

Database Atlas of Mutual Heritage toont gedeeld overzees erfgoed

# Handelsgebieden VOC en WIC online in beeld

Het team achter de Atlas of Mutual Heritage verzamelt en beschrijft al ruim 22 jaar historische beelden van vroegmoderne koloniale nederzettingen en presenteert deze in samenvatting op het web.

Maarten Heerlien



Eén van de afbeeldingen in de database Atlas of Mutual Heritage is de stad Melayu op Ternate. Het betreft een waterverftekening door Johannes Vingboons, uit circa 1666.

Credits: Atlas-Blaeu van der Hem, collectie Nationale Bibliotheek van Oostenrijk

Wat in de jaren negentig begon als stageproject bij het Rijksmuseum is uitgegroeid tot een belangrijke online bron voor onderzoek naar het koloniale verleden van Nederland. In de Atlas of Mutual Heritage (AMH) zijn gegevens en duizenden afbeeldingen van historische kaarten, tekeningen en prenten te vinden van locaties gerelateerd aan de Verenigde Oost-Indische Compagnie en West-Indische Compagnie. Het is daarmee de oudste nog operationele online onderzoeksdatabase voor geschiedenis in Nederland.

De AMH, een samenwerking van het Rijksmuseum, Nationaal Archief, Koninklijke Bibliotheek en de Rijksdienst voor het Cultureel Erfgoed, toont afbeeldingen uit collecties van 26 instellingen. Dat aan-

tal groeit nog steeds, aldus Tristan Mostert, docent Geschiedenis aan de Universiteit Leiden en één van de coördinatoren van de Atlas. “De afbeeldingen geven inzicht in plaatsen waar de VOC en WIC actief waren. Dat beeld is nooit compleet.”

## Gedeeld erfgoed

Sinds de start is de AMH zowel in technisch opzicht, als in haar doelstellingen en benadering, met de tijd meegegaan. Mostert: “Toen Martine Gosselink, nu hoofd Geschiedenis bij het Rijksmuseum, als stagiaire de eerste versie ontwikkelde,

gaf deze museumbezoekers een beeld van het dagelijks leven van VOC- en WIC-dienaren overzee. Met nieuwe partners veranderde de nadruk naar de notie van gedeeld erfgoed: we delen deze geschiedenis met de hedendaagse bewoners van de octrooigebieden. De geografische spreiding van de gebruikers van de Atlas illustreert dat ook.”

## Groei voor de toekomst

De grote duurzaamheid van de Atlas schrijft Mostert toe aan de organisatie ervan. “We zijn een klein, hecht team dat bijna alles zelf doet.

Naast het onder de aandacht brengen van ons platform bij de deelnemers en gebruikers, zoeken we voortdurend nieuwe content. Zo voegen we momenteel afbeeldingen uit de British Library en de Koninklijke Bibliotheek Brussel toe. Ook willen we content met betrekking tot expeditiereizen toevoegen. Daarnaast is actualisering van oudere records nodig, omdat soms terminologie is gebruikt die nu niet meer acceptabel wordt geacht. Het logische gevolg van een database met zo'n lange geschiedenis.”

[atlasofmutualheritage.nl](http://atlasofmutualheritage.nl)

ODISSEI Secure Supercomputer brengt relatie inkomen en buurt in kaart

## Analyse van 3,4 miljard datapunten

**Wat is de relatie tussen de buurt waarin je woont en je individuele inkomen? Aan deze onderzoeksvraag werkten Ana Petrović (TU Delft) en collega's met behulp van de ODISSEI Secure Supercomputer. Erica Renckens en Lucas van der Meer**

Het onderzoek van Ana Petrović (TU Delft), Maarten van Ham (TU Delft) en David Manley (University of Bristol) richt zich op de invloed van ruimtelijke ongelijkheid op individuele sociaaleconomische status. Hiervoor gebruiken ze gegevens van het CBS, onder andere geanonimiseerde datasets over inkomen per woonadres door de tijd heen. In een verkennende analyse hebben ze grids – concentrische cirkels om elke individueel persoon – gemaakt. De grids hebben ze voor verschillende schalen gemaakt, variërend van 100 meter tot 10 kilometer. Ana: “Voor elke schaal-

grootte hebben we twee contextuele kenmerken berekend: het aandeel bewoners met een niet-Westerse achtergrond en het aandeel bewoners met een laag inkomen. Deze kenmerken hebben we gelinkt aan individuele kenmerken zoals inkomen, gemeten in 17 verschillende jaren.”

## Computerkracht

Al met al had het onderzoek 3,4 miljard datapunten om te analyseren. In de Remote Access-omgeving van het CBS zou het doorrekenen van de dataset minimaal vier maanden hebben gekost. Voor dergelijke gevallen heeft ODISSEI, in samenwerking met CBS en SURFsara, de ODISSEI Secure Supercomputer ontwikkeld. De beveiligde CBS-omgeving werd gekopieerd naar de Cartesius supercomputer van SURFsara. Lykle Voort, ontwikkelaar bij SURFsara: “Voor de analyse van Ana waren computers met minimaal 64GB aan werkgeheugen nodig. Omdat het hier ging om heel veel

combinaties van gegevens, is een cluster ingezet van 25 nodes met 24 cores en 64GB aan werkgeheugen elk. In totaal zijn dat dus 600 processor cores. Ter vergelijking: een laptop heeft er doorgaans 2 of 4.” Ana: “Zonder de ODISSEI Secure Supercomputer hadden we ons onderzoek praktisch niet kunnen uitvoeren. We hebben zelfs een variabele extra en een langere periode kunnen berekenen.

We gaan de resultaten open access publiceren. In vervolgonderzoek willen we de ruimtelijke structuur

van de sociale omgeving, de trends van ruimtelijke segregatie op meerdere schalen en de gevolgen van deze processen voor individuen nog beter begrijpen. Ook ons vervolgonderzoek zal erg veel datapunten hebben en we hopen opnieuw van de ODISSEI Secure Supercomputer gebruik te kunnen maken.”

De ODISSEI Secure Supercomputer wordt naar verwachting voor de zomer voor de ODISSEI-onderzoeksgemeenschap geopend.

[odissei-data.nl](http://odissei-data.nl)

## Wint u de Nederlandse Dataprijs 2020?

Ook dit jaar wordt de Nederlandse Dataprijs uitgereikt. Een prijs voor een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te stellen voor anderen. Bent of kent u zo'n onderzoeker met een goed voorbeeld van gedeelde data? U kunt uzelf of een ander nomineren.

Kijk voor meer informatie op de website van RDNL. (HB)

[researchdata.nl](http://researchdata.nl)

## E-DATA & RESEARCH

Jaargang 14 | nummer 2

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

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

## INHOUD

- 3 KB archiveert oertijd Hollandse web
- 4 Ook deze datasets zijn sinds kort beschikbaar
- 4 Kadaster en NA brengen onteigening WOII in beeld
- 5



DANS biedt schatkamer aan online informatie

- 6 Vragenlijstenbank over taal en cultuur online
- 7 Flora en fauna in kaart van Indonesische archipel
- 7 Jong Talent Garcia-Martí aan het woord
- 8 Gastcolumnist Steijn geeft kijkje in zijn keuken

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



Scan deze QR-code met een smartphone om de website van E-data te bezoeken. [edata.nl](http://edata.nl)

## NPSO Innovatiedag: 13 presentaties, 1 winnaar

Joris Mulder

Op dinsdag 26 november organiseerde het Nederlandstalig Platform voor Survey Onderzoek (NPSO) de tweejaarlijkse Innovatiedag voor Nederlandse en Vlaamse onderzoekers. Een mooie dag voor dertien, met name startende, onderzoekers om hun werk over methodologie voor het gebruik en/of de analyse van surveys, big data en combinaties van diverse databronnen te presenteren.

Zo presenteerde Anne Elevelt hoe sensors in smartphones fysieke gezondheid meet. De JavaScript-tool SurveyMotion meet de kniebuigingen van de respondenten en voegt ze als objectieve metingen aan de surveydata toe. In een andere presentatie toonde Jonas Klingwort aan hoe sensoren op het Nederlandse wegennet ervoor zorgen dat het vrachtverkeer kan worden geteld. Hierdoor kan ook het vrachtverkeer op wegen zonder sensoren worden voorspeld. Met behulp van *web-scraping* wordt een model gebouwd om dit accuraat te kunnen voorspellen. Keynote-spreker Irene Salemin, directeur Innovatie, Ontwikkeling en Beheer bij het CBS, sprak over de digitale transformatie en liet zien hoe het combineren van verschillende databronnen de toekomst van dataverzameling vormt.

Alle gepresenteerde onderzoeken hebben als doel om op termijn te leiden tot innovatie in de waarneming, de dataverzameling en de verwerking en analyse van surveystatistiek. Aan het einde van de dag konden deelnemers stemmen op het meest innovatieve onderzoek. Jiayun Jin, PhD-student aan de Katholieke Universiteit Leuven, ontving deze NPSO-innovatieprijs. Zij doet onderzoek naar het combineren en toepassen van multivariate en bootstrap-technieken op numerieke en categorische indicatoren, zodat de kwaliteit van interviewresponsen beter kan worden beoordeeld. [npsos.net](http://npsos.net)

## De werkelijke waarde van de Hack-a-LOD

Richard Zijdeman

Eind november werd de zesde Hack-a-LOD georganiseerd in de Theaterloods bij Radio Kootwijk. Binnen 24 uur diene teams van data-, software- en domeinmensen een product neer te zetten, gebaseerd op tenminste twee Linked Open Data-datasets, datasets die online te bevragen zijn door mens én machine. De twee winnende ontwerpen waren zeer spectaculair: een soort Tamagochi-app om erfgoed in leven te houden en een Radio-Kootwijk-Tinder-app, waarmee metadata

## GEHOORD & BIJGEWOOND



Winnaars Ed de Heer, Martijn Arts en Robbert van 't Westende bedachten een historische radio-app op basis van ANP-radiobulletins. Credits: Jacqueline van der Kort, KB

over historische radiob berichten kon worden geliked door ze te swipen. Het weekend, althans zo lang voelt de 24 uur, is een bijzondere mix tussen een gezellig samenzijn en een onderlinge strijd tussen 'peers' om de eeuwige roem. Ja, dat is natuurlijk 'lokale' roem, maar toch merk je dat het winnen (of het helaas niet-winnen) van de prijzen een dingetje is. Ieder team barst van de skills, creativiteit en drang om iets bijzonders neer te zetten en iedereen wint daarom graag de publieks- dan wel de juryprijs.

Eerder op de dag inspireerde Plezant, de theatergroep van Radio Kootwijk, door de vraag centraal te stellen: "wat is werkelijk van waarde?" Toegepast op de Hack-a-LOD: waarom organiseer je zoiets? Wat is het nut ervan? Wat drijft iemand om een stel nerds (sorry collega's) bijeen te 'herden' en 24 uur aardig te doen, tegen een stel slaapdronken vakidioten, die ongeacht de werkelijke snelheid altijd zullen klagen over gebrekkige WIFI? Waarom een jaar lang zoeken naar een locatie en vervolgens naar bedrijven die allerlei diensten kunnen leveren op de meest onmogelijke plekken? Eigenlijk zouden we dat aan de organisatoren moeten vragen: KB, KNAW HuC, Stimuleringsfonds Creatieve Industrie en Beeld & Geluid. Misschien doen ze het vanwege de outreach. Of vanwege de throughput value: 50 deelnemers werken zo'n 20 uur à € 100 per uur, dat is toch al snel een ton per etmaal. Of misschien, heel misschien, is het gewoon om een bub-

bel te creëren. Een focus-bubbel, waarin we 24 uur even niet na hoeven te denken over institutionele belangen en gewoon kunnen doen wat het beste voor het veld is. Een community-bubbel, waarin we elkaar leren kennen en elkaar verbinden. Want ondanks de onderlinge competitie wordt er uitgewisseld, gecommuniceerd, geleerd, gedeeld en ontsloten. Een waarde-bubbel, waarin we, heel even, puur kunnen werken aan wat werkelijk van waarde is.

[hackalod.com](http://hackalod.com)

## PID-NL: workshop over gebruik van Persistent Identifiers in Nederland

Iris Vocking

20 november organiseerde RDA-NL in samenwerking met 4TU, Data-Cite, FREYA, ORCID en SURF een workshop gericht op Persistent Identifiers (PID's) voor zeventig geïnteresseerden uit Nederland en België, werkzaam bij universiteiten, onderzoekscentra, datacenters en onderzoeksbibliotheken.

Persistent Identifiers zijn een belangrijk onderdeel van Open

Science en FAIR-datapraktijken. PID's zorgen ervoor dat zaken binnen de onderzoeksworkflow vindbaar, citeerbaar, begrijpelijk en verbonden zijn.

Tijdens de workshop introduceerden Nederlandse experts het belang van PID's binnen goed Research Data Management (Maria Cruz, Esther Plomp). Speciale aandacht was er voor PID-typen zoals DOI en PID's van onderzoekers (Madeleine de Smaele, Peter Verhaar). Hylke Koers onderzocht de kracht van PID's door poëzie.

In de middag vonden hands-on sessies plaats met onderwerpen zoals: een algemene verkenning van het identifier-landschap, PID's voor digitaal cultureel erfgoed, PID's in de natuurwetenschappen, werken met de PID-graph, en hoe de RDA effectieve PID's ondersteunt.

De dag werd afgesloten met een borrel en netwerksessie. Remco van Veenendaal, één van de deelnemers: "Deze combinatie van het ontmoeten van oude en nieuwe collega's en inspirerende workshops maakte het een zeer nuttig evenement."

[rd-alliance.org](http://rd-alliance.org)



Een volle zaal tijdens de opening van PID-NL door Lisa de Leeuw (RDA-NL/DANS). Credits: Helena Cousijn

## KORT

### Data Nobelprijswinnaars in Dataverse

Afgelopen oktober kregen Abhijit Banerjee (Massachusetts Institute of Technology, MIT), Esther Duflo (MIT) en Michael Kremer (Harvard University) de Nobelprijs voor Economie 2019, voor hun werk op het gebied van wereldwijde armoedebestrijding. Deze drie economen zijn betrokken bij het Abdul Latif Jameel Poverty Action Lab (J-PAL). J-PAL is een wereldwijd onderzoekscentrum dat impactevaluaties uitvoert om de effectiviteit van sociale programma's op het gebied van armoedebestrijding te testen en te verbeteren. Al meer dan tien jaar is J-PAL toonaangevend in het transparant maken van onderzoeksresultaten door hun gegevens te publiceren in de Dataverse repository van de Universiteit van Harvard. (MW) [dataverse.harvard.edu](http://dataverse.harvard.edu)

### DANS-subsidie voor ICPSR Summer Program

DANS verleent ook dit jaar weer subsidie aan research masterstudenten en PhD's voor het bezoeken van het Summer Program in Quantitative Methods of Social Research van het Inter-university Consortium for Political and Social Research (ICPSR). De subsidie bedraagt 2.000 euro, aanvragen kunnen tot 1 april worden ingediend. Meer informatie staat op de website van DANS. (HB) [dans.knaw.nl](http://dans.knaw.nl)

### Tip: omgaan met digitale manuscripten

Steeds vaker maken onderzoekers gebruik van digitale foto's in plaats van het daadwerkelijke manuscript. Dit heeft gevolgen voor bijvoorbeeld citatie: bij exclusief gebruik van digitale foto's moet niet worden geciteerd naar het originele manuscript. In het boek *Among Digitized Manuscripts: Philology, Codicology, Paleography in a Digital World* legt Cornelis van Lit uit hoe het digitaliseren van manuscripten delen van de geesteswetenschappen verandert naar digital humanities. Van Lit introduceert een aantal termen voor de materialiteit van gedigitaliseerde manuscripten. Ook bespreekt hij een breed scala aan praktische vaardigheden voor onderzoekers, zoals het maken van vectorafbeeldingen, het omgaan met standaarden zoals IIF en TEI en het schrijven van Pythoncode om afbeeldingen te analyseren. (MJ)

[lwcvl.com](http://lwcvl.com)

COLOFON Uitgever: E-data & Research. Redactieadres: Anna van Saksenlaan 51, 2593 HW Den Haag, 070-3494450, [edata@dans.knaw.nl](mailto:edata@dans.knaw.nl), [edata.nl](http://edata.nl).

Hoofd-/eindredacteur: Heidi Berkhout. Redactie: Steven Claeysens, Evi de Cock, Maarten Heerlien, Mathilde Jansen, Lucas van der Meer, Erica Renckens, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning. Aan dit nummer werkten mee: Marika de Bruijne, Felix van Gelder, René van Horik, Cornelis van Lit, Joris Mulder, Nicoline van der Sij, Kim Smeenk, Sander Steijn, Kees Teszelszky, Iris Vocking, Richard Zijdeman. Opmaak: Colette Sloots, Haarlem.

Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Oplage: 6000 papier, 4800 digitaal. ISSN: 1872-0374.

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

### OVERNEMEN ARTIKELEN

Wilt u een artikel uit dit blad overnemen?

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

Door het verdwijnen van XS4ALL gaat de KB ook XS4ALL-homepages bewaren voor cultuurhistorisch onderzoek. *Kees Teszelszky*

10 januari 2019 werd bekend dat XS4ALL verdwijnt. Deze provider bood vanaf 1 mei 1994 haar diensten aan het publiek aan en is daarmee een van de eerste internetproviders van ons land en een van de vroegste particuliere providers ter wereld. Veel internetpioniers brachten hun eerste website uit via XS4ALL. Doordat XS4ALL nooit van naam of URL is veranderd en veel trouwe klanten kende, is een indrukwekkend deel van de homepages uit de periode 1994-2001 bewaard gebleven. De KB spoort deze homepages op en archiveert een representatief deel.

### Prachtige historische bron

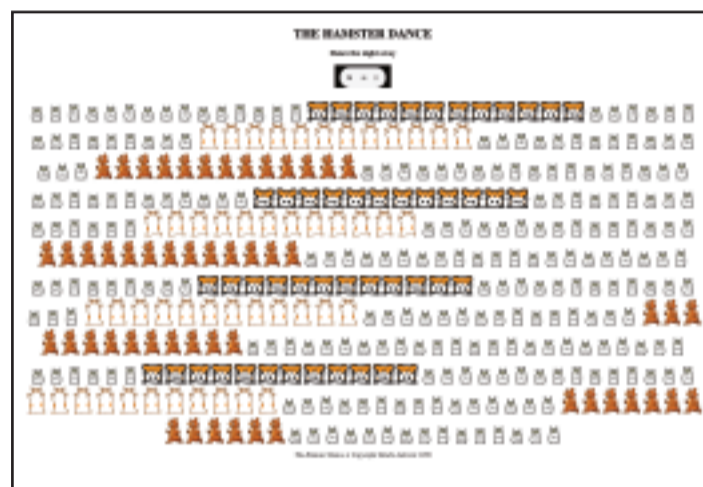
De vroegste homepages worden ook wel 'webincunabelen' of 'webwiegedrukken' genoemd: digitale publicaties met de kenmerken van papieren drukwerk. Het waren wilde experimenten met vormgeving en navigatie, tegelijkertijd beperkt door de technische mogelijkheden als opslag en internetsnelheid. De sites weerspiegelden ook alle modetrends en technische ontwikkelingen van de jaren negentig. Ze zijn daarom een prachtige historische bron van het ontstaan en de verandering van de digitale innovatie en digitaal geboren cultuur in Nederland.

### Zo authentiek mogelijk

Tot nu toe zijn deze sites niet actief verzameld door een Nederlandse erfgoedinstelling. Ook is een groot deel niet gearchiveerd door het Internet Archive, omdat ze maar beperkt verbonden waren met de rest van het internet door middel van hyperlinks en niet zijn gevonden door crawlers. Voor 2007 werden alleen websites van politieke partijen gearchi-

Archiveren van oertijd Hollandse web

# Koninklijke Bibliotheek redt homepages XS4ALL



Ongeveer 3.000 XS4ALL-homepages dateren van voor de eeuwwisseling. Drie digitale topstukken op een rij: de Stone-Age Computer (1993), het virtuele huis van Liesbet (1995) en de Nederlandse versie van de Hamster Dance, een van de eerste internetmemes. Ook de mijlpaal van 1000ste geregistreerde homepage bij XS4ALL (1996) wordt gearchiveerd.

veerd door het Documentatiecentrum Nederlandse Politieke Partijen in Groningen. Sinds 2007 houdt de KB zich bezig met webarchivering: het bewaren van cultuurhistorisch be-

langrijke websites voor toekomstig onderzoek. De KB archiveert de sites zo authentiek mogelijk, beschrijft ze in detail en probeert zoveel mogelijk historische contextinformatie

van de digitale topstukken terug te vinden. Het veiligstellen van een deel van dit erfgoed kan dankzij financiële bijdragen van het SIDN Fonds en Stichting Internet4all. *kb.nl*

EHRI brengt onderzoekers en bronnen samen

## Platform voor Holocaustonderzoek

**EHRI, een Europees platform met informatie over de Holocaust, biedt naast online toegang tot 325.000 beschrijvingen ook trainingen, workshops, conferenties en meer.** *René van Horik*

Het Joodse echtpaar Fientje en Jaap Weyel dook, om deportatie te voorkomen, vanaf september 1942 ruim twee en een half jaar onder in Aalten vlakbij de Duitse grens. Hun dagboek vormt een belangrijke historische bron voor onderzoek naar aspecten van de Jodenvervolgving en de Holocaust. Dit dagboek vormt een voorbeeld van online beschikbaar gesteld materiaal via EHRI.

### Duizenden bestanden

EHRI staat voor European Holocaust Research Infrastructure. EHRI maakt het voor onderzoekers mogelijk om vele duizenden bestanden te gebruiken voor Holocaust-studies. EHRI kwam van 2010 tot 2020 tot stand, gefinancierd door de EU. De ontwikkeling van deze gemeenschappelijke toegang tot archiefcollecties vereiste ontwikkeling en toepassing van richtlijnen en standaarden

op gebieden zoals archiefbeschrijvingen, termen en begrippen, gebruikslicenties en vertalingen. Afspraken over en toepassing van dataharmonisatie en -autorisatie speelden ook een belangrijke rol. Nu, tien jaar later, is EHRI onderdeel van de EU-roadmap van onderzoeksinfrastructuur en is het de betrokken organisaties uit meer dan 20 landen gelukt om naast het portaal ook trainingen, workshops, conferenties en fellowships neer te zetten.

Het belang van Holocaustonderzoek en -herdenking wordt nog steeds internationaal erkend. Zo heeft de Europese Raad in december 2018 een verklaring aangenomen over de strijd tegen het toenemende antisemitisme en de ontwikkeling van een gemeenschappelijke veiligheidsaanpak om de joodse gemeenschappen en instellingen beter te beschermen.

Het NIOD is een van de 161 Nederlandse instel-

#### blog.ehri-project.eu

Op [blog.ehri-project.eu](http://blog.ehri-project.eu) staan tientallen voorbeelden van toepassingen van digitale technieken en gebruik van gedigitaliseerde archiefbronnen om bestanden te digitaliseren, te documenteren en online beschikbaar te stellen.



Het echtpaar Weyel hield onder andere het aantal ondergedoken dagen bij. Credits: NIOD

lingen en een van de ruim 2.165 archiefinstellingen uit bijna 60 landen van wie 325.000 archiefbeschrijvingen toegankelijk zijn gemaakt via de EHRI-portal. Het dagboek van het echtpaar Weyel is hier een van. Onder leiding van het NIOD wordt de komende periode gewerkt aan een permanente onderzoeksinfrastructuur op het gebied van Holocaust-studies. *ehri-project.eu*

## Lancering online stappenplan voor FAIR software

De nieuwe online handleiding [fair-software.nl](http://fair-software.nl) helpt de Nederlandse onderzoeksgemeenschap met vijf aanbevelingen om software en code FAIR te maken. Zo zijn specifieke keuzes nodig bij het selecteren van een repository, het toevoegen van een licentie of het registreren van de software. De site is een initiatief van het Nederlands eScience Center en DANS, specialisten op het gebied van respectievelijk software en data en onderdeel van de Nederlandse nationale e-infrastructuur voor onderzoek. Ook andere partijen zijn welkom om zich bij dit initiatief aan te sluiten. (HB) *fair-software.nl*

#### Rectificatie

In E-data & Research van oktober 2019 staat het Nationaal Archief als knooppunt van het Netwerk Digitaal Erfgoed abusievelijk niet vermeld. Op [edata.nl](http://edata.nl) is een nieuwe versie geplaatst.

## SINDS KORT BESCHIKBAAR

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

## CentERdata

## • Redenen om minder vlees te eten

In juli en augustus 2018 hebben Wiebke Bleidorn en Chris Hopwood (UC Davis) het onderzoek 'Reasons to eat less meat' afgenomen in het LISS panel. De vragenlijst is ingevuld door 5.742 respondenten. Aan de panelleden werd gevraagd wat voor hen redenen zouden kunnen zijn om minder vlees te eten.



Credits: Gerhard Gellinger via Pixabay

Deze redenen konden te maken hebben met de eigen gezondheid, het effect op het milieu en het lijden van dieren. De resultaten lieten zien dat gezondheid de meest voorkomende reden is waarom mensen zouden overwegen om minder vlees te eten. De data van deze studie zijn beschikbaar via LISS Data Archive.

[lissdata.nl](http://lissdata.nl)

## Ook sinds kort beschikbaar:

## Studies LISS panel

- He, J., maart 2012, Response styles
- Stasiukynaite, R.; Adriaens, H.; Melenberg, B., juli 2012, Estimation of a behavioral model within an agent-based model of a financial market
- Thomaes, S.; Wildschut, T.; Sedikides, C., april 2014, Young Children Longing About Their Past: An Initial Investigation of the Developmental Origins of Nostalgia
- Van der Werf, M.; Doolaard, F.; Hilbert, L., april 2018, Social exclusion, control and financial scarcity
- Voorpostel, M.; Yahirun, J.; Lesnard, L.; Nazio, T., mei 2014, Attitudes Towards Alternative Partnership Arrangements

- Hulshof, D., december 2017, WTP Passenger Car Survey
- Vlasblom, J.D., november 2017, Opportunities in the labor market



Deze bestanden zijn kosteloos beschikbaar via [lissdata.nl](http://lissdata.nl). Bezoek deze site of scan de QR-code.

## DANS

## • Ook in EASY: Vrijwilligers

Door de Verenigde Naties is 5 december uitgeroepen tot 'Internationale Vrijwilligersdag', een dag om wereldwijd vrijwilligerswerk te promoten en om respect te tonen voor vrijwilligers. EASY bevat verscheidene datasets over vrijwilligers en hun werk. Zoals de dataset over 'Arbeidsomstandigheden in het vrijwilligerswerk' (DOI: 10.17026/dans-xjk-ryj2), waarin gegevens te vinden zijn over het beleid van organisaties voor vrijwilligerswerk met betrekking tot arbeidsomstandigheden. Een ander voorbeeld is de dataset 'Mantelzorg 2001: Informele hulpverlening', waarin gegevens te vinden zijn over mantelzorg en thuiszorg bij relatief serieuze zorgsituaties. DOI: 10.17026/dans-xrc-8bk4

## Ook sinds kort beschikbaar:

- De volgende datasets zijn open access beschikbaar via het online archiverings-systeem EASY van DANS:
- Dorst, M.C. (Gemeente Dordrecht) (2019): Lange Geldersekeade, gemeente Dordrecht. Een archeologische opgraving van de middeleeuwse stadsmuur en de St. Pontiaanstoren. DANS. DOI: 10.17026/dans-xak-9gwc.
  - Nunome, M.N. (Nagoya University) (2019): Assessment of genetic diversity and genetic relationships of farm quail populations using microsatellite DNA markers.

- DANS. DOI: 10.17026/dans-xf9-wkpg.
- Hendrix, R.H.J. (Maastricht University Medical Centre) (2019): A view on critical oxygen delivery values during cardiopulmonary bypass. DANS. DOI: 10.17026/dans-xw4-a6kc.
  - Dijk, dr. M. van (Wageningen Economic Research, Wageningen University & Research) (2019). FOODSECURE Scenario Driver Database. DANS. DOI: 10.17026/dans-zeh-fd4m.
  - Oosten, dr. R.M.R. van (Leiden University) (2019): Leiden's Pubscape, 1820-1894. Towards a better understanding of drink and labour. DANS. DOI: 10.17026/dans-z4q-8c2v.
  - Jeugd, dr. H.P. van der (NIOO-KNAW) (2019): Ring-recovery data of starlings in the Netherlands. DANS. DOI: 10.17026/dans-x87-3wu5.
  - Bouwens, ir. C.J.C. (NEPROM) (2019): woonbelevingsonderzoek bij bewoners van ZEN nieuwbouwwoningen. DANS. DOI: 10.17026/dans-z7k-vne5.
  - Summers (National Institutes of Health Clinical Center RN) (2018): PNEUMONIA DETECTION IN X-RAYS. DANS. DOI: 10.17026/dans-x5u-3zdf.
  - Goldbeck, F.G. (University Hospital Tübingen) (2019): Vitality in inpatients with depression. DANS. DOI: 10.17026/dans-zu7-rv3g.
  - Sociaal en Cultureel Planbureau - SCP (2013): Krimp in de Kinderopvang - KIK2013 v1. DANS. DOI: 10.17026/dans-z7a-dp9u.
  - Hudson, Dr. C. (Advanced Wellbeing Research Centre, United Kingdom) (2019): The influence of meteorology and air quality on parkrun athletic performance Meteorology, air quality and parkrun. DANS. DOI: 10.17026/dans-27m-9mkm.
  - Schöpfel, Dr. J. (GERiiCO Laboratory, University of Lille) (2019): Going Green: Publishing Academic Grey Literature in Laboratory Collections on HAL. DANS. DOI: 10.17026/dans-xsq-bus7.



Via [easy.dans.knaw.nl](http://easy.dans.knaw.nl) zijn deze bestanden beschikbaar. Bezoek deze site of scan de QR-code.

## AGENDA

## 11 - 12 maart • Berlijn

## Conferentie Open Science

Een uniek forum voor onderzoekers, bibliothecarissen, uitvoerders, beleidsmakers en andere belanghebbenden over Open Science-ontwikkelingen. [open-science-conference.eu](http://open-science-conference.eu)

## 16 - 19 maart • Groningen

## Conferentie Nederland Digitaal

Een conferentie gewijd aan het ontwikkelen en behouden van digitaal talent: van digitale vaardigheden naar soft skills tot innovatieve research & development en toponderzoek. [nederlanddigitaal.nl](http://nederlanddigitaal.nl)

## 17 - 18 maart • Groningen

## ICT.OPEN

Een event voor ICT-onderzoek in Nederland. [ictopen.nl](http://ictopen.nl)

## 18 - 20 maart • Melbourne

## 15e RDA bijeenkomst

Het thema van deze RDA-bijeenkomst is: Data for Real-World Impact. [rd-alliance.org/plenaries](http://rd-alliance.org/plenaries)

## 24 maart • Den Haag

## Voer voor Psychologen

Een event speciaal voor onderzoekers in de psychologie, met tools, tips en tricks voor RDM en data delen. [dans.knaw.nl/nl/actueel/agenda/voer-voor-psychologen](http://dans.knaw.nl/nl/actueel/agenda/voer-voor-psychologen)

## 26 - 30 mei • Barcelona

## QQML 2020

Twaalfde editie van de internationale conferentie Qualitative and Quantitative Methods in Libraries. [qqml.org/](http://qqml.org/)

## 27 - 31 mei • Gothenburg

## Jaarlijkse IASSIST-conferentie

Dit jaar met het thema: Data by Design: Building a Sustainable Data Culture. [iassist2020.org](http://iassist2020.org)

## 22 juni - 17 juli • Michigan

## ICPSR Summer Program

ICPSR organiseert ieder jaar een summer school, DANS stelt subsidies beschikbaar. [dans.knaw.nl](http://dans.knaw.nl)

## Unieke samenwerking Nationaal Archief en Kadaster

## Onteigening WOII zichtbaar door koppeling data

Tegen welke prijs werd onroerend goed van Joodse gedeporteerden tijdens de Tweede Wereldoorlog onteigend? 75 jaar na dato vinden nabestaanden én onderzoekers sneller antwoord op deze vraag.

Tim de Haan en Ed de Heer

Het Nationaal Archief beschikt over de dataset van de Verkaufsbücher; een administratie van de ogenschijnlijk legitieme verkoop van onroerend goed door de Duitse bezetter. Het bevat ruim 7.000 transacties en meer dan 13.000 adressen van eigenaren, kopers en verkopers. Door koppeling van deze dataset aan kadastrale informatie uit de Basisregistratie Adressen en Gebouwen (BAG) van het Kadaster, worden de destijds ont-



Per pand is bijgehouden wie de eigenaar was met zijn of haar woonadres. Naam en adresgegevens van de koper van het pand zijn ook geregistreerd.

eigende objecten letterlijk op de kaart gezet. Dit maakt inzichtelijk waar en op welke schaal in Neder-

land sprake was van deze roof. Binnen de digitale labomgeving van het Kadaster, Publieke Dienstverle-

ning Op de Kaart (PDOK), heeft een team van linked-open-data-specialisten in vier dagen een datastory ontwikkeld. Op een HTML-kaart zijn de verschillende panden geplot. Door op een marker te klikken, worden details uit de Verkaufsbücher en kadastrale details getoond, zoals oppervlakte van het onroerend goed en de gemiddelde WOZ-waarde. Ook is een koppeling met Google Streetview gemaakt. De gemaakte statistieken geven inzicht in ruim 7.000 transacties en beantwoordt vragen als: waar zijn de panden verkocht, kwamen de kopers uit dezelfde stad en hoeveel leverde de verkoop op? De datastory is te raadplegen op [labs.kadaster.nl/stories/verkaufsbuecher/](http://labs.kadaster.nl/stories/verkaufsbuecher/)

## Herstelbetalingen

Het onderwerp is actueel. Zo heeft

de gemeente Den Haag onlangs 2,6 miljoen euro gereserveerd voor schadevergoedingen aan Joodse slachtoffers die na de oorlog een aanslag kregen om alsnog hun erf-pacht te betalen. Met de gekoppelde datasets worden dit soort berekeningen makkelijker. In de gemeente Utrecht zijn bijvoorbeeld 81 panden met een oppervlakte van 13.559 m<sup>2</sup> verkocht. Bij 100 euro per vierkante meter is dat ruim 1,3 miljoen euro, te verdelen over eigenaren en nabestaanden.

[nationaalarchief.nl/onderzoeken/open-data/open-data-indexen](http://nationaalarchief.nl/onderzoeken/open-data/open-data-indexen)

Tim de Haan is medewerker Openbaarheid & Open Data en Ed de Heer is projectleider/adviseur Linked Open Data bij het Nationaal Archief.

Sharing is caring: onderzoeksdata delen als norm

# ‘DANS biedt schatkamer aan online onderzoeksinformatie’

**Data verdienen het om be-  
waard te worden. Want alleen  
door data te delen, kunnen ook  
andere onderzoekers kennis  
uit de data halen. Precies  
voor dat doel is DANS 15 jaar  
geleden opgericht.**

*Heidi Berkhout*

“Volgens Re3data.org, een wereldwijd register van repositories voor onderzoeksgegevens, telt ons land momenteel bijna 60 research data repositories, databases en metadatacatalogi. Dat is geweldig.” Aan het woord is Peter Doorn, tot 1 april directeur DANS. “Deze hoeveelheid databewaarplaatsen geeft onderzoekers altijd de mogelijkheid om data te delen. Maar er is een keerzijde; de meeste repositories uit de lijst garanderen geen lange termijn duurzaamheid van de opgeslagen data. En dat is nu juist van het grootste belang.”

## Wekkerfunctie

Doorn: “Er is de afgelopen 15 jaar veel veranderd. In 2005 was de bestuurlijke interesse in het toegankelijk maken van onderzoeksdata nog heel beperkt. Universiteiten voerden geen beleid daarover. Het idee van Open Access publicaties was wel al geboren, zo rond 2002. Maar de waarde van research data drong pas goed door in de jaren 2010. De term ‘Big Data’ stamt net als DANS uit 2005, maar was aanvankelijk alleen in beperkte kring een hot item. Voormalig vicevoorzitter van de EU Neelie Kroes was één van de eerste beleidsmakers die de waarde van onderzoeksdata inzag en breed verkondigde. Ze noemde het ‘het nieuwe goud’ en de ‘olie voor de wetenschap’; nu zijn dat afgesleten clichés. Tegenwoordig hebben alle Nederlandse universiteiten een afdeling Research data support en hanteren KNAW en NWO heldere principes voor databeheer en databeschikbaarheid. Ik vrees trouwens dat ook de datafraude-gevallen in het afgelopen decennium een wekkerfunctie hebben gehad.”

## Vertienvoudiging

In vijftien jaar is DANS uitgegroeid tot hét Nederlandse instituut voor permanente toegang tot digitale onderzoeksgegevens. DANS zorgt ervoor dat onderzoekers en dataprofessionals onderzoeksgegevens vindbaar, toegankelijk, interoperabel en herbruikbaar kunnen maken. Zo biedt DANS naast trainings- en advies-trajecten diverse diensten: DataverseNL, EASY en NARCIS. Doorn: “In 2019 groeiden deze diensten sneller dan ooit tevoren. Via

### Maak kans op subsidie

Ook data uit kleine projecten kunnen heel waardevol zijn voor hergebruik. Door de KDP-subsidies van DANS wordt dit mogelijk gemaakt, het afgelopen jaar zijn weer tien subsidies verstrekt. Informatie over de subsidieronde 2020 staat op de website van DANS.



*De directie van DANS: Ingrid Dillo (links), Peter Doorn (tot 1 april aanstaande, rechts) en Henk Harmsen (niet op de foto). Credits: Bart van Vliet*

## INTERVIEW

**‘Het verleden is  
en blijft een  
gigantische bron  
voor de toekomst’**

EASY (CoreTrustSeal-gecertificeerd) zijn inmiddels ruim 120.000 onderzoeksdatasets duurzaam opgeslagen, klaar voor hergebruik door diverse toegangscategorieën en persistent identifiers. Verdeeld over wetenschappelijke discipline zijn dit vooral archeologische datasets en datasets uit de geesteswetenschappen en sociaal-culturele wetenschappen. Bijna 40.000 datasets zijn in 2019 gedownload, een vertienvoudiging in tien jaar tijd.

NARCIS is dé nationale portal voor wie informatie zoekt over wetenschappers en hun werk. NARCIS biedt nu toegang tot 260.000 datasets van 18 Nederlandse en 2 internationale repositories en ook toegang tot bijvoorbeeld meer dan 2 miljoen digitale publicaties van 37 Nederlandse onderzoeksinstituten, waarvan 765.000 open access beschikbaar. DataverseNL biedt onderzoekers de mogelijkheid om onderzoeksgegevens al tijdens onderzoek op te slaan, te delen en te publiceren, het platform bevat bijna 1.300 datasets en telde 37.500 downloads.”

### Schat aan datasets

“De ransomware-aanval op de Universiteit Maastricht in de kerstperiode 2019 laat zien hoe rampzalig het is als onderzoekers niet meer bij hun data kunnen. Daarnaast geeft

het publiceren van data ook iets extra’s aan onderzoek, aan de invulling van de rol als wetenschapper en aan iemands zichtbaarheid – wat mede van belang is om nieuwe projecten gefinancierd te krijgen. Door data via een archief te delen ofwel te publiceren, wordt het onderzoeksproces beter verifieerbaar.” De bij DANS opgeslagen datasets zijn enorm divers. Zo zijn bijvoorbeeld de ervaringen van Nederlandse oorlogsveteranen digitaal vastgelegd in audiobestanden met bijbehorende transcripten. Als men wil weten wat het weer was bij Kaap de Goede Hoop op 6 januari 1763: het staat in een van de gearchieverde scheepsjournaals uit de periode 1750 tot 1850. Ook alle Nationale Kiezersonderzoeken zijn beschikbaar. Verder is veel archeologische data beschikbaar, zoals digitale kaarten van vindplaatsen en scans van analoge veldtekeningen, objectfoto’s en veldfoto’s en tabellen met determinaties van vondsten. 224 jaar na de eerste volkstelling in Nederland zijn er meer dan twee eeuwen aan data over bevolkingsgrootte, leeftijden, geslacht en beroepsgroepen online te raadplegen. En in 2019 stelde het Amsterdam UMC als eerste ziekenhuis in Europa geanonimiseerde data over intensive-carepatiënten beschikbaar via DANS. Dit zijn slechts enkele voorbeelden van de tienduizenden datasets die via EASY te vinden zijn. Beschikbaar gesteld door wetenschappelijke onderzoekers voor wetenschappelijke

### H. Wals nieuwe directeur DANS

1 april start Henk Wals als nieuwe directeur van DANS. Wals is sociaaleconomisch historicus. Hij studeerde en promoveerde aan de Universiteit van Amsterdam. Wals was eerder directeur Huygens ING en Internationaal Instituut voor Sociale Geschiedenis, ook KNAW-instituten. Peter Doorn blijft als adviseur bij DANS betrokken.

onderzoekers en anderen. Iedereen kan zoeken in EASY ([easy.dans.knaw.nl](http://easy.dans.knaw.nl)) en een groot deel van de datasets downloaden. Soms zijn er bijvoorbeeld privacyredenen die ervoor zorgen dat alleen geregistreerde gebruikers toegang krijgen. Bij de vermelding ‘Open Access’ is de dataset voor iedereen toegankelijk.

### Optimaal hergebruik van data

Ingrid Dillo, adjunct-directeur: “Nationaal en internationaal is het delen van data prioriteit, waarbij een verantwoorde, toegankelijke data-opslag basisvoorwaarde is om data een langere levensduur en verhoogde waarde te geven. Zo financiert de Europese Commissie projecten en infrastructures om de toegankelijkheid van onderzoeksresultaten te bevorderen, zoals: EOSC-hub, OpenAIRE, FAIRsFAIR en FREYA. DANS leidt FAIRsFAIR, een groot Europees project waarin gewerkt wordt aan praktische oplossingen voor het gebruiken van de FAIR-principes tijdens de levenscyclus van onderzoeksgegevens. Een ander voorbeeld van een onlangs gestart project is het project Social Sciences and Humanities Open Cloud (SSHOC), waarin DANS een van de projectpartners is. Binnen het nieuwe Europese financieringsprogramma Horizon Europe (2021-2027) wordt het ‘hergebruiken van onderzoeksdata’ genoemd als een van de speerpunten. Ook het internationaal erkende keurmerk voor repositories CoreTrustSeal is een voorbeeld van internationale samenwerking. In de Wetenschapsvisie van Minister Van Engelshoven (OCW) komt ‘open science’ veelvuldig voor. Ook NWO vindt dat onderzoeksresultaten die zijn betaald uit publieke middelen, wereldwijd vrij toegankelijk moeten zijn. De KNAW onderstreept hetzelfde belang met haar invulling van het thema Open Science.” Met de toename van belang van het delen van data stijgt ook het aantal spelers binnen het Nederlandse datalandschap. Binnen CLARIAH en ODISSEI werkt DANS mee aan een nationale data-infrastructuur voor respectievelijk de geesteswetenschappen en sociale wetenschappen. Onlangs is ook de FAIR Software Route gelanceerd door het eScience Center en DANS, een online stap-voor-stap handleiding voor ontwikkelaars, onderzoekers en datastewards om onderzoekssoftware FAIR te maken.

Doorn sluit af: “Na 15 jaar geef ik het roer aan mijn opvolger Henk Wals. Ik heb er alle vertrouwen in dat hij samen met mij en de medewerkers van DANS een aansprekende en toekomstgerichte visie uitwerkt op de steeds belangrijkere rol van DANS in het complexe datalandschap. Het verleden is en blijft een gigantische bron voor de toekomst.”



### Dit jaar bestaat DANS 15 jaar

Een overzicht van de geschiedenis en activiteiten voor dit jaar staat op de website van DANS: [dans.knaw.nl](http://dans.knaw.nl)

SHARE-data koppelbaar aan CBS-microdata

# Het leven van vijftigplussers

De koppeling tussen data van SHARE Nederland en data van CBS Microdata Services zorgt voor nog meer informatie over het leven van 50-plussers.

Evi de Cock

Geïnteresseerden in onderzoek naar de leefsituaties en veranderingen in het leven van 50-plussers, kunnen gebruik maken van gegevens uit de Survey of Health Ageing and Retirement in Europe (SHARE) databank.

## Uit 28 landen

Adriaan Kalwij, teamleider van SHARE in Nederland: "SHARE is een multidisciplinaire, cross-nationale databank met gegevens van personen die 50 jaar of ouder zijn, en hun partners, over onder meer hun lichamelijke en psychische gezondheid, sociale netwerk van vrienden en familie, en hun sociaal-



De databank van SHARE bevat gegevens over de leefsituatie van 50-plussers in Europa. Credits: brfcs, pixabay

economische situatie (zoals opleidingsniveau, arbeidsmarktparticipatie, inkomen en besparingen)." Om de twee jaar wordt een enquête uitgezet in 28 landen. In de eerste 5 golven zijn de data verzameld door middel van interviews bij de respondenten thuis en in golven 6 en 7 zijn de vragenlijsten in Nederland via het internet afgenomen. Ook

zijn er lichamelijke metingen uitgevoerd zoals handgreepkracht, loop-snelheid, zitten en opstaan en longcapaciteit.

## Voordelen

Sinds kort kunnen SHARE-NL-data gekoppeld worden aan bestanden binnen de CBS Microdata Services faciliteit. Kalwij: "Dit levert voor-

delen voor onderzoekers op. Zo kunnen gegevens die niet bevraagd zijn in de SHARE-enquête, maar die wel beschikbaar zijn in CBS-microbestanden, worden toegevoegd aan de SHARE-NL-data voor onderzoek. Denk bijvoorbeeld aan gezondheidsinformatie zoals het gebruik van medicijnen of informatie over ziekenhuisopnames. Daarnaast kunnen SHARE-respondenten, ook als ze gestopt zijn met SHARE, toch worden gevolgd over de tijd als bijvoorbeeld gegevens over iemands arbeidsmarktparticipatie bekend zijn binnen de CBS-faciliteit." SHARE-data kunnen kosteloos worden gebruikt, aan het gebruik van CBS Microdata Services zijn wel de reguliere kosten verbonden. [cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata](https://cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata)

Neem voor meer informatie contact op met het CBS via [microdata@cbs.nl](mailto:microdata@cbs.nl)

Toestemming voor het gebruik van SHARE-NL-data via de CBS Microdata Services faciliteit kan worden aangevraagd bij dr. Adriaan Kalwij: [A.S.Kalwij@uu.nl](mailto:A.S.Kalwij@uu.nl)

## KORT

### OPPA maakt cross-cultureel onderzoek mogelijk

Open Probability-based Panel Alliance (OPPA) is een samenwerking tussen panels in Nederland (LISS panel), Duitsland (GESIS panel), de Verenigde Staten (Understanding America Study) en Zuid-Korea (KAMOS). Deze samenwerking maakt het mogelijk om cross-cultureel onderzoek uit te voeren volgens vergelijkbare methodologische standaarden. In 2018 vond de OPPO Showcase Study-competitie plaats, waarbij onderzoekers paneltijd konden winnen. Winnaars Liliya Leopold en Giseline Kuipers (Universiteit van Amsterdam) en Norbert Schwarz (University of Southern California) hebben vorig jaar hun vragenlijst gratis laten afnemen in de vier panels. De data komen in 2020 open access beschikbaar voor niet-commercieel wetenschappelijk onderzoek via OPPO. Op de website van OPPO staat meer informatie, ook over de kosten. (EdC)

[openpanelalliance.org](https://openpanelalliance.org)

Vragenlijstenbank over Nederlandse taal en cultuur online

# 180.000 antwoorden op 15.000 vragen

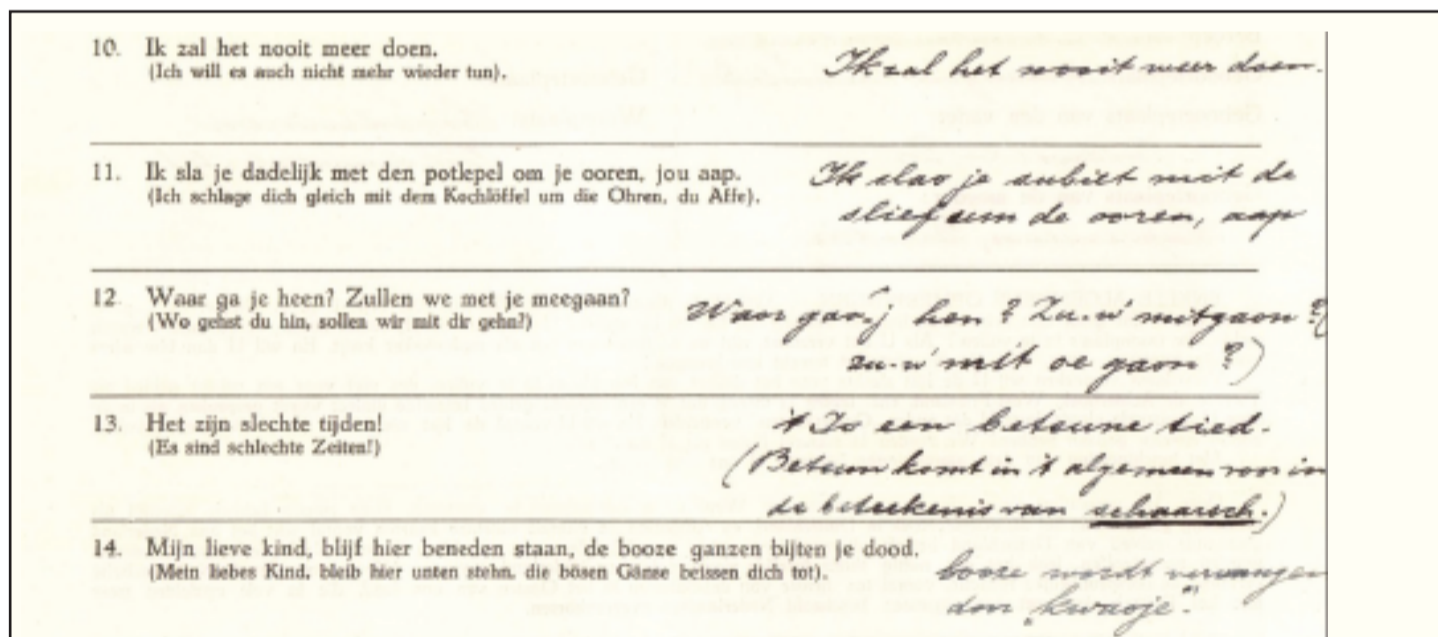
Hoe zijn de Nederlandse taal en cultuur in de 20e eeuw veranderd? De Vragenlijstenbank van het Meertens maakt het mogelijk dat te achterhalen.

Nicoline van der Sijs

Vanaf 1931 zijn door het Dialectenbureau, de voorganger van het Meertens Instituut, vragenlijsten rondgestuurd naar informanten verspreid over Nederland en later ook Vlaanderen. Het doel was de variatie in de Nederlandse taal en cultuur in kaart te brengen. De informanten moesten antwoorden geven op vragen als 'Hoe noemt men de doeken waarin de zuigelingen gewikkeld worden?' of 'Hechte men er belang aan of de eerste die men met nieuwjaar ontmoette een man of een vrouw was? In welk opzicht?'

## Van archiefdoos naar scan

Tussen 1931 en 2005 werden 227 vragenlijsten uitgestuurd met ruim 15.000 vragen over taalgebruik en culturele gewoontes. Het leverde 180.000 handgeschreven antwoorden op. Die antwoorden werden in meer dan 3.000 archiefdozen gestopt, keurig gesorteerd op vraag en plaats. In de jaren tachtig van de vorige eeuw werden de antwoorden op 148 vragenlijsten op microfiches gezet, en in de daaropvolgende jaren werden binnengekomen antwoorden van 16 vragenlijsten gescand, zodat in totaal de antwoorden van 164 vragenlijsten zijn gedigitaliseerd. Maar tot nu



Vragenlijst 3 uit 1934: de informanten werd gevraagd de zinnestjes in hun dialect te vertalen.

Credits: Vragenlijstenbank, samenvatting Nicoline van der Sijs, 2020, [meertens.knaw.nl/vragenlijstenbank](https://meertens.knaw.nl/vragenlijstenbank)

toe was alle informatie alleen beschikbaar binnen het Meertens Instituut.

## Vragen online doorzoekbaar

Sinds kort staat alle informatie online op de Vragenlijstenbank. Dit portaal bevat een overzicht van alle vragen die ooit zijn uitgezonden. Als je op 'hond' zoekt, vind je alle vragen waarin 'hond' voorkomt, als taalkundig begrip of in de volksgebruiken ('Welke leefwijze moet de zwangere vrouw in acht nemen? Bijvoorbeeld geen hond schoppen.'). Bij 25 vragenlijsten zijn per vraag scans van de handgeschreven antwoorden toegevoegd, dit betreft al 127.799 scans en dit aantal zal alleen

nog maar toenemen. De persoonsgegevens waarmee ieder antwoord begint, zijn automatisch onleesbaar gemaakt om te voldoen aan de AVG.

In de ideale wereld zijn ook de antwoorden digitaal doorzoekbaar. Enthousiaste vrijwilligers hebben hiertoe in het verleden de antwoorden van zes vragenlijsten overgetikt. Die overgetikte antwoorden zijn toegevoegd aan de Vragenlijstenbank. Via het VeleHandenplatform worden momenteel de antwoorden van nog eens 16 vragenlijsten overgetikt binnen het project 'Meertens-vragenlijsten Taal en Cultuur (1931-2005)'. Ook deze antwoorden worden toegevoegd. Zo kan straks ieder-

een genieten van de antwoorden op vragen als 'Geloof men, dat de heksen zich in ooevaars kunnen veranderen? Welke verhalen kent U hierover?'

[meertens.knaw.nl/vragenlijstenbank](https://meertens.knaw.nl/vragenlijstenbank)

Nicoline van der Sijs is projectleider bij het Meertens Instituut.

## Ook vrijwilliger worden?

Kijk voor meer informatie over het crowdsourcingproject 'Meertens-vragenlijsten Taal en Cultuur (1931-2005)' op: [velehanden.nl/projecten/bekijk/details/project/mei](https://velehanden.nl/projecten/bekijk/details/project/mei)

Wereldwijd meest complete online archief

# Natuurhistorie Indonesië online

Onderzoekers in het project 'Making Sense of Illustrated Handwritten Archives' maken een complex negentiende-eeuws onderzoeksarchief toegankelijk met behulp van kunstmatige intelligentie en semantische technologie.

Maarten Heerlien

Tussen 1820 en 1850 bracht de Natuurkundige Commissie voor Nederlands-Indië de flora, fauna en natuurlijke hulpbronnen van de Indonesische archipel in kaart. De onderzoekers stuurden duizenden zoölogische en botanische objecten terug naar Nederland en produceerden een omvangrijk corpus van aantekeningen en publicaties. Naturalis Biodiversity Center in Leiden, beheerder van de collectie, liet het geschreven corpus tussen 2008 en 2011 digitaliseren. Toch bleef het vanwege de complexiteit van de handschriften grotendeels ontoegankelijk.

## Waardevolle inblik

"Zonde," zegt Andreas Weber, universitair docent Science, Technology and Policy Studies (STePS) aan de Universiteit Twente, "de veld-



Pagina uit een bundel veldaantekeningen met een beschrijving en afbeelding van de vleermuisensoort *Rhinolophus larvatus*.

Credits: Naturalis Biodiversity Center, archief van de Natuurkundige Commissie voor Nederlands-Indië. Rechten: Public Domain Mark 1.0.

aantekeningen en dagboeken geven een waardevolle inblik in de dagelijkse praktijk van historisch natuuronderzoek. Bovendien vormt de collectie van de Natuurkundige Commissie waarschijnlijk wereldwijd het meest complete archief van de negentiende-eeuwse biodiversiteit van Indonesië, een belangrijke biodiversiteitshotspot."

Weber is als PostDoc verbonden aan 'Making Sense of Illustrated Handwritten Archives'.

In dit NWO Creatieve Industrie-project, met cofinanciering van Brill Scholarly Publishing, maken Naturalis, de universiteiten van Leiden, Groningen en Twente en Brill de handschriften van de Natuurkundige Commissie doorzoekbaar en koppelbaar aan de verzamelde objecten. Dit doen ze met behulp van automatische handschriftherkenning en semantische annotatie. Een uitdaging volgens Weber: "Dergelijke natuurhistorische archieven zijn

complexe visuele landschappen waar gerelateerde informatie op verschillende plekken en in verschillende vormen en talen te vinden is. Bijvoorbeeld als tekeningetjes, uiteenlopende notatiewijzen, enzovoort." Als oplossing wordt MONK gebruikt, een geavanceerde ICT-architectuur voor automatische handschriftherkenning, ontwikkeld door het team van Lambert Schomaker, hoogleraar Kunstmatige Intelligentie aan de Rijksuniversiteit Groningen. Weber: "MONK zoekt naar visuele patronen in de vorm van woorden en andere beeldelementen. Door bovendien een betekenis aan een woordbeeld te koppelen, leert het systeem woordklassen te identificeren. Daar kan het systeem bij vergelijkbare nieuwe manuscripten weer gebruik van maken."

## Koppeling

De koppeling tussen collectieobjecten en gerelateerde informatie in de handschriften wordt verder gefaciliteerd door de Semantic Field Book Annotator, een tool waarvoor Lise Stork, PhD bij het Leiden Institute of Advanced Computer Science (LIACS) in oktober 2019 de Young eScientist Award ontving. Eind 2020 lanceert Brill het platform Natuurkundige Commissie Online, waarin het archief open doorzoekbaar zal zijn gemaakt. Een afsluitend congres vindt plaats van 22 tot en met 24 november bij Naturalis.

[makingsenseproject.org](http://makingsenseproject.org)

Machine Learning toegepast op Citizen Science-data

# GeoHealth-onderzoek zet data op de kaart

Hoe groot is het risico op een tekenbeet? En wat is de impact van extreme weergebeurtenissen? In dit artikel aandacht voor het GeoHealth-onderzoek van jong talent Irene Garcia-Martí. *Marion Wittenberg*

## JONG TALENT

### 'Burgerobservaties combineren met omgevingsvariabelen'

Elk jaar worden meer dan 1 miljoen Nederlanders gebeten door een teek, van dit aantal worden ongeveer 27.000 mensen besmet met de ziekte van Lyme. Bij 1.000 tot 2.500 mensen leidt dit tot ernstige en langdurig klachten. Data scientist Irene Garcia-Martí promoveerde 27 september jongstleden bij de Universiteit van Twente op het proefschrift 'Mapping tick dynamics and tick bite risk using data-driven approaches and volunteered observations'.

## Voorspellen

Garcia-Martí combineerde burgerobservaties met een breed scala aan omgevingsvariabelen (weergegevens, remote sensing, officiële geo-data) en ontwikkelde door middel van machine learning, modellen die tekenactiviteit en het risico op tekenbeten op nationaal niveau kunnen voorspellen. De burgerobservaties zijn afkomstig van de Natuurkalender ([naturetoday.com](http://naturetoday.com)), een fenologisch waarnemingsnetwerk, waaraan burgers al sinds 1868 waarnemingen in de natuur doorgeven. In 2006

startte de Natuurkalender met het registreren van tekenbeten. Door de grote hoeveelheid aanmeldingen is in 2012 voor deze registratie een aparte website gelanceerd: [tekenradar.nl](http://tekenradar.nl). Daarnaast hebben tussen 2006 en 2016 vrijwilligers teken gevangen op 12 locaties in Nederland, gecoördineerd door onderzoekers van de Universiteit van Wageningen. Momenteel zijn Garcia-Martí's modellen nog statisch, een vervolgstap is om ze dynamisch te maken, waarmee fluctuaties in het tekenrisico door het jaar heen te voorspellen zijn.

## Blogposts over onderzoek

Inmiddels werkt Garcia-Martí bij het Koninklijk Nederlands Meteorologisch Instituut (KNMI), waar ze onderzoek doet naar de impact van extreme weergebeurtenissen. Ook voor haar huidige werk gebruikt ze burgerobservaties verzameld via de weerstations van



Na haar promotie zijn de gegevens en code van Garcia-Martí gearchiveerd in EASY zodat ook anderen kunnen bestuderen hoe de modellen zijn ontwikkeld. Het was de 100.000ste dataset in EASY (DOI: [10.17026/dans-zre-tggd](https://doi.org/10.17026/dans-zre-tggd)). Om dit te vieren, stuurde DANS haar een taart. Op de foto van links naar rechts: Marga Koelen (research support coördinator), Tom Veldkamp (decaan ITC Faculteit), beide van de Universiteit Twente en Irene Garcia-Martí. Credits: John Horn

het Weather Observations netwerk WOW-NL ([wow.knmi.nl](http://wow.knmi.nl)). Garcia-Martí en haar collega's onderzoeken hoe de data van WOW-NL kunnen helpen bij het bestuderen van lokale fenomenen zoals stedelijke hitte-eilanden. Belangrijk, omdat het juist de steden zijn die bij de huidige wereldwijde temperatuurstij-

ging warmer worden, wat gezondheidsproblemen van de bevolking met zich meebrengt. Irene Garcia-Martí houdt een blog bij over haar onderzoek op [irenegarciamarti.com](http://irenegarciamarti.com). Haar proefschrift is beschikbaar via: [library.itc.utwente.nl/papers\\_2019/phd/garciamarti.pdf](http://library.itc.utwente.nl/papers_2019/phd/garciamarti.pdf)

Platform voor transparantie in machine learning

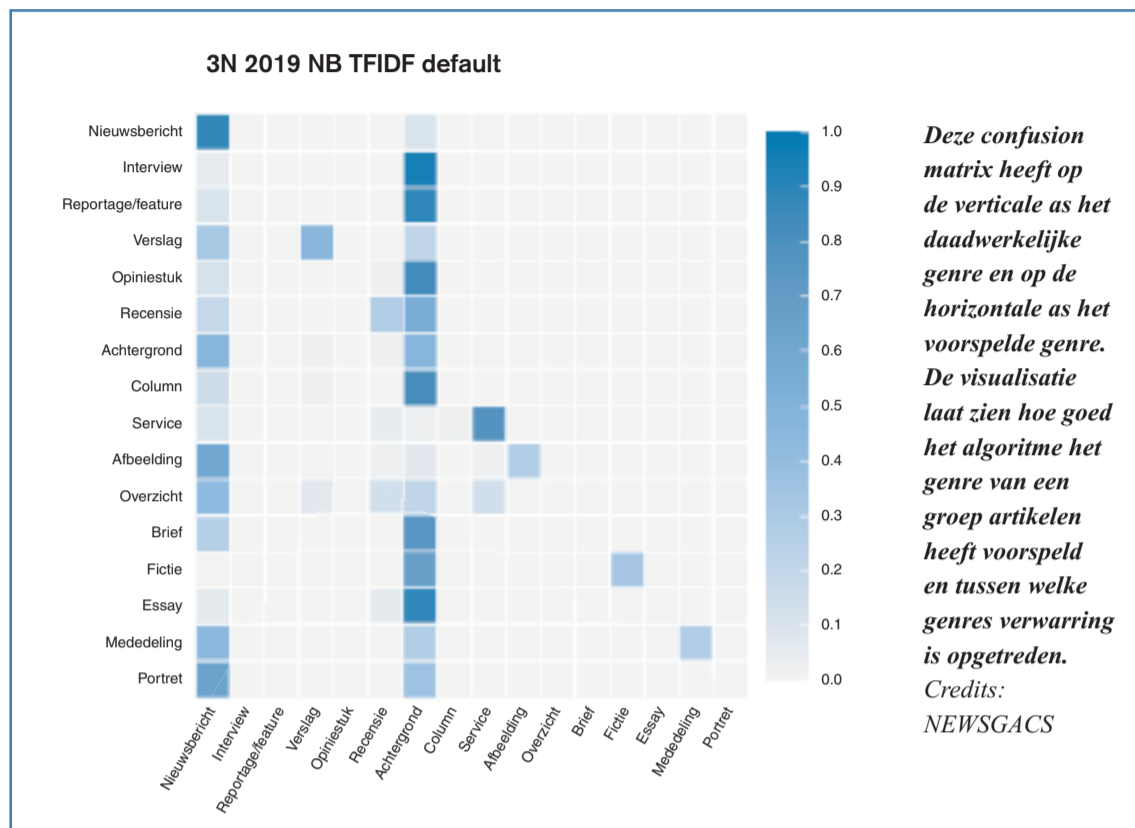
# NEWSGAC biedt een kijkje onder de motorkap

Met digitale methoden kan op grote schaal onderzocht worden hoe journalistiek zich in de twintigste eeuw heeft ontwikkeld. Het beste algoritme daarvoor kiezen, is echter een uitdaging.

Kim Smeenk

Digitale krantenarchieven maken het tegenwoordig mogelijk om op grote schaal onderzoek te doen naar ontwikkelingen in verslaggeving. Het NEWSGAC-project brengt journalistieke ontwikkelingen in beeld door de genres van krantenartikelen te classificeren. Een genre is bijvoorbeeld 'reportage', waarin verslaggeving op locatie belangrijk is. Een ander voorbeeld is 'opiniërende essay', een genre met de nadruk op de mening van de schrijver. Het classificeren gebeurt met machine learning. Onderzocht wordt of de journalistiek in de twintigste eeuw inderdaad steeds minder politiek en opiniërend werd, zoals vaak wordt aangenomen.

NEWSGAC wil niet alleen deze inhoudelijke vraag beantwoorden, maar ook een platform voor transparantie in machine learning bieden. Automatisch genres toekennen aan krantenartikelen is echter geen vanzelfsprekende klus. Tijdens de eerste experimenten bleken twee zaken van essentieel belang: algoritmes moeten met elkaar vergeleken kunnen worden en de onderliggende



opties en aannames moeten inzichtelijk zijn. NEWSGAC maakt via visualisaties inzichtelijk hoe algoritmes kiezen, welke 'features' belangrijk zijn bij de classificatie van een artikel.

## Onderwerp of genre?

Deze aanpak blijkt succesvol. Niet alleen kan de nauwkeurigheid van algoritmes vergeleken worden, ook de vraag of ze hun keuzes maken op de juiste gronden kan worden beantwoord. De meeste algoritmes

blijken bijvoorbeeld vooral goed in het onderscheiden van onderwerpen en minder in het onderscheiden van genres, terwijl juist de levensloop van genres centraal staat in het project. Doordat NEWSGAC dit proces inzichtelijk maakt, kon het beste algoritme gekozen worden. Dat algoritme classificeert nu ook miljoenen artikelen.

## ADAH-project

NEWSGAC is een van de vier ADAH-projecten (Accelerating

Scientific Discovery in the Arts and Humanities) van CLARIAH en het eScience Center. Andere partners zijn: Rijksuniversiteit Groningen (RUG), CWI, KB en Beeld & Geluid. Hoofdonderzoeker is prof. Marcel Broersma (RUG). Binnenkort wordt NEWSGAC geïntegreerd in de CLARIAH-infrastructuur zodat ook andere onderzoekers met andere vragen en datasets er gebruik van kunnen maken.

[sciencecenter.nl/project/newsgac](http://sciencecenter.nl/project/newsgac)

## GELEZEN

*Open a GLAM Lab*  
Digital Cultural Heritage  
Innovation Labs,  
Book Sprint, Doha, Qatar,  
23-27 September, 2019

Lotte Wilms, coördinator van het KB Lab van de Koninklijke Bibliotheek, deed eind vorig jaar in Qatar mee met een 'Book sprint'. In vijf dagen schreven 15 experts gezamenlijk een boek over het opzetten van een GLAM Lab. Een GLAM Lab is een innovation lab voor Galleries, Libraries, Archives and Museums. Naast de voordelen van een GLAM Lab behandelt het boek ook tips en manieren om een eigen GLAM Lab op te zetten met behulp van key points en case studies. Ook het KB Lab krijgt aandacht in het boek. In het KB Lab (lab.kn.nl) staan experimentele digitale tools en data gemaakt van, voor en met de digitale collecties van de Koninklijke Bibliotheek.

Het boek 'Open a GLAM Lab' is gratis te downloaden in .pdf en .epub bestand, via [glam-labs.io](http://glam-labs.io).

[hdl.handle.net/10576/12115](http://hdl.handle.net/10576/12115)



## COLUMN

### Kijkje in de keuken

Soms kan "weten hoe de worst gemaakt wordt" alle eetlust bederven. Zo heeft één middagje bijverdienen in de Pizza Hut op Amsterdam Centraal 15 jaar geleden me voorgoed genezen van elke behoefte om een pizza-slice te bestellen bij een fastfoodrestaurant. Geldt dat eigenlijk ook voor surveys? Zouden de bevindingen van het Sociaal en Cultureel Planbureau met nog meer smaak door het publiek worden verorberd als men een beter zicht had op wat er allemaal komt kijken bij het uitvoeren van een survey? Die vraag kwam afgelopen zomer bij mij op tijdens het keynotespektakel van mijn voormalige baas Ineke Stoop op het methodologiecongres ESRA. In vliegende vaart schetste Ineke de best practices in survey-onderzoek. Wie dacht na deze speedcursus survey-onderzoek in 2 minuten – overigens beloond met een ovationeel applaus – weer naar huis te kunnen, verkeerde zich. Want kloppen die best practices wel? En als ze kloppen, waarom worden ze dan niet altijd

toegepast? Als we niet weten of ze kloppen, hoe komen we daar dan achter? Als ze soms kloppen, wanneer dan en waarom juist dan? En als ze niet kloppen, wat kunnen we dan wel doen?

En grondige weging van de best practices onthulde dat survey-onderzoekers zich met talloze uitruilen en dilemma's geconfronteerd zien. Zo kunnen goede face-to-face interviewers mensen overtuigen om mee te doen aan onderzoek, aanmoedigen om (privé-)informatie te delen en respondenten helpen met complexe vragenlijsten. Maar ze kunnen ook in de verleiding komen

respondenten door het interview heen te jagen, vragen ongewenst te versimpelen of -in het ergste geval- data zelf te verzinnen. Valt dergelijk ongewenst interviewergedrag te voorkomen door met ervaren interviewers te werken? Misschien, maar staan die ervaren rotten ook open voor specifieke instructies of nieuwe inzichten? Ook zij kunnen en moeten bijleren. En hiermee heb ik slechts twee van de negentien best practices besproken.

Ineke's synthese van decennia aan theoretische kennis over- en praktische ervaring met survey-onderzoek bevatte gouden raad

voor iedereen die ooit overweegt een survey uit te voeren, én vormde een belangrijke onderzoeksagenda voor iedereen die eraan wil bijdragen dat surveys in de toekomst nog beter worden uitgevoerd.

Ongetwijfeld zal de rol van het survey binnen de sociale wetenschap door de toenemende beschikbaarheid van andere databronnen (big en organic data, maar ook kwalitatief onderzoek) in de komende jaren blijven veranderen. Maar surveys blijven nodig om te begrijpen hoe burgers de wereld om hen heen en hun eigen rol in die wereld ervaren en

waarden. Zolang we de keuzes en overwegingen bij het ontwerpen en uitvoeren van die surveys net zo consciëntieus en weloverwogen blijven maken als Ineke deed, durf ik iedereen een kijkje in de keuken te gunnen en de resulterende datamaaltijd met trots te serveren.

Sander Steijn

Sander Steijn is methodoloog bij het Sociaal en Cultureel Planbureau (SCP). Op het SCP adviseert hij in alle fases -dataverzameling, analyse en rapportage- van het (kwantitatieve) onderzoeksproces en werkt aan de langdurige (survey-)data infrastructuur van het SCP. Daarnaast is hij lid van het Core Scientific Team van het European Social Survey en namens de planbureaus en onderzoeksinstellingen lid van het Supervisory Board van ODISSEI.

Sander geeft de pen door aan Marieke Houben-van Herten, statistisch onderzoeker en project-leider bij het CBS.