



E-DATA & RESEARCH

Jaargang 15 | 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 Leidse taaldata via CLARIAH beschikbaar
- 3 Aan het woord: Research Support Champions SURF
- 4 Ook deze datasets zijn sinds kort beschikbaar
- 5



Interview: microdata zijn toekomst van onderzoek



6 Goudmijn aan data uit Russische staatsarchieven

7 Openjournals.nl faciliteert Diamond Open Access

8 Gastcolumnist Wielaard over het datameisje

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

Meertalige zoekmachine verenigt internationaal verhalennerfgoed

Zoeken naar weerwolven en zeemeerminnen

De zoekmachine ISEBEL bevat zo'n zeventigduizend sagen afkomstig uit Nederland, Duitsland en Denemarken. Verhalenonderzoeker Theo Meder van het Meertens Instituut hoopt dat andere landen volgen. *Mathilde Jansen*

Zoeken naar verhalen over weerwolven, zeemeerminnen of heksen in een collectie van Europese volksverhalen: dat was in de jaren 60 en 70 van de vorige eeuw al een van de ambities van het Meertens Instituut. Het was de tijd van Voskuil, waarin het toenmalige Volkskundebureau veel verhalen verzamelde. Voor die tijd een stap te ver, zo bleek, maar nu door *machine learning* toch eindelijk mogelijk. De techniek staat voor niets. Of toch? "Het was een helse klus", zegt Meder.

Zelf werkt Meder al sinds 1994 aan de Nederlandse Volksverhalenbank. In het door NWO gefinancierde project ISEBEL zijn de sagen uit deze databank nu gekoppeld aan een Duitse en een Deense volksverhalenbank, die je samen kunt doorzoeken. De meeste verhalen komen uit Nederland en Denemarken: die hebben er allebei zo'n dertigduizend geleverd. De overige komen uit Duitsland.

Volksgeloof

Sagen zijn iets anders dan sprookjes, legt Meder uit. Het zijn kortere vertellingen die staan voor een bepaalde vorm van volksgeloof. Daarin schuilt tevens een belangrijk verschil tussen de



Het project ISEBEL maakt het meertalig zoeken naar verhalen over weerwolven en andere erfgoedverhalen mogelijk. Credits: commons.wikimedia.org/wiki/File:Werwolf_von_Neuses.png

genoemde databanken. De Deense en Duitse databanken bevatten namelijk verhalen die zijn opgetekend rond 1900, toen mensen nog daadwerkelijk in heksen en weerwolven geloofden. Dat was niet meer het geval in de jaren 60 en 70 toen de meeste Nederlandse verhalen werden verzameld.

Eigen vertaalmodule

Alle zeventigduizend sagen met de hand vertalen was geen optie, dus dat moest automatisch gebeuren. En daar gooiden de ouderwetse spelling van het Fries en ook het Platduits roet in het eten. Google Translate kon met deze spellingsvarianten niet overweg. De Amerikaanse onderzoekers in het project ontwikkelden daarom een eigen vertaalmodule op basis van *machine learning*. Door het toevoegen van een lijst met domeinspecifieke woorden weet deze module inmiddels ook raad met woorden als 'duivelsdrek' of 'alruin'.

En dan wordt het eindelijk tijd om onderzoek te doen. Meder heeft al een kleinschalig onderzoek gedaan naar de verspreiding van verhalen over weerwolven. "Daar werd bijvoorbeeld van gezegd dat ze vooral zouden voorkomen in katholieke gebieden. Maar als je het Noord-Duitse gebied erbij pakt, blijkt dat niet te kloppen. Want ook daar komen veel weerwolfverhalen voor, maar dit gebied is van oudsher protestants. Dan blijkt het plausibeler om een verband te leggen tussen bosrijk gebied en dit soort verhalen, zoals een Duitse onderzoeker nu heeft gesuggereerd. Al met al verruimt ISEBEL dus ons blikveld."

isebel.eu

ISEBEL (afkorting van: Intelligent Search Engine for Belief Legends) is een samenwerking van het Meertens Instituut, de universiteit van Berkeley (VS) en de universiteit van Rostock (Duitsland).

Tal van mogelijkheden voor onderzoekers en beleidsmakers

Onderzoek met LISS panel data

Wat zijn kenmerken van VVD-kiezers? Hoe staat het met de mentale gezondheid van Nederlanders? Wie gebruikt de CoronaMelder app? Deze en andere vragen worden beantwoord met data uit het LISS panel. *Evi de Cock*

Sinds 2007 worden in het longitudinale LISS panel (Longitudinal Internet studies for the Social Sciences), beheerd door CentERdata, vragenlijsten voorgelegd aan ongeveer 7.500 panelleden van 16 jaar en ouder uit 5.000 huishoudens in Nederland. Het panel is gebaseerd op een willekeurige kanssteekproef uit het bevolkingsregister, getrokken door het Centraal Bureau voor de Statistiek (CBS). Zelfaanmelding voor het panel is niet mogelijk en huishoudens zonder internet krijgen apparatuur in bruikleen.

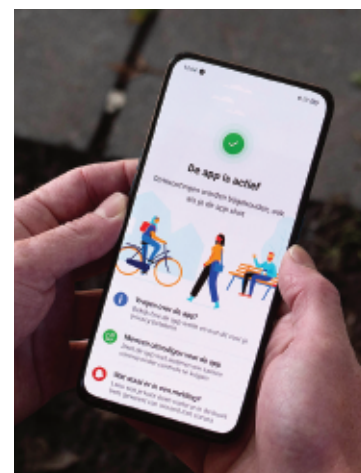
De door het LISS panel verzamelde data zijn beschikbaar via het LISS

Data Archive, al jaren een rijke bron van informatie om allerlei onderzoeksvragen te beantwoorden. De data zijn vrij toegankelijk voor onderzoekers en beleidsmakers na het ondertekenen van een statement.

Data hergebruiken

Onderzoeker Matthijs Rooduijn (Universiteit van Amsterdam) maakt bijvoorbeeld gebruik van LISS panel data voor de blog *Stuk Rood Vlees*. In de rubriek *Plaatjes van de electoraatjes* schrijft hij over de sociaal-demografische kenmerken en opvattingen van de kiezers van de tien grootste partijen. Zo blijken bijvoorbeeld VVD-stemmers steeds kosmopolitischer te zijn geworden. De data komen uit de LISS kernvragenlijst *Politics and Values* die jaarlijks wordt afgenomen.

Ook Peter van der Velden (CentERdata) gebruikt LISS panel data in zijn onderzoek naar de effecten van corona op de mentale gezondheid.



Het koppelen van gegevens levert rijke datasets op, bijvoorbeeld voor een onderzoek naar het gebruik van en opvattingen over de CoronaMelder app.

Credits: coronamelder.nl

Uit dit onderzoek kwam onder andere naar voren dat de mentale gezondheid van de Nederlandse bevolking in 2020 niet slechter was dan vóór de corona-uitbraak.

ODISSEI call LISS panel

In mei 2021 is de nieuwe ODISSEI call geopend voor LISS panel projecten. Naast reguliere voorstellen om data te verzamelen via het LISS panel nodigt ODISSEI onderzoekers uit om voorstellen in te dienen voor projecten die survey data combineren met CBS microdata. Er is in deze ronde financiering beschikbaar voor vier reguliere projecten van veertig vragen. Eén beurs zal daarnaast worden toegekend aan een project dat LISS panel data koppelt met CBS microdata. (SZ)

Lees verder op pagina 2, linksonder

Lunchlezing CLARIAH over transcriptie kranten

Suze Zijlstra

De Koninklijke Bibliotheek heeft veel zeventiende-eeuwse kranten gedigitaliseerd, maar de tekstherkenningstechniek OCR was niet goed genoeg om deze kranten goed doorzoekbaar te maken. Tijdens een CLARIAH lunchlezing op 8 april jongstleden spraken Nicolien van der Sijs (Radboud Universiteit) en Joris van Zundert (Huygens ING) over het project dat de transcriptie van deze kranten mogelijk maakte. Meer dan tweehonderd vrijwilligers werkten hieraan, wat een corpus van wel 20 miljoen woorden opleverde. Van Zundert lichtte eerst de achtergrond van het project binnen CLARIAH toe. CLARIAH heeft vanuit het werkpakket 'tekst' een oproep gedaan aan wetenschappers om pilotprojecten voor te stellen. Voor CLARIAH speelt de achterliggende vraag of het software kan ontwikkelen die breder toepasbaar is, of dat het kan leiden tot de ontwikkeling van bijvoorbeeld een tutorial

over al bestaande methoden en technieken. Van der Sijs droeg haar krantentranscriptieproject aan. Dit project bood verschillende uitdagingen op het gebied van digitale tekst: niet alleen wat betreft de transcriptie, maar ook wat betreft metadata en annotatie. Nu de zeventiende-eeuwse kranten zijn getranscribeerd, gaan ze in de toekomst kijken naar de toepassing van programma's als Transkribus of Calamari. Van der Sijs presenteerde hoe ze met de vrijwilligers te werk is gegaan. Het project draaide om kranten tussen 1618 en 1700, waarvan al enige metadata beschikbaar waren. Vrijwilligers gingen aan de slag toen Transkribus nog niet ver was ontwikkeld. De bronnenset is voor taalwetenschappers en historici uniek: het is een longitudinale collectie met een mooi aaneengesloten corpus. Je kan hier zowel op micro-

GEHOORD & BIJGEWOOND



Tijdens de meeting werden de mogelijkheden voor geo-analyses met CBS microdata verkend. Credits: Daria from TaskArmy.nl via Unsplash.

als op macroniveau onderzoek naar doen. Na de transcriptie zijn de metadata nog uitgebreid verbeterd. In de toekomst willen ze er uitgebreider onderzoek mee doen naar de herkomst van nieuws. Vanwege verschillende spellingsvarianten van plaatsnamen en het feit dat correspondenten niet altijd duidelijk waren over welke locatie ze schreven, is dit nog een mooie uitdaging.

CLARIAH.nl

ODISSEI workshop Integrating Microdata with Geospatial Approaches

Eva Heitbrink

Onderzoek met CBS-microdata biedt veel mogelijkheden voor interessant onderzoek, maar wat zijn de mogelijkheden voor geo-analyses met CBS microdata? Dat is waar de ODISSEI workshop van 22 april over ging.

De middag begon met een introductie van geo-analyse door Marco

Helbich (UU), die zelf onderzoek doet naar de invloed van omgevingsfactoren op mentale gezondheid met behulp van CBS microdata gekoppeld aan gps-data. Wat onderzoek met geo-data aantrekkelijk maakt, vertelt Helbich, is dat, vooral in Nederland, er vele geografische datasets gratis te gebruiken zijn. Deze datasets, gecombineerd met CBS microdata, kunnen voor de sociale wetenschappen nieuw en belangrijk onderzoek tweeweg brengen. Hierna presenteerde Hilde van Oirschot (CBS) en Deirdre Bosch (CBS) over de mogelijkheden van CBS microdata, de geografische datasets die CBS zelf beschikbaar heeft, en hoe met deze data aan de slag te gaan. Bij deze datasets blijft de bescherming van privacy de hoogste prioriteit voor CBS. Geografische datasets met daarin gebieden met weinig datapunten zijn dan ook niet beschikbaar.

Tijdens het volgende deel van de middag presenteerden onderzoekers hun projecten uitgevoerd met CBS microdata en geografische data. Als eerst presenteerde Bastian Ravesteijn (EUR) zijn Kansenkaartproject, waarin Ravesteijn en zijn team met gepseudonimiseerde data van CBS de plek waarin mensen opgroeien verbindt aan hun huidige socio-economische status. Zo kan men op de kaart zien dat mensen die opgroeien in een relatief arm huishouden, nu vaak ook een lagere socio-economische status hebben. Na Ravesteijn presenteerde Nienke Boderie (Erasmus MC) haar project *Do neighbourhoods affect health?*, waarin zij en haar team onderzoeken of de buurt waarin iemand woont effect heeft op zijn of haar (mentale) gezondheid. Hiervoor kijkt het team naar verhuisbewegingen binnen Rotterdam en gezondheidsgegevens, waarvoor zij gebruik

maken van gepseudonimiseerde CBS microdata. De middag sloot af met break-out rooms waarin de aanwezigen vragen konden stellen aan de presentatoren, onder andere over hun eigen ideeën voor onderzoek met microdata en geo-analyse.

odissei-data.nl

NPSO-lezingenmiddag over onzekerheden

Ricarda Braukmann

Op 6 mei organiseerde het Nederlandstalig platform voor survey-onderzoek (NPSO) een digitale lezingenmiddag over de presentatie van onzekerheden.

De middag trok zo'n 120 deelnemers en werd geopend door journalist Sanne Blauw. Als Correspondent Ontcijferen en in 'Het beste verkochte boek ooit' helpt Blauw het algemeen publiek om cijfers in de juiste context te plaatsen. Op deze bijeenkomst presenteerde zij haar visie op onzekerheid. Volgens Blauw worden we tijdens een pandemie continue geconfronteerd met onvolledige kennis en kunnen dit soort complexe problemen moeilijk juist worden ingeschat. Blauw citeert onderzoek naar het voorspellen van uitkomsten en pleit voor het durven twijfelen. Onzekerheden erkennen, jezelf bewust worden van je eigen vooroordelen en het veranderen van mening, leiden volgens haar tot de beste uitkomsten.

Edwin de Jonge, methodoloog bij het CBS, benadrukt dat onzekerheden een essentieel onderdeel zijn van onderzoek. Aan de hand van twee voorbeelden toont De Jonge hoe onzekerheden gepresenteerd kunnen worden. Ook benadrukt hij het belang van communicatie. "We denken vaak dat mensen betrouwbaarheidsintervallen niet kunnen begrijpen maar uit onderzoek blijkt dat ook leken onzekerheden best goed kunnen interpreteren."

De laatste presentatie door Marko Roos van het CBS ging verder in op de mogelijkheden van visualisaties. Roos en zijn team kregen in hun onderzoek naar mobiliteit te maken met grote hoeveelheden data. Door verplaatsingen tussen steden samen te vatten in zogenaamde 'donut-maps' kon het team van Roos op een slimme manier meer informatie weergeven door minder data te tonen.

De namiddag werd afgerond met een discussie rondom een aantal stellingen. De sprekers waren het eens dat het op de juiste manier communiceren over onzekerheden kan helpen om het begrip ervan te vergrootten. "Bij het weerbericht vinden we het ook niet meer gek als iemand zegt dat er morgen 80% kans is op regen".

npsso.net



Fragment uit Oprechte Haerlemsche courant, Haarlem, 1694/01/02 00:00:00, p. 1., urn=ddd:011227040, Delpher.

Vervolg van pagina 1

Data LISS panel inzetten

verzamelde gegevens kunnen vervolgens gekoppeld worden aan andere data in het LISS Data Archive. Dit levert een zeer rijke dataset op waarmee vele onderzoeksvragen beantwoord kunnen worden. Zo onderzoekt Nynke van der Laan (Tilburg University) in opdracht van het Ministerie van VWS het gebruik van en opvattingen over de Corona-Melder app. Voor deze vragenlijst werden panelleden geselecteerd die eerder hadden meegedaan aan de kernvragenlijst *Health* en aan een onderzoek naar de gevolgen van het coronavirus, zodat extra achtergrondinformatie gebruikt kon worden. De voorlopige resultaten laten

zien dat jongeren en mensen met basisschool of VMBO als hoogste opleidingsniveau de app naar verhouding minder gebruiken. De overheid wil daarom meer aandacht gaan besteden aan deze doelgroepen.

centerdata.nl/nl/databank/liss-panel-data

Vele mogelijkheden

Wil je meer informatie of ben je benieuwd of ook jouw vraagstuk beantwoord kan worden met gegevens uit het LISS panel? Neem dan contact op via info@lissdata.nl.

OVERNEMEN ARTIKELEN

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

Data toegankelijk voor alle geesteswetenschappers

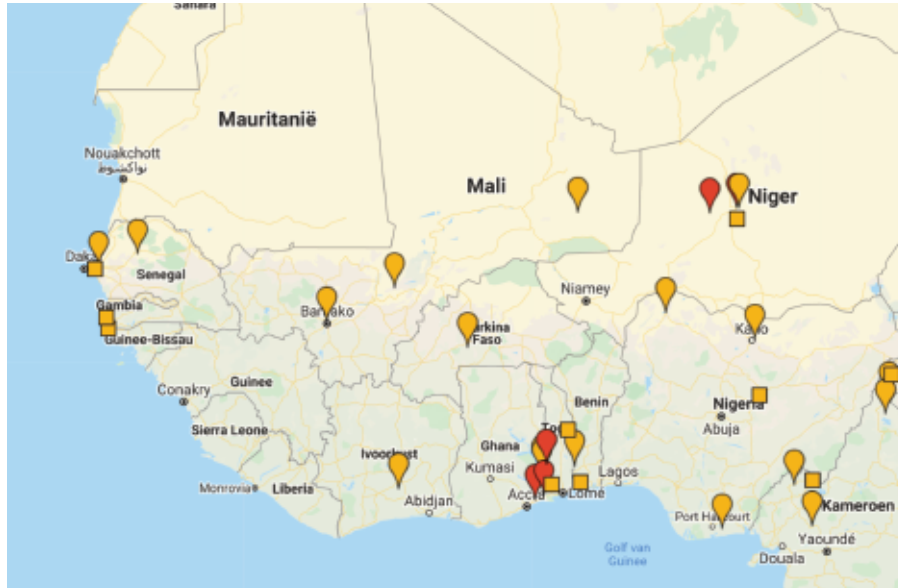
Leidse taaldata via CLARIAH beschikbaar

Het Leiden University Centre for Linguistics (LUCL) heeft in de loop der tijd vele datasets verzameld. Deze komen nu beschikbaar in de CLARIAH-infrastructuur. *Erica Renckens*

“De vele data die onderzoekers van het LUCL over de jaren heen hebben verzameld, bleven vaak onzichtbaar voor mensen buiten het instituut”, vertelt Sara Petrollino, taalkundige aan de Universiteit Leiden. “We wilden de informatie over deze data én de data zelf toegankelijk maken als een bron voor onderwijs en onderzoek. Vooral in deze lastige corona-tijd, waarin zelf data verzamelen vaak onmogelijk is, is dat heel belangrijk.”

In eerste instantie zetten student-assistenten de veelal analoge opnames en de bijbehorende metadata alleen om in digitale bestanden. “Daardoor werden de betrokken onderzoekers zich ook meer bewust van de moderne gang van zaken op het gebied van datamanagement en toegankelijkheid.”

De metadata van deze 140 datasets zijn doorzoekbaar via de database Leiden Language Data (Leiland). Daaronder bevinden zich onder andere zeventien datasets over het Nederlands, waaronder geschreven dialectdata, interviews over taalattitude en testresultaten van codeswitching met het



Fragment van het geografische overzicht met datasets, door de jaren heen verzameld.

Credits: Leiden University Centre for Linguistics (LUCL)

Papiaments. Maar er zijn ook datasets over meer exotische talen te vinden, zoals over het Akileh (Tanzania), het Alorese (Indonesië) en het Puquina (Peru).

Representatieve selectie

Een aantal van de LUCL-datasets komen nu ook beschikbaar in de CLARIAH-infrastructuur. Hiermee worden de data toegankelijk voor alle onderzoekers in de geesteswetenschappen. “We maken hiervoor een representatieve selectie met onder andere video-opnames van Afrikaanse, Zuid-Ame-

rikaanse en Austronesische talen, audio en geschreven data van leerders van het Nederlands, Spaans en Italiaans, en geschreven taaldata, zoals brieven en manuscripten”, vertelt Petrollino.

De komende twee jaar zullen de metadata van deze datasets worden omgezet in CMDI-formaat, dat gebruikt wordt in de CLARIAH-infrastructuur. De geselecteerde datasets – en wellicht meer, afhankelijk van de voortgang van het project – worden opgeslagen bij DANS, een van de CLARIAH Centers.

leiland.uliet.net

Drie winnaars SURF Research Support Championship 2021

SURF zet met het nieuwe initiatief Research Support Championship onderzoekssupporters in het zonnetje. Welk advies geven de winnaars aan hun collega's? *Josje Spinhoven*

Maria Kamp

“Vraag-gestuurde onderzoeksondersteuning”



In de categorie universiteiten won Maria Kamp, coördinator Research Support bij de Universiteit Twente. Zij werd door maar liefst 5 collega's genomineerd. Volgens het juryrapport ‘inspireert zij anderen met haar daad-

krachtige aanpak en haar samenwerkingsmentaliteit. Maria kruipt in de huid van de onderzoekers. Ze maakt complexe zaken begrijpelijk en neemt onderzoekers administratief werk uit handen, zodat zij zich kunnen richten op hun primaire taak.’

Volgens Maria is het doel van research support om effectieve, efficiënte en vooral vraaggestuurde onderzoeksondersteuning te bieden. “Zorg ervoor dat onderzoekers zich kunnen focussen op het doen van excellent onderzoek en het maken van een echte maatschappelijke impact.”

Mariëtte van Selm

“Beter achteraf om vergiffenis vragen, dan vooraf om toestemming”



‘Een onmisbare adviseur’, aldus het juryrapport over Mariëtte van Selm. Deze winnaar in de categorie hbo werkt als coördinator Research Data Management ondersteuning bij de Bibliotheek UvA-HvA. ‘Met niet aflatende

energie en pragmatische inzet helpt Mariëtte onderzoekers bij het managen van hun data. Ze luistert goed en brengt de juiste mensen op het juiste moment bij elkaar. Mariëtte vormt een waardevolle schakel tussen de wereld van praktijkgericht onderzoek en de academische wereld.’

Mariëttes tip voor andere onderzoeksondersteuners? “Soms kun je beter achteraf om vergiffenis vragen dan vooraf om toestemming.”

Bob van Dijk

“Goed communiceren”



Bob van Dijk, adviseur High Performance Computing bij Amsterdam UMC, kwam als winnaar uit de bus in de categorie umc & research. Het juryrapport: ‘Deze Research Support Champion slaat de brug tussen onderzoeker en

research IT door onderzoekers te begeleiden in technische oplossingen en hun supportvraag te vertalen naar technische eisen voor research IT, zoals rekencapaciteit en cloudomgevingen.’

Zijn advies aan collega's: “Het is het allerbelangrijkste dat je goed communiceert. Niet te lang wachten met antwoord geven op vragen. Ook als je het niet weet, kun je dat duidelijk vertellen. Ook erg belangrijk: neem nadat je advies hebt gegeven, niet aan dat het goed was, als je niets meer hoort. Vraag altijd om feedback. En ten slotte: zorg dat je collega's hebt waarop je kunt terugvallen. Zelfs als je een vrij unieke positie hebt binnen een organisatie - zoals in mijn geval - moet je ervoor zorgen dat er bij alles wat je doet iemand is die dat van je zou kunnen overnemen.”

Het volledige juryrapport staat op de website van SURF. Ook in 2022 wordt de prijs weer uitgereikt. surf.nl/en/support-champions

OpenAIRE: schatkamer met 146 miljoen onderzoeksoutput

Europa biedt vele initiatieven voor onderzoekers en dataprofessionals om publicaties, onderzoeksdata en kennis over researchdatamanagement met elkaar te delen. OpenAIRE is een van de voorbeelden. *Ricarda Braukmann en Heidi Berkhout*

OpenAIRE staat voor Open Access Infrastructure for Research in Europe. Via de website openaire.eu hebben onderzoekers toegang tot (open) wetenschappelijke publicaties, datasets en onderzoekssoftware, afkomstig uit meer dan 10.000 betrouwbare bronnen. DANS is een van die bronnen. OpenAIRE linkt de metadata records aan elkaar via de OpenAIRE Research Graph, bijvoorbeeld onderzoeksfinanciers aan publicaties. Elly Dijk, RDM Specialist bij DANS: “Voor onderzoekers is OpenAIRE een schatkamer met 146 miljoen onderzoeksoutput, waaronder open access publicaties, open onderzoeksdata, en onderzoekssoftware.”

Dijk vervolgt: “OpenAIRE heeft ook diensten ontwikkeld die onderzoekers helpen om zelf Open Science uit te oefenen. Voorbeelden zijn Argos (argos.openaire.eu) en

Amnesia (amnesia.openaire.eu).

Argos is een online tool voor Data Management Planning. Amnesia is een tool om onderzoeksdata te anonimiseren, zodat de data gedeeld kunnen worden volgens de richtlijnen van de General Data Protection Regulation (GDPR).”

National Open Access Desk

Hiernaast biedt OpenAIRE voor onderzoekers en voor datasupporters, repositorymanagers en onderzoekscoördinatoren andere diensten. Dijk: “Elk Europees land heeft een National Open Access Desk (NOAD) die Open Science in hun land bevordert. De afgelopen zes jaar was DANS met de bibliotheek van de Technische Universiteit Delft de NOAD voor Nederland. In dat kader hebben we bijvoorbeeld bijgedragen aan verschillende open science bijeenkomsten in Nederland, zoals de Open Day on Open Science georganiseerd door DANS.”

Trainingsmateriaal RDM

DANS was taakleider van de Taskforce RDM in OpenAIRE Advance, het meest recente OpenAIRE-project. Het doel van deze

Taskforce was om de kennis van de NOADS op het gebied van RDM te vergroten en om gezamenlijk nieuw trainingsmateriaal te ontwikkelen. Dijk: “Hoeveel kost databeheer? Hoe kunnen gevoelige data worden verwerkt? Hoe richt je researchdatamanagement op een instituut in? Op deze en andere vragen geven wij antwoord, in handleidingen, infographics, blogs en tijdens webinars en andere events. Ook laten we voorbeelden zien van het hergebruik van data en hoe de FAIR dataprincipes dit ondersteunen. We geven een overzicht van datamanagementplannen van Horizon 2020 projecten. Vele dataprofessionals zetten de door onze Taskforce ontwikkelde trainingsmaterialen in. Samen combineren we zo de theorie en de praktijk van Open Science.”

OpenAIRE AMKE

Dit jaar liep OpenAIRE Advance af. OpenAIRE is opgegaan in de rechtspersoon, onder Griekse wet, OpenAIRE AMKE. Dit biedt de mogelijkheid om de bovengenoemde diensten van OpenAIRE verder te ontwikkelen. DANS is ‘regular member’ van OpenAIRE AMKE.

openaire.eu

DANS lanceert Data Station per domein

DANS lanceert dit jaar domeinspecifieke Data Stations: een plek waar onderzoekers uit verschillende wetenschappelijke disciplines zelf hun data online kunnen opslaan, delen en publiceren, tijdens de onderzoeksperiode en daarna. De Data Stations zijn gebaseerd op Harvard University's open source Dataverse-technologie.



De Data Stations sluiten zoveel mogelijk aan bij de verschillende wensen van specifieke onderzoeksgemeenschappen. Sommige stations bieden bijvoorbeeld op verzoek van een community ook toegang tot elders opgeslagen datasets, terwijl ook metadata-schema's en vocabulaires kunnen verschillen. Elk Data Station krijgt een Data Station Manager, aanspreekpunt voor de betreffende onderzoeksgemeenschap. Naast de stations biedt DANS de diensten DataverseNL en Data Vault. DataverseNL is een platformdienst waarin universiteiten, hogescholen en onderzoeksinstituten hun eigen repository kunnen vestigen. Data in de Data Stations en DataverseNL komen automatisch in de Data Vault, een beveiligd, betrouwbaar en gecertificeerd datarepository voor de lange termijn opslag van data. (HB) dans.knaw.nl

CLARIAH-website vernieuwd

De website van CLARIAH, onderzoeksinfrastructuur voor de geestwetenschappen, is vernieuwd. Via deze site zijn alle CLARIAH-bronnen (data, tools, standaarden en workflows) te bereiken. Vanaf september worden deze bronnen aangeboden in de nieuwe portal Ineo. (ER) clariah.nl

NWO moedigt gebruik Persistent Identifiers aan

Onlangs heeft NWO het belang van Persistent Identifiers (PIDs) om duurzaam te kunnen verwijzen naar digitale objecten onderstreept. Naast de bekende Digital Object Identifier (DOI) voor artikelen en datasets, gaat NWO ook inzetten op het gebruik van ORCID IDs voor onderzoekers, Crossref Grant IDs voor NWO-subsidies en PIDs voor organisaties. Het gebruik van PIDs maakt het mogelijk om informatie aan elkaar te linken en op gestructureerde manier te verwerken. Volgens NWO is het van belang om samen met organisaties zoals SURF en DANS het gebruik van PIDs in de onderzoeksgemeenschap te stimuleren. (RB) nwo.nl

SINDS KORT BESCHIKBAAR

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

CentERdata

• Effect 'thuisonderwijs' tijdens coronacrisis

Vanwege de coronacrisis sloten de scholen in Nederland op 16 maart 2020 de deuren voor meerdere weken. Onderzoeker dr. Thijs Bol van de Universiteit van Amsterdam onderzocht daarom in april 2020 in het LISS panel het effect van 'thuisonderwijs' op sociale ongelijkheid. Een van de conclusies uit het onderzoek is dat de sluiting van de scholen leidt tot groeiende ongelijkheid onder leerlingen. Het blijkt dat hoogopgeleide ouders hun kind veel vaker helpen met het schoolwerk en zich beter in staat voelen om te helpen dan laagopgeleide ouders. De data van dit onderzoek zijn beschikbaar via het LISS Data Archive. dataarchive.lissdata.nl



Credits: Julia M Cameron, Pexels

Ook sinds kort beschikbaar:

Studies LISS panel

- Glas, I., Demant, F., maart 2020, National Freedom Study 2020
- Bouma, J., april 2020, State of the environment and environmental policy
- Reeskens, T., Halman, L., Luijckx, R., mei 2020, European Values Study 2017 - Additional measurement May 2020
- Reeskens, T., Halman, L., Luijckx, R., oktober 2020, European Values Study 2017 - Additional measurement October 2020
- CentERdata, november-december 2020,

Health - Wave 13

- CentERdata, september-oktober 2020, Economic Situation: Assets - Wave 7
- CentERdata, augustus-september 2020, Religion and Ethnicity - Wave 13
- CentERdata, oktober-november 2020, Social Integration and Leisure - Wave 13



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

DANS

• Collectie maritieme opgravingsdossiers Batavialand compleet

Stichting Batavialand heeft de papieren (scheeps)archeologische opgravingsdocumentatie van het voormalig maritiem depot van de Rijksdienst voor het Cultureel Erfgoed gedigitaliseerd. Deze bijzondere collectie is nu compleet toegankelijk via DANS. Het betreft hoofdzakelijk de dossiers van wrakken die werden aangetroffen in Flevoland zoals bij de ontginning van de Wieringermeer en de Noordoostpolder. Het betreft waardevolle en unieke brondocumentatie van wrakken uit de IJzertijd tot en met het begin van de 20ste eeuw. De wrakdossiers bevatten onder andere meldingsbriefjes, correspondentie, situatieschetsen, vondstenlijsten, interne rapporten en divers beeldmateriaal.

DOI: 10.17026/dans-x6z-3dnp

easy.dans.knaw.nl/ui/datasets/id/easy-dataset:155737

Ook sinds kort beschikbaar:

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

- Batavialand te Lelystad, Maritiem Archeologisch Depot (2019): Opgravingsdocumentatie scheepswrak Buytensorgh.
- DANS. DOI: 10.17026/dans-zv3-hpm5

- Buringh, Dr. Dr. Ir. E. (Utrecht University) (2020): European urban population, 700 - 2000. DANS. DOI: 10.17026/dans-xzy-u62q
- Cebo, Prof. Dr. DC (Daniel Cebo) (2017): Immunisation for HIV-Infected People. DANS. DOI: 10.17026/dans-z28-jb59
- Doornik, MSc. S.F.W. van (University of Groningen) (2019): Eating Disorder Symptoms - Relationships with Meaning in Life. DANS. DOI: 10.17026/dans-zh5-cq72
- Ebben, PhD. RHA (HAN University of Applied Sciences) (2019): Quality of solo ambulance care by physician assistants versus ambulance nurses for non-conveyed patients: An observational study in the Netherlands. DANS. DOI: 10.17026/dans-zqw-qbqp
- Hartevelt, dr. E.; Brug, prof. W. van der; de Lange, prof. S.L. de; Meer, prof. T.W.G. van der (University of Amsterdam) (2021): Replication data (with rounded standardized macro variables) for "Multiple roots of the populist radical right: support for the Dutch PVV in cities and the countryside". DANS. DOI: 10.17026/dans-25u-s3af
- Inist-CNRS (2018): OpenGrey. System for Information on Grey Literature in Europe. DANS. DOI: 10.17026/dans-xtf-47w5
- Prikaziuk, MSc E. (Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente) (2021): Google Earth Engine Sentinel-3 OLCI Level-1 Dataset Deviates from the Original Data: Causes and Consequences. DANS. DOI: 10.17026/dans-xb8-efke
- Smeets, dr. E. (KBA) (2021): Monitor Scholen Passend Onderwijs, 2019. DANS. DOI: 10.17026/dans-z37-auws
- Snijder, dr R. (OAPEN Foundation) (2021): Words Algorithm Collection : finding closely related open access books using text mining techniques. DANS. DOI: 10.17026/dans-xbm-qr5e



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

AGENDA

7 - 10 juni • Online

16th International Conference on Open Repositories
Het thema is dit jaar: 'Open for all'. or2021.openrepositories.org

10 - 11 juni • Online

IWSG 2021 Workshop
Dertiende internationale workshop over Science Gateways. iwsgateway.github.io/iwsg2021

14 en 17 juni • Online

Responsible Open Science
Workshop georganiseerd door ODISSEI, CESSDA-ERIC en DANS over de basisprincipes van RDM voor sociale wetenschappers. dans.knaw.nl

15 - 17 juni • Online

Dataverse Community Meeting 2021
Jaarlijks netwerkevent voor de Dataverse-community. projects.iq.harvard.edu/dcm2021

16 - 18 juni • Online

EOSC Symposium 2021
Een inclusief platform voor uitwisseling van doelstellingen, visies, activiteiten en resultaten. eosc.eu/events/eosc-symposium-2021

23 - 25 juni • Online

LIBER Annual Conference 2021
Het thema is dit jaar: 'Libraries and Open Knowledge: from vision to implementation'. liberconference.eu

25 juni • Online

SHARE-Nederland
SHARE-gebruikers zijn uitgenodigd voor discussie en kennisdeling. odissei-data.nl

1 juli • Online

Essentials 4 Data Support
Een introductiecursus voor diegenen die onderzoekers (willen) ondersteunen bij het opslaan, beheren, archiveren en delen van hun onderzoeksdata. datasupport.researchdata.nl

9 - 11 juli • Online

NASKO 2021: North American Symposium on Knowledge Organization
Het thema is dit jaar: 'Resilience, Resistance, and Reflection: Knowledge Organization at a Crossroads'. iskocus.org/nasko2021.php

13 - 16 juli • Online

IPERION HS 1st Doctoral Summer School
Theoretische en technische lezingen gegeven door een groot aantal internationale deskundigen op het gebied van erfgoedwetenschappen. iperionhs.eu/doctoral-summer-school

20 september • Online

COASP 2021
Conferentie over Open Access Scholarly Publishing. oaspa.org/conference

Kijk voor meer interessante events op edata.nl

Schoonhoven en Scherpenzeel (het CBS) aan het woord:

‘Microdata zijn de toekomst van ons onderzoek’

Ruurd Schoonhoven, Senior Relatiebeheerder Microdata Services bij het CBS, ging recent met pensioen. Annette Scherpenzeel begon in september 2020 bij het CBS als hoofd Microdata Services. E-data sprak hen.

Suze Zijlstra

Beiden hebben de nodige ervaring in de wetenschap en ervaring met data-analyse. Schoonhoven promoveerde in de biomedische fysica in Nijmegen. Hij werkte de eerste helft van zijn carrière als fysicus in het gehooronderzoek bij het UMC Leiden en begon in 2004 bij het Centraal Bureau voor de Statistiek (CBS). Scherpenzeel promoveerde in methoden en statistiek aan de Universiteit van Amsterdam. Hierna werkte ze als survey methodoloog en projectcoördinator bij onder andere CentERdata en de Survey of Health, Ageing and Retirement in Europe (SHARE).

Hoe heeft de microdata-afdeling zich door de jaren ontwikkeld?

Schoonhoven: “Eerst was het gebruik van CBS microdata in sociaalwetenschappelijk onderzoek nog vrij kleinschalig. De mogelijkheid bestond al in de jaren negentig, maar toen moest dergelijk onderzoek nog binnen het CBS zelf worden uitgevoerd. In 2006 is de Remote Access omgeving ontwikkeld, zodat onderzoekers vanuit een beveiligde ruimte in hun eigen instelling hun onderzoek konden doen.”

Schoonhoven vervolgt: “Er heeft sindsdien een flinke professionaliseringsslag plaatsgevonden. Waar het CBS vroeger vooral via enquêtes werkte, zijn sinds omstreeks 2000 registerbestanden steeds belangrijker geworden, met als kern voor persoonsstatistiek de Gemeentelijke Basisadministratie. Registerbestanden zijn integrale datasets met gegevens over de hele Nederlandse populatie. Dit maakt het aantrekkelijker om de data voor onderzoek te gebruiken. Bovendien kunnen wetenschappers sinds medio 2017 vanaf hun eigen computer de Remote Access omgeving betreden, uiteraard met een streng beveiligd inlogstelsel. Hierdoor is onderzoek met microdata nog toegankelijker geworden. Dit zien we terug in de hoeveelheid gebruikers: dit aantal is verdubbeld in de afgelopen vijf jaar.” Naarmate de hoeveelheid data toenam, ontstond de wens om meer rekenkracht te hebben, met dezelfde vertrouwelijkheid en veiligheid als de Remote Access omgeving. Schoonhoven: “Het CBS heeft toen samen met SURF de ODISSEI Secure Supercomputer ontwikkeld. Hierdoor kunnen onderzoekers gebruik maken van CBS microdata met de rekenkracht van de nationale supercomputer Cartesius. Deze samenwerking voelt heel vanzelfsprekend. Het CBS interesseerde zich al langer voor een samenwerking met universi-



Annette Scherpenzeel (links) is sinds september 2020 hoofd Microdata Services bij het CBS. Onlangs is Ruurd Schoonhoven, Senior Relatiebeheerder Microdata Services bij het CBS, met pensioen gegaan. Credits: Bart van Vliet

INTERVIEW

‘De toekomst van microdata is een combinatie met andere datavormen’

teiten, en ODISSEI is de nationale data-infrastructuur voor de sociale wetenschappen.”

Wat zijn de toekomstplannen van de microdata-afdeling?

Scherpenzeel: “Ik wilde graag bij het CBS werken vanuit wetenschappelijke interesse. Vroeger speelden enquêtes een grote rol in sociaalwetenschappelijk onderzoek, maar steeds minder respondenten willen daaraan meedoen. Ik zie dus een toekomst van groeiend gebruik van microdata gecombineerd met andere vormen van data. Deze baan sprak mij dus aan. En ik blijf het leuk vinden om een praktische functie uit te oefenen, dichtbij data en dichtbij gebruikers.”

Scherpenzeel: “Een uitdaging voor de microdata-afdeling zie ik op technisch gebied. Niet alle projecten hoeven meteen een supercomputer te gebruiken. Maar er zitten ook grenzen aan de servers van de normale Remote Access omgeving. Het is wellicht nodig om bijvoorbeeld servers met meer geheugen te gaan gebruiken. Dit is een nabije toekomstproject. Een tweede ontwikkeling is de groei in gebruikersaantallen, en in diversiteit van projecten. Wie laten we toe, hoe houden we

het veilig en beheersbaar? Onlangs is een externe evaluatie geweest. Hieruit bleek dat het gebruik van CBS microdata heel goed beveiligd is, maar natuurlijk zijn er altijd punten waarop we nog beter op verdere groei in de toekomst voorbereid kunnen zijn.

Een ander doel is internationalisering. Nederland is een koploper in het toegang geven tot registerdata. Daar wil ik graag een voorbeeldrol van maken, zodat we in de toekomst naar een geharmoniseerd data-model toe kunnen. De metadata moeten via een portal goed vindbaar zijn, zodat het voor onderzoekers in een internationale context makkelijker is de juiste data te vinden en deze te hergebruiken.”

Schoonhoven: “Er is veel aandacht voor de

len maken, zoals bij gezondheidswetenschappen. Ik zie hier zeker ook een rol voor ODISSEI, waarmee we natuurlijk al informatie-evenementen houden. Maar ik zou dat ook vanuit het CBS verder willen ontwikkelen. We kunnen onderzoekers meer gaan opzoeken met roadshows. Ook wil ik in de toekomst het onderzoek dat gedaan wordt met microdata meer naar buiten brengen. Bijvoorbeeld door resultaten van dat onderzoek op onze website te plaatsen en in onze nieuwsbrief voor gebruikers te delen.”

En wat zijn jullie eigen plannen?

Schoonhoven: “Ik heb geen specifieke plannen met mijn pensioen, maar ik weet wel waar ik meer tijd voor wil maken. Zo wil ik in ieder geval meer piano spelen. Ik verheug me daarnaast op een live Community Conference van ODISSEI in het najaar.” Scherpenzeel sluit af: “Ik werk natuurlijk nog niet zo lang bij het CBS, ik hoop dat wel nog minstens tien jaar te mogen doen. Ik heb echt ambitie om de mogelijkheden van microdata nog meer zichtbaar te maken, samen met het team.”

cbs.nl

‘Samen de mogelijkheden van microdata zichtbaar maken’

ODISSEI Secure Supercomputer. Minder spectaculair, maar minstens zo belangrijk, is de financiële ondersteuning die ODISSEI biedt voor microdata-projecten. De groep wetenschappers die van de ODISSEI microdata-korting of beurzen gebruikmaakt, is groot. Hierdoor wordt onderzoek met microdata een stuk toegankelijker. Hiernaast hoop ik ook dat het CBS in de toekomst de capaciteit van de computers kan uitbreiden op een gebruiksvriendelijke en flexibele manier.”

Scherpenzeel: “Bovendien is outreach belangrijk. Bij sociale en economische wetenschappers zijn de microdata ondertussen goed bekend. Ik zou de microdata wel in andere domeinen van wetenschap meer bekend wil-

Meer informatie

Het Centraal Bureau voor de Statistiek (CBS) verzamelt data op persoons- en bedrijfsniveau. Onder strenge voorwaarden kunnen wetenschappers gepseudonimiseerde microdata voor onderzoek gebruiken. Dit kon al enige tijd in de beveiligde Remote Access omgeving van het CBS. Tegenwoordig kunnen projecten die grotere rekenkracht nodig hebben, gebruikmaken van de ODISSEI Secure Supercomputer. Voor meer informatie kan contact worden opgenomen met CBS Microdata Services, via microdata@cbs.nl.

Statistieken maken vergelijkend onderzoek mogelijk

Goudmijn aan data gedolven uit Russische staatsarchieven

In tien jaar tijd verzamelde sociaal historicus Gijs Kessler van het Internationaal Instituut voor Sociale Geschiedenis (IISG) drie eeuwen aan data uit Russische archieven. Ze zijn nu beschikbaar via ristat.org. Mathilde Jansen

Het lijkt op het eerste gezicht mischien gek dat een Nederlands onderzoeksinstituut data verzamelt in Rusland, maar dat is het niet. Voor internationaal vergelijkend onderzoek – de kernactiviteit van het IISG – is Rusland een ontzettend interessante case, legt Kessler uit. Niet alleen omdat het zo’n groot land is. Kessler: “Hoewel Rusland niet tot de meest ontwikkelde landen ter wereld behoorde, zijn de statistieken kwalitatief vergelijkbaar met die uit West-Europa.”

Onderzoeksteam

Als Ruslandexpert – Kessler was 15 jaar gestationeerd in Moskou – werd hij geregeld door collega’s gevraagd om Russische data. Maar



Ristat.org biedt drie eeuwen aan data uit Russische archieven. Credits: Gijs Kessler, IISG

die waren bijna niet voorhanden. Daarom startte de onderzoeker een samenwerking met hoogleraar Andrei Markevich van the New Economic School (NES) in Moskou. Samen stuurden zij een onderzoeksteam aan dat sociaaleconomische data verzamelde in de staats-

archieven van Moskou en Sint-Petersburg. Historisch gezien heeft Rusland ontzettend veel statistieken. “Het was altijd een bureaucratisch land”, legt Kessler uit. “Het was in het belang van de staat om zoveel mogelijk gegevens van de bevolking te verzamelen.”

Maar in tegenstelling tot in westerse landen is er bijna geen wetenschappelijk onderzoek mee gedaan. Kessler: “Tot in de twintigste eeuw was de hele geschiedwetenschap in Rusland sterk geïdeologiseerd door het communisme. Daarbij werden heel weinig statistieken naar buiten

gebracht, omdat die het echte plaatje lieten zien. Het werd wel allemaal verzameld en opgeslagen in de archieven.”

Regionale diversiteit

Sinds 1991 zijn de archieven toegankelijk voor wetenschappers. Historici uit binnen- en buitenland zijn sindsdien bezig geweest om die archieven in kaart te brengen. Kessler en zijn collega’s verzamelden demografische data op basis van volkstellingen uit 5 steekjaren: 1795, 1858, 1897, 1959 en 2002. Daarbij zochten ze gegevens over industrie, landbouw, diensten, grondbezit, kapitaal en arbeid. Wat deze data ook uniek maakt, is dat ze verzameld zijn voor afzonderlijke regio’s. Dat maakt de data interessant voor historici, maar ook voor een breder publiek, met name in Rusland. “Er zijn zo ontzettend veel mythes over hoe Rusland zich heeft ontwikkeld. We willen de mensen ook instrumenten geven om zelf hun mening te vormen over hun eigen geschiedenis.”

ristat.org

Onder de paraplu van ICAI

Onderzoekslab Cultural AI van start

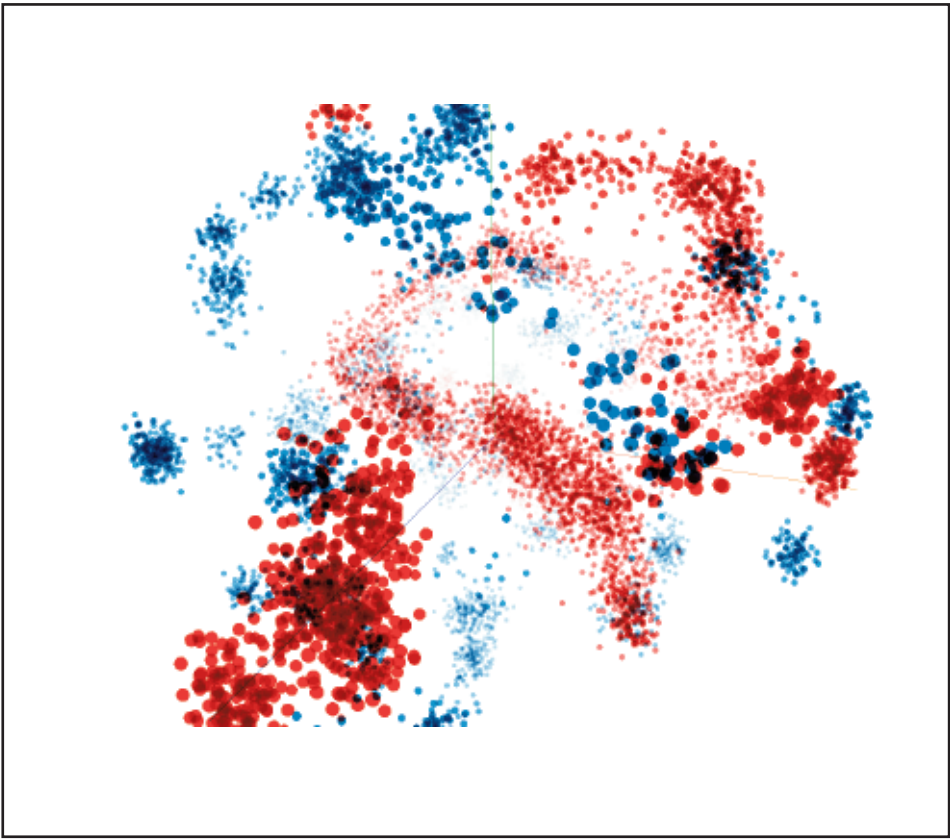
Onderzoekers en erfgoedprofessionals werken samen aan de ontwikkeling van AI-systemen met cultuur en menselijk gedrag als vertrekpunt. Maarten Heerlien

In de zoektocht van erfgoedinstellingen naar mogelijkheden om hun collecties en kennis toegankelijk en bruikbaar te maken voor een zo groot en divers mogelijk publiek, wordt steeds vaker naar kunstmatige intelligentie gekeken. Hoewel AI-toepassingen veelbelovend zijn en vaak interessante resultaten opleveren, heeft deze ontwikkeling een schaduwzijde. Cultuurhistorische collecties zijn doorgaans samengesteld en beschreven vanuit het enkelvoudige perspectief van de verzamelaar. De afgelopen jaren is aangetoond dat dergelijke eenzijdige trainingsdata bij AI’s kan leiden tot bevestiging en versterking van vooroordelen; resultaten die moeilijk te verenigen zijn met de gewenste meerstemmigheid in collecties. Tegen deze achtergrond startten acht onderzoeks- en erfgoedinstellingen Cultural AI. Dit nieuwe onderzoekslab, gekoppeld aan het Innovation Center for Artificial Intelligence (ICAI), neemt culturele inzichten en mense-

lijke uitingen mee in de ontwikkeling van beter gebalanceerde AI-systemen en algoritmes. Martijn Kleppe, Hoofd Onderzoek bij de Koninklijke Bibliotheek (KB), is één van de initiatiefnemers: “Cultural AI gaat om het begrijpen van de menselijke cultuur, maar ook om het benutten van diezelfde cultuur om AI-technologie te analyseren en te verbeteren. Inzichten uit de geestes- en computerwetenschappen komen hierin samen en versterken elkaar.”

Doelgerichte projecten

De kernpartners, KB, Beeld & Geluid, Rijksmuseum, KNAW Humanities Cluster, CWI, TNO, VU en UvA, werken in wisselende samenstelling al langer samen op de thema’s waarop Cultural AI de focus legt. Kleppe: “Door de samenwerking te structureren en voortdurend te evalueren, kunnen we doelgerichte projecten ontwikkelen. De partners zijn écht complementair. De erfgoedpartijen kunnen de meest uiteenlopende collecties inbrengen, van vroegmoderne kunst en publicaties tot born-digital materiaal zoals websites. Tegelijkertijd zijn onze onderzoeksteams multidisciplinair en mede dankzij de inbreng van TNO praktijkgericht.” Organisatorisch is Cultural AI ingericht op groei en biedt het lab verschillende mogelijke niveaus van betrokkenheid voor geïnteres-

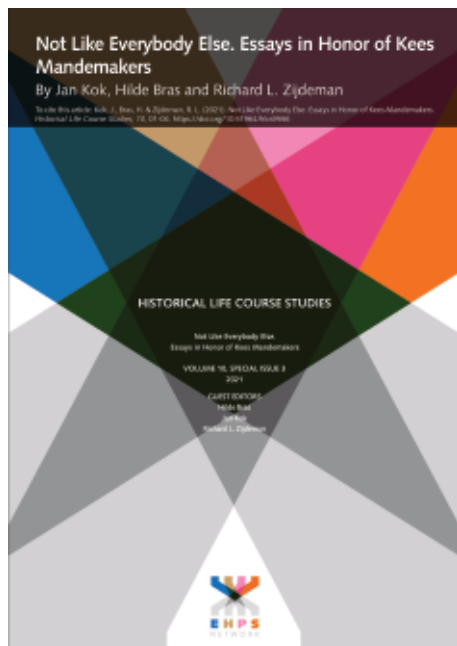


Eén van de projecten onder de vlag van Cultural AI is SABIO. Onderzoekers van het KNAW HUC en Nationaal Museum voor Wereldculturen bestuderen welke vooroordelen impliciet aanwezig zijn in de metadata van de collecties van het Nationaal Museum voor Wereldculturen en doen suggesties voor nieuwe mogelijke trefwoorden. SABIO illustreert de kern van Cultural AI: samen collecties meerstemmiger maken én op nieuwe manieren ontsluiten. De afbeelding is een projectie van de vectorruimte van termen uit de collectie van het Nationaal Museum voor Wereldculturen.

Credits: Valentin Vogelmann/SABIO project

seerde partijen. Naast de kernpartners, die actief bijdragen aan de onderzoeksprojecten en deze mede financieren, kunnen partijen zich committeren als associate partner. Deze partners delen mee in de onderzoeksresultaten in ruil voor projectondersteuning. Daarnaast kent Cultural AI affiliate partners, organisaties met

interesse in het lab die nog niet actief zijn in onderzoeksprojecten. Partijen die onderzoekers of projecten willen inbrengen, kunnen terecht bij de science directors van Cultural AI: Marieke van Erp, Laura Hollink en Jacco van Ossenbruggen. cultural-ai.nl



Plan S

Plan S is gelanceerd door wetenschapsfinanciers om de transitie naar open access te versnellen. De coalitiepartners eisen dat de publicaties die voortkomen uit door hen gefinancierd onderzoek, open access beschikbaar worden gesteld, zonder embargo.

Diamond Open Access

Met Diamond Open Access worden open access-publicaties van artikelen bedoeld waar geen kosten voor de auteur (article processing charges) aan zijn verbonden.

Via openaccess.nl zijn onafhankelijke, wetenschappelijke tijdschriften vrij beschikbaar, zoals: TSEG - The Low Countries Journal of Social and Economic History, Historical Life Course Studies en The Rijksmuseum Bulletin. Credits: openjournal.nl

Nieuw platform voor duurzame, vrij toegankelijke publicaties

Openjournals.nl faciliteert Diamond Open Access

Begin dit jaar werd openjournals.nl gelanceerd, een online platform voor duurzame open access-publicatie van onafhankelijke, wetenschappelijke tijdschriften. Maarten Heerlien

Het aandeel wetenschappelijke publicaties die vrij toegankelijk zijn neemt jaarlijks toe, mede gestimuleerd door het in 2018 gepresenteerde en begin 2021 bekrachtigde Plan S (zie kader). Maar de overstap van een betaalmodel naar open access is niet voor elk journal goed haalbaar. Publicaties die afhankelijk zijn van commerciële uitgevers worden vaak geconfronteerd met hoge article processing charges

(APC's), kosten die verhaald worden op de auteurs of het onderzoeksproject. Om deze journals binnen Nederland een duurzaam alternatief te bieden, is openjournals.nl ontwikkeld. "De behoefte aan een dergelijk platform binnen Nederland bestaat al langere tijd," aldus Sasha Goldstein-Sabbah, projectleider van openjournals.nl.

Hulp bij overstap naar OA

"Zes jaar geleden startte Jeroen Sondervan, projectleider Open Access aan de Universiteit Utrecht, een incubator om wetenschappelijke tijdschriften te helpen naar open access over te stappen. Dat project, Uopen Access, werd een succes maar de beschikbare middelen belemmerden een duurzame groei." Openjournals.nl, een samenwerking tussen KNAW, NWO en stichting OPuS (oapus.nl) met Sondervan als

een van de kwartiermakers, richt zicht specifiek op dat aspect.

Het non-APC open access model moet de beoogde duurzame groei van openjournals.nl mede mogelijk maken. Goldstein-Sabbah: "Niet de auteur of lezer betaalt, maar de partij die het tijdschrift publiceert. Doordat we geen winstoogmerk hebben, blijven deze jaarlijkse servicekosten beperkt. In ruil daarvoor krijgen tijdschriftredacties toegang tot het platform met technische ondersteuning, wordt elk artikel voorzien van een DOI, wordt het tijdschrift opgenomen in het Directory of Open Access Journals (doaj.org) en vindbaar via bijvoorbeeld Google Scholar. Met als belangrijkste doel dat kwaliteitsjournals op openjournals.nl hun behoeften vinden."

Als nationaal platform voor Diamond Open Access (zie kader) volgt openjournals.nl ini-

tiatieven uit Finland en Denemarken, die draaien op dezelfde open source software, het door het Public Knowledge Project ontwikkelde Open Journals Systems (OJS).

Drie jaar seed funding

Inmiddels zijn er acht tijdschriften beschikbaar via openjournals.nl, waaronder de titels die eerder via Uopen Access werden gepubliceerd. Meest recente aanwinst is het Rijksmuseum Bulletin, het eerste museale wetenschappelijke tijdschrift op het platform.

Openjournals.nl wordt gehost en technisch ondersteund door de afdeling Digitale Infrastructuur van het KNAW Humanities Cluster en ontvangt drie jaar lang seed funding van NWO. Daarna verwacht het platform financieel op eigen benen te kunnen staan.

openjournals.nl

Veilig, lokaal en verantwoord

Rondreizend algoritme analyseert data

Onderzoekers in Maastricht ontwikkelen onder leiding van Michel Dumontier een algoritme om gevoelige data te analyseren zonder dat de onderzoeker zelf de data te zien krijgt. Suze Zijlstra

Sociale wetenschappers staan voor de uitdaging om onderzoek te doen met grote datasets die gevoelige informatie bevatten. De computationele mogelijkheden om dergelijke datasets te analyseren, nemen toe. Daarom groeit de vraag naar innovatieve manieren om analyses veilig uit te voeren. Michel Dumontier, universiteitshoogleraar Data Science aan de Universiteit Maastricht, doet met zijn team onderzoek om aan die vraag te voldoen. Het project is deel van de ODISSEI Roadmap, gefinancierd door NWO.

Het team ontwikkelt Distributed Analytics Techniques voor de sociale wetenschappen. Het doel is om een algoritme te programmeren dat op verschillende data-locaties gevoelige data analyseert, zonder dat de onderzoeker zelf deze data inzielt.

Analyse op plek data

Dit uitgangspunt is gebaseerd op de Personal Health Train (PHT) uit de gezondheidswetenschap. Onderzoekers moesten data vanuit bijvoorbeeld ziekenhuizen op een centrale plaats verzamelen. De PHT gaat juist langs de ziekenhuizen om daar de analyse uit te voeren. Alleen de resultaten neemt het algoritme mee naar het volgende data-station. Op deze manier blijft het data-station (in dit voorbeeld het ziekenhuis) controle houden over de data.



Credits: Markus Spiske via Unsplash

Het Distributed Analytics Techniques project heeft in samenwerking met het Nationaal Regieorgaan Onderwijsresearch (NRO) een vergelijkbaar doel, vanuit de wens om onderzoek te doen aan de hand van leerlingengegevens op scholen. Het wil de data bij de scholen laten, waar het algoritme de analyse uitvoert. Ook dan krijgt de onderzoeker alleen de resultaten. Michel

Dumontier licht toe: "Het is ons doel om de gezondheid en het welzijn van leerlingen en docenten te onderzoeken door middel van een veilige, betrouwbare, en privacy-beschermende data-infrastructuur. ODISSEI biedt ons de ideale samenwerkingsomgeving om op een verantwoorde manier een nieuwe benadering van sociaalwetenschappelijk onderzoek met big data te

onderzoeken."

Het project heeft verschillende uitdagingen. Natuurlijk moet het garanderen dat het algoritme de analyse op een wetenschappelijk solide manier uitvoert. Het is daarbij belangrijk een goed metadata-systeem te ontwikkelen. Juridisch gezien is het project vanzelfsprekend gebonden aan de Europese privacywetgeving. Maar wie geeft toestemming om bepaalde data op deze nieuwe manier te gebruiken? Tot slot staat transparantie centraal. Het project moet vertrouwen inboezemen dat er hierdoor veilig met gevoelige data wordt omgesprongen. Uiteindelijk wil het team software ontwikkelen die ook voor andere projecten kan worden ingezet.

odissei-data.nl/en/work-streams/en-laboratory

Een netwerk voor trainers in de SSH

SSHOC stimuleert internationale uitwisseling van training en services

Het Europees gefinancierd project SSHOC faciliteert trainers in de sociale- en geesteswetenschappen met een community voor het uitwisselen van kennis en ervaring. Ellen Leenarts

SSHOC staat voor Social Sciences and Humanities Open Cloud en zoomt in op de sociale- en geesteswetenschappen en het stimuleren van samenwerking hierbinnen. Het sorteert voor op een interdisciplinaire samenwerking binnen de European Open Science Cloud. Verschillende Europese onderzoeksinfrastructuren en andere stakeholders zoals DARIAH, CLARIN, LIBER, CESSDA en DANS zijn partners in het project. Het project ondersteunt onder andere trainers op het gebied van Open Science en Research Data Management. Met workshops, train-the-trainer bootcamps en webinars wordt kennis gedeeld die in meerdere disciplines bruikbaar is. Bij inmiddels 150 vakgenoten staat het maandelijks online training



Credits: Chris Montgomery on Unsplash

Online trainingen

Bij het maandelijks online overleg delen professionals hun ervaringen over bijvoorbeeld beschikbare tools voor (interactieve) online trainingen en herbruikbare trainingsmaterialen. Geïnteresseerde trainers kunnen zich aansluiten bij het netwerk. Aanmelden kan via sshopencloud.eu/ssh-training-community.

community overleg in de agenda. De leden bepalen de onderwerpen en nodigen eventueel externe sprekers uit, bij interesse zijn leden bij het overleg en delen zij hun ervaringen. Onderwerpen die de revue passeren, zijn bijvoorbeeld: welke

tools zijn beschikbaar voor het geven van een (interactieve) online training welke herbruikbare trainingsmaterialen zijn er op het gebied van de bekostiging van research data management of het gebruik van gevoelige persoonsgegevens?

Online toolkit

DANS is binnen het project verantwoordelijk voor het vormgeven van het training netwerk, het organiseren van train-the-trainer bootcamps en het ontwikkelen van een service voor het vinden van herbruikbare trainingsmaterialen en tools. Deze ‘Training Discovery Toolkit’ is online beschikbaar via de website van het project. Ontwikkelt een partner een training, tool of ander handig hulpmiddel, dan kan dit worden opgenomen in de toolkit. Een voorbeeld is de DANS Data Game, geïnteresseerden kunnen het kwartet als pdf downloaden, per post toegestuurd krijgen en er is een online variant. Onlangs is dit spel aan de toolkit toegevoegd, waardoor nog meer mensen spelenderwijs een

indruk kunnen krijgen van het Europese onderzoeksdata-landschap. SSHOC loopt tot maart 2022. Met extra netwerkbijeenkomsten, en een gids van trainers in de sociale en -geesteswetenschappen, staan er nog een aantal mooie deliverables op de rol.

sshopencloud.eu/ssh-training-community



Data Game

Een van de beschikbare tools is de DANS Data Game. Dit spel is door DANS ontwikkeld, door het vindbaar te maken via de ‘Training Discovery Toolkit’ kunnen nog meer betrokkenen spelenderwijs een indruk krijgen van het Europese onderzoeksdata-landschap.

COLUMN

Het datameisje

In november 2019 won Max Verstappen de grand prix van Brazilië. Na de race ging Hannah Schmitz met Verstappen het podium op om de champagne te ontkurken. ‘Het datameisje’, aldus de Nederlandse TV-commentator wat neerbuigend. Wel een meisje met een masterdiploma van de prestigieuze Universiteit van Oxford op zak en 10 jaar ervaring als racestrateeg. Belangrijker dan dat: ze nam in Interlagos binnen een paar seconden een cruciale beslissing op basis van een enorme hoeveelheid data die beslissend bleek voor de winst van Verstappen. Voor de liefhebbers: iets met een zogeheten undercut door een bandenwissel en een ingewikkelde rekensom over de tijdswinst die een set nieuwe banden zou moeten brengen ten opzichte van concurrent Lewis Hamilton. Een gewaagde maar door data onderbouwde beslissing. Dit voorbeeld is belangrijk voor iedereen die serieus met data-analyse aan de slag wil, omdat het voorval inspireert over waarom we moeten investeren in

data(kwaliteit). Veel organisaties kampen met versnipperde en complexe IT-landschappen en daardoor met tekortschietende data-kwaliteit. In de onderzoekswereld is datakwaliteit ook bepaald niet vanzelfsprekend en onderzoekers streven naar beter Research Data Management om data beter te kunnen (her)gebruiken. Er is vaak geen eenheid in hoe data in systemen wordt geregistreerd, bewerkt en geaggregeerd. Ook bestaan er nog veel workarounds en zijn er vele aanbieders van overlappende ondersteunende producten en diensten.

De meest geweldige algoritmes zullen in zo’n omgeving leiden tot teleurstellende resultaten. Daarom investeren veel organisaties in data-kwaliteitsprojecten. Maar de teams die daaraan werken, weten maar zelden te komen tot structurele verbetering. De huidige werkwijze doet denken aan mensen die willen afvalen. Een paar maanden op dieet in combinatie met een abonnement op de sportschool levert aanspre-

kende resultaten op. Maar wie de levensstijl niet aanpast, weet diep van binnen al dat het effect tijdelijk zal zijn. Alleen als je handelt vanuit de overtuiging dat een andere levensstijl je een beter en fitter leven oplevert, zul je ook op termijn succesvol zijn.

In de Formule 1 is iedereen van nut en noodzaak doordrongen en is er binnen het hele team een honger naar goede en betrouwbare data. De teams zijn apetrots op hoe ze met data hun prestaties continu verbeteren. Het hele team beseft ook hoe belangrijk die data zijn in de concurrentie en iedereen



Foto Mirella Boot Fotografie

weet dan ook hoe belangrijk het is om de kwaliteit van die data op niveau te hebben. Juist omdat die data hen op cruciale momenten de expliciete inzichten levert om het verschil te maken. Het is voor hen ‘a way of life’. Dat zou in alle maatschappelijke sectoren zo moeten zijn. De concurrentie wordt immers uitgedaagd op het verwerken van een enorme rijkdom aan informatie tot de beste beslissingen. En die boodschap moet actief worden uitdragen, net zoals de teambaas van Red Bull Racing dat doet. Die stuurt het ‘datameisje’ het podium op om daarmee een niet mis te verstane boodschap af te geven: je wint races met data.

Nart Wielaard (www.nart.nl) is zelfstandig schrijver-adviseur, met name actief op het thema digitale transformatie. Hij schreef onder meer Wij Zijn Big Data, Vertrouwen in de slimme samenleving en The CFO in pole position. Hij geeft de column door aan Egge van der Poel.

GELEZEN

Humanities Data Analysis: Case Studies with Python
Folger Karsdorp, Mike Kestemont en Allen Riddell, Princeton University Press 2021

Lamyk Bekius
Humanities Data Analysis: Case Studies with Python van Karsdorp, Kestemont en Riddell biedt de uitkomst voor wie wil investeren om beter bekend te raken met data-analyse met Python. Het boek maakt studenten en onderzoekers wegwijs met Python voor de analyse van geesteswetenschappelijk materiaal. Enige voorkennis van Python wordt verondersteld, maar het boek frist de essentiële kennis daarvan wel op in de introductie. In dit hoofdstuk wordt de lezer tevens enthousiast gemaakt door een case study met data van Amerikaanse kookboeken, waarin de mogelijkheden voor de verkenning van data uiteen worden gezet. De introductie is samen met hoofdstukken over inlezen, parsing en manipuleren van verschillende bestandsformaten, het vector space model en analyse van tabelgegevens onderdeel van het eerste deel ‘Data Analysis Essentials’, waarin ook Pandas en NumPy worden geïntroduceerd. Daarop verder bouwend gaat het tweede deel ‘Advanced Data Analysis’ aan de hand van casussen in op respectievelijk beschrijvende statistiek, kansrekening, geospatiale analyse, stylometrie (specifiek auteursherkenning) en topic modeling. Het boek geeft niet alleen uitleg over code en statistiek, maar ook achtergrondinformatie en interpretatie van de data, en good practices in data-analyse. Door de intensieve onderdompeling in data-analyse met Python is het boek een uitstekend uitgangspunt om aan de slag te gaan met eigen data.
press.princeton.edu

CentERdata behaalt ISO 27001

Als er ergens veel met data en informatie gewerkt wordt, is dat wel bij CentERdata. Het willen behalen van de ISO 27001-certificering was dan ook een logische ambitie. ISO 27001 is een internationale standaard voor informatiebeveiliging. Binnen deze norm zijn de vereisten voor een informatiebeveiligingssysteem gedefinieerd en wordt het proces rondom informatiebeveiliging getoetst. Linda Uijtdewilligen (ISO coördinator) geeft namens CentERdata aan, trots te zijn. Uijtdewilligen: “Het helpt ons de informatiebeveiligingsrisico's te verkleinen en incidenten te voorkomen. Het geeft opdrachtgevers en partners de zekerheid dat CentERdata de informatiebeveiliging beheerst, optimaal heeft geïmplementeerd in haar beleid en processen en continu werkt aan verbeteringen.” (EdC)
centerdata.nl