

Tonen de sociale en geesteswetenschappen hun kracht of hun zwakte?

Effecten COVID-19 overal voelbaar

De coronacrisis heeft grote maatschappelijke gevolgen – ook voor de sociale en geesteswetenschappen. E-data & Research maakte een rondgang onder haar stakeholders.

Steven Claeysens en

Erica Renckens



Credits: image 241222 John Darkow

De bestrijding van een pandemie is in de eerste plaats een (bio)medische zaak; er staan tenslotte mensenlevens op het spel. Daarnaast is het ook een wijdvertakt experiment op ongeziene schaal, met een nog ongekende reikwijdte voor mens en maatschappij. Een lockdown, intelligent of minder intelligent, heeft ingrijpende gevolgen op sociaal, economisch, politiek, ecologisch, technologisch en mentaal vlak: *gefundenes Fressen* voor de sociale en geesteswetenschappen.

Digitaal onderzoek

“Onderzoekers zijn over elkaar gaan buitelen om COVID-relevant onderzoek te doen”, signaleert Pearl Dykstra (ODISSEI). Ze wijst onder meer op ssh-covid.nl, een website met experts uit de sociale en geesteswetenschappen, en op onderzoek uitgevoerd via Lifelines en het LISS-panel van CentERdata. Marcel Das (CentERdata) valt haar bij: “Tal van onderzoekers zijn aan de slag gegaan met maatschappelijke vraagstukken zoals: Welke effecten heeft de uitbraak op de arbeidsmarkt? Wat is het effect van thuisonderwijs op de onderwijsresultaten en sociale ongelijkheid? Hoe beïnvloedt de crisis de fysieke en mentale gezondheid?”

Ze signaleren beiden ook een grotere bereidheid om onderzoeksdata te delen. Andrea Scharnhorst (DANS) onderstreept dit belang en wijst op de vele open science grassroots-initiatieven: “Duizend vrijwilligers zorgden voor 50.000 vrij toegankelijke wetenschappelijke artikelen op coronawhy.org. Ze gebruiken nu technieken uit de computationele taalkunde, machine learning en andere AI-technieken om met deze data de medische wetenschap verder te helpen.”

Digitaal erfgoed

Volgens Jan Willem van Wessel (KB) gaat het te ver om te stellen dat we de coronacrisis nodig hadden

“Met data kunnen we de medische wetenschap verder helpen”

voor de verdere digitalisering van erfgoed. Wel bieden talrijke nieuwe initiatieven online alternatieven om historisch bronnenmateriaal in te zien. Ook het Rijksmuseum liet zich daarbij gelden, door bijvoorbeeld de scan on request-service opnieuw op te starten en drie nieuwe digitale formats (een *virtual tour*, *#Rijksmuseumfromhome* en online Rijksmuseum voor het onderwijs) te lanceren. Saskia Scheltjens (Rijksmuseum): “We zijn erin geslaagd om interne en externe nieuwe doelgroepen aan te boren en bepaalde diensten versneld verder uit te breiden. Het publiek reageerde met name massaal wanneer het mogelijk was om in interactie te gaan met het digitale aanbod. Als het nog niet duidelijk was dat *digital engagement* cruciaal is in de verdere ontwikkeling van digitaal erfgoed, dan is dat nu wel overduidelijk geworden.”

Geen zegen

Toch zitten veel onderzoekers in de knel tijdens deze coronacrisis, door zorgtaken of simpelweg omdat ze hun bronnen niet kunnen inzien. Lex Heerma van Voss (CLARIAH): “Mensen die voor hun werk afhankelijk zijn van archieven en biblio-

theken lopen vertraging op en komen nu in de problemen. Dat zie je vooral bij promovendi en post-docs, zij werken op tijdelijke contracten.” Ook Antal van den Bosch (KNAW Humanities Cluster) ziet negatieve gevolgen: “De crisis schudt zaken op, maar is in de eerste plaats een vloek, geen zegen, laat daar geen misverstand over zijn. Men kan blij zijn met het bruuske ‘iedereen werkt thuis’-experiment en de relatief positieve ervaringen daarmee, maar tegelijk went men maar slecht aan het afstandelijke *video-conferencing* en het virtuele gevoel van ‘niet echt’ aan het werk te zijn. Collecties in onze instituten zijn weer beperkt beschikbaar en de onbereikbaarheid van die collecties is zwaar gevoeld. Werk is simpel-

“Nieuwe initiatieven bieden online alternatieven voor onderzoek”

“We signaleren een grotere bereidheid om data te delen”

weg niet hetzelfde, maar een soort slap aftreksel geworden.” Tijdens een crisis toont de mens zijn veerkracht, inventiviteit en de grenzen daarvan. Zo vergaat het dezer dagen ook de wetenschap en het erfgoed – die beide immers mensenwerk zijn. Welke innovaties zullen beklijven en welke zullen verzwinden, weten alleen de (on)heilsprofeten. Van den Bosch: “De enige manier waarop we ons nu kunnen onderscheiden, is in het kiezen, waar mogelijk, voor het investeren van tijd in het helpen van onze maatschappij, die toch behoorlijk in de penarie zit. Dat kan prima met het doen van onderzoek. Maar net als iedereen kunnen onderzoekers ook vrijwilligerswerk gaan doen. Zingeving heeft des te meer zin gekregen in deze tijd.”

edata.nl

Benieuwd naar concrete onderzoeken naar de effecten van de coronacrisis? Op pagina 4 en 5 staan voorbeelden hoe tools en data worden ingezet in onze gezamenlijke strijd tegen de coronacrisis.



E-DATA &
RESEARCH

Jaargang 15 | 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
Claes de Vreese over
belang infrastructuur



4 / 5
Voorbeelden van inzet
data en tools bij virus

6
ODISSEI faciliteert
digitaal onderzoek

7
Eén basisregistratie voor
historische personen

7
Resampling methode
schat tekstverlies in

8
Gastcolumnist Hagoort:
“Samen leren van data”

Wilt u dit blad ook gratis per post of per mail ontvangen? Of tijdelijk op een ander adres? Mail de redactie: edata@dans.knaw.nl.



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

GEHOORD & BIJGEWOOND

ICTeSSH:
open discussie over SSH
in het digitale tijdperk

Lucas van der Meer

SSH-onderzoekers wijzen op de mogelijkheden voor onderzoek op basis van ICT. De toekomst van de wetenschap draait om multidisciplinaire samenwerking en het toepassen van nieuwe technologieën. Met die gedachten vond van 29 juni tot en met 1 juli het event *ICT enhanced Social Sciences and Humanities* plaats. Belanghebbenden kwamen samen voor een open discussie over het veranderende onderzoeksecosysteem op SSH-gebied in het digitale tijdperk.

Wie een belanghebbende is, was breed gedefinieerd: toonaangevende SSH-onderzoekers, computerwetenschappers, informatici, uitgevers, bibliothecarissen, leveranciers van ICT-tools, SSH-besluitvormers. Ook het programma was breed ingestoken: van lezingen over open science (wanneer gaat het nou eindelijk eens doorbreken; waarom is het zo moeilijk om beloning voor negatieve resultaten te krijgen?); big data (kunnen computers zelfstandig betekenis aan gegevens geven?); marktplaatsen om onderzoekstools te delen (bijvoorbeeld die van SSHOC); en hoe de wetenschap bijdraagt aan de *Sustainable Development Goals* van de Verenigde Naties. Vanwege corona vond de conferentie niet in Amsterdam plaats, maar digitaal. Daardoor was er plaats voor alle 457 deelnemers en was deelname kosteloos.

testminis.uns.ac.rs/ictessh2020

SURF lightning talk:
vertrouwelijke data delen

Lucas van der Meer

Bij het delen van vertrouwelijke gegevens komt veel kijken, denk bijvoorbeeld aan: contracten, toegangscontrole, pseudonimisering en veilige computercontainers. Er zijn initiatieven die het delen vergemakkelijken. Tijdens een door SURF georganiseerde *lightning talk* op 23 juni presenteerde een



Tijdens het online event 'Voer voor Psychologen' was prof. dr. Andrea Evers (Universiteit Leiden) een van de gasten bij host dr. Ricarda Braukmann (DANS). Zij lichtte toe hoe datamanagement in een faculteit georganiseerd kan worden. Credits: DANS & LCRDM

aantal van deze initiatieven zichzelf.

Eén van de initiatieven is de CBS Remote Access omgeving, gepresenteerd door Ruurd Schoonhoven (CBS). In deze strikt afgesloten omgeving kunnen onderzoekers (onder voorwaarden) gepseudonimiseerde microdata analyseren. Indien de controle op outputresultaten succesvol is, geeft CBS deze resultaten vrij. Door deze controles heeft de data-aanbieder zekerheid dat er geen onbedoelde data naar buiten lekt, en kan de onderzoeker toch een analyse uitvoeren op vertrouwelijke data. Ananya Choudhury (Maastrou) presenteerde een andere aanpak: de *Personal Health Train*. Een onderzoeker ontwikkelt een algoritme (in deze analogie: een trein) die elke belanghebbende (*station*) bezoekt, de data op locatie analyseert en met de resultaten, maar zonder de data, naar het volgende station gaat. Leon Gommans (UvA/Air France KLM Group) besprak een prototype van de *Digital Data Marketplace*, waarin data scientists algoritmes kunnen indienen op datasets van meerdere eigenaren. Slimme contracten beheren het uitzetten van gefedereerde leerprocessen. Ten slotte presenteerde

Freek Dijkstra (SURF) de *Data Exchange*. Hiermee kan een onderzoeker een data-aanbieder vragen een analyse te doen op een dataset. Na goedkeuring wordt de analyse in een beveiligde omgeving uitgevoerd, zodat er geen data lekken. Na afloop kan de data-aanbieder het resultaat bekijken voordat besloten wordt of dit vrijgegeven wordt aan de onderzoeker.

surf.nl

Data-event voor psychologen smaakt naar meer

Margriet Miedema (LCRDM)

Henk Wals, directeur van DANS, is in zijn openingswoorden duidelijk over DANS en de rationale achter deze middag: samen de kwaliteit van onderzoek verbeteren door hergebruik van data te faciliteren. In zijn keynote haakte prof. dr. Jelte Wicherts hierop aan: de randvoorwaarden voor wetenschappelijke publicaties en van onderzoeksfinanciers stimuleren de groei in het aantal open datasets. Steeds meer psychologen adopteren de Open Science praktijk door het delen van hun onderzoeksplannen en data. In de afsluitende *keynote* lichtte prof. dr. Andrea Evers toe hoe datamanagement in een faculteit georganiseerd kan

zijn. Zij pleit voor verantwoorde onderzoekspraktijken, omdat het essentieel is om de impact en kwaliteit van onderzoek te garanderen. In Open Science is samenwerken het sleutelwoord, voor vertrouwen in elkaar en in elkaars onderzoek. De drie workshops tussen de keynotes in waren concreet: ODISSEI biedt bijvoorbeeld het LISS panel en toegang tot data van onder andere het CBS. De workshop over GDPR en AV-middelen liet zien hoe de GDPR onderzoek complex maakt, maar hulp van een *Data Protection Officer* en een Ethische Commissie een juist verloop kan faciliteren. De derde workshop richtte zich op onderzoekers die werken met fMRI data en gaf praktische tips voor het delen van data en open code. Een online bijeenkomst blijkt een goede manier, zelfs met kleine haperingen, om in korte tijd zinnige informatie en inspiratie op te doen. De video's, presentaties en andere interessante downloads staan op de website van DANS. Daartussen ook de presentatie van Jacob Jolij, die helaas vanwege ziekte zijn workshop niet kon verzorgen. Dit smaakt naar meer.

dans.knaw.nl/nl/actueel/agenda/voer-voor-psychologen

Samenwerking Universiteit van Nederland, Stichting Open Spraaktechnologie en Nederlandse Taalunie

ASR zorgt voor ondertiteling online colleges

Op de website universiteitvannederland.nl staan korte colleges van Nederlandse wetenschappers. In samenwerking met Stichting Open Spraaktechnologie voorziet de Nederlandse Taalunie vijftig van deze video-opnames via automatische spraakherkenning van ondertiteling.

Doel hiervan is de filmpjes beschikbaar te ma-

ken voor het onderwijs Nederlands als Vreemde Taal (NVT) en voor dove en slechthorende gebruikers van het Nederlands. Recent bleek uit een inventarisatie van de Nederlandse Taalunie dat er in het NVT-onderwijs veel behoefte is aan hoogwaardige Nederlandstalige filmpjes met ondertiteling.

Nadat de automatisch gegenereerde transcripties

handmatig zijn gecorrigeerd, worden deze gebruikt om het systeem te verbeteren en om synchroon te laten lopen met de spraak in het opgenomen college. De ondertitelde filmpjes komen in de toekomst waarschijnlijk ook via het Instituut voor Nederlandse Taal beschikbaar. (ER)

universiteitvannederland.nl

AGENDA

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

19 - 20 oktober • Online
EOSC Symposium 2020

Conferentie over de laatste stappen naar de eerste versie van een *European Open Science Cloud*. eosccsecretariat.eu/events/eosc-symposium-2020

20 - 24 oktober • Dublin
i-Society 2020

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

21 - 24 oktober • Beijing
iPRES

Jaarlijkse conferentie over digitale preservatie is verplaatst naar 2021 ipres2020.cn/dct/page/1

22 oktober - 1 november • online
ASIS&T 2020 Annual Meeting

Het thema dit jaar is *Information for a Sustainable World: Addressing Society's Grand Challenges*. assist.org/am20

27 oktober • online

Webinar Digitale Transformatie

Webinar georganiseerd door de interessegroep Digitale transformatie naar aanleiding van het boek *Digitale trends*. knvi.nl/agenda

4 - 6 november • Wenen

Conference on Cultural Heritage and New Technologies

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

5 november • wereldwijd

World Digital Preservation Day

Het thema dit jaar is *At-Risk Digital Materials*. dpconline.org/events/world-digital-preservation-day

9 - 12 november • online

RDA 16th Plenary Meeting

Het thema is *Knowledge Ecology*. rd-alliance.org/node/66206

18 november • online

RDNL Nederlandse Dataprijs

RDNL reikt iedere twee jaar de Nederlandse Dataprijs uit aan een onderzoeker of onderzoeksgroep die onderzoeksdata beschikbaar heeft gesteld. researchdata.nl/diensten/dataprijs

1 - 2 december • online

EDDI20

De 12e jaarlijkse Europese DDI-gebruikers conferentie. eddi20.sciencesconf.org

30 januari • Utrecht

CLIN30

Computational Linguistics in the Netherlands. clin30.sites.uu.nl

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

Hoofd-/eindredacteur: Heidi Berkhout. Redactie: Steven Claeyssens, Evi de Cock, Maarten Heerlien, Mathilde Jansen, Lucas van der Meer, Erica Renckens, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning. Aan dit nummer werkten mee: Valentijn Gilissen, Marjan Grootveld, Karin Hagoort, Margriet Miedema, Iris Muste, Marcel Ras, Steamwork Graphics. Opmaak: Colette Sloots, Haarlem. Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Oplage: 5200 papier, 5000 digitaal. ISSN: 1872-0374. We hebben getracht alle belanghebbenden met betrekking tot het gebruikte beeldmateriaal te benaderen.

Degenen die menen rechten te kunnen doen gelden, kunnen zich tot ons wenden. Toezending papieren en/of digitale versie is kosteloos aan relaties van de stakeholders en studenten in de alfa- en gamma-richtingen. E-data & Research liever (tijdelijk) op een ander adres ontvangen? Geef wijzigingen aan ons door via edata@dans.knaw.nl.

OVERNEMEN ARTIKELLEN

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.

Claes de Vreese, oud-voorzitter van het SSH-beraad:

“PDI-SSH investeert in digitale infrastructuur”

Claes de Vreese (UvA) is een van de initiatiefnemers van het Platform Digitale Infrastructuur SSH. Recent financierde het platform twaalf infrastructuur.

Lucas van der Meer

Claes (spreek uit als “Clees”; hij is van oorsprong Deens) de Vreese is hoogleraar Politieke Communicatie aan de Universiteit van Amsterdam, waar hij onderzoek doet naar hoe in traditionele en sociale media over politiek wordt gesproken en wat de invloed daarvan op publieke opinie en kiesgedrag is. Zijn afgelopen drie jaar stonden ook in het teken van het SSH-beraad, een vertegenwoordiging van KNAW, NWO en alle decanen in de sociale wetenschappen, economische wetenschappen en bedrijfskunde, rechtsgeleerdheid en geesteswetenschappen. Voor deze vier velden wordt de overkoepelende naam ‘Social Sciences and Humanities’ (SSH) gebruikt.

Eerste SSH-sectorplan

Als voorzitter van het SSH-beraad hield De Vreese zich onder andere bezig met de adviezen van de commissie Van Rijn en de zogenaamde sectorplannen. Claes: “Het ministerie van Onderwijs, Cultuur en Wetenschap stelde zo’n drie jaar geleden € 10 miljoen extra basisfinanciering beschikbaar voor de SSH. Het ministerie vroeg ons hiervoor een sectorplan op te stellen”. In een dergelijk plan stelt een sector zelf prioriteiten vast voor de komende periode en creëert zo een gezamenlijke strategie. Dit stimuleert samenwerking tussen verschillende universiteiten en zorgt voor breed gedragen keuzes. Het ministerie stelt vervolgens de middelen ter beschikking voor de uitvoering van het plan. Bij een succesvolle uitvoering dalen de middelen ook structureel in.

In de bèta en techniek werkt men al aan het derde sectorplan; voor de SSH was dit instrument nieuw. Claes: “Het SSH-beraad concludeerde al gauw dat het uitsmeren van de beschikbare middelen over het erg brede SSH-domein geen zoden aan de dijk zou zetten. We besloten om een gedetailleerd plan te schrijven voor een van de vier velden, rechtsgeleerdheid, en een globaler plan voor de sociale en geesteswetenschappen als geheel.” Met name het SSH-brede plan



Claes de Vreese: “Het is uniek dat een vertegenwoordiging van het veld zelf besloot welke faciliteiten te financieren” foto: Claes de Vreese

INTERVIEW

“Samenwerking leidt tot betere onderzoeksresultaten”

zorgde voor eenheid in een toch wat versnipperd veld, vindt Claes. Voor het plan koos het veld voor de richting digitalisering. Een deel van de beschikbare gelden wordt dientengevolge nu besteed aan het opzetten, opschalen en samenbrengen van digitale infrastructuur. Hiervoor is het Platform Digitale Infrastructuur SSH (PDI-SSH) opgericht, ondergebracht bij de Erasmus Universiteit Rotterdam.

Digitale infrastructuur

Claes: “Omdat PDI-SSH nieuw is, kozen we voor een brede insteek. Een ‘digitale infrastructuur’ kan gaan over de verzameling van nieuwe digitale data, over digitalisering van archieven en over structuren om bestaande onderzoeksgegevens met elkaar te verbinden.

In alle gevallen wordt een infrastructuur gebouwd, die onderzoekers kunnen gebruiken.”

De initiatiefnemers beogen een faciliterende rol voor het platform, waarbij ook ruimte is voor de continuïteit van faciliteiten die elders moeilijk aan financiering kunnen komen. Claes: “PDI-SSH is een aanvulling op bestaande infrastructuur als ODISSEI en CLARIAH, zij zijn ook bij de besluitvorming over de toekenning van de middelen betrokken.”

Twaalf voorstellen

De besluitvorming van de eerste *call for proposals* is recent afgerond, waarbij de besluitvorming geschiedde door het veld. Claes: “Per sector boog een vertegenwoordiging uit het veld zich over de voorstellen van de openbare call. We vonden deze unieke bottom-up benadering het best recht doen aan de diversiteit van het veld.”

Er konden twaalf voorstellen gehonoreerd worden, die uit alle vier de sectoren van het SSH-domein komen. Claes: “Er zijn prachtige voorstellen gehonoreerd. Julia Noordegraaf (UvA) bijvoorbeeld ontwikkelt een infrastructuur die SSH-onderzoekers in staat stelt om publieke discussies op sociale media te analyseren. In het project werken sociale en geesteswetenschappers samen en de infrastructuur sluit aan bij die van ODISSEI en CLARIAH. In een ander project inte-

greert Chantal Kemner (UU) datacollecties over de ontwikkeling van kinderen, waarbij de uitdaging zit in de integratie van metadata. Weer een ander project gaat over de inhuur van proefpersonen. Martin Tanis (VU) kijkt of dit via een digitale infrastructuur eenvoudiger, goedkoper en beter kan.”

Betere resultaten

Op de vraag waarom gekozen is om PDI-SSH zowel de sociale als geesteswetenschappen te laten omvatten, geeft Claes twee antwoorden: “De pragmatische reden is dat het platform dan gelijkloopt met de domeinen van de *European Research Council* en NWO. De idealistische reden is dat ik geloof dat samenwerking tussen die twee velden leidt tot betere onderzoeksresultaten. Voor bijvoorbeeld automatische analyse van sociale mediagegevens is kennis nodig uit zowel de communicatie- als taalwetenschap. We hebben nu ook de financiering om die samenwerking te stimuleren. Een mogelijk nadeel is echter de diversiteit tussen de verschillende sectoren binnen het domein, waardoor je de focus kunt verliezen.” Lachend: “Gelukkig betekent een samenwerking niet dat je nooit iets buiten die samenwerking mag doen.”

Faciliteiten voor elkaar

Een van de uitdaging voor PDI-SSH is om te zorgen dat de infrastructuur ook gebruikt gaan worden. Claes: “Een van de vereisten voor financiering was het opstellen van concrete *use cases* die de behoefte aantonen. Via het SSH-beraad, waarin alle decanen vertegenwoordigd zijn, gaan we het gebruik van de faciliteiten stimuleren. Ook hopen we op bijvoorbeeld een samenwerking met NWO, waarbij zij voor subsidies bonuspunten geven aan indieners die de infrastructuur gebruiken. Het is in ieders belang dat belastinggeld goed besteed wordt.” Claes sluit af: “In het verleden lag er te veel nadruk op individuele dataverzamelingen. Het is prachtig dat er nu een beweging gaande is om faciliteiten voor elkaar te ontwikkelen. In de wetenschap moet daar meer erkenning en waardering voor komen. Onderzoek is tegenwoordig echt een teamsport, de tijd is rijp om dat zichtbaar te maken”.

pdi-ssh.nl

Het platform is voornemens in 2021 een nieuwe *call for proposals* te openen. Meer informatie volgt via pdi-ssh.nl.

KORT

DANS kondigt nieuwe visie aan

Een totaal vernieuwd digitaal repository, met een *Vault* voor *long-term-preservation* en op verschillende wetenschappelijke communities toegesloten *Data Stations*; dat is iets wat gebruikers binnenkort van DANS kunnen verwachten. Maar de nieuwe visie van het grootste Nederlandse repository voor onderzoeksdata behelst meer. Als nationaal expertisecentrum voor onderzoeksdata zal DANS onderzoekers én dataprofessionals trainen en ondersteunen, het datalandschap monitoren en samenwerking bevorderen. DANS zal hiermee de komende jaren een krachtige bijdrage leveren aan de toename van *Findable, Accessible, Interoperable en Reusable* (FAIR) onderzoeksdata in Nederland. (HB) dans.knaw.nl

Nominaties Nederlandse Dataprijz 2020 bekend

Onlangs heeft een vakkundige jury zich over ruim 40 inzendingen gebogen voor de Nederlandse Dataprijz 2020, stuk voor stuk interessante datasets. Wie de winnaars zijn, wordt op 18 november bekend gemaakt. De Dataprijz wordt om het jaar uitgereikt aan een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor aanvullend of nieuw onderzoek. De winnaars van de Dataprijz krijgen € 5.000 om de dataset (nog) toegankelijker te maken en een sculptuur.

De Dataprijz is een initiatief van RDNL (4TU.ResearchData, DANS en SURFsara). Partners van deze editie zijn: DTL, Netwerk Digitaal Erfgoed, Nederlands eScience Center, Landelijk Coördinatiepunt Research Data Management (LCRDM), Open AIRE, en de Werkgroep Research Data van het samenwerkingsverband van universiteitsbibliotheken en de Koninklijke Bibliotheek. (HB) researchdata.nl

Online DANS Data Stories: GDPR

In een korte video vertelt Emilie Kraaikamp (DANS, adviseur juridische zaken) over het delen van persoonlijke gegevens uit wetenschappelijk onderzoek. Daarnaast bespreekt Emilie praktische zaken over het archiveren van onderzoeksgegevens bij DANS en het delen van deze gegevens. De video *Reusability and the GDPR: Archiving your data with DANS* staat op het YouTube-kanaal van DANS. (HB)

YouTube: [DANSDataArchiving](https://www.youtube.com/watch?v=...)

De coronacrisis heeft enorme gevolgen – ook voor sociale en geesteswetenschappers. Deze zeven artikelen illustreren de inzet van hun data en onderzoek in onze gezamenlijke strijd tegen de crisis. Kijk voor meer informatie over deze onderzoeken op edata.nl.

Europese Commissie vraagt hulp aan RDA

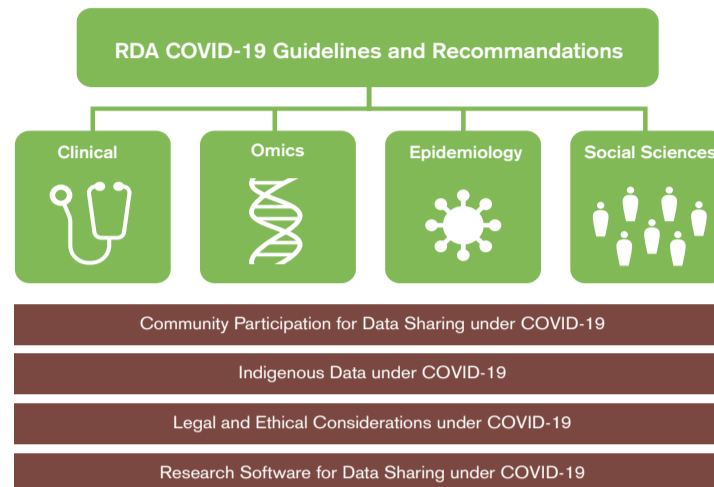
600 professionals delen kennis in COVID-19 werkgroep

De COVID-19 pandemie bewijst eens te meer dat het delen van onderzoeksgegevens noodzakelijk is. De Europese Commissie vroeg de Research Data Alliance (RDA) om hier aanbevelingen en richtlijnen voor op te stellen.

Marion Wittenberg



Eind maart ontving de RDA een rechtstreeks verzoek van de Europese Commissie om deze richtlijnen op te stellen. Ingrid Dillo, adjunct-directeur DANS, co-chair van de RDA Council en mede-voorzitter van de COVID-19 werkgroep: “Meer dan zeshonderd data-professionals en domeinexperts van



Deze infographic behandelt de hoofdlijn van de uitdagingen en aanbevelingen van de COVID-19 werkgroep. Credits: RDA

over de hele wereld meldden zich aan. Onder de paraplu van de RDA hebben ze een uitgebreide reeks gedetailleerde richtlijnen voor onderzoekers en datastewards opgesteld in combinatie met meer algemene aanbevelingen voor beleidsmakers en financiers van onderzoek.”

Eind juli publiceerde de RDA de definitieve versie voor vier COVID-19 gerelateerde onderzoeksgebieden: klinisch onderzoek, omics, epidemiologie en sociale wetenschappen. Ook wordt er in het document uitgebreid aandacht besteed aan een viertal overkoepelende thema's: juridische en ethische aspecten, onderzoeksoftware, gemeenschapsparticipatie en gegevens van oorspronkelijke bewoners. Daarnaast is er een zeer uitgebreide, ondersteunende Zotero library opgezet waarin alle onderliggende bronnen en referenties zijn opgenomen. Een infographic biedt een overzicht op hoofdlijnen van de uitdagingen en aanbevelingen.

www.rd-alliance.org/value-rda-covid-19-0

Hulp voor medische gemeenschap bij ontsluiten van informatie

CoronaWhy bestrijdt virus met AI

1.300 vrijwilligers sloten zich aan bij CoronaWhy, een initiatief om de medische gemeenschap te helpen bij het ontsluiten van informatie over COVID-19.

Marion Wittenberg



Eén van de vrijwilligers is Slava Tykhonov, senior informatiekundige bij DANS. Hij licht toe: “Met Kunstmatige Intelligentie (AI)

maken we met CoronaWhy honderdduizenden wetenschappelijke artikelen over de familie van coronavirussen, van decennia terug tot zeer recent, doorzoekbaar. We gebruiken verschillende medische classificaties om met *Natural Language Processing* de metadata van de artikelen te verrijken. Daarnaast werken we aan het linken van artikelen aan specifieke onderzoeksvragen. *Deep Learning* gebruiken we om de affiliaties van alle

auteurs te herkennen en een database te maken met *Global Research Identifiers*, *Geonames* en dergelijke.”

Juiste plek

Artur Kiulian, een AI-onderzoeker uit Los Angeles, is de initiatiefnemer achter CoronaWhy. Hij realiseerde zich dat medische onderzoekscentra overspoeld worden door beschikbare data en informatie, en steun zouden kunnen gebruiken van

AI. Ruim 1.300 specialisten op het gebied van AI en data science van over de hele wereld sloten zich bij hem aan. Samen helpen zij de medische gemeenschap om de analyse van beschikbare gegevens die relevant zijn voor de COVID-19-uitbraak te verbeteren en te zorgen dat bevindingen terecht komen bij degenen die ze nodig hebben.

coronawhy.org

Onderzoeksgegevens online beschikbaar

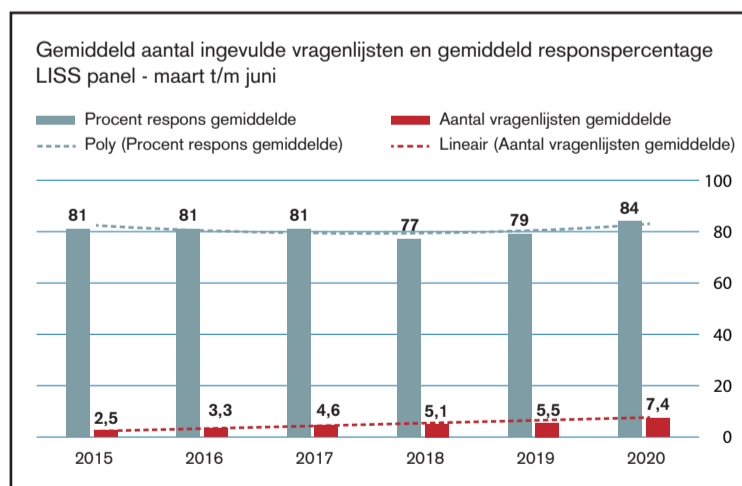
LISS panel veelvuldig ingezet

Het LISS panel is veelvuldig gebruikt om de effecten van de uitbraak van het coronavirus te onderzoeken.

Evi de Cock



LISS staat voor *Longitudinal Internet studies for the Social Sciences*. Eén van de gebruikers van het LISS panel is onderzoeker Hans-Martin von Gaudecker (Universität Bonn). Met hulp van het LISS panel kon hij al in april aantonen dat thuiswerkers dubbel voordeel hadden. Hun werkuren namen maar licht af omdat ze veel vanuit huis konden werken en tegelijkertijd hadden ze een laag risico op besmetting met het virus. Het ging hierbij vooral om veelverdieners met een vervolgoopleiding. Laagverdieners zagen een veel grotere afname in hun werkuren. Herhaalde



metingen lieten zien dat in juni de thuiswerkuren afnemen voor veelverdieners en de kloof tussen thuiswerkers en niet-thuiswerkers kleiner werd. Dit kwam onder andere door heropening van de horeca en het verspreiden van de recessie over de hele economie. Alle onderzoeksresultaten en vragenlijsten van Von Gau-

decker's onderzoek zijn beschikbaar gesteld via <https://covid-19-impactlab.readthedocs.io/en/latest/>. Ook andere onderzoekers maakten gebruik van het LISS panel. Voorbeelden zijn onderzoeken over sociale contacten, eenzaamheid (bij ouderen), financiële situatie en media-gebruik, kinderopvang, vertrouwen

in de overheid, (subjectieve) kans op besmetting, navolgen van de maatregelen, beleving medische zorg. In veel gevallen kan de data vergeleken worden met in het verleden verzamelde gegevens. Alle gegevens worden beschikbaar gesteld voor onderzoekers en beleidsmakers via lissdata.nl

Credits: CentERdata

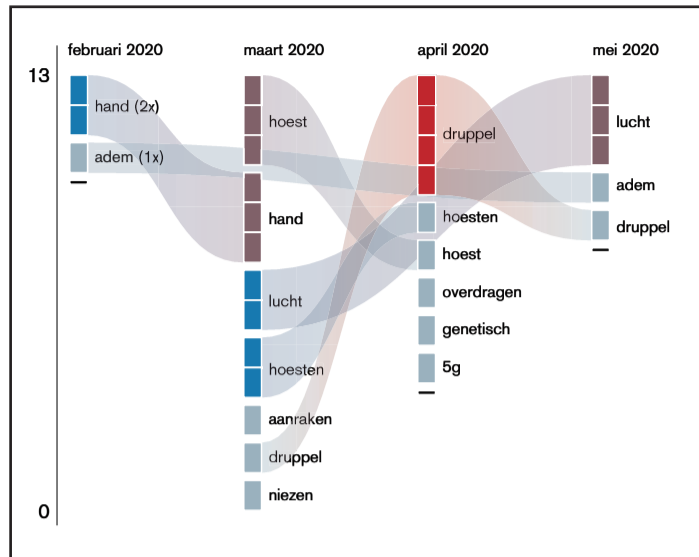
Toegang tot de Nederlandse multimediale archieven

Inzet data uit Media Suite

Twee mediawetenschappers gebruikten de Media Suite om berichtgeving over COVID-19 in nieuwsprogramma's te vergelijken met die over aids/hiv (1980-2000) en SARS (2002-2004).
Erica Renkens



De CLARIAH Media Suite geeft onderzoekers toegang tot de Nederlandse multimediale archieven van onder andere het Nederlands Instituut voor Beeld en Geluid, het EYE Film-museum, DANS en de Koninklijke Bibliotheek. Emillie de Keulenaar en Ivan Kisjes (Universiteit van Amsterdam) gebruikten de tool voor onderzoek naar berichtgeving over COVID-19. "Bij alle drie de epidemieën zagen we dat er op de Nederlandse televisie ook te-



genstrijdige informatie aan bod kwam. Tijdens de huidige COVID-pandemie gebeurt dat echter veel meer dan bij de andere twee", aldus De Keulenaar. "We zien dat dat vooral gebeurt om te *debunk*en: mythes worden doorgeprikt.

Die taak hebben televisiemakers blijkbaar op zich genomen."

Nieuwe toepassingen

De Media Suite is continu in ontwikkeling. De Keulenaar: "In de toekomst is het hopelijk ook moge-

*Tussen de meest genoemde besmettingsroutes van COVID-19 (februari-mei 2020) staat ook '5g', met tussen haakjes hoe vaak in die maand het woord voorkwam in combinatie met woorden als 'overdracht'. Het onderwerp '5g' komt tijdens deze virusuitbraak aan bod op televisie in het kader van het 'debunk*en van misinformatie. Aan de hand van de gekleurde banen zijn de toe- en afname van termen te volgen.

Credits: mediasuitedatastories. clariah.nl/covid-19/

lijk om via de aangeboden archieven de interactie tussen sociale media en televisie te onderzoeken; dat lijkt me ook erg interessant in de context van COVID-19." mediasuitedatastories.clariah.nl/covid-19/

Onderzoeken en corrigeren

Wetenschap anno '20

Tijdens een pandemie communiceren onderzoekers snel voorlopige resultaten. Nog voor de peer-review staan de krantenkoppen al te drogen: 'Immuniteit is een illusie!'. Maar klopt dat wel?

Karlijn Roex



Zeker tijdens een crisis is het een opgave om de onzin van de waarheid te onderscheiden. Van sensationele krantenkoppen tot aan gekopieerde misverstanden, van overhaaste tot aan robuuste onderzoeken. Deze informatievloed noemen we in de vakliteratuur een infodemie.

Wetenschap bevaagd

Karlijn Roex (UvA) en Giovanni Colavizza (UvA) onderzoeken hoe wetenschappers bijleren en communiceren over Covid-19. Hoe gaan zij om met de infodemie? Roex en Colavizza ontvingen deze zomer bijna 600 volledig ingevulde vragenlijsten. Een kwart van alle deelnemers gaf aan dat zij wetenschappelijk onderzoek verrichten naar de Covid-19 pandemie. Per discipline waren de percentages als volgt: 34% van de medische en gezondheidswetenschappers, 34% van de sociale wetenschappers, 13% van de exacte en natuurwetenschappers en 16% van de geesteswetenschappers.

Misinformatie corrigeren

Wetenschap biedt de middelen om informatie kritisch te evalueren. Hoe reageren wetenschappers als zij misverstanden aantreffen over de Covid-19 pandemie? Zowel hoogleraren als promovendi zijn daarin actief. Iets meer dan een kwart (26%) van de Covid-onderzoekers gaf aan dat ze recent hadden gereageerd op inaccurate informatie op sociale media of online fora. Van de wetenschappers die geen onderzoek deden naar de pandemie bedroeg dit aantal nog steeds 17%.

Verder onderzoek

Het is slechts een greep uit het vele informatiegedrag dat is onderzocht. Roex en Colavizza gaan nu zoeken naar verbanden om beter te kunnen begrijpen wanneer bepaald informatiegedrag plaatsvindt en waarom. In oktober lanceren ze een tweede enquête om veranderingen over tijd te onderzoeken. Met die kennis kunnen wetenschappers en experts beter communiceren tijdens uitbraken of andere crises.

covid19.humanities.uva.nl

"De COVID-19 pandemie bewijst eens te meer dat het delen van gegevens uit onderzoek noodzakelijk is"

Hoe bewaar je een pandemie?

#coronacollectie van erfgoedinstellingen

Kort na het uitbreken van de COVID-19-crisis deed het Netwerk Digitaal Erfgoed (NDE) een oproep aan Nederlandse erfgoedinstellingen om mee te bouwen aan een #coronacollectie.

Marcel Ras



Veel communicatie, discussie en nieuwsgaring verloopt via digitale media, zoals websites, chats, Facebookgroepen, tweets, YouTube-kanalen, memes en digitale foto's. Door

deze bronnen duurzaam te bewaren en toegankelijk te maken, blijven data en herinneringen levend. Maar hoe bewaar je een pandemie?

Samenwerking

De oproep van het NDE was een aansporing voor erfgoedinstellingen om samen te werken aan een digitale erfgoedcollectie over het coronavirus en de gevolgen daarvan in Nederland. Door de NDE werden voorbeelden en hulpmiddelen aangereikt. Zo leverde de inzet van de hotspotmonitor door

een aantal archieven digitale collecties op, die een beeld geven van de samenleving in een stad, regio of provincie tijdens de crisis. Een ander voorbeeld was de coördinatie van het Nationale Register Webarchieven bij het archiveren van websites. Ook werd een wekelijks vragenuur gestart over sociale media-archivering. Meer informatie en voorbeelden staan op de website van het Netwerk Digitaal Erfgoed.

netwerkdigitaalerfgoed.nl/actueel/coronacollectie

Draagvlak coronamaatregelen meten met tweets

Twitterdata input voor RIVM

Wat vindt Nederland van de coronamaatregelen? De Corona Gedragsunit van het RIVM onderzoekt ons gedrag en welbevinden met data uit interviews, enquêtes en Twitter.

Mathilde Jansen



Vragen over gedrag en welbevinden zijn gemakkelijk te stellen met gerichte (online) interviews en enquêtes. Naast deze standaardmethodes gebruikt de Corona Gedragsunit ook Twitter voor analyses over coronamaatregelen. De analyses zijn van



Illustratie van de top 4-vragen in mei 2020. Credits: Antal van den Bosch

Antal van den Bosch, hoogleraar taal en kunstmatige intelligentie en directeur van het Meertens Instituut.

Top-10

Zijn team, bestaande uit onderzoekers van het RIVM, de Universiteit Leiden en de Radboud Universiteit, meet met Twitter het

vertrouwen in de besluitvorming en het draagvlak voor maatregelen zoals mondkapjes.

"We maakten onder meer iedere week een top-10 van meest gestelde vragen. Bij sommige vragen zie je dan een stijgende lijn: vragen over de aanpassingen in de horeca stond eind mei op nummer 1, maar die zagen we de

weken ervoor al opklimmen. Hetzelfde zagen we met de uitbraken in slachthuizen. Voordat de problematiek in Nederland speelde, werd al over gevallen in Amerika en Duitsland getweet, en bijvoorbeeld gevraagd of vlees uit zulke slachthuizen besmet was."

rivm.nl/gedragsonderzoek

SINDS KORT BESCHIKBAAR

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

CentERdata

• Slachtoffers in de moderne samenleving

In 2018 is de longitudinale studie *Victims in Modern Society* (VICTIMS) gestart, in opdracht van en in samenwerking met het Fonds Slachtofferhulp. Voor deze studie wordt jaarlijks bij het gehele LISS panel een vragenlijst afgenomen over met name potentiële trauma's (zoals criminaliteit, ongevallen, verlies dierbaren), mentale gezondheid (MHI-5) en PTSS-symptomen (PCL5 8-item), zelfredzaamheid (CSE7), sociale steun (SSL-D) en erkenning (SAQ). In onder meer *PLOS ONE*, *Journal of Anxiety Disorders*, *Journal of Interpersonal Violence*, en *Psychiatry Research* zijn resultaten van de eerste metingen gepresenteerd. Inmiddels zijn de data van 2018 toegankelijk in het LISS data archive.

lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Struminskaya, B, Keusch, F. maart 2019, Smart devices – Part 1
- Struminskaya, B, Keusch, F. april 2019, Smart devices – Part 2
- Cloin, M, van Bakel, H., van Deurzen, I., april 2019, Parenthood
- Van der Kolk, H., Hakhverdian, A., Rooduijn, M., Zaslove, A., april 2019, Political Discontent
- Hageaars, M., Hageaars, J., mei 2019, Events in your life
- Breugelmans, S., Zeelenberg, M., Hoyer, K.,

mei 2019, Motives for greed and self-interest

- Van der Voet, J., juli 2019, How primary schools perform
- Von Gaudecker, H., van Rooij, M., Mastrogiacomo, M., maart 2020, Effects of the Outbreak of COVID-19 – wave 1
- Van der Velden, P., maart 2020, Questionnaire on the corona virus in the LISS panel
- CentERdata, april-mei 2019, Work and Schooling – Wave 12
- CentERdata, augustus-september 2019, Religion and Ethnicity – Wave 12
- CentERdata, mei-juni 2020, Personality – Wave 12



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

DANS

• Ook in EASY: Vriendschap

30 juli is door de Verenigde Naties uitgeroepen tot de internationale dag van de vriendschap. Het idee achter deze dag is dat vriendschap tussen personen, landen, culturen en individuen kan leiden tot vrede en bruggen bouwt tussen gemeenschappen. In het digitale data-archief EASY staan ook datasets en publicaties over vriendschap. Een van de voorbeelden is de dataset “Vriendschap en sociale relaties in multi-etnisch samengestelde schoolklassen 1986”, een dataset met informatie over multi-etnische vriendschappen en sociale relaties van

schoolkinderen, vergeleken met de etnische perceptie van leraren en de inhoud van het curriculum. Een andere in EASY beschikbare dataset is “Gescheiden netwerken 1987-1990: De relaties met vrienden en verwanten na echtscheiding”. Deze dataset bevat informatie over onder andere vriendschap na echtscheidingen. Heeft u vragen over deze of andere datasets? Neem dan contact op met een van de datamanagers van DANS.

easy.dans.knaw.nl



Vriendschap Credits: Klimkin via Pixabay

Ook sinds kort beschikbaar:

De volgende datasets zijn open access beschikbaar via het online archiverings-systeem EASY van DANS:

- Heine, Dr. F.A. (Tilburg University); Strobel, Dr. M.A. (Maastricht University) (2014): Data: Reward and Punishment in a Team Contest. DANS. DOI: 10.17026/dans-zmp-8vyv
- Humayun, Dr. M.I. (Hansa Luftbild) (2020): Its4land - Publish and Share platform. DANS. DOI: 10.17026/dans-xka-sj4w
- Reckien, Dr. rer. nat. D (Faculty of Geo-Information Science and Earth Observation

(ITC), University of Twente) (2017): European Local Climate Plans - EURO-LCPs.

DANS. DOI: 10.17026/dans-zj7-ejmr

- Peerdeman, Dr. K.J. (Leiden University) (2019): Pre-consultation information about one's physician can affect trust and treatment outcome expectations.

DANS. DOI: 10.17026/dans-zp9-wwm7

- Salomé, Dr SLF Le Franc (University Hospital of Rennes) (2020): HANDS - Influence of visuo-proprioceptive illusion of movement and motor imagery of the wrist on EEG cortical excitability in healthy participants.

DANS. DOI: 10.17026/dans-xt5-ncj2

- Schouten, MSc. M.J.E. (Arkin) (2020): Ehealth depression and problematic alcohol use.

DANS. DOI: 10.17026/dans-xfm-g98b

- Sociaal en Cultureel Planbureau – SCP (2018): Mediatijdbestedingsonderzoek 2018 - MTBO2018 (v1).

DANS. DOI: 10.17026/dans-2x2-8sah

- Subías, D. de León (IDDS Archeologie) (2020): Spaarne 1-35, Haarlem, Gemeente Haarlem. Opgraven landbodems, variant archeologische begeleiding.

DANS. DOI: 10.17026/dans-xtk-n76b

- Wilderink, MSc L (Windesheim University of Applied Sciences; Vrije Universiteit Amsterdam) (2019): Key elements of a successful integrated community-based approach aimed at reducing socioeconomic health inequalities in the Netherlands: a qualitative study.

DANS. DOI: 10.17026/dans-zkf-46xc



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

ODISSEI creëert met investering NWO nieuwe vormen van samenwerking

Digitale mogelijkheden veranderen sociaalwetenschappelijk onderzoek

ODISSEI, een grote wetenschappelijke infrastructuur van Nederland, heeft onlangs financiering van NWO ontvangen. Lucas van der Meer

De samenleving staat voor een groot aantal uitdagingen, zoals klimaatverandering, migratie en toenemende sociale ongelijkheid. Wetenschappelijk directeur van ODISSEI, Pearl Dykstra: “We leven in een zeer onvoorspelbare en complexe wereld. Om bestaande onzekerheden en onderlinge afhankelijkheden te doorgronden, dienen sociale wetenschappers toegang te hebben tot gekoppelde gegevens, geavanceerde computertechnieken en analytische expertise. Uiteraard met inachtneming van de hoogste veiligheidsnormen ten opzichte van persoonsgegevens.”

Precies met dat doel is eind 2016 ODISSEI opgericht. ODISSEI staat voor *Open Data Infrastructure for*

Social Science and Economic Innovations. Via ODISSEI hebben onderzoekers toegang tot grootschalige, longitudinale datacollecties, evenals innovatieve en diverse vormen van data. Deze kunnen gekoppeld worden aan administratieve gegevens van het Centraal Bureau voor de Statistiek. Het samenwerkingsverband telt inmiddels 40 partnerorganisaties.

Uitbouw infrastructuur

Met de investering vanuit de Nederlandse Roadmap voor Grootschalige Wetenschappelijke Infrastructuren zal ODISSEI haar infrastructuur uitbouwen. Zo wordt de *Data Facility* een cluster van systemen om de vindbaarheid, toegankelijkheid, interoperabiliteit en herbruikbaarheid (FAIR) van gegevens te vergroten. Het omvat ook een veilige, krachtige computeromgeving met koppelbare datasets en geavanceerde analysetools op de nationale supercomputer



Het ODISSEI-team toast op de investering die het van NWO ontving.

Credits: ODISSEI

van SURF. Als tweede zal het *Observatory* de gegevensverzameling optimaliseren, integreren en continuïteit daarin waarborgen. Hierdoor is het mogelijk om trends te bestuderen. Het *Laboratory* werkt met

behulp van de financiering verder aan de ontwikkeling van innovatieve databronnen, analysetechnieken en methodologieën. Deze mogelijkheden moeten het repertoire van sociale wetenschappers uitbreiden,

waardoor onderzoek met 'traditionele' en 'nieuwe' datasets mogelijk is. Als laatste zal de *Hub* het onderwijs- en opleidingsprogramma van ODISSEI coördineren. Hiermee kunnen onderzoekers zich bekwamen in geavanceerde analysetechnieken en instrumenten voor computationele sociale wetenschappen. Door gebruik te maken van de digitale onderzoeksmogelijkheden probeert ODISSEI de manier te veranderen waarop sociaalwetenschappelijk onderzoek wordt uitgevoerd. Onafhankelijke beoordelaars denken dat dit Nederland in staat stelt om wereldleider in de sociale wetenschappen te blijven.

Dykstra sluit af: “ODISSEI creëert nieuwe vormen van samenwerking tussen de sociale en computationele wetenschappen. Die samenwerking is cruciaal, willen we de grote sociale problemen van onze tijd in al hun complexiteit doorgronden.”

odissei-data.nl

Eén basisregistratie voor historische personen

Miljoenen geboorte-, overlijdens- en huwelijksaktes gelinkt

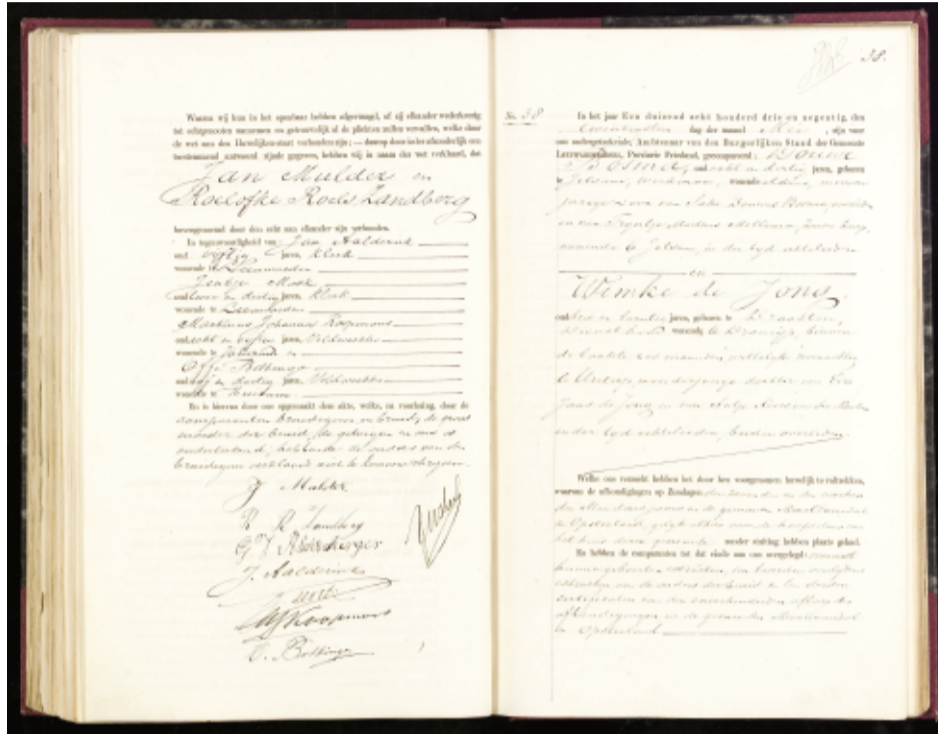
Vanaf 1812 houdt men bij wie wordt geboren, huwt en overlijdt. Deze akten zijn nu gelinkt beschikbaar voor de wetenschap en amateurgeneologen via wiewaswie.nl.

Mathilde Jansen

De aktes van de Burgerlijke Stand zijn een schat aan data voor iedereen die meer wil weten over zijn voorouders. Dat de aktes ook voor demografisch en ander historisch onderzoek relevant zijn, bedacht Kees Mandemakers (IISG) al enkele jaren geleden. Hij bracht een samenwerking tot stand tussen het Centraal Bureau voor Genealogie (CBG) en het IISG. Het CBG leverde de aktes – die vanaf de jaren negentig waren gedigitaliseerd en getranscribeerd door vele vrijwilligers – en onderzoekers van onder andere het IISG ontwikkelden een systeem om de aktes automatisch te linken.

Minder tijdrovend

Onlangs werd via CLARIAH PLUS een nieuwe methode ontwikkeld waardoor het berekenen van een linkset veel minder tijdrovend werd. Auke Rijpma, projectleider bij CLARIAH PLUS en onderzoeker aan de Universiteit Utrecht en het IISG: “De regelsets die door het



Online is deze huwelijksakte van Douwe Bosma, registratiedatum 1893, nu ook gelinkt aan andere informatie, zoals informatie over de vader en moeder van de bruid en van de bruidegom. Credits: wiewaswie.nl/nl/detail/7174744

team van Mandemakers zijn ontwikkeld, gebruiken we nog steeds. Joe Raad van de Vrije Universiteit heeft een nieuwe implementatie gedaan, gericht op efficiëntie. Nu kost het linken ongeveer een dag.”

De methode werkt op basis van regelsets, of-

tewel *rule based linking*. De namen in twee verschillende aktes mogen bijvoorbeeld aan elkaar gelinkt worden wanneer ze maximaal twee karakters van elkaar verschillen, mits de naam vijf letters of meer bevat. Namen hoeven niet identiek te zijn, dit heeft bijvoorbeeld te

maken met de variatie in schrijfwijze.

Overlinking

Door de grote hoeveelheid aktes gaat het linken nog niet altijd goed. Om de mismatches eruit te halen, kijken de onderzoekers naar ‘overlinking’: “Een geboorteakte moet uniek gelinkt zijn aan de aktes van de ouders. Als een akte vaker gelinkt wordt, gaat er ergens iets mis. Daarnaast maken we een evaluatiedataset waarin we een paar duizend links handmatig aanleggen. Vervolgens maken we een vergelijk met de geautomatiseerde links.”

44 miljoen aktes

Voorlopig zijn alle 12 miljoen huwelijksaktes gelinkt, daar komen nog 20 miljoen geboorteaktes en 12 miljoen overlijdensaktes bij. De gelinkte data zijn om diverse redenen belangrijk: het biedt onderzoekers bijvoorbeeld de mogelijkheid om sociale ongelijkheid te zoeken, zowel in termen van waar je in de maatschappij terecht komt, als hoe je sociale omgeving eruit ziet. “Het is een kwestie van dezelfde procedure opnieuw implementeren. We verwachten dan ook dat dit nog sneller gaat. Ons streven is om nog dit jaar een besloten release voor geïnteresseerde onderzoekers te hebben, en medio 2021 een volwaardige open release.” De gelinkte data zijn al zichtbaar op wiewaswie.nl en komen via druid.datalegend.net beschikbaar voor onderzoekers.

druid.datalegend.net

Schatting van tekstverlies met methodes uit ecodiversiteit

Het Atlantis van de ridderepiek

Letterkundigen duiden literatuur uit de middeleeuwen aan als ‘de verloren gegane wereld Atlantis’ of als ‘wrakhout van een aangespoelde vloot’. De grote vraag is: hoe groot was die vloot? Folgert Karsdorp en Mike Kestemont rekenden het uit.

Mathilde Jansen

Een eerdere ruwe schatting van historisch letterkundige Frits van Oostrom telde ruim honderd verschillende ridderepische teksten. Hij baseerde dit aantal op bestaande bronnen die verwijzen naar teksten die we niet kennen, de tekstgetuigen. Folgert Karsdorp (Meertens Instituut) en Mike Kestemont (Universiteit Antwerpen) besloten om de vraag te beantwoorden met een *resampling methode*, een methode nog niet eerder toegepast binnen historisch letterkundig onderzoek.

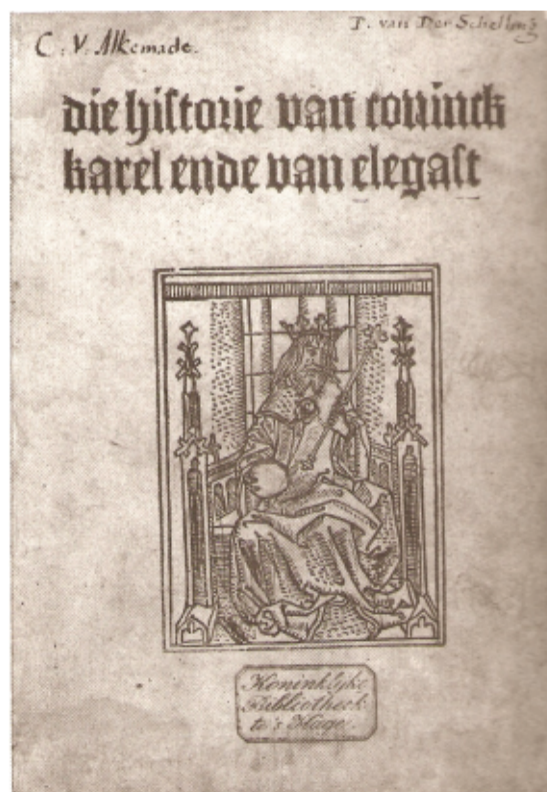
Soortenrijkdom

Karsdorp licht toe: “Biologen bepalen aan de hand van observaties het werkelijke aantal diersoorten. Hierbij speelt een negatieve bias:

het aantal getelde diersoorten is een onderschatting van de totale soortenrijkdom. Hetzelfde geldt voor middeleeuwse teksten. Door deze overeenkomst kwamen we op het idee om gebruik te maken van de statistische methoden die biologen gebruiken om de totale soortenrijkdom te bepalen.” De onderzoekers lieten zich in hun methodiek inspireren door Alex Mesoudi, hoogleraar Culturele Evolutie aan de Universiteit van Exeter (UK), die al in 2011 een pleidooi hield voor een Darwiniaanse studie van de ontwikkelingsgang van culturele fenomenen zoals literatuur.

Nieuwe conclusie

Karsdorp en Kestemont gingen bij hun berekening uit van 74 unieke teksten en 164 tekstgetuigen. De eerste toegepaste methode was *Jackknife resampling* (knipmesmethode). Hierbij wordt elke tekstgetuige tijdelijk uit de data verwijderd. Op basis van deze methode werd een schatting gemaakt van (minimaal) 113 tot (maximaal) 128 teksten. De tweede methode was de methode-Chao, die op een verregaande ma-



nier gebruikmaakt van *resampling*. De schatting met behulp van deze methode kwam op 106 tot 219 teksten. Karsdorp: “Op basis van de uitkomsten is de schatting van Van Oostrom niet onmogelijk, maar wel erg conservatief. Volgens de laatste

Karel ende Elegast is een ridderverhaal met een hoofdrol voor Karel de Grote, geschreven omstreeks 1270. Dankzij meerdere drukken uit de 15de en 16de eeuw is dit de enige Middelnederlandse Karelroman waarvan de volledige tekst beschikbaar is.

Credits: KB

methode is het immers ook mogelijk dat we met 219 originele teksten te maken hebben gehad. Dat zou betekenen dat zelfs maar de helft van het totaal aantal teksten bewaard is gebleven.”

DOI: 10.2143/SDL.61.3.3287540

IISG-collectie doorzoekbaar in meerdere talen

Het IISG heeft veel boeken, tijdschriften en archiefmateriaal in latijnse talen en alfabetten, zoals Amhaars, Arabisch, Bengali, Farsi en Tamil. Om dat materiaal te catalogiseren, is altijd gebruikgemaakt van transliteratie – het ‘converteren’ van de niet-westerse karakters naar het westerse alfabet.

Catalogussystemen ondersteunden vroeger alleen het westerse alfabet. Bij transliteratie kan echter veel fout gaan. Bovendien was het materiaal op deze manier slecht vindbaar voor mensen uit de taalgebieden waar de collecties vandaan komen. Daarom werden tijdelijk vijf *native speakers* aangesteld die auteursnamen, titels en archiefbeschrijvingen aanleverden van tijdschriften en boeken die al in de catalogus zitten. Het resultaat staat op de website van het IISG. (MJ)

iisg.amsterdam/nl/collecties

Een vergelijking van verschillende methodes voor dataverzameling

Nationaal Kiezersonderzoek in de toekomst online?

Online dataverzameling is kostenbesparend, maar hoe zit het met de datakwaliteit? Onderzoekers van de Universiteit van Amsterdam onderzochten dit voor het Nationaal Kiezersonderzoek. Evi de Cock

Het Nationaal Kiezersonderzoek (NKO), georganiseerd door Stichting Kiezersonderzoek Nederland (SKON), is een surveyonderzoek dat sinds 1971 na elke Tweede Kamerverkiezingen plaatsvindt. Via huisbezoeken door interviewers worden willekeurig geselecteerde Nederlandse kiezers gevraagd naar hun politieke opvattingen, percepties en gedragingen. In 2017 voegde het NKO een online vragenlijst toe, gestuurd naar een nieuwe, willekeurige steekproef en naar respondenten van het LISS panel, een lang-

lopend *probability based* panel van ongeveer 4.500 huishoudens. Roderik Rekker, Tom van der Meer en Wouter van der Brug onderzochten of de online vragenlijsten leidden tot andere data ten opzichte van huisbezoeken.

Nauwelijks andere data

Rekker: "Wat bleek: de online vragenlijsten leverden nauwelijks andere data en soms zelfs een betere datakwaliteit op. Alleen de oudste groep kiezers leek wat lastiger te bereiken via internet.

Als het NKO in de toekomst volledig online wordt afgenomen, zijn vergelijkingen over de tijd met eerdere NKO's nog steeds mogelijk. Onderzoekers moeten dan, met de data van 2017, wel goed controleren of de verandering van methode in het geval van hun specifieke vraagstelling niet toch een trendbreuk heeft veroorzaakt", zegt Rekker.

Ook de resultaten met betrekking tot het gebruik van het LISS panel waren positief. Rekker: "Het benaderen van respondenten uit het LISS panel leidde slechts in beperkte



Het Nationaal Kiezersonderzoek meet sinds 1971 na elke Tweede Kamerverkiezingen politieke opvattingen, percepties en gedragingen.

Credits: Andreas Breiting, Pixabay

mate tot andere resultaten dan het trekken van een verse steekproef, maar de LISS steekproef bleek wel minder representatief dan de verse steekproef."

Het NKO hoopt in de toekomst het werven via een verse willekeurige steekproef te combineren met afname in het LISS panel. Op deze manier kan de representativiteit van

het NKO gewaarborgd blijven, terwijl er ook geprofiteerd wordt van de voordelen van een panel waarin bijvoorbeeld dezelfde kiezers over langere tijd gevolgd kunnen worden.

Rapport:

DOI: 10.13140/RG.2.2.17927.42400

Data:

DOI: 10.17026/dans-xby-5dhs

COLUMN

Samen leren van data

In de zorg, zeker in de psychiatrie, wordt diagnostiek en behandeling op basis van kennis, ervaring en richtlijnen uit wetenschappelijk onderzoek bedreven. Maar de praktijk wijkt nogal eens af van de bevindingen in het lab. Zo is er altijd sprake van selectiebias bij onderzoek, zoals het feit dat ernstig zieke, suïcidale of minder wilsbekwame patiënten en patiënten met co-morbiditeit niet worden geïncludeerd in onderzoek.

De verplichting om op systematische wijze gegevens te verzamelen en te registreren, om de kwaliteit van zorg te monitoren, te evalueren en te verbeteren, zou dit gat tussen kennis uit onderzoek en kennis uit de praktijk kunnen dichten. Maar hoe?

De sleutel ligt mijns inziens in een vruchtbare samenwerking tussen mensen van verschillende domeinen. De data scientist is nodig om de data te ontsluiten, te processen, te analyseren en te interpreteren. Psychiaters en verpleegkundigen kennen de klinische vraagstukken, kunnen de

kwaliteit van de data duiden en begrijpen de inhoud. Uiteraard zijn zorgdata niet vrij toegankelijk en vindt vanwege de hoge sensitiviteit de verwerking en analyse met grote zorgvuldigheid plaats.

Binnen de afdeling Psychiatrie van UMC Utrecht werkt een data science team nauw samen met de klinici en patiënten. En wat blijkt: het voordeel werkt twee kanten op. Enerzijds komen de relevante vraagstukken uit de praktijk bij de data scientists en tegelijkertijd openen inzichten uit de data de deur voor nog niet ge-

voerde discussies. Samen wordt verkend wat de mogelijkheden zijn om kennis laagdrempelig uit te wisselen, *tooling* gezamenlijk in te zetten en uitkomsten en algoritmes uit te wisselen om verder te verbeteren.

Instelling overstijgend is onze samenwerking met Parnassia Groep, St Antonius en GGZ Eindhoven in het *Compute Visits Data consortium*. De data blijven binnen de instellingen, maar de eerste algoritmes zijn succesvol uitgewisseld. Ook zijn we onderdeel geworden van het initiatief *Perso-*

nal Health Train dat gelijke doelen nastreeft vanuit de principes van *federated learning*. Om dit type toegepast dataonderzoek verder te brengen, is wetenschappelijk onderzoek nodig. Technieken die kunnen worden toegepast, zijn legio, waarbij met name *text mining* veelbelovend is. Immers, zeker in de psychiatrie, is er sprake van grote hoeveelheden geschreven tekst. Om deze beter te benutten, zijn steeds meer *state of the art* methodes voorhanden.

Om die kennis vervolgens daadwerkelijk in de praktijk te benutten, is de samenwerking tussen data experts en zorg experts cruciaal. Samen zorgen we voor een betere kwaliteit van de zorg van morgen.

Karin Hagoort

Karin Hagoort is teamleider Innovatie en onderzoeker op de afdeling Psychiatrie van het UMC Utrecht.

Karin geeft de column door aan: prof.dr. Sander Klous.



GELEZEN

Verkenning en optimalisering Nationaal Datalandschap
Marjan Grootveld

"Laat duizend bloemen bloeien onder deskundige leiding van een hovenier". Zo zou je de *Verkenning Nationaal Datalandschap* kunnen samenvatten. Een projectgroep namens het Nationaal Programma Open Science (NPOS) heeft bij onderzoekinstellingen en aanbieders van datadiensten – zoals opslag en archivering van data, research softwareontwikkeling en kennisoverdracht inzake FAIR data – geïnventariseerd welke partijen en structuren een rol spelen in het Nederlandse datalandschap. De verkenning was gericht op de organisatie van optimaal hergebruik van onderzoeksdata en gaat dus voorbij aan zaken als de kwaliteit of de huidige mate van herbruikbaarheid van data.

De Verkenning beschrijft een rijk maar gefragmenteerd datalandschap, al identificeert het ook een klein aantal 'gebruikelijke partners', zoals DANS, 4TU.Research, SURF, het UKB-bibliothekenverband en het LCRDM-netwerk. Niettemin is de eerste stap op de voorgestelde roadmap voor FAIR data het inrichten van een nationaal datacoördinatie- en expertisecentrum: de hovenier. Dit centrum kan bijvoorbeeld de aanstaande *Data Competence Centers* ondersteunen en een nationale dienstencatalogus coördineren – met de kanttekening dat internationaal onderzoek dikwijls zijn datadiensten elders afneemt. Of het centrum uiteindelijk meer of minder zicht op de bloemen geeft, zal afhangen van zijn plek in het datalandschap en het gezag dat het heeft of verwerft. In verdere aanbevelingen knoopt de Verkenning aan bij andere NPOS-doelen, zoals waardering voor Open Science.
opensecience.nl/projecten/projecte-verkenning-en-optimalisering-nationaal-datalandschap

Ben jij FAIR-Aware? Nieuwe tool voor FAIR-data

Met de nieuwe tool *FAIR-Aware* kunnen onderzoekers zelf beoordelen, hoe FAIR hun data zijn, voordat de data gedeponereerd worden. De online *self-assessment tool* doorloopt middels 10 vragen met de gebruikers alle aspecten van de FAIR-principes. Dankzij tips bij elke vraag, leert de gebruiker de vraagstukken beter begrijpen en weten zij hoe ze hun data FAIR of FAIR-der kunnen maken. De tool is ontwikkeld binnen het Europese FAIRs-FAIR-project, DANS is een van de partners. (HB)

fairsfair.eu