



Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

E-data & Research 1602

Bosschaert, Stephanie; Berkhout, Heidi

published in

E-data & Research
2022

document version

Publisher's PDF, also known as Version of record

document license

CC BY

[Link to publication in KNAW Research Portal](#)

citation for published version (APA)

Bosschaert, S. (Gasteditor), & Berkhout, H. (editor) (2022). E-data & Research 1602. *E-data & Research*, 16(2).
<https://edata.nl/wp-content/uploads/2022/02/1602.pdf>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

pure@knaw.nl

CLARIAH Fellowships faciliteren gebruik digitale infrastructuur

Historische spelletjes-database in de maak

Met een FAIR-georganiseerde database en financiering van CLARIAH hopen historici meer te weten te komen over de ontwikkeling van strategische spellen.

Erica Renckens

“We hebben in het Digital Ludeme Project (DLP) een database samengesteld met meer dan duizend strategische spellen die belangrijk waren voor de historische ontwikkeling daarvan over de hele wereld,” vertelt Carlos Utrilla Guerrero, datawetenschapper aan het Institute of Data Science (Maastricht University). “Daarmee kunnen we de spellen niet alleen reconstrueren, maar het is potentieel ook de meest uitgebreide en complete spellendatabase ter wereld.”

Als voorbeeld van een historisch spel noemt hij Senet, het oudste bekende bordspel dat drieduizend jaar voor Christus al in Egypte werd gespeeld. “In de database vind je daarvan onder andere een korte beschrijving, de regels, de vindplaatsen en de bronnen.”

Grootschalige data-analyse

Om de bruikbaarheid en de duurzaamheid van de DLP-database te maximaliseren, past Utrilla Guerrero de komende tijd de FAIR-principes toe. Daarbij onderzoekt hij ook de mogelijkheid om Semantic Web- en Linked Data-technieken in te zetten. Hij ontving hiervoor een van de vijftien CLARIAH Fellowships voor onderzoek of onderwijs, die het gebruik van de digitale infrastructuur voor de geesteswetenschappen aanmoedigen en in de toekomst vergemakkelijken.

“We gaan in de digitale geesteswetenschappen een nieuw tijdperk tegemoet van ongekende grootschalige data-analyses, mogelijk gemaakt



Een Senet-spel uit de 14e eeuw voor Christus. De speelstukken konden in de la worden opgeborgen. In de DLP-database staat van dit spel onder andere een korte beschrijving, de regels, de vindplaatsen en de bronnen.

Credits: Brooklyn Museum, Charles Edwin Wilbour Fund, via Wikimedia Commons

door enorme dataverzamelingen,” aldus Utrilla Guerrero. “Maar daarvoor moeten die collecties wel gestructureerd en onderling verbonden zijn. Dat is een grote uitdaging waarbij we effectieve methoden voor data-integratie en kennismanagement moeten onderzoeken en toepassen. Dat gaan we nu doen voor de DLP-database.”

Toekomstig onderzoek

De implementatie van de FAIR-principes in de DLP-database is niet alleen van belang voor de huidige generatie wetenschappers, stelt de onderzoeker. “Ook toekomstige wetenschappers en machines moeten de data kunnen vinden, begrij-

pen en gebruiken. Daarom is het belangrijk dat we nu al nadenken over het hergebruik van onze data.”

Utrilla Guerrero hoopt dan ook dat meer geesteswetenschappers de FAIR-principes gaan implementeren. “Dit Fellowship zal laten zien hoe datamanagement de toegang tot de historische spellen vergroot en verdiept, nieuwe relaties tussen spellen aan het licht brengt en kruiswijzingen tussen materialen aanbrengt, zodat grondige en systematische data-analyse nu en in de toekomst mogelijk zijn.”

ludeme.eu

FAIR-Aware als eerste stap naar een FAIR dataset

Eerste hulp bij FAIR data

FAIR-Aware is een tool die verschillende mogelijkheden biedt voor onderzoekers en trainers om het bewustzijn rondom FAIR data te vergroten.

Linus Čepinskas, Maaike Verburg en Ilona von Stein

FAIR-Aware is een online zelfevaluatie tool die gebruikers helpt te begrijpen wat de FAIR-principes betekenen en hoe ze in de praktijk te brengen zijn. De tool is door DANS ontwikkeld binnen het Europese FAIRsFAIR-project dat zich richt op het ondersteunen van FAIR-data praktijken in Europa. Door middel van tien simpele vragen en uitge-

breide toelichting leren gebruikers over het belang van verschillende aspecten van FAIR en hoe deze te implementeren zijn.

400 gebruikers

FAIR-Aware, gelanceerd in juni 2020, heeft inmiddels bijna 400 gebruikers geholpen met het evalueren en stimuleren van hun kennis over FAIR. Deze maand komt het FAIRsFAIR-project tot een einde, maar FAIR-Aware zal verder worden onderhouden door DANS. Na ruim anderhalf jaar valt er terug te blikken op een aantal mooie mijlpalen. Zo is samen met Ecole nationale des chartes (URFIST de Paris) en DataSuds data repository (IRD) een Franstalige versie van FAIR-Aware tot stand gekomen om de Franstalige wetenschappelijke gemeenschap te

ondersteunen. Dankzij Institut de l'information scientifique et technique (INIST), werd de tool in de zomer van 2021 op de educatieve DoRANum portal gehost.

Community feedback

De ontwikkeling van FAIR-Aware is iteratief en continu geweest. Het projectteam heeft met succes uitgebreide gebruikersfeedback over de metrische gegevens en de gebruikersinterface verwerkt. Zo biedt de tool nu, naast de gewijzigde structuur en formaat, meer discipline-specifieke voorbeelden in de begeleidende teksten. Dankzij de nieuwe woordenlijst zullen ook gebruikers

met beperkte kennis van FAIR makkelijker door de tool kunnen navigeren.

Ook is de tool uitgebreid met een functionaliteit om trainers te ondersteunen. Geïnteresseerden in het beoordelen van het ‘FAIR-bewustzijn’ van anderen kunnen eenvoudig de resultaten van hun gebruikers downloaden en de antwoorden analyseren met behulp van een kant-en-klaar sjabloon en grafieken. Om gebruik te maken van deze functionaliteit, kunnen trainers contact opnemen met de DANS medewerkers van het FAIR-Aware ontwikkelings-team (fair-aware@dans.knaw.nl).

fairaware.dans.knaw.nl



E-DATA & RESEARCH

Jaargang 16 | 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: CBS, Centerdata, CLARIAH, DANS, KNAW Humanities Cluster, ODISSEI en het Rijksmuseum.

INHOUD

3 **Leer de kennis van Open Science in vier uur**

4 **Deze datasets zijn sinds kort beschikbaar**

5 **Marcel Das, directeur Centerdata: “Voor en door onderzoekers”**



6 **Diamantbewerkers tot leven gebracht**

6 **Kennisgraaf als oplossing voor snel doorgronden datasets**

7 **Gastcolumnist Daniel Kapitan droomt over open data en algoritmes**

E-data & Research liever (tijdelijk) op een ander adres ontvangen? Geef wijzigingen door via edata@dans.knaw.nl.



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

DANS Data Trail Workshop - FAIR data assessment tools: an evaluation

Ellen Leenarts

Op 23 november 2021 vond de derde editie van de DANS Data Trail plaats – een online workshop reeks waarbij verschillende thema's rondom Research Data Management en Open Science aan bod komen. In deze bijeenkomst stond de evaluatie van FAIR data tools op de agenda. De workshop werd georganiseerd door DANS in samenwerking met Maastricht University en het RIVM.

Als eerste werden we verwelkomd door Femmy Admiraal (Data Station Manager Humanities, DANS) die de DANS data services introduceerde. Vervolgens vertelde Ammar Ammar (Maastricht University) over de evaluatie van tien services waarmee de FAIRness van data beoordeeld kan worden. De tools die hij met collega's bij het RIVM heeft geëvalueerd, zijn online en offline te vinden, hebben de vorm van vragenlijsten of zijn semi-automatisch in te vullen. In de evaluatie werd gekeken naar de gebruiksvriendelijkheid, hoe breed toepasbaar de tool is, de benodigde voorkennis en tijd, en de output en adviezen die de tool genereert. Naar aanleiding van Ammars presentatie stelde het publiek vragen, bijvoorbeeld over domein gerelateerde adviezen en standaarden en wat de rol van data stewards zou kunnen zijn. Ammar gaf aan binnenkort de resultaten van de evaluatie te publiceren.

Maaïke Verburg en Linas Čepinskas, beiden RDM-specialisten bij DANS, introduceerden vervolgens de vernieuwde FAIR-Aware tool die gebruikt kan worden om meer bewustzijn te genereren omtrent FAIR data. Deze tool werd grotendeels ontwikkeld na de evaluatie van Ammar en collega's. In de respons op de tool kwam naar voren dat een prioritering van de te ondernemen stappen voor een onderzoeker nuttig zou zijn. Ook werd de wens voor meer specifieke praktijkvoorbeelden uitgesproken. In het algemeen werd zeer enthousiast op de tool, en op deze editie van de DANS Data Trail, gereageerd.

Dit event terugkijken kan; de opname is beschikbaar via de website van DANS.

dans.knaw.nl/nl/agenda/dans-data-trail-workshop-fair-data-assessment-tools-an-evaluation/

Slotconferentie Programma Netwerk Oorlogsbronnen

Tessa Free

De imposante Koningin Máxima Zaal in Koninklijk Instituut voor de Tropen was op 2 november 2021 decor voor de feestelijke afsluiting

GEHOORD & BIJGEWOOND



Credits: DANS

van het programma Netwerk Oorlogsbronnen en tegelijk de start van een nieuw begin voor het netwerk. Met 150 mensen in de zaal en honderden (terug)kijkers via de livestream werd teruggeblikt, vooruitgeblikt met toekomstdromen en vooral ook stilgestaan bij de collectieve prestatie van de afgelopen zes jaar. Het platform Oorlogsbronnen.nl brengt bronnen over en uit de Tweede Wereldoorlog in een geautomatiseerd proces inhoudelijk bij elkaar. De website bevatte tijdens



Credits: Netwerk Oorlogsbronnen

de lancering onder meer 12,5 miljoen bronnen en 244.419 persoonlijke tijdlijnen. "Kostbare bouwstenen om de herinnering aan de Tweede Wereldoorlog levend te houden", aldus demissionair staatssecretaris Paul Blokhuis in zijn toespraak.

Ook werd tijdens de slotconferentie stilgestaan bij wat 'om ons heen' gebeurt. Fotograaf en World Press Photo winnaar Roger Cremers brengt al jaren het toeristische aspect rond oorlogsplekken in beeld. Hij nam de zaal mee langs zijn tocht en vertelde over de bezoeker van Auschwitz met de tekst "Laugh, that's an order!" op zijn T-shirt, een volgepakt bus transport richting museum Birkenau en een Russisch pretpark waar je je kunt verkleden als SS'er.

Keynote spreker en onafhankelijk strateeg op het terrein van digitale transformatie in de culturele sector Michael Peter Edson ging in op de

sociale en digitale gemeenschappen waarin mensen zich anno 2021 bevinden. En met welke onderwerpen en op welke platforms culturele instellingen tegelijkertijd bezig zijn. Edson confronteerde de aanwezigen door hardop de vraag te stellen welke rol culturele instellingen kunnen spelen, of zouden moeten spelen, in het maatschappelijke debat en de vorming van een rechtvaardige en duurzame toekomst.

Het einde van het programma betekent niet het einde van Netwerk

Oorlogsbronnen. Het netwerk gaat door, in een andere vorm en met nieuwe doelstellingen, maar vanuit dezelfde ambitie: analoog erfgoed op een digitale en innovatieve manier beschikbaar maken.

De slotconferentie terugkijken kan via:

<https://youtu.be/kV5BoQWqzRY>

EOSC Future Open Days

Elly Dijk

Van 23 tot 25 november 2021 vonden de EOSC Future Open Days plaats, waarin de deelnemers werden uitgenodigd om te discussiëren over de ontwikkelingen van de European Open Science Cloud (EOSC).

Binnen de EOSC wordt het voor onderzoekers gemakkelijker om onderzoeksdata, publicaties, software en diensten te vinden, delen en combineren, ongeacht disciplines of landsgrenzen. In EOSC Future, een Horizon 2020-project dat in april

2021 van start ging, wordt een EOSC-platform ('System of Systems') ontwikkeld. De verwachting is dat dit platform in 2023 operationeel is. De eerste dag van de EOSC Future Open Days was gericht op wat onderzoekers van de portal kunnen verwachten. Voordelen zijn onder andere dat onderzoekers toegang krijgen tot nieuwe typen bronnen (datasets, publicaties, software), die beter gestructureerd zijn (metadata) en vindbaar worden door slimme zoektechnologie. Ook bevat de portal een dashboard, trainingmodules en een helpdesk. De e-infrastructures en Science Clusters zijn nauw bij EOSC Future betrokken, net als de bredere onderzoeksgemeenschap. Op de tweede dag werd ingegaan op het perspectief van service providers. De EOSC Research Graph, die de bronnen aggregereert, speelt hierbij een belangrijke rol. Zowel bronnen van individuele providers als catalogi worden opgenomen. De EOSC Open Science Observatory gebruikt meerdere indicatoren, zoals Open Science beleid, praktijk en impact.

Op de derde dag kwam onder andere de EOSC Future Knowledge Hub ter sprake. Deze Hub zal bestaan uit een Training Catalogue en een Learning Platform met cursussen en curricula. De Training Catalogue is een 'Catalogue of Catalogues'. DANS is, als lid van OpenAIRE AMKE, betrokken bij de ontwikkeling van een uitgebreid trainingsprogramma om de EOSC-stakeholders te ondersteunen in het gebruik van data, diensten en software.

eoscfuture.eu

ODISSEI Community Conference

Godelieve de Wijer

Op 18 november 2021 vond de ODISSEI Community Conference plaats voor deelnemers van ODISSEI en andere geïnteresseerden in computationele sociale wetenschappen.



Credits: MG Fotografie (Michel Groen)

Het programma bood hen, online en op locatie, de gelegenheid om gebruikers en bouwers van ODISSEI te ontmoeten en onderzoeksideeën uit te wisselen.

Na opening door ODISSEI's wetenschappelijk directeur Pearl Dykstra vond de presentatie van keynote spreker Frank Pijpers (CBS en UvA) plaats, waarin hij het belang van connecties tussen personen en connecties tussen ondernemingen voor de nationale statistiek toelichtte. Na de presentatie van Pijpers ging het programma verder met drie interactieve sessies. In de eerste sessie werden verscheidene onderzoeksprojecten van deelnemers van ODISSEI besproken. Marike Knoef (LU), Frank Takes (LU en POPNET) en Bastian Ravesteijn (EUR en KansensKaart.nl) presenteerden innovaties in onderzoek die onder andere mogelijk zijn gemaakt door de ODISSEI Secure Supercomputer (OSSC), LISS Panel data of de nieuwe onderzoeksinfrastructuur POPNET.

In de tweede sessie stond databeveiliging in Open Science centraal. Bas van der Klaauw (VU) en Erik-Jan van Kesteren (UU en SoDa) bespraken met respondent Melanie Imming (Imming Impact) de problematiek van het delen van onderzoeksresultaten die in een beveiligde omgeving zijn bevonden. De sprekers discussieerden over nieuwe mogelijkheden om onderzoek in de sociale wetenschappen meer FAIR te maken. De laatste sessie richtte zich op de relatie tussen de sociale wetenschappen en de geesteswetenschappen, specifiek over het linken van grote datasets in deze velden. Sprekers Rick Mourits (RU) en Richard Zijdemans (IISG) bespraken met respondent Suze Zijlstra (ODISSEI) hoe de sociale wetenschappen en geesteswetenschappen niet alleen bij gelegenheid overlappen, maar zelfs elkaars onderzoek zouden kunnen ondersteunen en versterken.

Er is de mogelijkheid om de conferentie terug te zien via

vimeo.com/647365199/9530c003f6

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.

Maatwerk en remote access mogelijk

Werken met CBS-microdata

Naast StatLine biedt het CBS op meerdere manieren toegang tot data over de Nederlandse economie en samenleving. De mogelijkheden op een rij.

Astrid de Kock

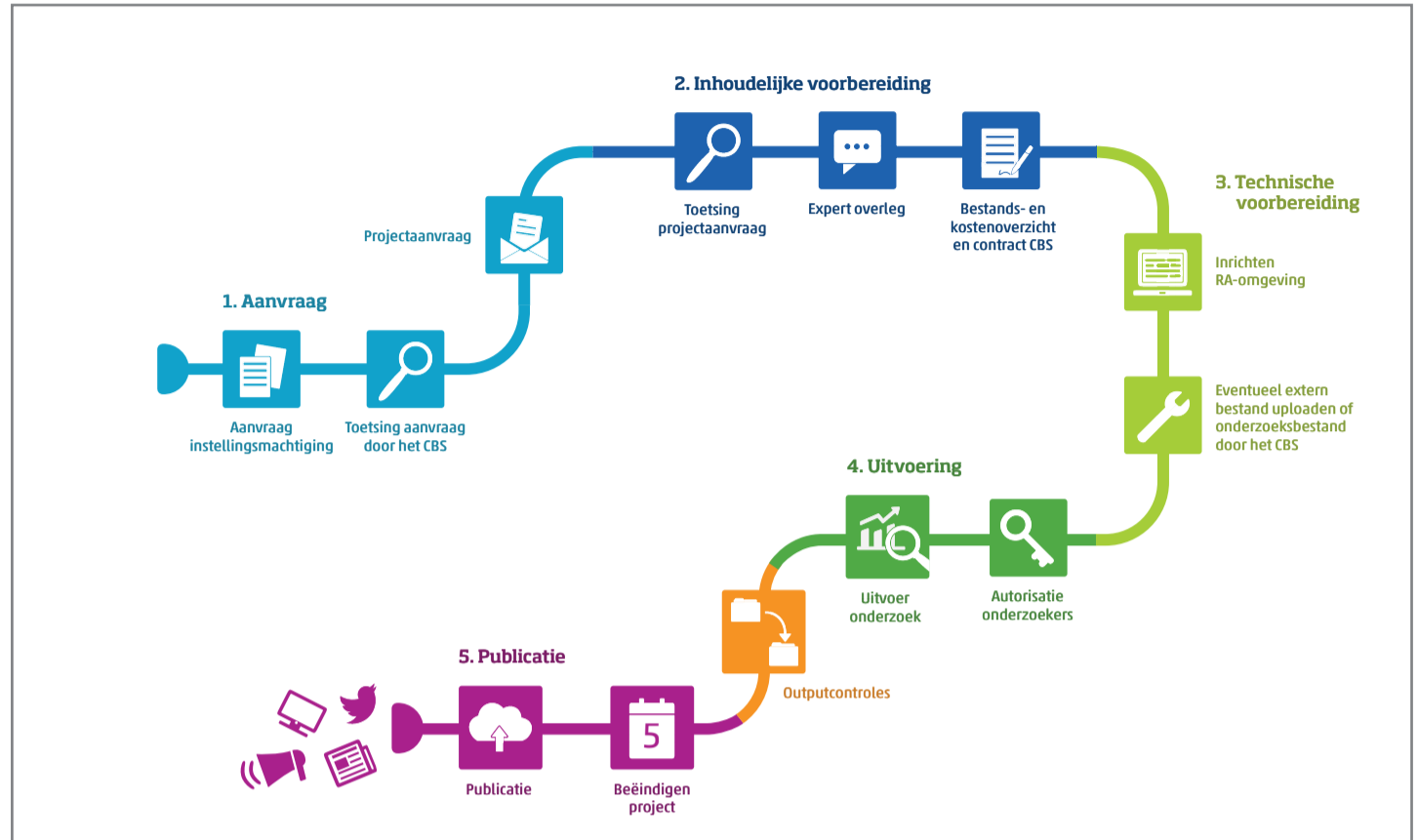
StatLine is de databank van het CBS. Van inflatie tot bevolkingsontwikkeling, van energie naar industrie, alle informatie is gerubriceerd en beschikbaar voor iedereen via open-data.cbs.nl.

Wellicht minder bekend is dat het CBS naast StatLine ook toegang tot ander statistisch materiaal biedt. En dat het in veel gevallen zelfs mogelijk is om eigen data te koppelen aan die van het CBS.

Diensten op een rij

Zo kunnen instellingen die beschikken over een 'instellingsmachtiging', afgegeven door de Directeur Generaal van het CBS, toegang krijgen tot CBS-onderzoeksbestanden, benaderbaar via een beveiligde Remote Access-omgeving. Op cbs.nl staat een 'Catalogus microdata' met per thema de meest recente documentatierapporten van de beschikbare microbestanden.

Het CBS biedt daarnaast aanvullende statistische diensten, bijvoorbeeld ten behoeve van



Het stroomschema 'Werken met CBS-microdata' laat zien welke stappen nodig zijn van 'Aanvraag' tot 'Publicatie'. Credits: CBS

beleids- en besluitvorming. Het team Beleidsstatistiek beschikt over de nodige ervaring en expertise op dit gebied. Zo is onlangs de Monitor laaggeletterdheid opgezet en verzorgt het CBS periodiek een publicatie over daklozen in Nederland. Meer voorbeel-

den staan in het 'Overzicht aanvullend statistisch onderzoek per jaar' op cbs.nl.

Meer informatie

Heeft u vragen over het werken met microdata? Neem dan contact op via microdata@cbs.nl.

asd@cbs.nl.

Wilt u onderzoek laten uitvoeren door het CBS? Mail dan naar asd@cbs.nl.

Nieuwsgierig naar de onlangs beschikbaar gekomen datasets? Bekijk dan de rubriek 'Sinds kort Beschikbaar' op pagina 4 van dit blad.

Basiskennis Open Science in vier uur

Nieuwe, gratis online cursus EOSC Synergy

Hoe werkt Open Science?

Wat is FAIR-data en waarom is dit belangrijk? De nieuwe online cursus 'Bringing synergy to research and data management in Europe' helpt geïnteresseerden verder op weg. Linas Čepinskas

Onlangs heeft DANS in samenwerking met partners van het EOSC Synergy-project een inleidende cursus over Open Science, de European Open Science Cloud (EOSC) en FAIR data ontwikkeld. De acht modules bevatten video's, quizzen en praktijkgerichte opdrachten, opgesteld door professionals, bedoeld om geïnteresseerden op weg te helpen met innovatieve diensten en benaderingen voor het beheer van onderzoeksgegevens om Open Science en FAIR-data te faciliteren.

Basisbegrip

Onderzoekers, studenten, trainers en data professionals die op zoek zijn naar inleidend lesmateriaal kunnen deze cursus op hun eigen tempo volgen. Aan het einde van de cursus hebben ze een basisbegrip van Open Science en de European Open Science Cloud, en zijn ze bekend met de belangrijkste concepten, praktische tools

In de cursus wordt ingegaan op de voordelen van Open Science.

Credits: Gaelen Pinnock, wikimedia.org/wiki/File:UCT_RDM_Why-Open-Science.png

en best practices voor FAIR-data en datamanagement.

Flexibel leren

De cursus is opgedeeld in verschillende modules die los van elkaar bestudeerd kunnen worden. Het duurt gemiddeld 30 minuten om een module te voltooien. Wie de gehele cursus wil volgen zal er zo'n vier uur aan besteden.

Open materialen

De inhoud van de cursus is beschikbaar onder een open licentie en kan zo gedeeld en hergebruikt worden. Iedereen is uitgenodigd om eventuele opmerkingen, vragen of suggesties via de cursuswebsite te delen.

learn.eosc-synergy.eu/bringing-synergy-to-better-data-management-and-research-in-europe

KORT

CLARIAH call voor maatschappelijke projecten

Nederlandse onderzoekers kunnen tot 1 maart bij CLARIAH een voorstel indienen voor een onderzoeksproject met een commercieel bedrijf of een maatschappelijke partner. De projecten in deze call laten zien dat de CLARIAH-infrastructuur niet alleen bruikbaar is voor wetenschappelijk onderzoek, maar ook een commerciële of maatschappelijke impact kan hebben. Zo werd in een eerdere call automatische taalanalyse ingezet in de psychiatrie en werd gewerkt aan ondersteuning voor docenten in het basisonderwijs bij het beoordelen van de spreekvaardigheid van leerlingen door middel van automatische spraakherkenning. (ER) clariah.nl/nl/call-voor-maatschappelijke-commerciële-projecten

Start Nederlandse collectie binnen World Historical Gazetteer

Plaatsnamen en plaatsbepalingen, in al hun spellings-, taal-, en andere varianten, vormen een continue uitdaging voor historici. De World Historical Gazetteer heeft tot doel om historische plaatsen centraal te ontsluiten via een allesomvattende databank van historische plaatsnamen. Onlangs is het KNAW Humanities Cluster een partnerschap aangegaan met de University of Pittsburgh, de gastinstelling van de World Historical Gazetteer. Samen verbeteren ze whgazetteer.org en start de aanleg van een collectie met wereldwijde plaatsnamen gerelateerd aan de Nederlandse geschiedenis. Deze Dutch Collection is momenteel in ontwikkeling, maar enkele eerste datasets worden reeds aangeboden. (RS) whgazetteer.org

SINDS KORT BESCHIKBAAR

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

CBS

Het CBS stelt bestanden beschikbaar voor statistisch onderzoek. Een volledig overzicht van beschikbare bestanden staat in de microcatalogus op de CBS-website:

<https://www.cbs.nl/nl-nl/onze-diensten/-maatwerk-en-microdata/microdata-zelf-onderzoek-doen/catalogus-microdata>

Ook een link naar de bestanden die de afgelopen periode beschikbaar zijn gekomen staat online:

<https://www.cbs.nl/nl-nl/onze-diensten/-maatwerk-en-microdata/microdata-zelf-onderzoek-doen/overzicht-van-alle-bestanden>

Centerdata

• **Overheidsgegevens, hoe breng je die het beste over?**

In juli 2020 werd de vragenlijst 'Overheidsgegevens weergeven' in het LISS panel afgenomen.

Met dit vignettenonderzoek deed Annemarie Balvert (promovenda Tilburg University) onderzoek naar wat voor overheden de 'beste manier' is om ingewikkelde data weer te geven in hun communicatie naar burgers. In de vragenlijst kregen LISS panelleden fictieve gemeentelijke data over energie- en watergebruik voorgelegd waarbij de vorm een kaart, een tekstuele samenvatting, of een datatabel betrof. Hierover werden ervarings- en begripsvragen gesteld. Een voorlopige conclusie uit het onderzoek is dat deze verschillende vormen van data geschikt zijn voor verschillende doelen van overheidscommunicatie, en dat burgers afhankelijk van hun visies en vaardigheden er bovendien verschillend op reageren. De data van dit onderzoek zijn beschikbaar via het LISS Data Archive.

[lissdata.nl](https://www.lissdata.nl)

DOI: 10.17026/dans-26s-pn6p

Ook sinds kort beschikbaar:

Studies LISS panel

• Van Soest, A., Knoef, M., Muns, S., december 2019, februari 2020, Employers about hiring employees and the role of the retirement policy.

DOI: 10.17026/dans-259-gduw

• Van der Velden, P., maart - april 2020, Victims in Modern Society 2020 - Follow-up measurement.

DOI: 10.17026/dans-27v-paa3

• Von Gaudecker, H.M., mei 2020, Effects of the Outbreak of COVID-19 - part 3.

DOI: 10.17026/dans-x8h-cuwg

• Von Gaudecker, H.M., juni 2020, Effects of the Outbreak of COVID-19 - part 4.

DOI: 10.17026/dans-x4z-eny2

• Yerkes, M., Besamusca, J., van der Zwan, R., André, S., Remery, C., Kruijven, P., Beckers, D., Geurts, S., juli 2020, Gender inequalities in times of the COVID-19 pandemic - Wave 2.

DOI: 10.17026/dans-x46-f8zj

• Von Gaudecker, H.M., september 2020, Effects of the Outbreak of COVID-19 - part 5.

DOI: 10.17026/dans-xuw-pgjt

• Fouarge, D., Künn, A., oktober 2020, Developing for a lifetime.

DOI: 10.17026/dans-zca-ctum

• Bleidorn, W., december 2020, Life Event Study.

DOI: 10.17026/dans-xft-dsbj

• Smeekes, A., december 2020, Nostalgia 2019 - Wave 3.

DOI: 10.17026/dans-zz2-ccv

• Von Gaudecker, H.M., december 2020, Effects of the Outbreak of COVID-19 - part 6.

DOI: 10.17026/dans-zva-nzq

• Centerdata, april - mei 2021, Work and Schooling - Wave 14.

DOI: 10.17026/dans-zvq-wnzp

• Centerdata, juni - juli 2021, Economic Situation: Income - Wave 14.

DOI: 10.17026/dans-zmt-qkd9

• Centerdata, juli - augustus 2021, Economic Situation: Housing - Wave 14.

DOI: 10.17026/dans-z65-jfss

• Centerdata, augustus - september 2021, Religion and Ethnicity - Wave 14.

DOI: 10.17026/dans-2cy-bqjv



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

Ook sinds kort beschikbaar:

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

• Centraal Bureau voor de statistiek (2021): Schoolverlatersonderzoek 2020.

DANS. DOI: 10.17026/dans-zat-5ag8

• Davis, Mr CJ (Leiden University) (2020): International Criminal Law Prosecution Appeals Briefs Database.

DANS. DOI: 10.17026/dans-xwq-ka6y

• Farace, Dr. D. (GreyNet International) (2021): Exploring Next Generation Grey.

DANS. DOI: 10.17026/dans-xrg-2gf6

• Jaarsveld, dr. C.H.M. van (Radboud University); Reukers, D.F.M. (Radboud University) et al (2021): Impact of Q-fever on physical and psychosocial functioning until 8 years after Coxiella burnetii infection.

DANS. DOI: 10.17026/dans-zpa-fkph

• Janssen, Dr. B.D. (Utrecht University) (2021): DIOPTRA-L; Digital Opinions on Translated Literature.

DANS. DOI: 10.17026/dans-zv4-5phn

• Liu, W (University of Groningen) (2021): Dataset for: 'Annual CO2 budget estimation

from chamber-based flux measurement on intensively drained peat meadows: effect of gap-filling strategies'.

DANS. DOI: 10.17026/dans-xtq-ajzs

• Petram, Dr. LO (Huygens Institute) (2021): War dummies: structured data on organized armed confrontations with Dutch involvement, 1566-1812.

DANS. DOI: 10.17026/dans-2cn-smxt

• Pierik, dr. HJ (Utrecht University/Cultural Heritage Agency) (2021): Digitaal basisbestand Laat-Holocene kustontwikkeling (KustGIS).

DANS. DOI: 10.17026/dans-x46-b7d2

• Poorter, Prof. Dr L. (Wageningen University & Research (WUR)) (2021): Functional recovery of secondary tropical forests.

DANS. DOI: 10.17026/dans-zz5-hf3s

• Pouwels, Dr. B.J. (Bureau Pouwels) (2021): Verantwoording van het veldwerk: Bedrijvenmonitor Topvrouwen 2020.

DANS. DOI: 10.17026/dans-xc3-5c7u

• Scholtens, MA J.(Commissariaat voor de Media) (2021): Representatie van vrouwen in Nederlandse non-fictie televisieprogramma's 2019 en 2021.

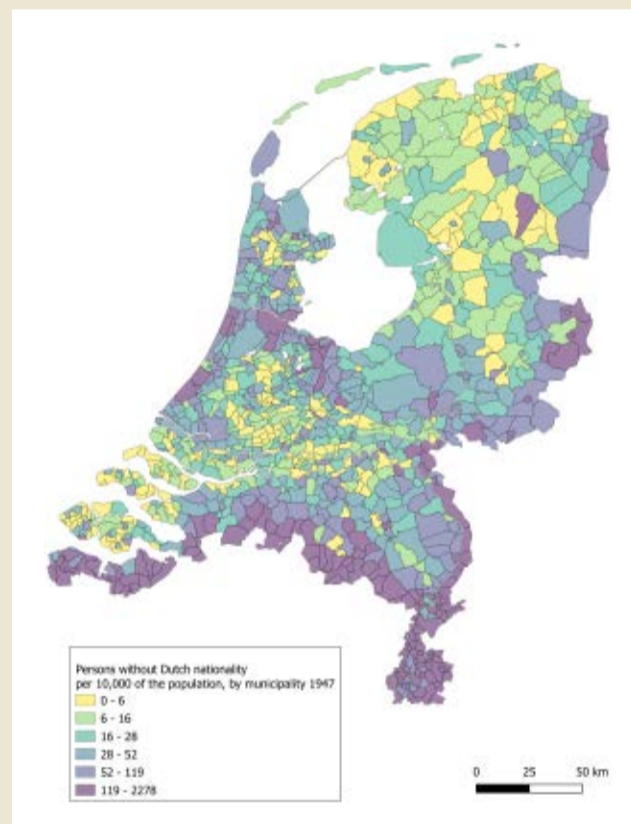
DANS. DOI: 10.17026/dans-z8r-uz4t

• Sun, MSc. X (Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente) (2021): Building outline extraction of Enschede, the Netherlands using aerial images and digital surface models.

DANS. DOI: 10.17026/dans-248-n3nd

• Vicq, A. de; Gelderblom, Prof. dr. O.; Jonker, Prof. dr. J. (2021): Dutch Banking Database, 1880-1940.

DANS. DOI: 10.17026/dans-xre-kfd



Credits: doi: 10.1163/24523666-bja10018

DANS

• **FAIR data van de volkstelling 1947**

Via DANS is een omvangrijke collectie data per gemeente van de Nederlandse Volkstelling 1947 beschikbaar. Onlangs verscheen hierover het artikel 'Detailed Tables from the Dutch Census 1947: Experiences and Lessons Learned in Publishing a Large Dataset' in het Research Data Journal for the Humanities and Social Sciences. Het artikel beschrijft hoe de gegevens (scans van zo'n 30.000 bladen werden met data-entry overgenomen in Excel-bladen) op zo'n manier zijn gepubliceerd, dat nieuw historisch onderzoek mogelijk is. Richtlijn was om de totale dataset FAIR te maken, met speciale aandacht voor de R van Reusable. Het artikel sluit af met een aantal mogelijkheden voor verder gebruik van de dataset.

DOI artikel: 10.1163/24523666-bja10018

DOI dataset: 10.17026/dans-zs3-cf4m



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

KORT

DANS ontwikkelt controlled vocabulary service

Binnen het EU Horizon 2020-project Social Sciences & Humanities Open Cloud (SSHOC) heeft DANS een controlled vocabulary service ontwikkeld voor de dataverse software. Data-verse is open source software voor data repositories, ontwikkeld door het instituut voor kwantitatieve sociale wetenschappen (IQSS) van Harvard University. Inmiddels wordt de software wereldwijd door meer dan 70 installaties gebruikt, waaronder Data-verseNL en de nieuwe DANS Data Stations, en draagt een internationale groep ontwikkelaars bij aan de code. Met de functionaliteit die DANS ontwikkeld heeft, is het mogelijk om externe thesauri en andere begrippenlijsten te koppelen aan de metadatatelden binnen dataverse, wat de kwaliteit van de metadata ten goede komt. De functionaliteit is opgenomen in versie 5.7 van de dataverse software. (MW)

<https://github.com/IQSS/dataverse/releases/tag/v5.7>

Data visualiseert hertogdom Brabant

Hertogdombrabant.nl toont de (geografische) samenstelling van het hertogdom Brabant in de late middeleeuwen. De site ontsluit door middel van historische GIS-kaarten allerlei historische statistieken, zoals bevolkingsomvang, belastingen, politieke representatie en religieuze instellingen. De achterliggende data zijn vrij te downloaden via de IISH Data Collection website. Dit is het resultaat van een samenwerking en financiële bijdragen van diverse partners, zoals het NWO-project Imagining a territory. Constructions and representations of late medieval Brabant (dr. Mario Damen, UvA), het Internationaal Instituut voor Sociale Geschiedenis (dr. Rombert Stapel) en de Fryske Akademy (prof.dr. Hans Mol). (RS)

[hertogdombrabant.nl](https://hdl.handle.net/10622/UOKBYL)
hdl.handle.net/10622/UOKBYL

CoreTrustSeal voor LISS Data Archive

Het LISS Data Archive heeft het CoreTrustSeal (CTS) keurmerk ontvangen. Het CTS is een internationaal keurmerk voor duurzame, betrouwbare en toegankelijke data repositories. Via het LISS Data Archive worden de data van het LISS panel beschikbaar gemaakt voor wetenschappers. De repository ondersteunt de FAIR principes voor vindbare, toegankelijke, uitwisselbare en herbruikbare onderzoeksdata. Steeds vaker vereisen subsidiegevers en tijdschriften dat onderzoeksdata worden opgeslagen in een betrouwbare dataopslagplaats, een zogenaamde 'Trustworthy Digital Repository'. Met de CTS-certificering toont het LISS Data Archive aan dat de repository aan deze eis voldoet. (EdC)

<https://www.lissdata.nl/>

Prof. dr. Marcel Das, directeur van Centerdata:

‘Wij geloven in onderzoek voor en door onderzoekers’

Op 20 december 2021 bestond Centerdata 25 jaar. Een mooi moment voor een blik op de hoogtepunten. Dat doen we met de directeur van Centerdata, prof. dr. Marcel Das.

Sanne Jansen

Wanneer was jouw start bij Centerdata?

“In het vierde jaar na de oprichting kwam ik binnen, dat was in september 2000. De oprichter Arie Kapteyn vertrok voor werk naar Amerika. Zelf werkte ik als onderzoeker bij CentER Applied Research, een instituut voor contractonderzoek van de Katholieke Universiteit Brabant (KUB), zoals Tilburg University eerst heette.”

Wat voor organisatie was het toen?

“In die tijd werkten er ongeveer tien mensen. We waren een hecht team. We moesten de markt nog veroveren en de boer op om projecten binnen te halen. Ik heb zeker in mijn beginperiode veel hulp gekregen van Arie Kapteyn: Centerdata was echt zijn kindje. In het begin was het ook maar de vraag of Centerdata overleefde. Dat vond ik een geweldige uitdaging.”

“De oorsprong van Centerdata ligt bij het Telepanel, toentertijd een innovatief panel. Interviews werden afgenomen zonder tussenkomst van een interviewer. De Universiteit van Amsterdam, waar het Telepanel was ondergebracht, wilde de stekker eruit trekken. Om dat te voorkomen, werd het panel door Arie Kapteyn voortgezet in Tilburg. Arie analyseerde in dat panel het spaargedrag van Nederlandse huishoudens (het Spaaronderzoek). Dat markeerde de oprichting van Stichting Centerdata op 20 december 1996. Na deze overname kreeg het panel de naam CentER-panel.”

Hoe is Centerdata veranderd?

“We zijn gegroeid naar circa 50 mensen. Allerlei technologische ontwikkelingen hebben de organisatie beïnvloed. Zo waren de mogelijkheden van internet in 2000 veel beperkter, waardoor onderzoek doen er echt anders uitzag. De laatste jaren verandert er veel rondom databeveiliging en privacy. Deze onderwerpen hebben we altijd serieus genomen, alleen nu moet het meer gedocumenteerd worden.”

Wat is hetzelfde gebleven?

“De teamspirit is onveranderd. We zijn nog steeds een flexibele organisatie. Bij ons zijn de lijntjes kort. Arie Kapteyn was er 25 jaar geleden al van overtuigd dat het goed was om dataverzameling over te laten aan wetenschappers. Wij geloven ook in onderzoek door onderzoekers, voor onderzoekers. Dat betekent dat er grondig wordt meegedacht welke vragen je stelt en hoe je ze stelt.”

Wat maakt Centerdata uniek?

“Ons hechte team en het feit dat er bij ons



INTERVIEW

‘Centerdata denkt mee: welke vragen stel je en hoe stel je ze’

diverse onderzoeksdisciplines samen komen: survey-, beleids- en consumentenonderzoek. Ook kan je bij ons terecht voor toepassing van data science technieken en softwareontwikkeling op maat.”

Wat zijn hoogtepunten uit 25 jaar Centerdata?

“Ons eerste hoogtepunt was in 2001, het eerste boekjaar waarin alle kosten werden gedekt. Heel belangrijk voor het voortbestaan van het instituut. In 2002 zijn we gevraagd software te ontwikkelen voor SHARE (Survey of Health, Ageing and Retirement in Europe), een internationaal onderzoek naar vergrijzing. Hierdoor gingen onze IT’ers ook extern een rol van betekenis spelen. En nu - 20 jaar later - zijn we nog steeds verantwoordelijk voor alle software voor SHARE. We hebben diverse geavanceerde applicaties gebouwd. Onze software wordt ook in andere grote Europese studies gebruikt voor de ontwikkeling van vragenlijsten en support van veldwerk. Dit maakt dat we een internationale speler zijn geworden op het gebied van toolontwikkeling.”

“In 2003 hebben we met de inzet van een NWO-middelgroot subsidie een succesvolle

update kunnen uitvoeren aan het Spaaronderzoek. Ook haakte dat jaar De Nederlandsche Bank (DNB) aan bij het Spaaronderzoek. Ik ben er trots op dat DNB zich tot de dag van vandaag aan ons verbindt.”

“In 2005 kwam de afdeling Kwantitatieve Analyse erbij, met de focus op de onderzoeksgebieden arbeidsmarkt, onderwijs en pensioenen. Later is daar onderzoek naar consumentenvraagstukken aan toegevoegd. En nog wat jaren later het werkveld data science waarbij met de nieuwste technieken ongestructureerde data wordt uitgeplozen.”

“Ik kan nog wel even doorgaan. Zo kregen we in 2006 goedkeuring voor een omvangrijke subsidie waardoor we het LISS panel hebben kunnen bouwen. Dat was echt een knaller. We waren bijzonder trots dat onze aanvraag werd gehonoreerd in concurrentie met de aanvragen uit de bètawetenschappen.”

“De oprichting van ODISSEI in 2016 is ook een hoogtepunt te noemen. Centerdata speelt in dit platform een belangrijke rol. Via ODISSEI hebben onderzoekers toegang tot grootschalige, longitudinale dataverzamelingen en innovatieve en diverse vormen van data.”

“Ik had een persoonlijk hoogtepunt in 2009. In dat jaar ben ik benoemd tot hoogleraar Econometrie en Dataverzameling aan Tilburg University. Ik beschouw de inrichting van die leerstoel als erkenning voor het vak van dataverzameling.”

Waar ben je trots op?

“Ik ben heel trots op de mensen die ons team vormen. Ook ben ik nog steeds trots dat wij in 2006 als klein instituut in staat waren om zo’n grote subsidie binnen te halen waardoor we het LISS panel konden oprichten. Dat heeft de data-infrastructuur voor de wetenschap een boost gegeven. En ik ben trots op het feit dat we het initiatief van Arie Kapteyn, onze foun-

Directeur Marcel Das is vooral trots op het hechte team. Ook roemt hij het feit dat bij Centerdata diverse onderzoeksdisciplines samenkomen en dat Centerdata data science en softwareontwikkeling op maat biedt. Credits: Bart van Vlijmen

ding father, verder hebben uitgebouwd tot een waardevol instituut.”

Wat zijn jouw dromen voor Centerdata?

“Ik hoop dat Centerdata een belangrijke rol blijft spelen op het gebied van wetenschappelijk verantwoorde dataverzameling, softwareontwikkeling en beleidsonderzoek met innovatieve experimenten en state-of-the art technieken, nationaal en internationaal. En dat we daarbij een partij blijven die maatschappelijke betrokkenheid als een kernwaarde beschouwt.”

<https://www.centerdata.nl/>

Marcel Das

Prof. dr. Marcel Das studeerde cum laude af in de Algemene Econometrie / Statistiek (Tilburg University) en is aansluitend gepromoveerd aan dezelfde universiteit. Hij werkte als onderzoeker bij CentER Applied Research, heeft een groot aantal onderzoeksprojecten geleid en publiceerde wetenschappelijke artikelen. Sinds 2000 is hij directeur van Centerdata; sinds 2009 hoogleraar Econometrie en Dataverzameling aan Tilburg University. Ook is hij lid van de Management Board van ODISSEI en van verschillende wetenschappelijke adviesraden van omvangrijke (inter)nationale data-infrastructuren. Vragen stellen aan Marcel kan via marcel.das@centerdata.nl.

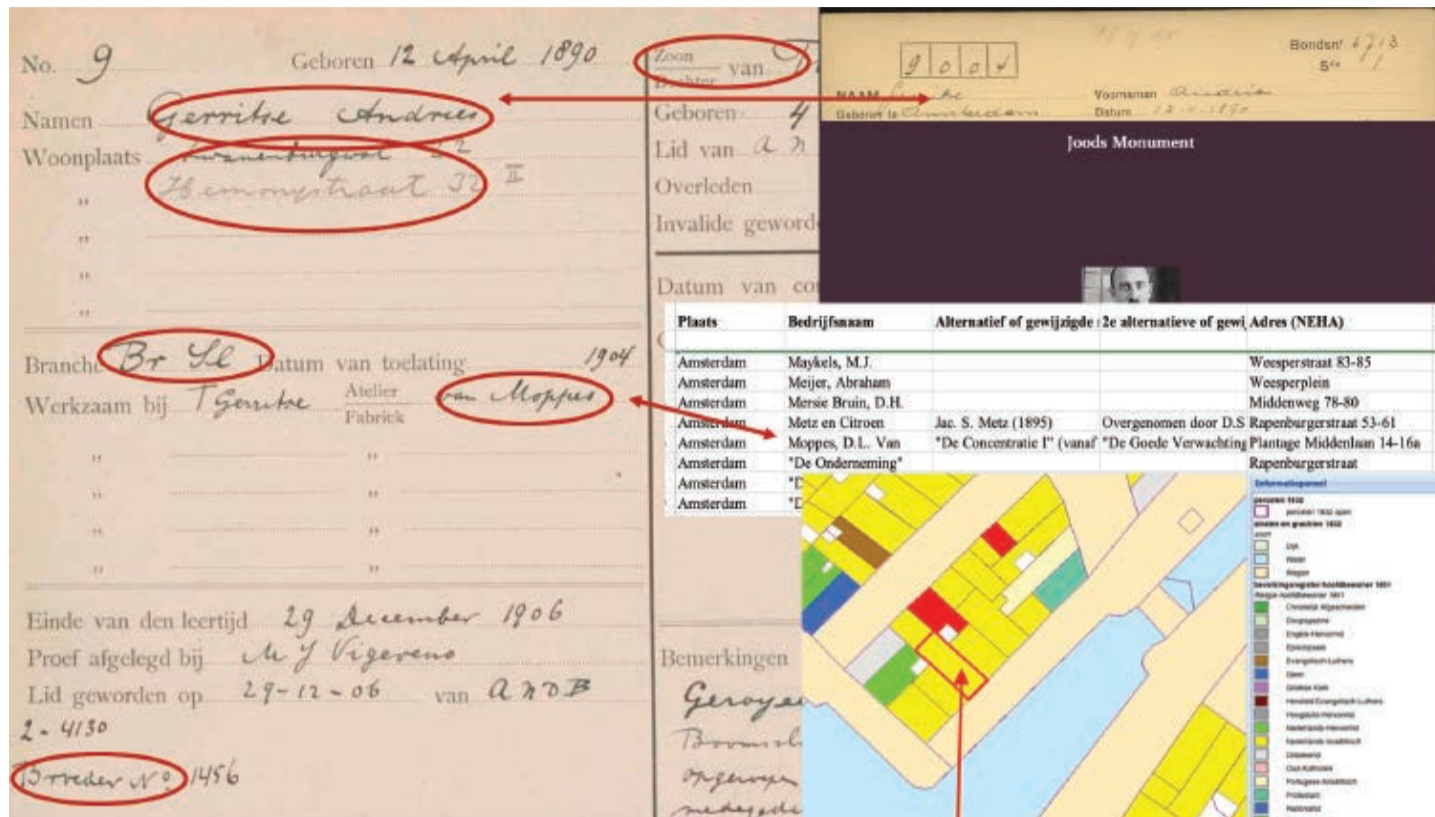
Linked data geven context aan diamantbewerkerbondarchief

Diamantbewerkerbond tot leven gebracht

Het IISG heeft de gedigitaliseerde lidmaatschapskaarten van diamantbewerkerbond naar linked data omgezet en verbonden met andere bronnen.

Karin Hofmeester en Eric de Ruijter

Het archief van de Algemene Nederlandse Diamantbewerkerbond (ANDB) is in 2017 via Metamorfoze gedigitaliseerd. Hierdoor werd het beter behouden en makkelijker online te raadplegen. Het is een bijzonder archief omdat het 28.000 leden- en leerlingkaarten bevat met gegevens over familieleden in het vak, adressen, migratie en bonduitkeringen. De ANDB was een succesvolle bond die als eerste de achturige werkdag en betaalde vakantie wist te realiseren. Doel van het project was op basis van de ANDB-kaarten, aangevuld met automatisch gelinkte gegevens uit andere bronnen, elke diamantbewerker tot leven te brengen.



Deze afbeelding laat een ANDB-ledenkaart zien met links naar andere materialen. Credits: diamantbewerkerbond.nl/afbeelding-andb-lod.png

Van ruw naar bruikbaar

Als eerste stap zijn de kaarten naar een ruwe dataset omgezet door vrijwilligers van Vele Handen. Vervolgens moest hier bruikbare data van gemaakt worden, dat bleek een tijdrovende klus: schrijfwijzen van adressen en namen varieerden aanzienlijk en ze waren voor de vrijwilligers soms slecht te lezen of werden niet juist overgenomen. Met de tool OpenRefine was het mogelijk een redelijk schoon CSV-bestand te creëren. Dit is naar linked data

omgezet op basis van schema.org, met gebruik van de Time- en Organization-ontologieën voor het modelleren van de vakbond en de tijdsaspecten van woonadres en lidmaatschap.

Linked dataset

Vervolgens konden links worden aangebracht met de Oorlogslevens-website omdat zeventig procent van de ANDB-leden Joods was, met HisGis-punten voor Amsterdam om adressen

te lokaliseren, en met de Diamantencyclopedie als LOD om hun werkzaamheden te verklaren. Het resultaat is een linked dataset, gepubliceerd op het Druid Linked Data Platform van het IISG. Daar kunnen onderzoekers zelf SPARQL-queries opstellen, wordt met datastories het datamodel gevisualiseerd en interessante zoekresultaten weergegeven. Hiernaast kan het grote publiek via een laagdrempelige Drupal-website op basis van Sponkedesk de data doorzoeken. De website toont

de gegevens per ANDB-lid, inclusief een kaartje met de opeenvolgende woonadressen en links naar andere bronnen.

Of de diamantbewerkerbond tot leven zijn gekomen? Onlangs meldde een mevrouw dat ze via de website de lidmaatschapskaart van haar vermoorde opa had gevonden. Ze was ontroerd iets tastbaars van hem te vinden dat iets liet zien over zijn werkzame leven en zijn rol in de ANDB.

diamantbewerkerbond.nl

Transparant, toegankelijk, herbruikbaar en duurzaam

Kennisgraaf voor VOC-datasets

Het aantal digitale bronnen dat beschikbaar is voor onderzoek is enorm. Om datasets snel te kunnen doorgronden, vormen kennisgrafen een oplossing. Mathilde Jansen

Binnen de digital humanities werken onderzoekers vaak met meerdere gedigitaliseerde datasets. Daarbij zijn ze veel tijd kwijt met het verzamelen, interpreteren en gelijktrekken van verschillende bronnen. Niet alleen is dit een tijdrovende bezigheid, het zorgt er ook voor dat bronnen op verschillende manieren geïnterpreteerd worden. Om deze problemen te ondervangen, kunnen kennisgrafen worden ingezet. In een kennisgraaf worden data zodanig gemodelleerd en gekoppeld, dat ze voor mensen en computers snel te doorgronden zijn.

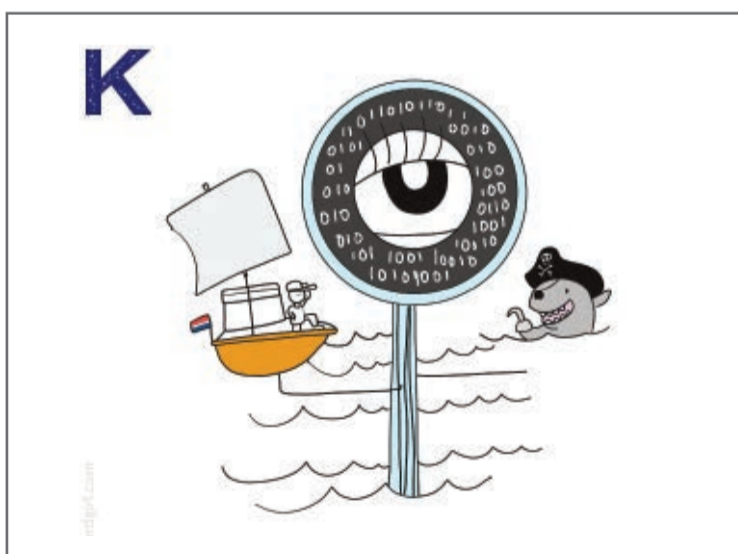
Datasets van de VOC

Het idee van kennisgrafen is niet nieuw. De kennisgrafen die reeds ontwikkeld zijn, hebben echter de nodige beperkingen. Daarom ontwikkelde masterstudent Stijn

Schouten onder begeleiding van Victor de Boer (Vrije Universiteit Amsterdam), Lodewijk Petram (Huygens Instituut) en Marieke van Erp (KNAW HuC) een kennisgraaf die gestoeld is op de principes transparantie, toegankelijkheid, herbruikbaarheid en duurzaamheid. De kennisgraaf werd in eerste instantie ontwikkeld voor vier grote VOC-datasets, onder beheer van het Nationaal Archief en het Huygens Instituut voor Nederlandse Geschiedenis.

Kennis in brokjes

Om een kennisgraaf te ontwikkelen, wordt gebruikgemaakt van een ontologie: een datamodel waarin alle concepten en entiteiten en hun onderlinge verhoudingen zijn vastgelegd. In de VOC-datasets zijn die entiteiten bijvoorbeeld personen (opvarenden), scheepslading en plaatsen waartussen gereisd wordt. Al deze entiteiten krijgen een unieke code. Ook de relaties worden uitgedrukt in unieke codes via Resource Description Framework (RDF). RDF is ontwikkeld in het Semantic Web-onderzoeksveld en het meest gangbare format voor



Afbeelding gemaakt tijdens de elfde editie van de internationale conferentie over Knowledge Capture #kcap2021.

Credits: www.whiteboardgirl.com

Linked Data. Hierin worden brokjes kennis opgeslagen als 'triples', die bestaan uit een subject, object en predicaat, bijvoorbeeld: De VOC (subject) verhandelde (predicaat) specerijen (subject). Deze manier van modelleren geeft een grote flexibiliteit wat betreft het toevoegen en koppelen van nieuwe bronnen en kennis. Schouten: "Ons gebruikersonder-

zoek laat zien dat deze kennisgraaf bruikbaar is. Bovendien zorgen transparantie en toegankelijkheid ervoor dat onderzoekers voort kunnen bouwen op deze kennisgraaf, maar ook kunnen teruggaan in het proces om bijvoorbeeld andere beslissingen te nemen."

<https://dutchshipsandsailors.nl>

DOI:<https://doi.org/10.1145/3460210.3493548>

Project 'Coordinate' van start

Om het welzijn van kinderen in Europa goed te kunnen monitoren, is longitudinale surveydata belangrijk.

In het COORDINATE-project (Cohort community Research and Development Infrastructure Network for Access Throughout Europe) wordt de basis gelegd om een cohortstudie op te zetten. Zo wordt de wetenschappelijke infrastructuur opgezet en een pilot-vragenlijst voor ouders en kinderen gelanceerd in vijf landen.

Centerdata is onder andere betrokken bij de tooling en ontwikkeling van de survey en surveysoftware en bij het opzetten van datadiseminatie en remote data-access.

Het project wil ook het gebruik van bestaande data stimuleren. Zo zullen Centerdata en andere partijen gastonderzoekers ontvangen die onderzoek willen doen met hun data. Op de projectwebsite kan men zich inschrijven voor aankomende webinars.

coordinate-network.eu

Unieke database

Multigenerationeel onderzoek naar Surinaamse bevolking

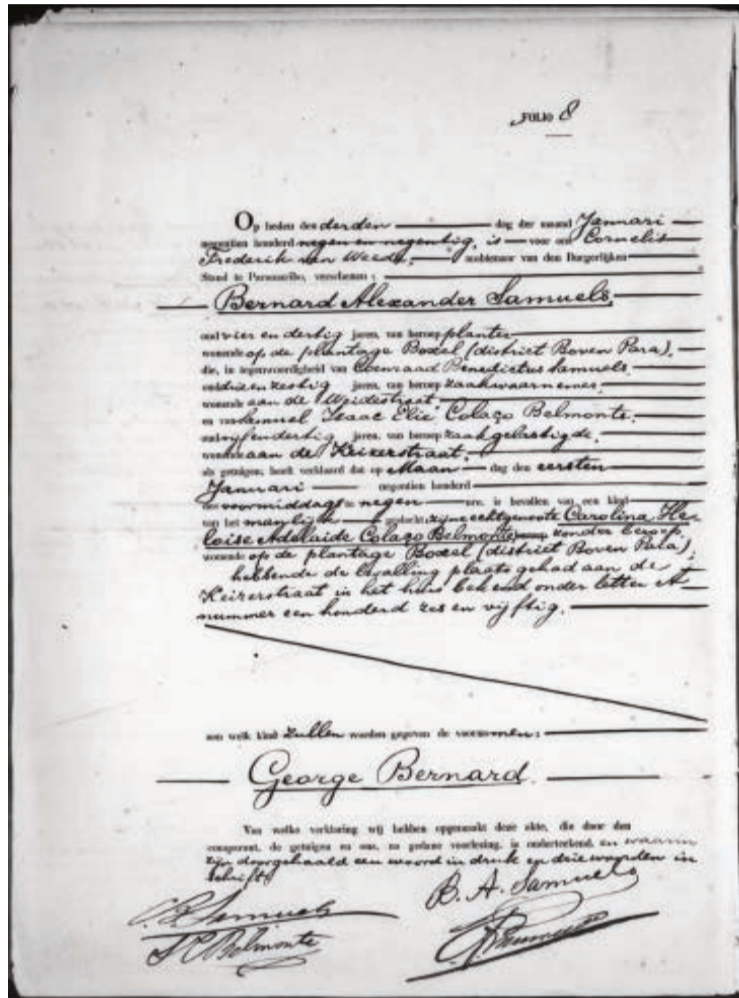
Multigenerationeel onderzoek naar de bevolking van Suriname wordt mogelijk door de koppeling van slavenregisters-data met de burgerlijke stand.

Suze Zijlstra

Lange tijd was het moeilijk om grootschalig historisch onderzoek te doen naar de bevolking van Suriname, omdat er nog geen goede database beschikbaar was met gegevens uit de burgerlijke stand. Het project 'Legacies of bondage: toward a database of Surinamese life courses in a multigenerational perspective (1830-1950)' brengt hier verandering in. Het bouwt voort op het project 'Maak de Surinaamse Slavenregisters openbaar' dat met veel vrijwilligers een database opzette over de slaafgemaakte mensen in Suriname en Curaçao tussen 1830 en de formele afschaffing van de slavernij in 1863.

Historische complexiteit

Bijna vierhonderd vrijwilligers dragen sinds afgelopen september bij aan het nieuwe project. Via het on-



Dit voorbeeld laat een Surinaamse geboorteakte zien. Vrijwilligers zorgen voor het digitaliseren van dit bestand via hetvolk.org.

Credits: Scan Nationaal Archief.

line platform hetvolk.org voeren zij gegevens uit de burgerlijke stand van Suriname van 1828 tot 1950 in. Uitgaande van zo'n 800 scans per dag is de verwachting dat het invoeren ongeveer 2,5 jaar gaat duren. Hiermee kunnen de levens in kaart worden gebracht van vrije mensen en contractarbeiders. De combinatie met de al bestaande slavenregisters-database maakt het mogelijk om straks multigenerationeel onderzoek te doen. Projectleider Coen van Galen (Radboud Universiteit): "Met deze database kunnen we onderzoeken hoe levensomstandigheden van slaafgemaakte mensen doorwerken in de generaties na de afschaffing van slavernij. Dit kan je nergens anders in de wereld doen."

De database komt openbaar beschikbaar. Bovendien worden de data zo veel mogelijk aan elkaar gekoppeld. Dat is ingewikkeld, vertelt Van Galen: "Mensen die in slavernij leefden, hadden geen achternaam, die kregen ze pas bij hun vrijlating." Een bijkomende uitdaging is de naamgeving van contractarbeiders, vervolgt hij: "Zij worden in een Europees stamien van het bevolkingsregister geperst, terwijl bijvoorbeeld

Hindoestaanse arbeiders vaak alleen een voornaam hadden. Europese ambtenaren gebruikten dan het contractnummer, later voerden zij vaak achternamen in op basis van de naam van de vader. Bij dergelijke verschuivingen gaat het dus om veel meer dan alleen verschillende schrijfwijzen."

Vergelijkend potentieel

Het project is een samenwerking tussen universiteiten en archieven in Suriname en Nederland. In 2020 heeft het financiering ontvangen van PDI-SSH (Platform Digitale Infrastructuur Social Sciences & Humanities). Ook in 2021 heeft PDI-SSH een vergelijkbaar project voor de bevolkingsregisters van Curaçao gehonoreerd. Voor onderzoek is dit cruciaal, licht Van Galen toe: "De levensomstandigheden in deze twee samenlevingen lopen uiteen. Waar in Suriname veel plantages waren, was Curaçao meer gericht op handel. In de toekomst maken deze projecten dus uniek vergelijkend onderzoek mogelijk."

ru.nl/rich/our-research/research-groups/radboud-group-historical-demography-family-history/

Menno Rasch, directeur Digitale Infrastructuur KNAW Humanities Cluster:

'Meerwaarde van IT'ers voor onderzoek is groot'

Sinds juni 2021 is Menno Rasch directeur van de afdeling Digitale Infrastructuur bij het KNAW Humanities Cluster. E-data sprak hem over de samenwerking tussen IT'ers en onderzoekers. Thijs van der Veen

Menno Rasch studeerde Informatica en Filosofie in Utrecht. Bij het HuC zit hij helemaal op z'n plek. "Ik heb affiniteit met de humanities. En ik mag werken aan een infrastructuur die naadloos aansluit bij wat onderzoekers nodig hebben. Het ontwikkelen van een digitale infrastructuur en research tools maakt nieuw en beter onderzoek mogelijk, dat je ook nog sneller kunt uitvoeren."

Specifieke wensen

Samen met zijn team probeert hij specifieke wensen van een onderzoeker zo generiek mogelijk op te lossen. "Dat is dé manier om herbruikbaarheid te vergroten. Wanneer we met een generieke basis in korte tijd iets goeds kunnen neerzetten, dan kunnen we de overgebleven projecttijd gebruiken om specifieke diensten te ontwikkelen." Hiervoor zijn veel gesprekken tussen IT'ers en onderzoekers nodig. Rasch: "Samen creëren zij mooie dingen,

wat kan leiden tot baanbrekend onderzoek. IT'ers kunnen echt een grote impact hebben op het onderzoek."

"Binnen het HuC wordt strategisch slim gehandeld. De instituten opereren niet als eilandjes, maar bundelen krachten. De IT'ers van de instituten zijn bij elkaar gezet, men maakt gebruik van elkaars infrastructuur. Het werkt efficiënter en effectiever. Uiteindelijk groeien we door naar een gemeenschappelijke infrastructuur. Je zou de infrastructuur zelfs kunnen opschalen naar de rest van de humanities."

Nationale infrastructuur

Het HuC heeft niet de financiering om de infrastructuur voor de Humanities in Nederland in haar eentje te realiseren. "Dat zou een heel overspannen ambitie zijn, denk ik. Je kan je afvragen of de drie instituten zijn opgericht om een landelijke infrastructuur te bouwen. Het is al heel wat om een stukje van dat geld bij elkaar te leggen om zo een bijdrage te leveren aan een nationale infrastructuur."

Volgens Rasch kan het financieringsvraagstuk alleen nationaal worden opgelost. "De krachtenbundeling van drie KNAW-instituten in het HuC kan uiteraard als voorbeeld dienen.



Menno Rasch

Ook de samenwerking tussen CLARIAH en ODISSEI biedt nieuwe kansen. Omdat het zicht groeiende is dat er geen harde grenzen meer zijn tussen collectie, data en infrastructuur, ligt een verdere samenwerking ook echt voor de hand."

huc.knaw.nl/digital-infrastructure

AGENDA

Maandelijks • online

SoDa Data Drop-in

Informele vragensessies voor datage-relateerde onderzoeksvragen.
odissee-data.nl

8 - 10 maart • online

Conferentie Open Science

Een forum voor onderzoekers, bibliothecarissen, beleidsmakers en andere belanghebbenden over Open Science-ontwikkelingen.
open-science-conference.eu

10 - 12 mei • online

SURF Research Week

Nationale conferentie over onderzoek, ICT en innovatie.
surf.nl/agenda/surf-research-week

20 - 23 juni • online

RDA-bijeenkomst

Het thema van deze 19e RDA-bijeenkomst is: 'Data to improve our world'.
rd-alliance.org/plenaries

7 - 10 juni • online

IASSIST 2022

Jaarlijkse conferentie voor dataprofessionals over datamanagement.
iassist2022.org

13 - 16 juni • online

17th International Digital Curation conference

De 17e editie van de internationale Digital Curation conferentie.
dcc.ac.uk/events/idcc2022

De nieuwe nationale supercomputer, in beheer van SURF, is onlangs onthuld. In de toekomst zal deze ook worden gebruikt voor de ODISSEI Secure Supercomputer (OSSC).

Suze Zijlstra

Afgelopen september is de nieuwe nationale supercomputer Snellius in het Amsterdam Science Park officieel geopend door koningin Máxima. De door SURF beheerde supercomputer heeft al drie keer zoveel rekenkracht als de vorige supercomputer Cartesius. Door geplande uitbreidingen zal deze rekenkracht in de komende jaren nog verder toenemen. Snellius is nu al in staat om 6 miljard berekeningen per seconde te maken. Het is een van de meest krachtige high-performance en computing-systemen in Nederland. Naast verbeterde prestaties heeft SURF bij de keuze voor de nieuwe supercomputer ook veel aandacht besteed aan verlaging van het energiegebruik.

OSSC

Momenteel draait de ODISSEI Secure Supercomputer (OSSC) nog op de oude supercomputer Cartesius, maar die zal in de nabije toekomst worden overgezet naar Snellius. Voor gebruikers van de OSSC, die op grote schaal in een streng beveiligde omgeving data kan analyseren, heeft deze overstap verschillende voordelen. De snellere nodes zorgen voor de genoemde toename in rekenkracht.



Afgelopen september is de nieuwe nationale supercomputer Snellius in het Amsterdam Science Park officieel geopend door koningin Máxima. Credits: Vera Duivenvoorden

Ongekende rekenkracht maakt vernieuwend sociaalwetenschappelijk onderzoek mogelijk

Nieuwe supercomputer Snellius onthuld

In de nieuwe omgeving hebben gebruikers toegang tot de nieuwste hardware technologie en mogelijkheden. Bovendien is het nog beter mogelijk om binnen de beveiligde omgeving data te traceren en is het gemakkelijker om data te migreren en te verplaatsen.

Omgevingsfactoren

De nieuwe supercomputer is ook zeer geschikt voor machine learning. Tot slot zal de nieuwe supercomputer waarschijnlijk tot meer gemak leiden voor de OSSC-

gebruikers die in deze omgeving hun onderzoek willen opzetten.

Een van de OSSC-projecten die op de nieuwe supercomputer zal worden overgezet, is een onderzoeksproject van het RIVM. Dit project richt zich op de invloed van milieurisico's en omgevingsfactoren op de gezondheid van mensen. Milieurisico's zijn bijvoorbeeld luchtverontreiniging en geluid, maar omgevingsfactoren kunnen ook positieve factoren zijn, zoals veel natuur in de omgeving. Voor dit onderzoek worden verschil-

lende enorme datasets gebruikt, waaronder registerdata en milieudata van het RIVM. De grote hoeveelheid data, die binnen de streng beveiligde omgeving kunnen worden gekoppeld, vragen om de grote rekenkracht die de Snellius biedt. Hiermee kan worden geanalyseerd hoe verschillende factoren invloed hebben op de gezondheid van inwoners van Nederland.

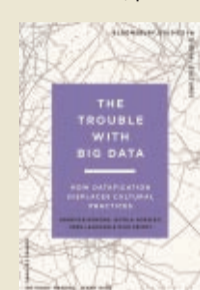
<https://visualization.surf.nl/snellius-virtual-tour>

GELEZEN

The Trouble with Big Data - How Datafication Displaces Cultural Practices

Jennifer Edmond, Nicola Horsley, Jörg Lehmann, Mike Priddy

Een van de grootste drijfveren van technologische ontwikkelingen en innovatie is big data. Terwijl de term big data inclusief en alomvattend klinkt, zijn big data-aanpakken vaak selectief en worden bronnen uitgesloten die niet effectief gestructureerd, gerepresenteerd of zelfs gedigitaliseerd kunnen worden. 'The Trouble with Big Data' onderzoekt de uitdagingen die onze maatschappij tegemoet gaat met big data. Het boek kijkt naar big data vanuit een cultuurperspectief, in plaats van sociale, politieke en economische



trends, in de woorden die we gebruiken, de waarden die onze interacties beïnvloeden en de aannames die ons bewegen. Voortbouwend op onderzoek

uitgevoerd binnen het Knowledge Complexity (KPLEX)-project, waarin DANS een van de partners was, richten de auteurs zich op gebieden zoals data en taal, data en zingeving, data en macht, data en onzichtbaarheid, en aggregatie van big data. Hoe culturele praktijken verdrongen worden maar tegelijkertijd weerstand bieden tegen massale 'datafication', kan leerzaam zijn voor de kritische benadering van big data-onderzoek en -innovatie. Het boek is open access beschikbaar via het Bloomsbury Open Programme, gefinancierd door het Trinity College Dublin, DARIAH-EU en de Europese Commissie.

kplex-project.com
doi.org/10.5040/9781350239654

COLUMN

Dromen over open data en algoritmes

Toen ik in 1991 begon aan mijn studie in Leiden, begon iemand in Finland met het schrijven van een nieuw operating systeem. Gewoon voor de lol. De broncode was vanaf het begin openbaar. Anderen waren van harte welkom om mee te doen. Vandaag de dag draait 96% van de grootste servers wereldwijd op Linux. In 1999 ging Red Hat naar de beurs. Zij bewezen daarmee dat het onmogelijke mogelijk is: een bedrijf bouwen op open source software en toch geld verdienen.

Net zoals mijn voorganger in deze column, droom ik van een wereld waarin data niet wordt gezien als een commodity maar als een common pool resource. Een

wereld waarin de broncode van algoritmes openbaar is en daarmee toetsbaar voor peer review. Kortom, een wereld waarin de hacking cultuur samenkomt met die van Elinor Ostrom. Helaas staan tussen droom en daad wetten in de weg, en praktische bezwaren.

Of kan het toch? Droom eens met mij mee: we leggen voor de hele wereld een openbare database aan van beelden van borstkankercellen van zogenaamde HER2-positieve borstkanker. Deze vorm van kanker komt bij 19 op de 100.000 vrouwen voor en is met relatief goedkope beeldvormende diagnostiek vast te stellen. Omdat de database enkel fo-

to's van cellen bevat met een label kwaadaardig/goedaardig zijn er ook weinig belemmeringen wat betreft privacy. Deze database wordt vervolgens publiek beschikbaar gesteld zodat iedereen zelf een algoritme kan ontwikkelen. De enige verplichting is dat je het resultaat - het getrainde model, de broncode, de methodische verantwoording - ook weer deelt.

Ik stel mij voor dat een dergelijke aanpak binnen afzienbare termijn goede en robuuste algoritmes zal opleveren. Vanwege de volledige openheid van het ontwikkelproces, kan ook meteen worden voldaan aan een aantal belangrijke eisen van transparantie, uitlegbaarheid en reproduceer-

baarheid voor hoog-risico toepassingen van algoritmes. Commerciële partijen kunnen zich via een open source business model richten op dienstverlening op het gebied van certificering, integratie met bestaande software systemen en het beheren van de levenscyclus van de modellen.

De werkelijkheid: op 30 september 2021 ontving de Hippo Found-

dation de Duitse AI startup prijs 2021. Bovenstaand voorbeeld is letterlijk de eerste aandoening die zij via viktorianezero.org aan het bouwen zijn. Gelukkig ben ik niet de enige die durft te dromen.

Daniel Kapitan

Daniel Kapitan (1973) is opgeleid als gepromoveerd fysicus aan Oxford University en inmiddels meer dan 20 jaar actief op het gebied van data science. Recentelijk heeft hij de dataverbinders opgericht, een coöperatie van zelfstandige professionals die zich richten op mensgericht gebruik van data. Daarnaast is hij Academic Director bij JADS.

Hij geeft de column door aan Joop Arends.

