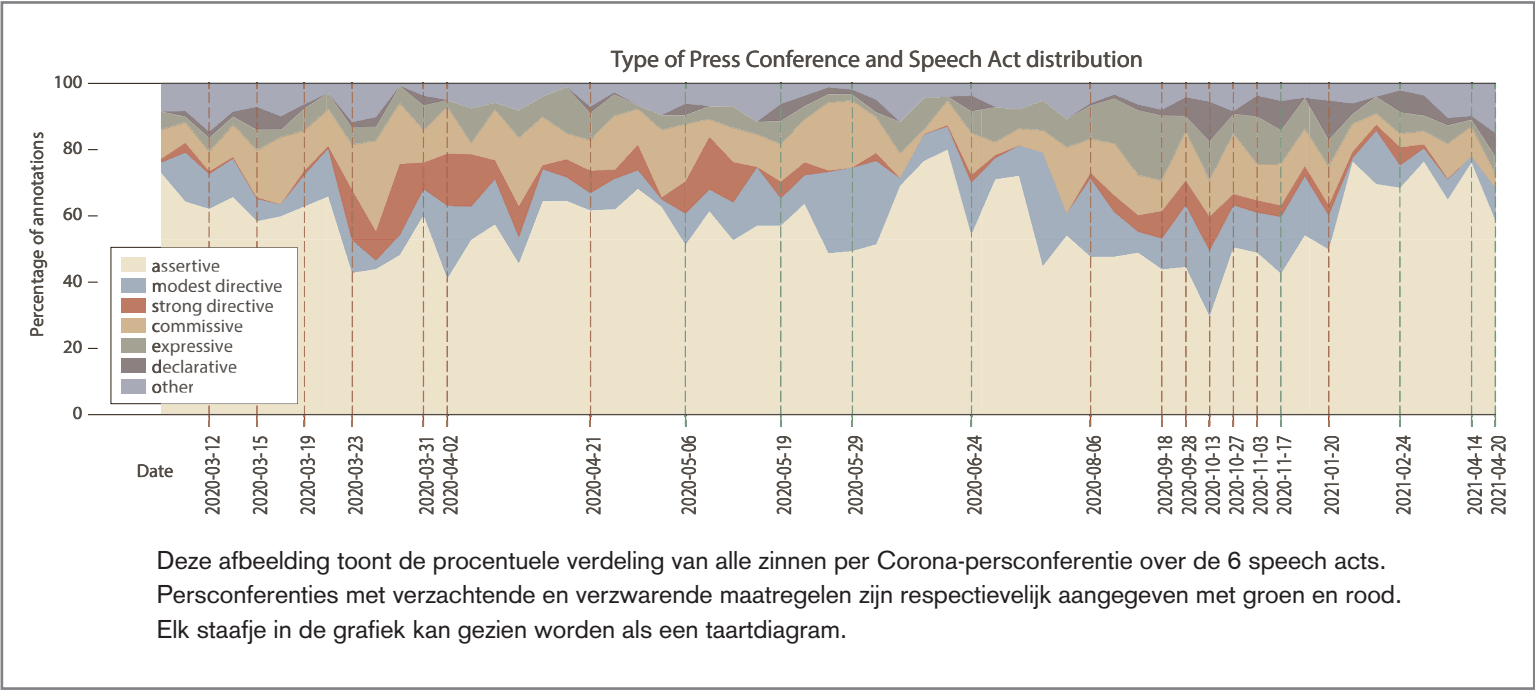
Mogelijkheden

Dykstra is trots op de resultaten die de afgelopen vijf jaar zijn bereikt: “Steeds meer onderzoekers gaan zich bekwaamen in het werken met registerdata en in het werken met een supercomputer. Er is een computational shift gaande, en ODISSEI maakt dit mogelijk.”

In de toekomst hoopt ODISSEI uit te breiden door verdere samenwerking met aanverwante velden. De pandemie heeft bijvoorbeeld aangetoond hoe volksgezondheid nauw samenhangt met sociale processen. Door samenwerking met gezondheidswetenschappers kunnen sociale wetenschappers beter begrijpen hoe virussen zich in een samenleving verspreiden, hoe sociale ongelijkheid zich uit op het gebied van gezondheid, en hoe de gezondheid van individuen kan verbeteren door passende veranderingen. Ook kijkt Dykstra uit naar meer samenwerking met de geesteswetenschappelijke infrastructuur CLARIAH: “Het is cruciaal dat we elkaar opzoeken en samen optrekken. We hebben gedeelde belangen en kunnen bovendien veel van elkaar leren.”

odissei-data.nl

Dataset over persconferenties Corona online beschikbaar



Tussen maart 2020 en april 2021 werden 54 persconferenties over het Coronavirus georganiseerd door de Rijksoverheid. Onderzoeker Maarten Marx en zijn team (Universiteit van Amsterdam) hebben alle zinnen uit deze persconferenties handmatig geannoteerd op aanwezigheid van declaraties of toezeggingen (Searle’s speech acts) en op gedrags-

beïnvloedingstechnieken, zoals het complimenteren of bang maken van burgers. Ook voegden de onderzoekers lexicale en grammaticale informatie per zin toe. *Benieuwd naar alle data uit het onderzoek? De data zijn vrij online beschikbaar via DOI: 10.17026/dans-2af-rwmr*

KORT

Centerdata viert 25-jarig jubileum

Op 20 december bestaat Centerdata 25 jaar. Centerdata is een onafhankelijk non-profit onderzoeksinstituut, gevestigd op de campus van Tilburg University, met als missie het beantwoorden van onderzoeksvragen op het gebied van mens en maatschappij. Geïnteresseerden zijn van harte welkom bij de jubileumviering met lezingen en workshops op 25 november. Meer informatie hierover volgt via de website van Centerdata. In het februari-nummer van E-data & Research zal uitgebreid stil worden gestaan bij het jubileum van Centerdata. (EdC) centerdata.nl



FAIR en CARE omgaan met data

CARE bevordert inheemse datasoevereiniteit

Voor inheemse volkeren kan het vrij en open delen van data een keerzijde hebben. De CARE Principles for Indigenous Data Governance geven richting aan een oplossing. Maarten Heerlien

Als gevolg van de Open Science-beweging worden steeds meer dataset die zijn gefinancierd met publieke middelen open beschikbaar gesteld. Hoewel dat het hergebruik van deze data bevordert, waarschuwen inheemse bevolkingsgroepen wereldwijd de laatste jaren dat het open delen van inheemse data een spanningsveld creëert. Enerzijds steunen zij open

data in principe en praktijk, maar anderzijds streven inheemse volkeren de bescherming na van hun eigen rechten en belangen ten aanzien van datasets waarin informatie is vastgelegd over bijvoorbeeld hun cultuur, taal, individuen, gemeenschappen en grondgebied. Als antwoord hierop publiceerde de Global Indigenous Data Alliance (GIDA) in 2019 de CARE Principles for Indigenous Data Governance.

Cruciale rol

CARE moet waarborgen dat inheemse data kunnen worden hergebruikt terwijl het zelfbeschikkingsrecht van inheemse volkeren, vastgelegd in de Verklaring over de

Rechten van Inheemse Volkeren (UNDRIP), wordt gerespecteerd. Toegang tot, en benutting van inheemse data, speelt een cruciale rol in de uitvoering van dit recht en daaruit voortvloeiende rechten, zoals het recht op zelfbestuur. De term ‘inheemse data’ is hierin een breed begrip en omvat alle soorten vastgelegde informatie en kennis die expliciet in relatie kan worden gebracht met een inheems volk. Hierbij doet zich het probleem voor dat dergelijke data vaak zijn samengesteld door overheden en andere instellingen van buitenaf, wat voor een ongelijke machtsverhouding zorgt en de inheemse datasoevereiniteit, de zeggenschap van een be-

volkingsgroep over de eigen data, bemoeilijkt.

Vier hoofdprincipes

CARE staat voor *Collective benefit*, *Authority to control*, *Responsibility* en *Ethics*: vier hoofdprincipes die richting geven aan het vergroten van de betrokkenheid van inheemse bevolkingsgroepen bij beleid en uitvoering omtrent data governance en het respecteren van de belangen van deze groepen. Elk hoofdprincipe is nader gespecificeerd in drie subprincipes; een opzet die verwijst naar de FAIR data principes. CARE wordt dan ook gepresenteerd als complementair aan FAIR: waar de laatste vooral richting geeft aan het

wegnemen van technische barrières om data te delen, benadert CARE dit vanuit een mens- en doelgericht perspectief. Momenteel wordt binnen de Research Data Alliance (RDA), waarvan de International Indigenous Data Sovereignty Interest Group in 2018 de eerste versie van de CARE principes schreef, aan een model gewerkt om FAIR met CARE te operationaliseren onder de noemer *Be FAIR and CARE*.

gida-global.org/care



Nieuwe inzet van tools en datasets in onderzoek en onderwijs

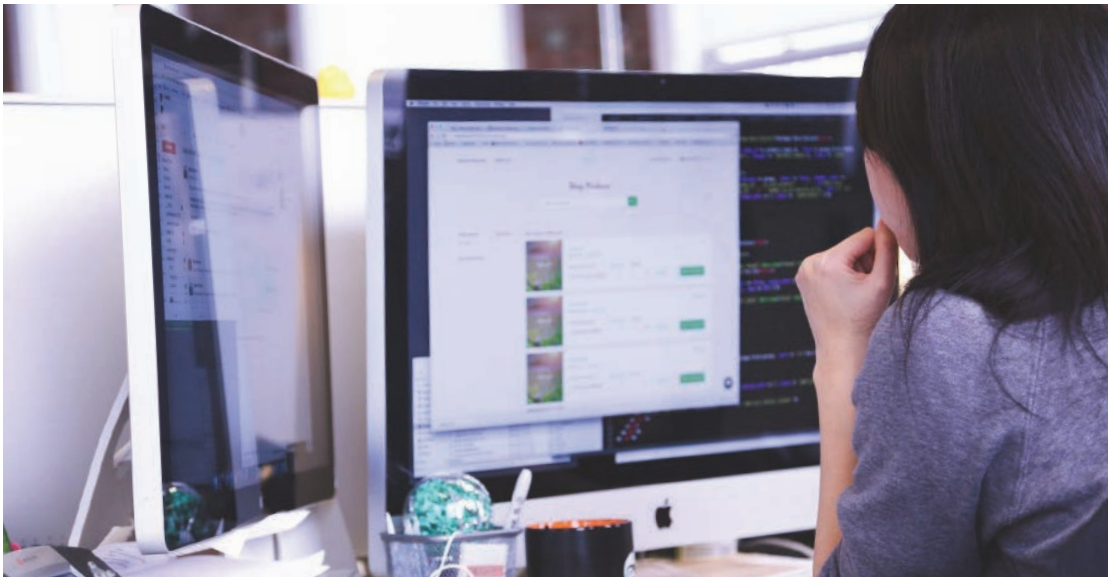
CLARIAH verwelkomt vijftien research en teaching fellows

Maar liefst 15 nieuwe fellows zullen komend jaar CLARIAH-tools en -datasets inzetten en doorontwikkelen binnen hun onderzoek en onderwijs. *Sebastiaan Fluitsma*

Uit de CLARIAH Fellowship Call 2021 zijn 15 aanvragen gehonoreerd. Het gaat om 8 research fellows en 7 teaching fellows, verdeeld over de domeinen Taalkunde, Mediastudies en Sociaaleconomische Geschiedenis. Het doel van de CLARIAH Fellowships is tweeledig. Enerzijds wordt het gebruik van al ontwikkelde digitale tools, datasets en infrastructuur gestimuleerd. Anderzijds kunnen via dat gebruik de concrete ervaringen ermee binnen onderzoek en onderwijs worden verzameld. Hiermee kunnen de digitale resources, het ondersteunend materiaal en de manier waarop CLARIAH ze aanbiedt, verder worden verbeterd.

Research fellows

In de research fellowships voeren de onderzoekers een klein project uit op basis van een expliciete onderzoeksvraag met behulp van



CLARIAH stimuleert met het Fellowships het gebruik van digitale tools, datasets en infrastructuur en verzamelt hiermee concrete ervaringen binnen onderzoek en onderwijs.

Credits: foto StartupStockPhotos via Pixabay

bestaande CLARIAH-data en/of -software. Deze onderzoeksprojecten leveren uiteindelijk op: een onderzoeksartikel, een presentatie op een conferentie of workshop, een lijst met vereisten rond de gebruikte resource (zoals gebruikersbehoeften, wensenlijst, bugs) en een webtekst waarin de gebruikte methode en aanpak worden uitgelegd, gericht op een specifieke gebruikersgroep.

Teaching fellows

In de teaching fellowships gebruikt

een docent aan een Nederlandse universiteit CLARIAH-data en/of -software in een bachelor-, master of research mastercursus, voor een specifieke categorie studenten. Bijvoorbeeld een groep met gelijke achtergrondkennis. De onderwijsprojecten leveren uiteindelijk op: educatief materiaal en een video-college die aangeboden kunnen worden in het CLARIAH-resourcesportal 'Ineo' (verwachte livegang eind 2021), een lijst van vereisten (zoals gebruikersbehoeften, wensenlijst, bugs) en een blogpost

over hoe de studenten het project hebben ervaren. Het educatieve materiaal bestaat uit een gedetailleerde en geïllustreerde set gebruikersinstructies met daarbij, waar mogelijk, aanvullende materialen zoals oefeningen en trainingsbestanden.

De nieuwe CLARIAH-fellows zijn inmiddels bijna allemaal gestart. Op clariah.nl en via de CLARIAH-nieuwsbrief verschijnen regelmatig updates.
clariah.nl/fellowships

GELEZEN

The age of surveillance capitalism. The fight for a human future at the new frontier of power.
Shoshana Zuboff. 2019.

Dit boek draait om de grote tech-bedrijven die zich in onze dagelijkse levens nestelen door 'surveillance'-technologie in social media en apparaten toe te passen. Door het vergaand verzamelen van gebruikersdata proberen bedrijven niet alleen in kaart te brengen hoe mensen zich gedragen, maar ook te voorspellen en uiteindelijk te sturen wat zij doen. Dit alles om meer winst te maken met advertenties.

Met resultaten uit literatuuronderzoek, eigen onderzoek naar organisaties en interviews met werknemers toont Zuboff hoe deze technologie een bedreiging is voor de moderne democratie en de menselijke vrijheid. Het boek omvat het ontstaan van surveillance capitalism tot en met de huidige staat ervan. Hoewel Zuboff het belang aangeeft van het weerstaan van deze vorm van kapitalisme, geeft zij geen duidelijke oplossing hoe dit daadwerkelijk te doen.

Het boek is dan ook een begin: om de controle van techbedrijven (en de mensen aan de top hiervan) te weerstaan, is het nodig om te begrijpen hoe deze bedrijven werken. Het flinke boek leest niet makkelijk weg, maar het slaagt er erg goed in de lezer te doordringen van de vele manieren waarop onze data worden gezocht en misbruikt. (SZ)

en.wikipedia.org/wiki/The_Age_of_Surveillance_Capitalism

COLUMN

Datadroom in gezondheidszorg: delen als default

De COVID-pandemie is een schier onuitputtelijke bron voor reflecties op onze samenleving en organisatievormen. Als een paal boven water staat de absolute noodzaak van samenwerken: van het verspreiden van patiënten door heel Nederland (en daarbuiten) tot en met het delen van kennis over effectieve behandelingen voor patiënten thuis en in het ziekenhuis. Dit alles kon door intensieve samenwerking en snelle kennisdeling. Nu steeds meer mensen in Nederland (eindelijk) goed gevaccineerd zijn en de zwaarste tijd achter ons lijkt te liggen, kunnen we trots zijn op deze samenwerking. Toch is er nog weinig reden tot groots optimisme: de vaccinatiegraad wereldwijd biedt op geen enkele manier bescherming tegen het optreden van mutaties. Wat dit virus nog voor ons in petto heeft, is ongewis, maar een hogere vaccinatiegraad biedt sowieso betere bescherming. "Nobody is safe until everybody is safe", aldus de WHO. Alle reden om aan snelle kennisdeling te doen dus. Maar farmaceutische bedrijven zijn erg terughoudend

en houden vast aan hun intellectuele eigendomsrechten. Hoewel de vaccins grotendeels, en soms zelfs geheel, zijn ontwikkeld met overheidsgeld en via Europese subsidies. Volgens deze bedrijven kan het echt niet sneller. Met alleen de kennis van een patent ben je er namelijk nog niet: de knowhow om het vaccin te produceren, is cruciaal en kennelijk lastig te verspreiden.

Aanalogoos aan bovenstaand voorbeeld zal er in de toekomst sprake zijn van kennis die veel waarde in zich draagt, bijvoorbeeld een gepatenteerd AI-algoritme. Het is niet ondenkbaar dat dit algoritme grote gezondheidsvoordelen oplevert voor groepen mensen met toegang tot

dit algoritme. Hoe maatschappelijk geaccepteerd zouden wij het vinden als een handvol commerciële bedrijven bepaalt wie wel en wie niet toegang heeft tot dit algoritme? Voor dit algoritme is er vanzelfsprekend geen enkele praktische beperking dit zo breed mogelijk toe te passen. Waar een vaccin nog in een fabrieksomgeving gemaakt wordt, kan het algoritme simpelweg van de ene computer naar de andere gekopieerd worden.



ritme simpelweg van de ene computer naar de andere gekopieerd worden.

Kan dit anders? Ja. In de wetenschap is er een groeiende beweging die zogenaamde FAIR data principes hanteert (zie ook go-fair.org/fair-principles). FAIR staat hier voor het feit dat data Findable, Accessible, Interoperable en Reusable is. Gepubliceerde artikelen dienen de onderliggende data FAIR beschikbaar te stellen. De implicaties van deze FAIR-uitgangspunten voor een individuele onderzoeker en zijn/haar onderzoeksgroep zijn groot: waar men vroeger met eigen, unieke datasets publicaties (en status) scoorde, komen deze data nu ten dienste van het grotere geheel.

Dat is voor het individu spannend, maar voor het systeem niet onoverkomelijk. Als we de prikkels voor presteren weghalen bij het individu en veel meer waarde toekennen voor bijdragen aan het geheel zal deze spanning weg-ebben.

Een snelle, gedeelde leeromgeving maakt ons het beste weerbaar tegen een volgende wereldwijde uitdaging. Of het een pandemie is of iets anders, kennis kan levens redden. Delen is dus wel zo fair.

Egge van der Poel

Egge van der Poel (1983) is gepromoveerd in de experimentele deeltjesfysica bij CERN en deed daarnaast een Bachelor Filosofie. Hij zet zich in verschillende rollen in om het leervermogen in de zorg te vergroten en te versnellen, opdat we slimmer gebruikmaken van het nog niet benut potentieel aan data en menselijk talent. Dat doet hij als zelfstandig Data Scientist, als Academic Director bij JADS en als Executive Professor bij TIAS. Egge geeft de column door aan Daniel Kapitan.