

Online Amsterdam Time Machine en het oude Amsterdamse dialect

# Van het Kalverstraats tot het Duvelshoeks

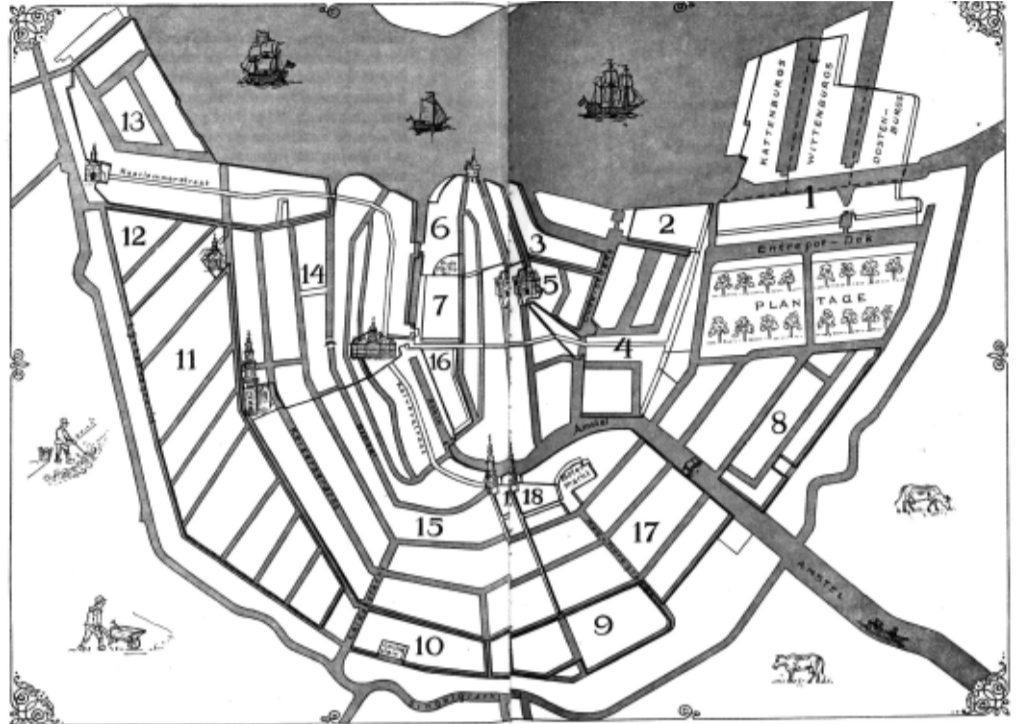
Het Amsterdam van de 19e eeuw telde bijna 20 verschillende sociolecten. Met het project Amsterdam Time Machine maakt het Meertens Instituut een reconstructie van de taal van deze tijd.

Nicoline van der Sijs

Sinds H.G. Wells de menselijke fantasie prikkelde met het idee van tijdreizen, hebben velen van ons ervan gedroomd om naar een andere tijd te gaan en ons onder te dompelen in de beelden, kleuren, vormen en klanken van die periode. Inmiddels ligt dit binnen onze mogelijkheden, hoewel anders dan de manier waarop Wells het zich eind negentiende eeuw had voorgesteld. In onze tijd is de computer de tijdmachine, en hij laat ons niet daadwerkelijk naar een vierde dimensie reizen, maar presenteert een virtuele reconstructie van het verleden.

## Digitale reconstructie

Voor die reconstructie hebben we gekozen voor het Amsterdam van de negentiende eeuw. De Amsterdam Time Machine (ATM), gefinancierd door CLARIAH en gecoördineerd door Julia Noordegraaf (UvA), is gebaseerd op een gedetailleerde kaart met kadaster- en bevolkingsgegevens, die aan elkaar worden gekoppeld via HisGIS van Hans Mol (Fryske Akademy). Het Meertens Instituut wil binnen ATM een reconstructie maken van de negentiende-eeuwse Amsterdamse dialecten. Daardoor hopen we het realiteitsgehalte te kunnen testen van de 19 verschillende Amsterdamse 'tongvallen' die volgens een boek van de taalkundige Johan Winkler uit 1874 aan het begin van de negentiende eeuw werden gesproken. Met 'tongvallen' worden sociolecten bedoeld, lopend van het Kalverstraats, 'het beste en welluidendste amsterdamsch' tot



In 1874 telde Amsterdam 19 dialecten; dialectologe Jo Daan heeft in 1949 in 'Hij zegt wat. Grepen uit de Amsterdamse volkstaal' de lokatie van 18 daarvan gereconstrueerd.

Credits Jo Daan

het Duvelshoeks, dat 'in zijn platste platheid, doormengd [was] met tal van woorden uit de dieve- en bedelaarstaal'.

Van lang niet al deze tongvallen geeft Winkler voorbeelden. Binnen ATM gaan we de gegevens aanvullen met zoveel mogelijk gelocaliseerde data. We beschikken bijvoorbeeld over een groot aantal bronnen van het Amsterdamse Bargoens. Ook hebben we veel informatie over Jiddische leenwoorden in het Amsterdams, die globaal gelocaliseerd kunnen worden in het 'Jodenhoeks'.

## Waar is het origineel?

Prachtig materiaal levert ook een vragenlijst op die W.W. van Lennep en J.A. Alberdingk Thijm in 1877 aan een groot aantal Amsterdammers hebben voorgelegd. De meeste vragen betreffen de uitspraak. Zo bleek 'ij' in 'mijn' of 'ijs' op vier

verschillende manieren te worden uitgesproken: als 'èè', 'aa', 'ei' en 'aai'. Die laatste uitspraak is sinds eind twintigste eeuw heel verbreid geraakt en valt onder het zogenaamde Poldernederlands. De door Van Lennep en Thijm gepubliceerde samenvatting van de antwoorden zijn heel waardevol. Maar wat zouden we graag de originele antwoorden met gegevens van de informanten willen terugvinden. Vooralsnog zijn we daarin niet geslaagd, maar wie weet heeft u als lezers van E-data een tip waar ze zijn te vinden...

Eind september / begin oktober vond de Time Machine 2018 conferentie plaats in het Tech Convention Center in Lausanne, Zwitserland. Meer informatie over dit event staat online: [conference.timemachine.eu](http://conference.timemachine.eu).  
[create.humanities.uva.nl](http://create.humanities.uva.nl)

## RDA-wegwijzer voor sociale wetenschappers

# 'Samen voor Open Science'

Meer dan 90 groepen zijn actief binnen de Research Data Alliance (RDA). DANS zoekt uit welke groepen relevant zijn voor sociale wetenschappers. *Ricarda Braukmann*

De Research Data Alliance is een internationale en op leden gebaseerde organisatie. Zij zet zich sinds 2013 in om uitwisseling en hergebruik van data te stimuleren en de bestaande infrastructuur uit te breiden.

Inmiddels telt de RDA meer dan 7000 leden uit 137 verschillende landen. RDA-leden, voornamelijk onderzoekers en data science professionals, zijn actief in meer dan 90 groepen en vormen de kern van het werk van de Alliance. Zij werken

aan verschillende uitdagingen op het gebied van data sharing en open science. Er is dus veel te beleven, maar voor nieuwkomers kan het lastig zijn om een goed overzicht te krijgen van alle ontwikkelingen binnen de RDA.

## Gids naar groepen

Om RDA-nieuwkomers uit de sociale wetenschappen op weg te helpen, heeft DANS onlangs een RDA-introductie samengesteld. De introductie bevat een overzicht van de verschillende RDA-groepen en een wegwijzer die aangeeft welke groepen het meest relevant zijn voor de sociale wetenschappen.

Zo is er bijvoorbeeld een RDA-groep die zich bezighoudt met datamanagement training, een onderwerp

dat steeds meer onder de aandacht komt bij onderzoekers en financiers. Ook het werk van de Data-Discoverygroep, die onder meer discipline-brede tips geeft voor het zoeken en vinden van onderzoeksdata, zou interessant kunnen zijn voor een sociale wetenschapper.

In de sociale wetenschappen wordt vaak gewerkt met proefpersonen en persoonsgegevens. RDA-groepen gericht op privacy en ethische aspecten van het delen van data worden daarom als zeer relevant geacht voor onderzoekers uit dit vakgebied. Sociaalwetenschappelijke onderzoekers met een medische focus kunnen zich aansluiten bij een RDA-groep specifiek gericht op gezondheidsdata, en omdat big data een steeds grotere rol speelt in de sociale

wetenschappen is ook de RDA-groep rondom big data meegenomen in de wegwijzer.

## De rol van DANS

De volledige introductie met alle RDA-groepen en hun relevantie voor de sociale wetenschappen is door DANS samengesteld in het kader van het RDA Europe 4.0 project (Horizon 2020, kenmerk 777388). DANS heeft in dit project de rol als nationale RDA-node en als ambassadeur voor de sociale en geesteswetenschappen.

Het rapport en de bijbehorende dataset zijn open access beschikbaar via Zenodo ([doi.org/10.5281/zenodo.1401105](https://doi.org/10.5281/zenodo.1401105)).

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



E-DATA & RESEARCH

Jaargang 13 | 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, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

## INHOUD

2 Project van CentERdata toont potentie ODF

3 In gesprek met Lora Aroyo over context

4 Virtual reality biedt wereld aan nieuwe data

5 Online databank PAN toont al 11.500 objecten

5 De voordelen van persistent identifier ORCID

6 Sinds kort beschikbaar en de actuele agenda

7 In gesprek met jong talent Chris Dijkshoorn

8 'Den lezer heil', aldus Sanders in laatste column

Kom ook 28 november naar de prijsuitreiking van de Nederlandse Dataprijs zie pagina 2

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)

## Digital Humanities 2018: Puentes/Bridges

Steven Claeysens

Deze zomer werd het jaarlijkse grote Digital Humanities-congres in Mexico City georganiseerd. In de hoop culturele, sociale en geesteswetenschappelijke bruggenbouw tussen onderzoekers van over de hele wereld te stimuleren, was het gastland voor het eerst een Latijns-Amerikaans land. Met dezelfde intentie wierp men een extra taalbarrière op door het congres officieel tweetalig (Engels en Spaans) te maken.

De openingslezing werd gehouden door Janet Chávez Santiago, Mexicaans taalkundige gespecialiseerd in het Zapoteeks, een van de oudste nog gesproken talen in Mexico. In haar introductie van Chávez Santiago herinnerde Elisabeth Burr de aanwezigen eraan hoe op het web tijdens de eerste jaren alleen plaats was voor het Engels of hooguit andere talen in het Latijnse alfabet. Chávez Santiago onderzoekt nu hoe digitale media ingezet kunnen worden voor en door sprekers van inheemse talen. Zo werkt ze samen met sprekers van het Zapoteeks aan een online woordenboek van de taal.

In haar slotlezing deed Schuyler

## GEHOORD & BIJGEWOOND



De rekenkracht van ODF maakt het mogelijk om sociale netwerken te onderzoeken. Credits Istock

Esprit verslag van de gevolgen van de klimaatverandering voor de Caraïben. Esprit is oprichtster en directeur van het Create Caribbean Research Institute, het eerste Digital Humanities-lab op de Engelstalige Caraïben. Ze verhaalde over de vernietigende passage van orkaan Maria in 2017, hoe dit het lab verwoestte en het eiland Dominica ontwrichtte. Tegenwoordig onderzoekt ze met haar studenten hoe wetenschap en kunst met digitale midde-

len een gemeenschap die de tol van de klimaatverandering betaalt kan bijstaan.

Deze editie zou de boeken in kunnen gaan als het jaar waarin de toepassing van 3D-technieken in de humaniora in een stroomversnelling kwam.

Benieuwd wat het volgend jaar brengt? Noteer alvast 9 tot en met 12 juli 2019 in de agenda. Locatie: TivoliVredenburg, Utrecht.

dh2018.adho.org

## ODISSEI Data Facility voor complexere vragen

Marion Wittenberg

27 juni presenteerde ODISSEI in Utrecht het prototype van het nieuwe dataplatform ODF. ODF wordt ontwikkeld door SURFsara en het Centraal Bureau voor de Statistiek (CBS) en biedt de mogelijkheid om in een veilige High Performance Computing omgeving verschillende soorten complexe data aan elkaar te linken.

John Kartopawiro en Ruurd Schoonhoven (CBS) legden in hun presentatie de ins en outs van dit nieuwe dataplatform voor de sociale en economische wetenschappen uit. De pilot is gestart met de koppeling van data van het Nederlands Tweeling Register (NTR) aan CBS-data; Eveline de Zeeuw liet zien dat met ODF het verband tussen verstedelijking en schizofrenie veel beter onderzocht kan worden.

Daniel Oberski (Universiteit Utrecht) hield een gloedvol pleidooi waarom computationele analyses essentieel zijn voor de hedendaagse sociale wetenschappen. Na de pauze kwam een serie van onderzoeks-

problemen voor het voetlicht die met ODF aangepakt zouden kunnen worden.

Edwin de Jonge (CBS) ging in op de complexiteit van sociale netwerken. In Nederland wonen 17 miljoen mensen die samen 39 miljard relaties onderhouden. De rekenkracht van ODF maakt het mogelijk deze netwerken te onderzoeken. Marcel Das (CentERdata) zou ODF graag willen gebruiken voor de analyse van Accelerometer data die CentERdata met het LISS-panel verzamelt. ODF maakt het mogelijk om deze data met 'Deep Learning' methoden te analyseren (red: zie artikel onderaan deze pagina). Nico Pieterse (Planbureau voor de Leefomgeving) ziet ook veel potentie in ODF omdat het hiermee mogelijk wordt gedetailleerde klimaatgegevens te koppelen aan zowel gegevens over de gebouwde omgeving als gezondheidsdata, om zo lokale klimaatrisico's in kaart te brengen. Andere sprekers gingen in op de mogelijkheden van koppeling van geodata, biobanken en grote surveys. Aan enthousiasme en ideeën was er tijdens deze drukbezochte bijeenkomst geen gebrek. Een goede basis om de plannen omtrent de ODISSEI Data Facility verder uit te werken. [odissei-data.nl](http://odissei-data.nl)

## Hoogfrequente accelerometerdata in de sociale wetenschappen

# Patroonherkenning in beweging

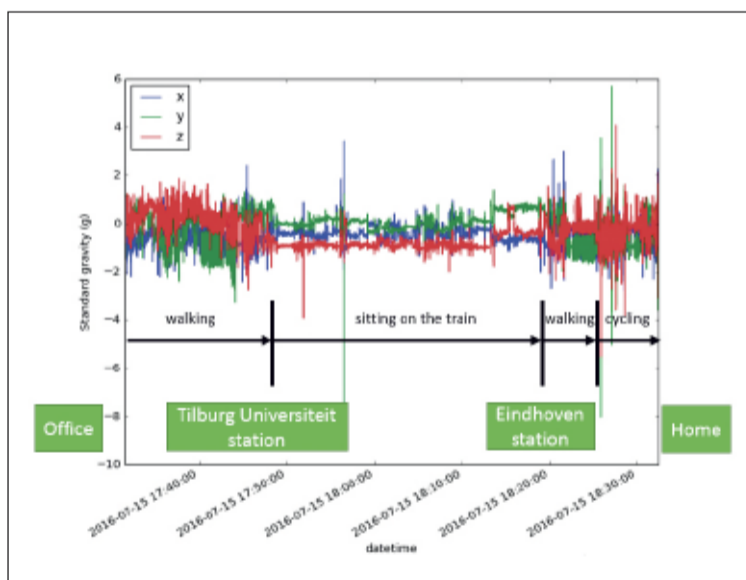
Het accelerometerproject toont de potentie van de ODISSEI Data Facility (ODF) om hoogfrequente data te analyseren en te combineren met survey- en registerdata.

Marije Oudejans

Tijdens het accelerometerproject hebben ruim 1.000 LISS panelleden 8 dagen continu een accelerometer om hun pols gedragen. Deze meter registreerde met 60 metingen per seconde hun fysieke activiteit. Dit resulteerde in een grote hoeveelheid hoogfrequente data op het gebied van activiteit gemeten op een x-y-z as. Tijdens dezelfde periode rapporteerden panelleden ook zelf hun dagelijkse bezigheden en mate van inspanning. Zo kon een vergelijking worden gemaakt tussen objectief gemeten activiteit en subjectief gerapporteerde activiteit (Kapteyn et al., 2018).

### Datakoppeling

Het koppelen met bestaande registerdata was voorheen al mogelijk bij het CBS door middel van de remote access faciliteit. Ook de High Performance Computing (HPC)



**ActivityPatternRecognition: Door het bijhouden van activiteiten (lopen, fietsen, zitten, etc.) en het dragen van de accelerometer wordt duidelijk welk patroon bij welke activiteit hoort en wordt een baseline gecreëerd. Met Machine Learning en patroonherkenning kunnen vervolgens de specifieke patronen en activiteiten herkend worden uit de volledige, ruwe data.** Credits CentERdata

clusters van SURFsara waren al toegankelijk voor de benodigde rekenkracht. "Echter, de combinatie van deze twee maakt de ODF juist zo uniek", vertelt Joris Mulder, onderzoeker bij CentERdata en projectleider van het accelerometerproject. "De faciliteit maakt het bijvoor-

beeld mogelijk om de accelerometerdata te koppelen aan vele bestaande (register)data van het CBS in een beveiligde omgeving, zodat privacy van gevoelige gegevens gewaarborgd blijft. Daarnaast maakt de beschikbaarheid van HPC binnen de ODF-omgeving het toepassen

van data science-technieken mogelijk, zoals machine learning, deep learning en (ongestructureerde) patroonherkenning.

### Vollediger beeld

Zo kunnen nu relaties tussen (en variaties in) gewicht en patronen van fysieke activiteit, persoonlijkheidskenmerken, sociaal-economische positie en gezondheid beter in beeld worden gebracht. Dit wordt gedaan door de hoogfrequente, niet-geaggregeerde data van de accelerometers te combineren met andere sensordata, zoals gewicht, vetpercentage en spiermassa van het weegschaalproject en andere data uit longitudinale LISS panelstudies op het gebied van persoonlijkheid, gezondheid, inkomen en bezit en microdata van het CBS. Dit kan tot nieuwe inzichten leiden op het gebied van relaties tussen beweging, gezondheid, persoonlijkheid en sociaal-economische status. Inzichten die in eerste instantie niet zo voor de hand liggend zijn, maar binnen de ODF omgeving nu wel geïdentificeerd kunnen worden."

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



## Wie worden de winnaars van de Dataprijs 2018?

Uit de bijna 50 ingezonden datasets voor de Dataprijs 2018 zijn onlangs door de jury, per categorie, 3 genomineerden gekozen die kans maken op € 5.000 voor het toegankelijk(er) maken van hun onderzoek. Benieuwd naar wie de genomineerden zijn? Ze staan op de website van RDNL, de initiatiefnemer van de Dataprijs. Rondom de prijsuitreiking op 28 november wordt een interessante dag georganiseerd door RDNL, LCRDM, UKB werkgroep Research Data, NFU Data4Life-Sciences en het Nationaal Platform Open Science voor onderzoekers en data stewards. Aanmelden kan online. (HB)

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

### OVERNEMEN ARTIKELEN

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

Lora Aroyo, professor in de computerwetenschappen:

# ‘Laat gebruikers actief vertellen en collecties verrijken’

“Erfgoedinstellingen beschikken over een goudmijn aan data waarmee ze hun maatschappelijke relevantie kunnen vergroten.” E-data interviewt Lora Aroyo, professor in de computerwetenschappen.

Maarten Heerlien

Begin oktober was Lora Aroyo keynotespreker tijdens de International Conference of Art Libraries. In haar kernboodschap richting bibliotheken, archieven en musea is Aroyo stellig: “gebruik je publiek om meer kennis en inzichten uit je collectie te halen. Zo kunnen we erfgoed interpreteren vanuit steeds meer perspectieven”. Samen met het Rijksmuseum, Beeld & Geluid en andere erfgoedinstellingen werkt Aroyo samen om collecties en objecten te verrijken met behulp van semantische technieken en machine learning.

## User data als sleutel

Aroyo: “Amazon zette door het gebruik van user data de retailmarkt op zijn kop. Erfgoedinstellingen hebben de sleutel in handen om datzelfde te doen. Door de userlogs van hun website, collectieportaal en andere online uitingen structureel te analyseren en trends daarin te identificeren, kunnen instellingen voortdurend anticiperen op de behoeften en voorkeuren van hun bezoekers. Zoektermen die online bezoekers invoeren bijvoorbeeld, geven inzicht in de diversiteit aan terminologie die ze daarbij bezigen. Met zulke inzichten kunnen instellingen hun eigen thesauri verrijken en zo hun zoekingen en aanbevelingen beter afstemmen op de vocabulaires van hun gebruikers.”

## Bredere context

Veel bezoekers zijn bovendien kundige erfgoedliefhebbers die staan te springen om actief en vrijwillig hun eigen kennis toe te voegen aan collecties en objecten waarvoor zij een passie hebben. Aroyo: “Het is voor musea essentieel om doelgroepen, of dat nou professionals, niche-experts of leken zijn, de ruimte en de middelen te bieden om collecties te verrijken met hun eigen annotaties en te gebruiken om hun verhaal te vertellen. Elk van die doelgroepen doet dat vanuit zijn eigen perspectief. Het vastleggen en leren begrijpen van al die verschillende perspectieven is waardevol want daardoor ontdek je wat elk van die groepen kan meebrengen om je collectie in een bredere context te plaatsen. Om zo meer en meer gestructureerde informatie te verzamelen over objecten en deze in te zetten om gebruikers continu nieuwe en rijkere ingangen op collecties te bieden. De uiteenlopende thematische experimenten die we deden rondom vogels, mode en bijbels, toonden aan hoeveel potentie dit heeft. Het bewerkstelligt inclusiviteit en creativiteit, belangrijke thema's voor erfgoedinstellingen.”

Crowdsourcing en nichesourcing, termen



In haar kernboodschap richting bibliotheken, archieven en musea is Aroyo stellig: ‘gebruik je publiek om meer kennis en inzichten uit je collectie te halen’. Credits: Nacho Guevara

waarmee Aroyo een combinatie van input vanuit de grote menigte met specialistische input van niche-experts aanduidt, zijn veelbelovende methoden om verschillende perspectieven op collecties te verzamelen. “Je zult over veel onderwerpen heel uiteenlopende uitspraken terugkrijgen van de crowd. Daaronder zullen zich onherroepelijk een aantal, in de ogen van de meerderheid, vreemde zienswijzen bevinden, waarvan je je niet kunt voorstellen dat er mensen zijn die zo tegen een onderwerp of object aankijken. Die afwijkingen moet je koesteren, want dan ga je inzien hoe controversieel een onderwerp of object eigenlijk is. En controversie zet mensen aan het denken, het bewustzijn van het bestaan van een tegenbewering is het begin van nieuwe kennis!”

## Patronen in Crowdtruth

De kans bestaat dat erfgoedinstellingen verdrienvoudigen in de zee van controversiële zienswijzen die dat mogelijk gaat opleveren. Aroyo: “Met dat gegeven in het achterhoofd hebben we Crowdtruth.org ontwikkeld. Ik doe onderzoek naar human-in-the-loop machine learning. Dat houdt kortweg in dat je computers traint om patronen te herkennen in grote hoeveelheden data op basis van verzamelingen door mensen geannoteerde voorbeelden, zogeheten ‘ground truths’. Samen met Chris Welty, één van de ontwikkelaars van IBM’s supercomputer Watson, onderzocht ik de methoden die onderzoekers doorgaans gebruiken om zulke ground truths te genereren. We constateerden dat dit vaak gebeurt op basis van een zevental misvattingen, bijvoorbeeld de aannames dat er op elke vraag slechts één

## INTERVIEW

### ‘Plaats je collectie in een bredere context’

juist antwoord mogelijk is en dat de input van de professional meer waard is dan die van de leek. Ook zijn het doorgaans momentopnames en worden veranderende inzichten over langere periodes niet ingecalculleerd. In CrowdTruth vangen we het hele spectrum aan mogelijke antwoorden op een vraag, ook door de tijd heen. We wegen hoeveel steun elk van die antwoorden geniet en stellen die weging

voortdurend bij op basis van nieuwe input. Voor erfgoedinstellingen kan deze methodiek doorlopend inzichtelijk maken hoe verschillende doelgroepen tegen collecties aankijken, hoe die zienswijzen zich tot elkaar verhouden en in hoeverre dat door de tijd heen verandert. Als ze dat in de infrastructuur van de organisatie inbedden, hebben instellingen een krachtig hulpmiddel om op innovatieve manieren de interactie met hun publiek aan te gaan.” Aroyo erkent overigens dat erfgoedinstellingen vaak onvoldoende expertise in huis hebben om dit soort datagedreven processen te kunnen beheren. “Datawetenschappers zoeken voortdurend naar partners om hun theorieën en methoden in praktijk te testen. Veel interessanter wordt het wanneer een erfgoedinstelling op óns afstapt met hun onderzoeksvragen. Daarmee wordt het voor beide partijen relevanter en dat vergroot de kans dat onderzoeksresultaten terugvloeiën naar de maatschappij. En dat is voor iedereen een gunstig perspectief.”

[lora-aroyo.org](http://lora-aroyo.org)

#### Lora Aroyo

Lora Aroyo studeerde computerwetenschap aan de universiteit van Sofia te Bulgarije en specialiseerde zich in kunstmatige intelligentie. In 2001 promoveerde ze aan de Universiteit Twente op het onderwerp *Task-based Approach to Information Handling Support for Web-based Education*. In de peri-

ode 2001-2006 bekleedde ze posities aan de Universiteit van Osaka, het CWI, de Universiteit Eindhoven. Vanaf 2006 was Aroyo werkzaam aan de Computer science faculteit van de Vrije Universiteit Amsterdam. Aldaar werd ze in 2016 benoemd tot professor en gaf ze leiding aan de onderzoeksgroep User-Centric Data Science.

Daarnaast is Aroyo onder andere als visiting researcher betrokken bij Columbia Data Science en CornellTech in New York en bij het Center for Advanced Studies van IBM Nederland en is ze Track Leader Creatieve Industrie bij Amsterdam Data Science. Ook is ze hoofdonderzoeker bij de New Yorkse startup Tagasauris Inc.

VR-lab biedt mogelijkheden voor de wetenschap

# Nieuwe onderzoeken in een virtuele wereld

Onderzoeker David Peeters managet het Virtual Reality-lab van het Nijmeegse Max Planck Instituut. Het lab biedt veel onderzoeksmogelijkheden.

Erica Renckens

“Dit VR-lab hebben we sinds midden 2015”, vertelt David Peeters, postdoc en manager van de CAVE (cave automatic virtual environment), zoals het VR-lab officieel heet. “Hiervoor gebruikten we een head-mounted display, zo’n headset die je nu ook veel in winkels ziet. In de CAVE kun je bepaalde soorten data, zoals hersensignalen, veel makkelijker meten. Verder kan de proefpersoon hier via de bril nog zijn eigen lichaam zien, wat de ervaring nog realistischer maakt.”

Zo goed als alle onderzoeksinstellingen in de sociale wetenschappen beschikken tegenwoordig over een VR-lab, al gaat het dan meestal om headsets. Peeters: “VR wordt op allerlei manieren in de sociale wetenschappen toegepast. Bijvoorbeeld tijdens therapie voor de behandeling van fobieën of trauma’s. Of als methode om bijvoorbeeld etnische vooroordelen te bestuderen. In VR kun je variabelen manipuleren die in het echte leven niet te controleren zijn, dat is heel handig.”

## Zeldzaam fenomeen

In de geesteswetenschappen is VR als onderzoeksmethode een veel zeldzamer fenomeen. Peeters: “Er is wel onderzoek gedaan naar bijvoorbeeld het switchen tussen talen bij tweetaligen. In de CAVE kun je dan twee verschillende avatars tegenover de proefpersoon zet-



*In de kelder van het Max Planck Instituut voor Psycholinguïstiek (MPI) speelt een proefpersoon de rol van ober in een restaurant. De hele omgeving is echter nep: het restaurant en de gasten bestaan alleen in virtual reality (VR). In werkelijkheid zit de proefpersoon op een stoel met om zich heen drie grote schermen. Daarop wordt een ruimte geprojecteerd, die dankzij een 3D-bril tot leven komt. Foto Bert Beelen*

ten, die elk een andere taal spreken. Dat is een veel natuurlijker situatie dan waarin dat doorgaans experimenteel onderzocht wordt.”

## Meerwaarde methode

VR heeft volgens Peeters duidelijk meerwaarde als onderzoeksmethode: “Sommige effecten die eerder werden gevonden, blijken wellicht toch niet te bestaan. Als je woorden met een negatieve betekenis bijvoorbeeld een voor een op een scherm ziet, verwerk je deze op een dieper niveau dan woorden met een positieve of neutrale betekenis. In VR zien we dat effect echter niet meer, dus in het alle-

daagse leven speelt dat waarschijnlijk nauwelijks.”

Onlangs is Peeters een project gestart met Roel Willems, onderzoeker aan de Faculteit Letteren van de Radboud Universiteit. “We gaan de perceptie van poëzie gecombineerd met VR bestuderen”, vertelt Peeters. “Een grafisch vormgever gaat bij vier gedichten een toepasselijke, beeldende VR-omgeving ontwerpen waarin je het gedicht krijgt voorgelezen. We zijn benieuwd welk effect dit heeft op de waardering van de gedichten.”

[mpi.nl/resources/labs/vr-lab](http://mpi.nl/resources/labs/vr-lab)

## Utrecht Young Academy lanceert podcast ‘The Road to Open Science’

De Road to Open Science podcast is een initiatief van de Utrecht Young Academy, ondersteund door de Universiteitsbibliotheek Utrecht. In de podcast wordt aandacht besteed aan wat open science is en wat het kan betekenen voor onderzoek. Diverse aspecten komen aan bod, variërend van praktische zaken tot de beleidskant en onderliggende theorieën. Geïnteresseerden worden uitgenodigd mee te praten via de website van de Open Science Community Utrecht, waar ook de afleveringen staan.

Ook via @R2OSpodcast op twitter kan men op de hoogte blijven van toekomstige afleveringen en handige verwijzingen. [openscience-utrecht.com/oscu-podcast](http://openscience-utrecht.com/oscu-podcast)

## SONORO-project van start op Curaçao

In het SONORO-project (NWO-project 2017-2020) onderzoeken lokale wetenschappers de relatie tussen financial literacy en health literacy op Curaçao.

De data worden verzameld bij huishoudens die samen een goede afspiegeling vormen van de bevolking. De personen- en adresregistratie op het eiland is echter niet up-to-date.

Als alternatief is gekozen voor een steekproef van adrescoördinaten. Met hulp van het Case Control Management System (CCMS) van CentERdata en Google My Maps gaan interviewers naar gemarkeerde adrescoördinaten. Met het CCMS in vier talen op hun tablets, vragen ze daar naar gegevens van de bewoners. De database staat op een beveiligde server van CentERdata. (CV)

[sonoro.community](http://sonoro.community)

Meer aandacht voor omgang met data

# Een nieuwe gedragscode wetenschap

Per 1 oktober is de nieuwe Nederlandse Gedragscode Wetenschappelijke Integriteit ingegaan.

Marika de Bruijne

De nieuwe gedragscode beschrijft de principes van wetenschappelijke integriteit, normen voor goede onderzoekspraktijken, procedures voor het geval de normen niet worden nageleefd en de verantwoordelijkheden van instellingen.

“Bovendien besteedt de nieuwe gedragscode meer aandacht aan de omgang met onderzoeksdata. Dit is niet verrassend, het gebruik en beheer van onderzoeksdata is steeds belangrijker geworden.” Aan het woord is Lex Bouter, lid van de commissie die de nieuwe gedragscode opgesteld heeft. “De nieuwe code sluit aan bij internationale ontwikkelingen en is bruikbaar voor zowel fundamenteel als toegepast



gesteld door de besturen van KNAW, VSNU, NFO, NWO, Vereniging Hogescholen en TO2-federatie, kreeg de opdracht. In de herfst van 2017 had de commissie een conceptversie klaar. Bouter: “Daarna zijn er consultatiebijeenkomsten en interviews gehouden met internationale experts en een openbare internetconsultatieronde. Met de opbrengst van deze consultaties is de code verder aangescherpt binnen de kaders van de ALLEA-code, de Europese gedragscode voor integer onderzoek.”

## In een notendop

- De nieuwe gedragscode is van toepassing op het publieke en het publiek-private wetenschappelijk onderzoek in Nederland;
- De code definieert vijf principes (eerlijkheid, zorgvuldigheid, transparantie, onafhankelijkheid, verantwoordelijkheid) van wetenschappelijke integriteit, 61 normen voor

goede onderzoekspraktijken en 21

zorgplichten voor de instellingen;

- Deze zorgplichten voor de instellingen geven aan dat onderzoeksorganisaties verantwoordelijk zijn voor het creëren van een werkomgeving waarbinnen goede onderzoekspraktijken worden bevorderd en geborgd;

- De code maakt onderscheid tussen schendingen van de wetenschappelijke integriteit, bedenkelijk gedrag en lichte tekortkomingen;

- In het laatste hoofdstuk staat beschreven hoe een instelling om moet gaan met klachten over mogelijke schendingen van de wetenschappelijke integriteit;

- De code laat aan de ene kant ruimte aan de instellingen om tot een gebalanceerd oordeel te komen over potentiële schendingen van de wetenschappelijke integriteit, maar noemt de wegingscriteria die daarbij een rol spelen expliciet.

## Normen voor goede onderzoekspraktijken

Enkele voorbeelden van normen voor goede onderzoekspraktijken:

- Beschrijf eerlijk, zorgvuldig en zo transparant mogelijk de verzamelde en/of gebruikte data;

- Beheer de verzamelde data zorgvuldig en bewaar de ruwe en de bewerkte versies gedurende een voor de discipline en methodologie passende termijn;

- Werk eraan mee dat data waarvoor dat gepast is, overeenkomstig de FAIR-beginselen vindbaar, toegankelijk, interoperabel en herbruikbaar zijn (Findable, Accessible, Interoperable, Re-usable).

De nieuwe Gedragscode is beschikbaar op [vsnu.nl](http://vsnu.nl)

Portable Antiquities of the Netherlands (PAN) toont al 11.500 objecten online

# Wetenschappers en amateurs slaan de handen ineen

**Metaalvondsten uit privéverzamelingen worden binnen het PAN-project geïnventariseerd, op een systematische manier beschreven en beschikbaar gesteld voor wetenschappelijk en erfgoedonderzoek.**

Hella Hollander

De Vrije Universiteit (VU), projectleider van het PAN-project, heeft een databank ontwikkeld waarmee vondsten van metaaldetectorhobbyisten ontsloten kunnen worden. De vondsten worden professioneel gefotografeerd, gedetailleerd beschreven en door experts geïdentificeerd met behulp van een uitgebreide versie van het Archeologische Basis Register (ABR).

## Online toegang

De collectie (inmiddels zijn 41.000 objecten gemeld, waarvan er 11.500 online te vinden zijn) toont kledingspelden, munten, haarpinnen, kunstvoorwerpen en wapens. Soms zijn deze voorwerpen compleet, vaker gaat het echter om een klein onderdeel. De PAN-publiekswaarsite bevat voor iedere vondst een uitgebreide beschrijving, foto en een determinatie van het referentietype. De exacte locatie en gegevens van de vinder zijn niet publiekelijk toegankelijk (om de vindplaatsen



*Laatmiddeleeuwse draaisleutel, gevonden in de buurt van Cuijk.  
Credits Addie Keizer, Stijn Heeren*

te beschermen), maar wel beschikbaar voor onderzoekers.

## Amateurs als hulpbron

Sinds 2016 is het gebruik van metaaldetectoren legaal, mits dit niet op een archeologische site gebeurt waar nog onderzoek plaatsvindt. Wel is er toestemming nodig van de landeigenaar en de verplichting de vondsten te melden. Levert dit geen risico op voor professio-

nele archeologen? Stijn Heeren en Nico Roymans, initiatiefnemers van het project, geven in 2017 in een artikel in Trouw aan waarom dit niet het geval is:

“Hobbyisten komen op plekken waar beroepsarcheologen niet komen en juist die vindplaatsen zijn interessant voor de wetenschap”. Heeren: “Archeologen hebben maar een beperkt aantal opgravingen per jaar, dan kun je één mooie vondst doen, maar of die in de regio veel voorkomt, valt lastig te zeggen. Met PAN kunnen we veel beter inschatten wat in een streek thuishoort, wat zeldzaamheden zijn en wat is geïmporteerd van ver weg”. Er is nog een reden dat een databank als PAN hoog tijd werd. “De metaaldetector kwam eind jaren zeventig breed beschikbaar. De jonge mannen van toen zijn nu op leeftijd en beginnen te overlijden. We moeten ze spre-

ken, zodat ze op de kaart - soms bij benadering - hun vondstlocaties kunnen aanwijzen.”

## Koppelen van data

Via thesauri, Linked-Open-Data-technieken en datamining van archeologische rapporten zal geprobeerd worden collecties aan elkaar te koppelen. Koppelingen met het Portable Antiquities Scheme (PAS) in Engeland en met nog op te richten databases als die van Denemarken zijn de volgende stappen. DANS zorgt voor de duurzame archivering van de PAN-databank en voor ontsluiting via ARIADNE en Europeana, zodat deze Nederlandse vondsten ook in Europees verband getoond en onderzocht kunnen worden.

[www.portable-antiquities.nl/pan/#/public](http://www.portable-antiquities.nl/pan/#/public)

*PAN is finalist Nederlandse Dataprijz!*

Zie [researchdata.nl/nieuws-en-agenda/nieuws](http://researchdata.nl/nieuws-en-agenda/nieuws)

## Persistent digital identifiers voor onderzoekers

# ‘Met open identifiers wint iedereen’

**Een ORCID (Open Researcher and Contributor ID) is een persoonlijke identifier die zorgt voor verbinding tussen onderzoekers, hun onderzoek, affiliaties en kan leiden tot onderzoeksbeurzen en innovatie. De voordelen op een rij. Iris Vocking**

John Doove, programmamanager Open Access bij SURFmarket: “Voor onderzoekers ligt de waardepropositie van ORCID onder meer in het terugbrengen van administratieve lasten. Onderzoekers steken veel tijd in het rapporteren over hun onderzoek. Dat gaat ten koste van het daadwerkelijk uitvoeren ervan. ORCID helpt door als ‘linking pin’ op te treden. Hierdoor wordt informatie over de onderzoeker automatisch getransporteerd tussen verschillende systemen, waardoor de onderzoeker of zijn/haar ondersteuner de informatie maar één keer hoeft in te voeren. Boven-

dien is door het gebruiken van ORCID duidelijk wie wat publiceert en zorgt het voor een goede informatiehuishouding bij instellingen.” Dat laatste is ook weer belangrijk voor financiers. Doove: “Wereldwijd sluiten steeds meer financiers zich aan bij de ORCID-community. Met

**“De auteurs-identifier ORCID lijkt de facto standaard te worden”**

ORCID kunnen financiers een stukje geholpen worden met de verantwoording en monitoring van hun onderzoeksgelden.”

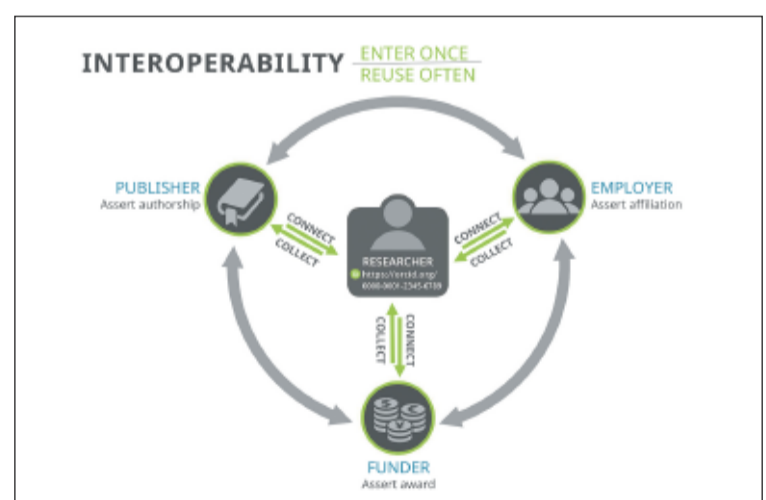
## Collectief abonnement

“De afgelopen jaren hebben acht instellingen een collectieve ORCID-

licentieabonnement afgenomen. Tijdens deze pilot is bekeken welke voordelen een licentie oplevert en is kennis over het werken met ORCID gedeeld. Er zijn altijd koplopers met dit soort dingen, zoals de Universiteit Leiden. Zij werken al richting beleid en zien de voordelen van ORCID in. Zo kunnen zij dus goed delen wat werkt voor onderzoekers en bestuurders. Door deze kennisuitwisseling verrijken we elkaar en daarbij kunnen we, omdat we als groep een licentie hebben, ook nieuwe manieren onderzoeken om ORCID te gebruiken.”

## Koppelen

Ook DANS is betrokken bij het project. Aan de persoonspagina binnen NARCIS kunnen ORCID's worden toegevoegd. Chris Baars, supervisor digital services bij DANS: “ORCID worden steeds vaker gebruikt in NARCIS. Binnen het Europese FREYA-project wordt gekeken naar het koppelen van identifiers, waar-



**Met het Collect & Connect-programma wil ORCID het integratieproces en het delen van gebruikerservaring internationaal stroomlijnen.**

*Credits orcid.org*

onder ORCID. Daarnaast wordt binnen FREYA gekeken naar zogenaamde ‘emerging PID-types’, bijvoorbeeld identifiers voor financiering, organisaties of onderzoeksfaciliteiten.”

Momenteel hebben bijna 7.500 onderzoekers een ORCID geregistreerd

binnen NARCIS en zijn ruim 17.500 Nederlandse ORCID's aangevraagd. Ook de persoons-identifiers ISNI (International Standard Name Identifier) en DAI (Digital Author Identifier) worden ondersteund door NARCIS.

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

## SINDS KORT BESCHIKBAAR

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

## CentERdata

## • ISSP 2015: werkoriëntaties

In oktober 2016 is de module werkoriëntaties van het International Social Survey Programme (ISSP) afgenomen bij het LISS panel. De vragenlijst werd afgenomen in opdracht van onderzoekers (H.B.G. Ganzeboom) aan de Vrije Universiteit Amsterdam en maakt onderdeel uit van een wereldwijd onderzoek over werkoriëntaties.

Er werden onder andere vragen gesteld over de aspecten die iemand belangrijk vindt in een baan, of iemand discriminatie heeft ervaren op het werk, over arbeidsomstandigheden en familie- en gezinsleven. Voor mensen die op dit moment niet werken, gingen de vragen over de laatste baan. Ten slotte werden er ook vragen gesteld over de opleiding en het beroep van de ouders.

Ruim 1.200 panelleden hebben de vragenlijst volledig ingevuld. Het databestand ISSP 2015 – Work Orientation is beschikbaar via LISS Data Archive.

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

Voor meer informatie over het International Social Survey Programme:

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

## Ook sinds kort beschikbaar:

## Studies LISS panel

- Suetens, S.; Cettolin, E., december 2014 - maart 2015, Trust relations in an ethnically diverse society: Trustworthiness of Dutch natives
- CentERdata, september 2017 - oktober 2017, Family and Household - Wave 10
- Veirman, M. de; Jong, J. de; DNB, januari 2016, Buying a house
- Ganzeboom, H.B.G., oktober 2016, ISSP 2015 - Work Orientation



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

## Huygens ING

## • Historische recepten uit KB Kranten

1945 - 1995

Met machine learning hebben onderzoekers van het KNAW HuC Digital Humanities Lab en de UvA automatisch 27.000 recepten geïdentificeerd in kranten van de Koninklijke Bibliotheek uit 1945-1995. De recepten zijn geïdentificeerd (bijvoorbeeld 'vegetarisch' of 'Italiaans') en de ingrediënten zijn apart beschikbaar. Met deze dataset hopen we historisch onderzoek naar eetcultuur in Nederland te vergemakkelijken. De dataset en beschrijving van de werkwijze zijn beschikbaar via:

[github.com/DHLab-nl/historical-recipe-web](https://github.com/DHLab-nl/historical-recipe-web)



Via [huygens.knaw.nl](http://huygens.knaw.nl) is dit bestand en andere bestanden beschikbaar. Bezoek deze site of scan de QR-code.

## DANS

## • Holoceen stuifzand in Nederland

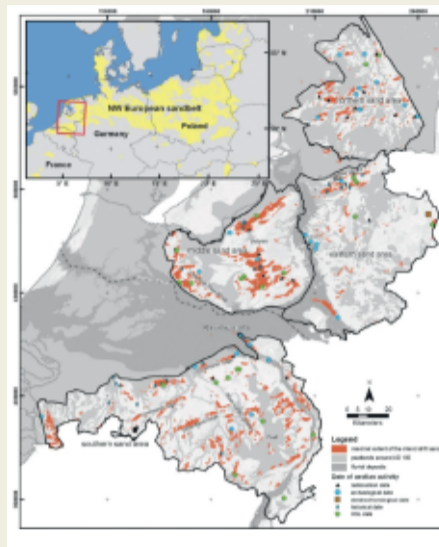
Deze dataset van H. J. Pierik van de Universiteit Utrecht geeft een nieuw landelijk overzicht van stuifzandvoorkomen en activiteit voor de periode omstreeks 5000 voor Christus tot circa 1700 na Christus in Nederland. De dataset is gebaseerd op diverse overzichtsstudies, losse veldstudies en nieuwe gegevens, waaruit afgeleid kan worden wanneer en waar de mens het stuifzand zou hebben kunnen veroorzaken en wanneer het stuifzand mogelijk een bedreiging vormde voor de landbouw of nederzettingen. Dit overzicht is nuttig om het samenspel tussen mens en landschap beter te begrijpen. Daarnaast kan deze dataset fungeren als historische context bij huidige stuifzanden die als waardevolle natuurgebieden worden beschouwd.

[doi.org/10.17026/dans-z2x-phq5](https://doi.org/10.17026/dans-z2x-phq5)  
[narcis.nl](http://narcis.nl)

## Ook sinds kort beschikbaar:

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

- Cohen, Dr K.M. (Universiteit Utrecht) (2017): Landschapskaarten en hoogte-



Credits H.J. Pierik

modellen naar periode en diepte voor archeologisch gebruik in Holoceen-afgedekte delen van Nederland. DANS.

DOI: 10.17026/dans-zck-y7ww

- Defilet, Drs. M.P. (Gemeente Arnhem)

(2018): Hoogstedelaan-Klingelbeekseweg.

Het archeologische verleden van een buurtschap in Arnhem-Noord. Resultaten van de opgravingen uitgevoerd van december 2011 tot en met maart 2012. DANS.

DOI: 10.17026/dans-zvs-74bk

- Derks, Dr. A.M.J. (VUHbs archeologie);

Kerckhove, Drs. J. Van (VUHbs archeologie);

Hoff, Drs. P.G. (VUHbs archeologie) (2008):

Nieuw archeologisch onderzoek rond de Grote Kerk van Elst, gemeente Overbetuwe (2002-2003). DANS.

DOI: 10.17026/dans-zrp-6rnn

- Dielemans, L. (Erfgoed gemeente Utrecht)

(2018): Utrecht in de prehistorie. HFD01:

Een archeologische opgraving van sporen uit de steen-, brons- en ijzertijd langs de Hoofddijk, de Uithof, Utrecht. DANS.

DOI: 10.17026/dans-zap-58ty

- Gerritsen, S. (Archeologie West-Friesland)

(2018): Wonen aan een wiel. DANS.

DOI: 10.17026/dans-xnu-j5je

- Haarlem, Dr W.M. van (Allard Pierson

Museum) (2018): Funerary Culture and

Bioarchaeology of Tell Ibrahim Awad II: from

the Early Dynastic Period to the Late Old

Kingdom. DANS.

DOI: 10.17026/dans-xm9-2dcc

- Jonge, MA L. de (Gemeente Breda)

(2017): Breda, Op de Dreef. DANS.

DOI: 10.17026/dans-xf2-rc3c

- KNMI (Koninklijk Nederlands

Meteorologisch Instituut; Royal Netherlands

Meteorological Institute); Koek, ing. F.B.

(KNMI) (2003): Thematische collectie:

CLIWOC. DANS.

DOI: 10.17026/dans-ze8-d3c8

- Koizumi, Y. (Nagaoka University of

Technology); Watabe, Dr. K. (Nagaoka

University of Technology) (2018): p2p

simulation with redundant data. DANS.

DOI: 10.17026/dans-2cq-x9gw

- Loopik, J. (ADC ArchoProjecten); Dijkstra,

J. (ADC ArchoProjecten) (2018): Bewogen

geschiedenis van twee kloosters. DANS.

DOI: 10.17026/dans-295-99dq

- Zwaag, Dr W van der (Spinoza Centre for

Neuroimaging) (2018): A comparison of 7T

imaging with 8- and 2-channel transmit

systems. DANS.

DOI: 10.17026/dans-zft-hgda



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

## AGENDA

22 - 28 oktober • overall

## Open Access Week

Dit jaar is het thema 'designing equitable foundations for open knowledge'.  
[openaccessweek.org](http://openaccessweek.org)

10 - 14 november • Vancouver

## ASIS&amp;T

## jaarlijkse bijeenkomst

Met de nieuwste ontwikkelingen op het gebied van informatica.  
[assist.org/am18](http://assist.org/am18)

12 - 13 november • Praag

## EHRI Workshop

Met als thema: 'Exploring Refugee Data'.

[ehri-project.eu/ehri-workshops](http://ehri-project.eu/ehri-workshops)

28 november • Den Haag

## Uitreiking Dataprijs 2018

Uitreiking van de Nederlandse Dataprijs aan onderzoekers die onderzoeksdata delen voor aanvullend of nieuw onderzoek. Rondom de prijsuitreiking organiseren RDNL, LCRDM, UKB werkgroep Research Data, NFO Data4LifeSciences en het Nationaal Platform Open Science een interessante dag voor onderzoekers en data stewards.

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

29 november • overall

## International Digital Preservation Day (IDPD)

Volg #IDPD18 op Twitter.

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

4 - 5 december • Berlijn

## EDDI18

Tiende jaarlijkse conferentie voor DDI-gebruikers.

[eddi-conferences.eu](http://eddi-conferences.eu)

7 - 9 december • Galway

## EADH 2018

Het thema van EADH 2018 is 'Data in Digital Humanities'.

[eadh2018eadh.wordpress.com](http://eadh2018eadh.wordpress.com)

10 - 13 december • Seattle

## IEEE BigData 2018

Deze conferentie gaat over: Volume, Velocity, Variety, Value and Veracity.

[cci.drexel.edu/bigdata/bigdata2018/index.html](http://cci.drexel.edu/bigdata/bigdata2018/index.html)

13 december • Amsterdam

## KNVI Jaarcongres

Een congres voor informatieprofessionals georganiseerd door de vakvereniging KNVI.

[congres.knvi.info](http://congres.knvi.info)

4 - 7 februari • Melbourne

## 14e IDCC

Met als thema 'Collaborations and Partnerships: addressing the big digital challenges together'.

[dcc.ac.uk/events/idcc19](http://dcc.ac.uk/events/idcc19)

14 februari • Den Haag

## Open dag on open science

DANS organiseert voor de tweede keer een open dag over open science.

[dans.knaw.nl](http://dans.knaw.nl)

## Koning bezoekt Humanities Cluster en NIAS

Zijne Majesteit Koning Willem-Alexander bracht 29 augustus een werkbezoek aan het KNAW Humanities Cluster en het NIAS (het Netherlands Institute for Advanced Study in the Humanities and Social Sciences). Hij sprak met wetenschappers die een toelichting gaven op de digitale ontwikkelingen in hun geesteswetenschappelijk onderzoek. Het KNAW Humanities Cluster, een alliantie tussen het Meertens Instituut, het Huygens ING en het Internationaal Instituut voor Sociale Geschiedenis, hebben als doel om geesteswetenschappelijk onderzoek te verbeteren door continue innovatie op het terrein van digitale infrastructuur en Digital Humanities. (MM)



*Elli Bleeker sprak over het succesvolle Text As a Graph (de vervanger van XML) en het belang van digitale infrastructuur voor het onderzoek.*

*foto Milette Raats*

Meer data in CLARIAH's Media Suite dankzij nieuwe archiefovereenkomst

# Archief Publieke Omroep toegankelijk voor wetenschappers

**Een lang gekoesterde wens van mediawetenschappers is uitgekomen: sinds kort kunnen zij gearchiveerde radio- en televisie-uitzendingen van de Publieke Omroep online raadplegen.** *Erica Renckens*

Zo'n dertig geesteswetenschappers kwamen de eerste week van juli bijeen in het Nederlands Instituut voor Beeld en Geluid voor de CLARIAH Media Studies Summer School. In teams werkten zij aan onderzoeksvragen waarin media centraal stonden. Hierbij gebruikten zij de Media Suite, een platform waarin wetenschappers onderzoek kunnen doen met audiovisueel materiaal en tekstuele mediabronnen. Zo analyseerden zij wie er wel en niet aan het woord komt in het vluchtelingendebat en of de verhouding mannen en vrouwen in het tv-programma Buitenhof is

veranderd. Voor het eerst konden zij ook op afstand de gearchiveerde uitzendingen van de Nederlandse Publieke Omroep raadplegen, dankzij een nieuwe archiefovereenkomst die online toegang exclusief voor wetenschappelijk gebruik toestaat. "De data op het platform zijn alleen toegankelijk met een account van een Nederlandse universiteit of onderzoeksinstituut", vertelt Eva Baaren, Adviseur Digital Humanities bij Beeld en Geluid.

## Delen van data

Naast de collecties bij Beeld en Geluid biedt de door CLARIAH ontwikkelde Media Suite ook de mogelijkheid om met kranten van de Koninklijke Bibliotheek, de Desmet-collectie van EYE Filmmuseum en de Oral History-collecties van DANS onderzoek te doen, al dan niet gecombineerd. "De Media Suite is gebouwd op de groei, zodat ook collecties van

andere instellingen kunnen worden toegevoegd", aldus Baaren. "Onderzoekers kunnen straks ook hun eigen data importeren. Iemand kan dan zelf Twitter-data scrapen en die in samenhang met andere media analyseren."

## Volop in ontwikkeling

De Media Suite beschikt over mogelijkheden data te doorzoeken, visualiseren, annoteren en analyseren. Daarnaast is er een persoonlijke gebruikersomgeving om resultaten te bewaren en annotaties te bewerken. Aan de tools wordt nog verder gebouwd. Baaren: "Onlangs hebben we een groot deel van het radio- en televisiemateriaal door ons spraakherkenningsysteem gehaald en de transcripties als data in de Media Suite toegevoegd. Uiteindelijk willen we dat gebruikers dit ook zelf met hun data kunnen doen."

De Media Suite is dus nog volop in ontwik-

keling – kan de gemiddelde geesteswetenschapper er wel al mee uit de voeten? "Het is zeker gebruiksvriendelijk", stelt Baaren. "We hebben ingezet op transparantie en duurzaamheid. Maar zoals met elke onderzoeksmethode moet je leren wat er kan en wat niet. Voor iemand met ervaring in de digitale geesteswetenschappen is het zeer goed te doen. Nieuwe gebruikers zullen we met onze kennis ondersteunen."

## Vierde online versie

Eind 2018 wordt de vierde versie van de Media Suite opgeleverd, de laatste binnen CLARIAH-Core. Hiermee komt de voortgang van het project echter niet tot stilstand: ook in het vervolproject CLARIAH PLUS zal het platform verder doorontwikkeld worden.

[mediasuite.clariah.nl/](http://mediasuite.clariah.nl/)

**Hoe zorgen semantische verbanden ervoor dat gebruikers online museumcollecties beter kunnen verkennen? Over deze vraag boog promovendus Chris Dijkshoorn zich de afgelopen jaren bij het Rijksmuseum.**

*Maarten Heerlien*

Dijkshoorn, die zijn onderzoek deed in het kader van het COMMIT/-project SEALINC-Media (Socially Enriched Access to Cultural Media), analyseerde de zoektermen van gebruikers van Rijksstudio, het online collectieplatform van het museum en onderzocht hoe semantische links naar gestandaardiseerde vocabulaires bijdragen aan de diversiteit van zoekresultaten. Hij werd daarbij begeleid door Guus Schreiber en Lora Aroyo van de onderzoeksgroep User-centric Data Science aan de Vrije Universiteit (VU).

Dijkshoorn: "Het resultaat was een verbreding en verdieping van zoekresultaten door de relaties tussen collectieobjecten en thesaurustermen enerzijds en tussen thesaurustermen onderling anderzijds. Het succes van deze relaties hing sterk af van de rijkheid van de objectdata bij het Rijksmuseum zelf. Om deze te vergroten, lanceerden we Accurator, een online platform om specialistische kennis bij niche-experts te verzamelen en hiermee objectdata te verrijken."

## Schat aan informatie

Terwijl Dijkshoorn de laatste hand legt aan zijn proefschrift, werkt hij al als databaseerder bij de afdeling Research Services van het Rijksmuseum. De resultaten uit zijn onderzoek gaat hij nu vertalen naar de praktijk. "Het Rijksmuseum is een Linked Data-uitdaging op zichzelf. Bij de collectie heeft het museum nog zoveel andere bronnen die relevant zijn bij de interpretatie ervan. Honderden meters documentatie, een enorme kunstbibliotheek, terabytes onderzoeksdata, enzovoort. Hoe verbind je dat allemaal aan elkaar op een betekenisvolle wijze die interne en

zoekresultaten toegankelijker

# 'Het Rijksmuseum is een Linked Data-uitdaging op zichzelf'



"We lanceerden Accurator, een online platform voor niche-experts" foto Maarten Heerlien

## JONG TALENT

externe gebruikers zo volledig mogelijk ondersteunt in hun zoektocht naar informatie?"

## Twee benen

Omdat Research Services nog niet de capaciteit heeft om intern grootschalige R&D-activiteiten te ontplooiën op het gebied van

datamanagement en content search, helpt het dat Chris met één been in de onderzoekswereld staat en met het andere in het museum. Het eerste onderzoeksvoorstel is dan ook al de deur uit én toegekend. "Samen met computerwetenschapper Victor de Boer en kunsthistorica Ingrid Vermeulen, beide van

**'Hoe kun je bronnen zodanig aan elkaar verbinden, dat het gebruikers ondersteunt in hun zoektocht naar informatie?'**

de VU, gaan we van een negentiende-eeuwse schilderijencollectie, die inmiddels is uitgewaaid over museale collecties wereldwijd, onderzoeken of we de samenstelling en herkomst kunnen reconstrueren. Een ideale Linked Data-uitdaging, want daarmee kunnen we de vele heterogene datasets die we hierbij nodig hebben normaliseren om ze veel efficiënter te bevragen."

[chrisdijkshoorn.nl](mailto:chrisdijkshoorn.nl)

Persistent identifiers voor het koppelen van onderzoeksresultaten

# De online kracht van PID's

'404 Not Found' is hoogstwaarschijnlijk een van de meest irritante foutmeldingen die je op internet kunt tegenkomen. Persistent identifiers (PID's) zorgen voor een oplossing. *Ricarda Braukmann*

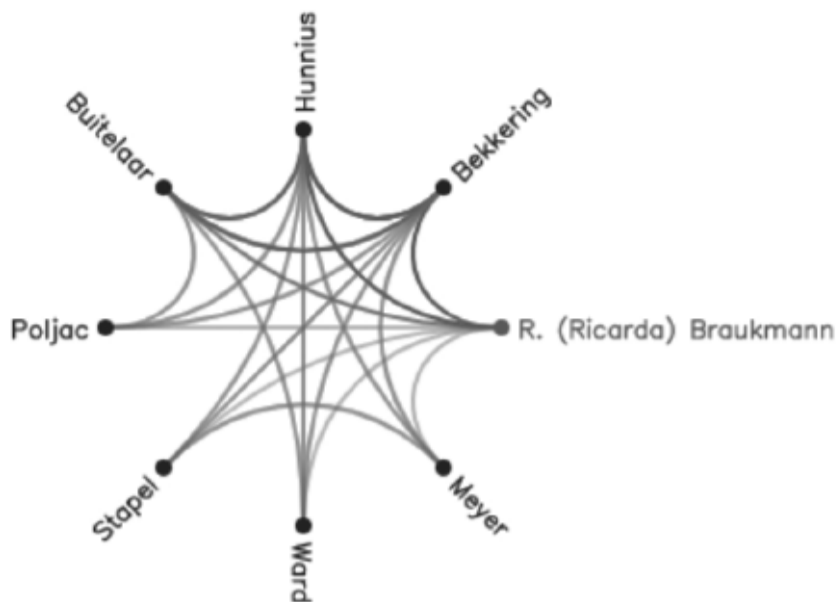
Linkrot is op elke website vervelend, maar het wordt echt problematisch als het om belangrijke digitale bronnen gaat, zoals wetenschappelijke artikelen of datasets. Daarvan mag je verwachten dat ze gedurende langere tijd behouden en toegankelijk blijven. Als oplossing voor dit probleem zijn persistent identifiers (PID's) ontwikkeld. Ze zorgen ervoor dat digitale objecten beschikbaar en ongewijzigd blijven.

## De wereld van de PID's

Er bestaan veel verschillende soorten PID's, met als bekendste waarschijnlijk de DOI's (Digital Object Identifiers) die op grote schaal gebruikt worden voor verwijzingen naar wetenschappelijke artikelen en datasets. Zoals de DOI wordt gebruikt voor publicaties, wordt met de ORCID (Open Researcher and Contributor ID) een persistente identiteit toegekend aan personen, meestal onderzoekers. Dit is handig als je een van de velen bent die J.M. van Dijk heten, of als je van naam wilt veranderen bij huwelijk of echtscheiding. Je ID blijft hetzelfde en uniek, zodat jij en je werk altijd ondubbelzinnig geïdentificeerd kunnen worden.

## Het FREYA-project

Het gebruik van PID's voor artikelen, data en onderzoekers is inmiddels een standaardproces in de wetenschappelijke wereld, maar in theorie bieden PID's eindeloos veel meer mogelijkheden. Ze kunnen ook worden toegewezen aan instellingen, geldstromen, studies, fysieke monsters of instrumenten. Een van de doelen van het FREYA-project is de bestudering van deze nieuwe PID-soorten en de bevordering van hun ontwikkeling en inge-



*NARCIS maakt ook gebruik van PID's om het netwerk van een onderzoeker, in dit geval dr. Ricarda Braukmann, te visualiseren in een samenwerkingsdiagram. Het diagram toont de verbanden tussen onderzoeker Braukmann en zijn mede-auteurs op basis van de PID's in de metadata van hun publicaties.*

bruikneming. FREYA is een driejarig Europees gefinancierd project waarin PID-aanbieders (Crossref, DataCite en ORCID), onderzoeksorganisaties (ANDS, British Library, CERN, DANS, EMBL-EBI, Pangaea en STFC) en uitgevers (PLOS, Hindawi) hun krachten hebben gebundeld voor de promotie en innovatie van PID's.

## Naar PID-diagrammen

Naast de ontwikkeling van nieuwe soorten PID's werkt het team van FREYA ook aan de verbinding en integratie van verschillende PID's in PID-diagrammen. Een PID-diagram kan bijvoorbeeld dienen om informatie te visualiseren en interessante inzichten te verschaffen in de verbanden tussen verschillende onderzoeksobjecten. Dankzij het werk van FREYA kan bijvoorbeeld de wetenschapsportal NARCIS nu verschillende onderzoeksresultaten aan elkaar koppelen op basis van

hun PID's, zodat het portaal zijn bezoekers nog waardevollere informatie kan aanbieden. Als bijvoorbeeld in de metadata van een digitaal object (zoals een artikel of een dataset) de PID van een gerelateerd object is opgenomen, legt NARCIS een verband en haalt het aanvullende informatie op over het gerelateerde object (zoals het type of de titel), die in het portaal wordt weergegeven. PID-diagrammen (zoals ook bijvoorbeeld het samenwerkingsdiagrammen in de afbeelding) en andere PID-services zijn volledig afhankelijk van grootschalig gebruik van PID's zoals ORCID's. Daarom blijft het vergroten van de bekendheid en het belang van PID's bij de hele onderzoeksgemeenschap een van de belangrijkste opgaven van het FREYA-project.

[narcis.nl](http://narcis.nl) / [project-freya.eu/en](http://project-freya.eu/en)

*Dr. Ricarda Braukmann is programmaleider sociale wetenschappen bij DANS*

## GELEZEN

*Plan S: wetenschappelijke artikelen vanaf 2020 Open Access*

Heidi Berkhout

cOAlition S, een Europese groep van onderzoeksfinanciers, wil dat wetenschappelijk onderzoek dat is betaald met belastinggeld vanaf 2020 gratis beschikbaar komt voor het publiek. NWO, de Nederlandse Organisatie voor Wetenschappelijk Onderzoek, is een van de partners.

Publiceren voor wetenschappers is weliswaar gratis, maar de artikelen verdwijnen vervolgens achter een betaalmuur. Universiteiten moeten veel geld betalen voor de abonnementen die deze betaalmuur vormen. NWO wil hier vanaf 2020 verandering in aanbrengen door bij onderzoeksprojecten die zij financieren te eisen dat de publicatie gratis toegankelijk moet zijn voor het publiek. "Wetenschap hoort niet achter een betaalmuur, maar moet voor iedereen vrij toegankelijk zijn", stelt NWO-voorzitter Stan Gielen. "In heel Europa voeren universiteiten moeizame onderhandelingen met uitgevers. Ondertussen maken ze exorbitante winsten op werkzaamheden die vooral door wetenschappers worden verricht. Dat moet afgelopen zijn. Het is hoog tijd voor uitgevers om hun businessmodel om te gooien."

De Europese Commissie geeft aan het plan van harte te ondersteunen. Ook de European Research Council (ERC) onderschrijft het plan.

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

## COLUMN

### Den lezer heil!

Dit is mijn laatste column voor E-Data, want ik vind dat ik mezelf een beetje begin te herhalen. Maar juist omdat het mijn laatste bijdrage is, vind ik dat ik mezelf enigszins mag herhalen.

Onlangs gaf ik bij de Universiteitsbibliotheek Amsterdam een workshop voor boekwetenschappers. Onderwerp: hoe je als boekwetenschapper optimaal gebruik kunt maken van internet.

Het zijn, zoals u weet, gouden tijden voor historici en boekwetenschappers. Nooit eerder hadden zij zo'n enorme rijkdom aan bronnen tot hun beschikking. Delpher, Digibron, Google Books, de Digitale Bibliotheek voor de Nederlandse Letteren, archieven.nl – het kan allemaal niet op. Ik geef vaker lezingen over slimmer zoeken op internet, maar meestal is het grotendeels eenrichtingsverkeer: ik praat, de toehoorders luisteren. Ditmaal was het anders. Het ging om een workshop van drie uur, waarbij de deelnemers – na een verhaal van mij over zoekstrategieën –

opdrachten moesten uitvoeren. Terwijl zij bezig waren, liep ik rond om op hun computerschermen te kunnen kijken.

Dat was heel leerzaam.

Een zoekstrategie die mij zelf heel veel heeft opgeleverd, noem ik kortweg 'perifeer zoeken'. Wie bijvoorbeeld onderzoek doet naar mensen met een zwarte huidskleur, moet zich niet beperken tot zoektermen als neger, negerin en zwarte(n). Daarnaast zou je in Delpher ook zeker dit soort zoekopdrachten moeten proberen: 'dikke lippen' PROX kroeshaar. Je maakt dan gebruik van stereotypingen en clichés waaraan weinig schrijvers weerstand hebben kunnen bieden.

Nadat ik dit had uitgelegd, gaf ik mijn boekwetenschappelijke toehoorders de opdracht om te zoeken naar een portret/beschrijving van een boekengeleerde of boekwurm.

Ik noem geen namen, maar er zaten hooggeleerde, bekende boekwetenschappers bij. Misschien had ik het niet duidelijk uitgelegd, maar het verbaasde mij nogal dat ik op diverse schermen de zoekopdracht 'portret boekengeleerde' en 'portret boekwurm' zag verschijnen. Terwijl ik nog had geroepen: 'Denk dus vooral



foto Leo van Velzen

aan clichés en stereotypingen!' Wat is een essentiële eigenschap van een boekenwurm? Dat hij of zij véél boeken heeft. Zoveel boeken dat die al snel op stapels terecht zullen komen. Stapels die vervolgens het risico lopen om stof te vergaren. Als ik zelf op zoek zou gaan naar een beschrijving/portret van een boekenwurm, zou ik dus zeker zoeken naar bijvoorbeeld 'stapels boeken' in de buurt van 'stoffig'. Of bijvoorbeeld 'boekenkasten' in de buurt van 'marokijn' (een bij bibliofielen geliefde leersoort).

Wie zoekt op clichés en stereotypingen, vindt clichés en stereotypingen. Methodologisch lijkt dat een bezwaar, maar in de praktijk valt dat reuze mee. Mijn ervaring is dat je met dit soort zoekopdrachten teksten vindt die je zoekrepertoire enorm verrijken. Je stuit op spellingsvarianten als boeck-worm en synoniemen als boek(en)wrotter. Je vindt clichés en stereotypingen waar je wellicht zelf niet meteen aan had gedacht, zoals boeken-

étagère. Kortom: binnen de kortste keren heb je een prachtig pallet aan zoektermen waar je véél meer mee vindt dan met een zoekopdracht als "portret boekengeleerde".

Voor mij is dit inmiddels zo vanzelfsprekend dat veel van de zoekopdrachten die ik op de schermen van de toehoorders zag verschijnen, mij nogal verbaasden. Ik bedoel: ze waren eerder direct (bibliomaan bijvoorbeeld) dan associatief.

Leerzaam was het wel. Net zoals ik het leerzaam vond om voor E-Data te schrijven. Ik heb dat sinds 2011 met veel plezier gedaan. Nu wordt het, in mijn ogen, tijd voor een andere columnist. Die ik met belangstelling zal volgen, want ik blijf dit tijdschrift natuurlijk lezen. Om met een oude groet te eindigen: den lezer heil!

*Ewoud Sanders*

*Taalhistoricus en journalist. Sanders is vaste medewerker van onder meer NRC Handelsblad en Onze Taal.*