



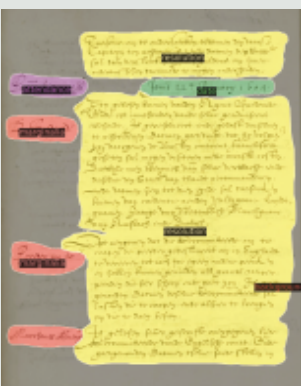
E-DATA & RESEARCH

Jaargang 17 | nummer 3

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

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: CBS, Centerdata, CLARIAH, DANS, KNAW Humanities Cluster, ODISSEI en RKD.

INHOUD



3
Open source transcriptie-software

4
Beschikbare databestanden

5
Kennismaken met TDCC

6
Vijf jaar na invoering van de AVG

8
Joris van Eijnatten over digitalisering van de geesteswetenschappen

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

Het onderzoeksplatform Van Gogh Worldwide

Collecties verbonden via Linked Open Data

Van Gogh Worldwide is hét wetenschappelijke onderzoeksplatform voor kunsthistorisch en technisch onderzoek naar de werken van Vincent van Gogh (1853-1890). *Jessie Fransen, Marjon van Schendel en Lian Wintermans*

Het doel van het platform is om kunsthistorisch onderzoek naar het werk van Vincent van Gogh toegankelijker en duurzamer te maken voor wetenschappers en onderzoekers, door onderzoeksdata in relatie tot elkaar beschikbaar te stellen. Van Gogh Worldwide is het belangrijkste platform voor het verbinden, presenteren en delen van kennis en onderzoek naar het werk van Van Gogh. Naast de objectgegevens van elk kunstwerk bevat het platform ook de herkomst-, tentoonstellings-, veiling- en literatuurgeschiedenis naast materiaaltechnische documenten of referenties.

Gezamenlijk initiatief

Van Gogh Worldwide is een gezamenlijk initiatief van de founding partners, het RKD – Nederlands Instituut voor Kunstgeschiedenis, het Van Gogh Museum en het Kröller-Müller Museum, en wordt gerealiseerd met een groot aantal partners over de hele wereld, waaronder musea, onderzoeksinstituten en particulieren. Dit digitale onderzoeksplatform is eind 2020 gelanceerd. Tijdens de eerste fase van het project zijn Nederlandse partners, musea en particuliere collecties met een Van Gogh in hun collectie, aangesloten. Momenteel bevinden we ons in de tweede fase en worden de kunstwerken uit internationale collecties aan het platform toegevoegd. In september



Vincent van Gogh, A Peasant Woman Digging in Front of Her Cottage, c. 1885.

Credits: The Art Institute of Chicago

2023 verwachten we data van alle bekende werken te hebben opgenomen.

Linked Data

Van Gogh Worldwide is het eerste platform binnen de museale en culturele wereld dat de data van kunstwerken publiceert in lijn met de Nederlandse Nationale Strategie Digitaal Erfgoed en de DERA, wat inhoudt dat de eigenaar van de gegevens verantwoordelijk blijft voor de kwaliteit van de gegevens op het platform en dat de gegevens worden aangeboden als Linked Open Data. Voor Van Gogh Worldwide is gekozen voor The Linked Art Data Model, dat speciaal van toepassing is op de kunsthistorische branche. De gegevens die deelnemers aanleveren worden als Linked Data gepubliceerd en met

elkaar verbonden door het gebruik van thesauri (zoals de Getty Art and Architecture Thesaurus (AAT), de Getty Thesaurus of Geographic Names (TGN) en RKDartists). Door het gebruik van Linked Data en de thesauri is alle informatie op het platform betrouwbaar en altijd up to date. Voor de deelnemers die zelf (nog) geen Linked Data kunnen aanleveren is een speciale template ontwikkeld, waarin de data van de werken opgenomen kunnen worden, die vervolgens worden omgezet in Linked Data. Door de gegevens van musea te publiceren via The Linked Art Data Model heeft Van Gogh Worldwide een wereldwijde samenwerking tot stand gebracht met meerdere partners die zich verder zal ontwikkelen in de nabije toekomst.

<https://www.vangoghworldwide.org>

Sneller sociaalwetenschappelijk onderzoek verbeteren:

Het potentieel van benchmarks

Sociale wetenschappers hebben veel methoden om hun theorieën te testen, maar hoe weet je welke de beste is? Het ODISSEI Social Science benchmarking team verkent nieuwe manieren om te vergelijken. *Paulina Pankowska*

Sociale wetenschappers willen de maatschappij verklaren. Voor elk sociaal fenomeen hebben ze veel theorieën ontwikkeld om sociale mechanismen uit te leggen. Traditioneel werden deze theorieën getest door deze in statistische modellen te vertalen en toe te passen op verschillende databronnen. Deze benadering heeft echter beperkingen. De voornaamste beperking is dat het kan leiden tot verschillende methoden en modellen. Hoewel deze modellen allemaal plausibel zijn en goed onderbouwd, leiden ze vaak tot verschillende resultaten. Momenteel is er geen framework om deze modellen te vergelijken. De vraag welk model beter werkt onder welke omstandigheden is dus nog onbeantwoord. Daardoor is het

moeilijk om conflicterende theorieën te evalueren.

Het Social Science benchmarking team van ODISSEI, de infrastructuur voor sociale wetenschappen, verkent hoe 'benchmarks' kunnen worden gebruikt als zo'n standaard referentiekader. Benchmarks hebben een groot potentieel voor het beantwoorden van al langer bestaande vraagstukken in het veld en kunnen het veld daarmee vooruit brengen.

Gestandaardiseerd framework

Een benchmark is een gestandaardiseerd framework dat directe vergelijkingen mogelijk maakt van verschillende modellen die hetzelfde onderzoeksprobleem proberen te beantwoorden. Gewoonlijk wordt een 'benchmark challenge' georganiseerd waarbij deelnemers proberen om een bepaalde uitkomst te voorspellen aan de hand van een specifieke dataset. De verschillende modellen die door deelnemers worden ingediend, worden vervolgens geëvalueerd en vergeleken

(met vooraf bepaalde criteria) aan de hand van een deel van de data die de deelnemers niet tot hun beschikking hadden. Het gebruik van benchmarks heeft al geleid tot doorbraken in veel wetenschappelijke velden, waaronder computer- en datawetenschap, natuurkunde, biomedische wetenschap en de geesteswetenschappen.

Veelbelovend middel

Tot nu toe is het gebruik van benchmarks in de sociale wetenschappen beperkt geweest. Er zijn echter wel enige challenges georganiseerd in de afgelopen jaren. Het ODISSEI benchmarking team gebruikt de lessen van een grootschalige challenge uit de Verenigde Staten, in combinatie met ervaringen opgedaan tijdens de challenge die uitgevoerd is tijdens de SICSS-ODISSEI Summer School 2022, om beter inzicht te krijgen in het gebruik van benchmarking in de sociale wetenschappen. Het team formuleert ook aanbe-

Vervolg op pagina 2

GEHOORD & BIJGEWOOND

CLARIAH Tech Day

Emma Verbree

Data en digitale onderzoeksmethoden hebben de laatste jaren een vaste plek verworven binnen de geesteswetenschappen. Om de beoogde gebruikers, de geesteswetenschappers, hierin bij te staan, organiseert CLARIAH enkele keren per jaar een zogenaamde Tech Day. De ene keer vindt die online plaats, de andere keer fysiek, zoals afgelopen 30 maart in het Internationaal Instituut voor Sociale Geschiedenis in Amsterdam. Naast gebruikers van de CLARIAH-onderzoeksinfrastructuur waren op deze dag ook andere geesteswetenschappers en ontwikkelaars aanwezig om te praten over de ontwikkelingen in de digitale geesteswetenschappen. Deskundigen gaven verschillende presentaties over tools, databronnen en methoden die geesteswetenschappers kunnen helpen bij hun onderzoek. De informatie over bijvoorbeeld werken met FAIR-datasets, geo-data en levensloopdata was niet alleen afkomstig van CLARIAH. Er was ook ruimte voor andere samenwerkingverbanden, bijvoorbeeld over het werken met ontoegankelijke data (wegens copyright of privacybescherming). Zo deelde SANE (Secure Analysis Environment) hoe het mogelijk is om analyses uit te voeren op beschermde data in ‘tinker mode’, waarbij de gebruiker de data kan zien, of in ‘blind mode’, waarbij de onderzoeker de data niet ziet en de analyse moet testen met voorbeelddata. Ook was er ruimte voor Twi-XL, dat toegang verschaft tot sociale mediadata zonder de ruwe posts te delen (wat verboden is). De deelname van dit soort projecten aan CLARIAH’s Tech Day vormen een mooie gelegenheid om te laten zien welke ontwikkelingen er plaatsvinden binnen de onderzoeksinfrastructuur voor de geesteswetenschappen. Ook maakt het een productieve discussie met andere projecten en ontwikkelaars mogelijk. De volgende CLARIAH



Sfeerimpressie workshop Metadata in een van de historische zalen van het Trippenhuis. Credits: DANS

Tech Day zal in het teken staan van data; deze Tech & Data Day vindt plaats op donderdag 29 juni.
<https://www.clariah.nl/events>

National Tripartite Event: Netherlands

Marjan Grootveld

In de European Open Science Cloud ofwel EOSC kunnen onderzoekers grensoverschrijdend samenwerken en gebruik maken van data, software en IT-diensten, waar die ook ‘staan’. Deze visie wordt langzaam concreter, waarbij men rekening houdt met wat de lidstaten zelf te bieden hebben. Dit gebeurt bijvoorbeeld in ‘tripartiete nationale bijeenkomsten’. Op 11 april was de beurt aan Nederland. Vertegenwoordigers van de Europese Commissie, de EOSC Association (EOSC-A) en Nederlandse instellingen ontmoetten elkaar, op uitnodiging van SURF en de Ministeries van Onderwijs, Cultuur en Wetenschap (OCW) en Economische Zaken en Klimaat (EZK). Javier Lopez Albacete, EC Policy Officer van de Open Science Unit, schetste op basis van een recente Europese studie dat het bewustzijn

van FAIR (Findable, Accessible, Interoperable en Reusable) bij onderzoekers nog vrij laag is. Volker Beckmann, die in de Strategic Board van EOSC-A Frankrijk vertegenwoordigt, vertelde dat Frankrijk de digitale transitie die de EOSC ondersteunt noodzakelijk vindt om de ‘return on investment’ van onderzoek te verbeteren: onderzoeksresultaten moeten vanaf het begin FAIR worden gemaakt (‘FAIR by design’), omdat het immers veel te kostbaar is om met terugwerkende kracht te doen. Tegelijkertijd maakte Chantal Kemner, hoogleraar aan de Universiteit Utrecht, duidelijk dat bepaalde FAIR-aspecten pas tijdens een project duidelijk worden. Een van de resultaten van de YOUTH-studie – winnaar van de Nederlandse Dataprijs 2022 – is de YODA-omgeving om data van langlopend onderzoek bij kinderen veilig te bewaren. Zij bepleitte financiering voor duurzaamheid: ook na afloop van het project komen er toegangsverzoeken en moeten de data goed beschikbaar blijven. De aanwezigen hadden verschillende ideeën over financiële en andere verantwoordelijkheden van instellingen,

ationale gremia, financiers en EOSC. Een centrale vraag was hoe top-down beleid en financiering bottom-up activiteiten zoals Open Science Communities kunnen versterken. Als belangrijke Nederlandse ontwikkelingen werden het regieorgaan Open Science NL en de lokale en thematische Digital Competence Centers (DCC’s) gepresenteerd. Bij lokale DCC’s is duurzaamheid, in de vorm van structurele financiering van capabele data stewards, eveneens een uitdaging.

Meer informatie over het even:
<https://eosc.eu/events/national-tripartite-event-netherlands>
European Data Landscape Study:
<https://dans.knaw.nl/nl/nieuws/european-research-data-landscape/>
Nederlandse Dataprijs 2022:
<https://researchdata.nl/en/winners-dutch-data-prize-2022/>

Workshop metadata (standaarden) voor FAIR erfgoedonderzoek

Cees Hof

In de historische zalen van het Trippenhuis in Amsterdam vond op 14 maart de tweede data-workshop van

het Nederlandse E-RIHS netwerk plaats. De workshop, gecoördineerd door DANS en het Netwerk Digitaal Erfgoed (NDE), richtte zich op de rol van metadata in de erfgoedwetenschappen. De bijna 20 deelnemers verdiepten zich in de ins and outs van metadata in (digitale) erfgoedcollecties en wetenschappelijke data-repositories, en hoe deze te gebruiken in interdisciplinair erfgoedonderzoek. Een nieuw wetenschappelijk netwerk zoals E-RIHS (European Research Infrastructure for Heritage Science), behoeft uiteraard enige toelichting. Jan van ’t Hof, Hoofd Monumenten en Collecties bij de Rijksdienst Cultureel Erfgoed, en coördinator van de Nederlandse tak van E-RIHS, heette de deelnemers daarom (digitaal) welkom en benadrukte het unieke interdisciplinaire karakter van deze tak van wetenschap. In de erfgoedwetenschappen moeten onderzoekers immers met data werken variërend van moleculaire pigmentdata uit een van Gogh schilderij tot LiDAR satellietbeelden van Romeinse legerkampen. Deze data moeten dan ook nog eens FAIR (Findable, Accessible, Interoperable en Reusable) worden gemaakt, volgens de laatste Europese en nationale onderzoeksrichtlijnen. Bovendien heb je in erfgoedonderzoek ook nog te maken met alle gegevens over de erfgoedobjecten zelf, al dan niet gedigitaliseerd. Uitdagingen te over dus, vooral voor alle data stewards en datamanagers, waar deze workshop vooral voor bedoeld was.

Spreekers en trainers afkomstig van DANS, NDE, de KB (Koninklijke Bibliotheek) en het RKD voorzagen de workshop van interactieve sessies over generieke en meer domein specifieke metadata, het gebruik van de Art & Architecture Thesaurus (AAT), de werking van het Nederlandse Termennetwerk en over wat Preserveringsmetadata nu precies zijn en hoe deze te gebruiken. In een grotere oefening gingen de deelnemers aan de slag met het gebruik van metadata voor hun zelfgekozen data use cases. Dit resulteerde in een aantal heel concrete en inzichtelijk schema’s over uiteenlopende onderzoeksdata, van oral history tot XRF (X-ray fluorescence) data. Onder het toeziende oog van de oude meesters aan de wanden van de Rembrandtzaal, leerden de deelnemers van elkaars data-problemen en uiteraard (nieuwe) oplossingen. Meer informatie over E-RIHS en de data workshop(s) is beschikbaar op: <https://e-rihs.nl>

Benchmarks

Vervolg van pagina 1

velingen voor het ontwerpen en organiseren van toekomstige benchmarks in de sociale wetenschappen. De eerste resultaten en bevindingen bevestigen dat benchmarks een veelbelovend middel zijn om de sociale wetenschappen sterk te ontwikkelen.
<https://odissei-data.nl/en/benchmarking/>

ISO 9001-certificaat voor Centerdata

Nadat Centerdata in 2021 het ISO 27001-certificaat (dé internationale standaard voor informatiebeveiliging) en vorig jaar NEN 7510 (informatiebeveiliging binnen de gezondheidszorg) behaalde, sloten ze recent ook de audit voor het kwaliteitsmanagementsysteem positief af. Daarmee is Centerdata nu ook gecertificeerd voor ISO 9001:2015. Het behalen van het ISO 9001-certificaat bevestigt dat Centerdata een solide kwa-



liteitsbeleid heeft, dat de processen goed geregeld zijn en dat een continue verbetercyclus gevolgd wordt. Opdrachtgevers weten zo met nog meer zekerheid dat ze bij Centerdata kwaliteit krijgen. Naast het behalen van het ISO 9001-certificaat, behoudt Centerdata ook het ISO 27001-certificaat en NEN 7510. Voor ISO 27001 was het de tweede controle audit en voor NEN 7510 de eerste controle audit. (MV)
www.centerdata.nl/

Loghi maakt gescande historische teksten leesbaar en zoekbaar

Transcriptie-software open source beschikbaar

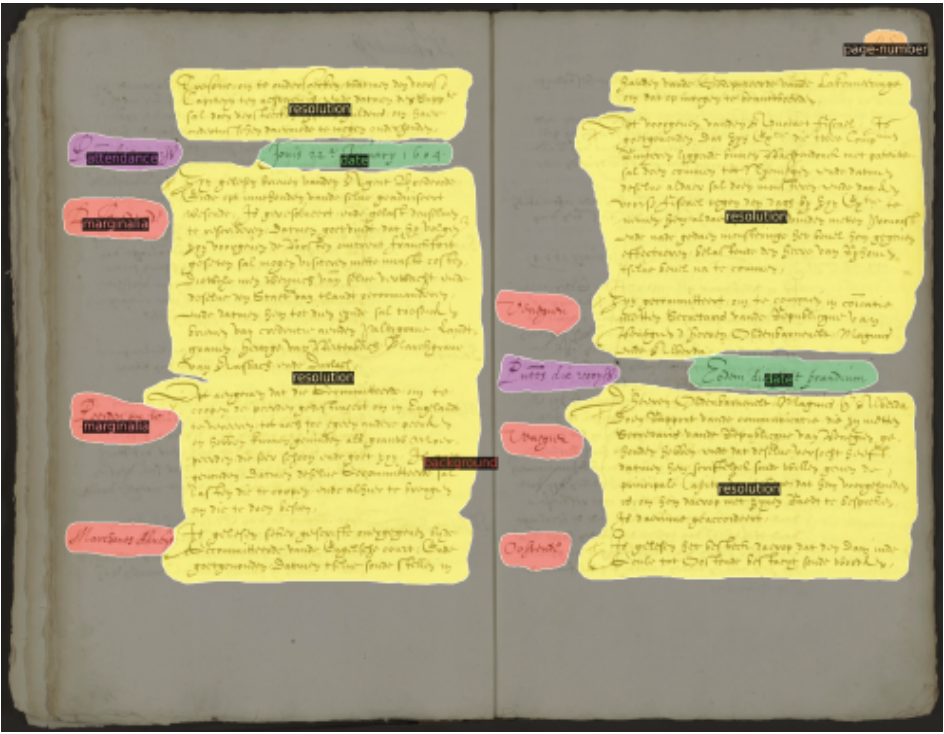
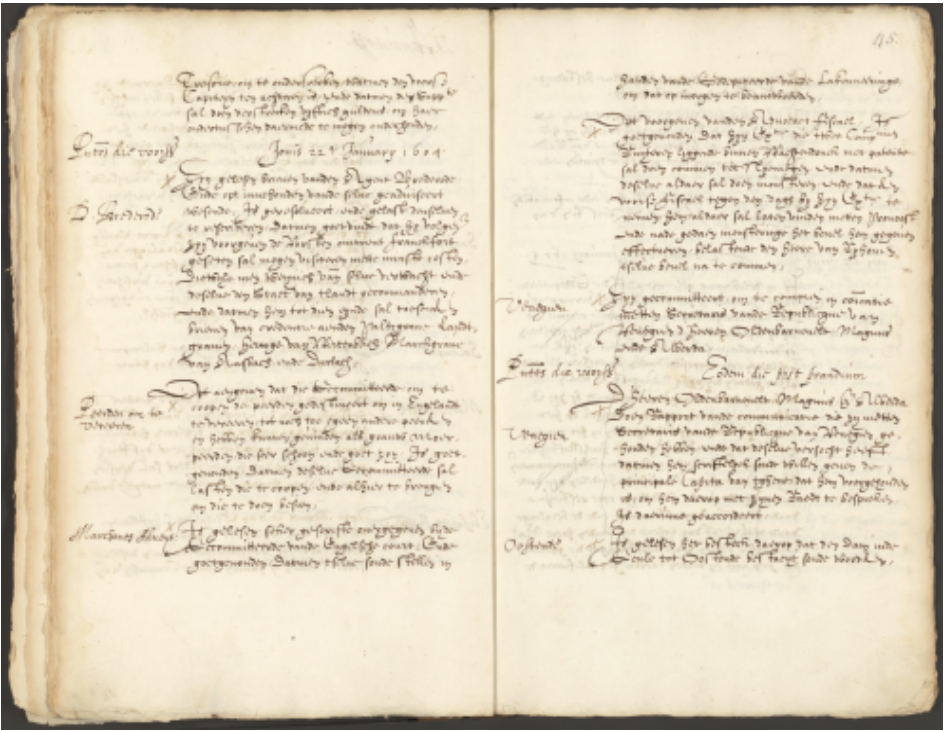
Het KNAW Humanities Cluster ontwikkelde de transcriptie-software Loghi, die tot minstens 96 procent correcte transcripties geeft van oude teksten.

Mathilde Jansen

Het ontcijferen van oude handgeschreven teksten is voor mensen al niet makkelijk, laat staan voor computers. Toch zijn er sinds kort transcriptieprogramma's op de markt die allerlei soorten handschriften om kunnen zetten naar gedigitaliseerde tekst. Voor die software moet je echter vaak betalen; het programma Loghi is transcriptie-software die volledig open source is en dus voor iedereen vrij te gebruiken. Het is toepasbaar op teksten uit de 16e tot en met de 20e eeuw. Het programma werd ontwikkeld door Rutger van Koert van het KNAW Humanities Cluster (HuC).

Plaatjes van 3x3 pixels

Net als andere transcriptie-software werkt Loghi op basis van machine learning. Dat betekent dat de software getraind is op basis van handschriften waar al transcripties van bestonden. Het computermodel maakt voor de transcriptie twee belangrijke stappen, legt Van Koert uit. "Eerst moet het model bepalen waar de tekst staat. Dat is voor ons logisch, maar voor de computer niet. In een volgende stap worden de tekstregels eruit geknipt en via een piramide-achtige structuur geanalyseerd, beginnend bij stukjes van 3 bij 3 pixels. Vanaf dat allerkleinste niveau worden transcripties opgebouwd op basis van het getrainde model." De software is vooralsnog toegepast in een aantal grote projecten, zoals REPUBLIC waarin de resoluties van de Staten-Generaal



toegankelijk worden gemaakt. "Die handschriften zijn geschreven door professionele schrijvers, in zogenaamd netschrift", zegt Van Koert. Dat betekent dat het handschrift over de hele linie erg consistent is. "Daar gaat Loghi nog maar in 1,8 procent van de gevallen de mist in, dat is heel weinig." De bronnen van dit project liggen in het Nationaal Archief (NA) in Den Haag. Van Koert is daarom anderhalf jaar bij het NA gedetacheerd geweest. Ook het IJSBERG-project van het Nationaal Archief heeft veel input geleverd. "Daarvoor zijn 7706 scans getranscribeerd, voornamelijk teksten over de VOC en notariële teksten. Die laatste teksten zijn opgetekend door notarissen en veel slordiger. Daar is de foutmarge net onder de 4 procent."

Tekst op z'n kop

Ook andere organisaties kunnen Loghi inzetten voor het ontcijferen van hun eigen tekstbronnen. Zij kunnen het algemene model gebruiken, maar deze ook doortraineren op hun eigen dataset, zodat de foutmarge nog meer omlaag gaat. "De handschriftherkenning wordt alleen nog maar beter als je er meer data in stopt", zegt Van Koert. "Eigenlijk ligt de grootste hobbel in dit onderzoek nog in die eerste stap van de machine learning. En dat is het bepalen waar de tekst staat. 17e-eeuwse brieven bijvoorbeeld zijn vaak tot envelop gevouwen en dan is de buitenkant ook beschreven. Als je die openvouwt staat een deel van de tekst op z'n kop. Dat is nog een grote uitdaging."

<https://github.com/knaw-huc/loghi>

Boven: Origineel.
Bron: Nationaal Archief, Inventaris van het archief van de Staten-Generaal, (1431) 1576-1796, nummer archiefinventaris 1.01.02, inventarisnummer 3152
Onder: Visualisatie automatisch gedetecteerde regio's.

Zoekmachine voor geopenbaarde overheidsdocumenten ontwikkeld

Struinen door Woo-dossiers

Overheidsdocumenten die zijn geopenbaard via een Woo-verzoek zijn vaak slecht doorzoekbaar. Informaticus Maarten Marx brengt daar verandering in.

Erica Renckens

Eén pdf van honderden pagina's, waarin tientallen documenten lukraak achter elkaar zijn geplakt, met teksten die onleesbaar zijn voor een computer. Zo ontvangen onderzoeksjournalisten meestal de dossiers die ze opvragen bij overheidsinstellingen in het kader van de Wet Open Overheid (Woo, de opvolger van de Wob). Niet alleen praktisch onbruikbaar voor de journalisten, maar ook voor wetenschappers die de data willen gebruiken voor hun onderzoek.

Dat kan beter, bedacht informaticus Maarten Marx van de Universiteit van Amsterdam. Met zijn onderzoeksgroep ontwikkelde hij WooGle, een zoekmachine voor geopenbaarde Woo-dossiers van gemeenten, provincies en ministeries.

Kunstmatige intelligentie

"Als je zo'n pdf opent, kan je de teksten zelf prima lezen", vertelt Marx. "Het probleem merk je pas als je in de tekst wil zoeken of wil laten voorlezen. Dat gaat simpelweg niet, doordat het eigenlijk alleen scans van de documenten zijn." Het team van Marx gebruikt daarom Optical Character Recognition (OCR) om de letters en woorden op de scans te herkennen. Het resultaat daarvan is nog niet perfect, maar wel al goed bruikbaar. "We proberen het nog te verbeteren, bijvoorbeeld met de kunstmatige intelligentie van ChatGPT. Die

kan de laatste fouten eruit halen, maar het gaat om zulke grote hoeveelheden tekst dat de kosten dan wel hoog oplopen. Daar zoeken we nog een oplossing voor." Marx en zijn collega's gebruiken AI ook voor het automatisch herkennen van de grenzen tussen verschillende documenten binnen een pdf. Momenteel doorzoekt WooGle ruim 1,1 miljoen pagina's afkomstig uit bijna 20.000 dossiers van 55 verschillende bestuursorganen. Die lopen uiteen van gemeenten, provincies en ministeries tot de Kansspelautoriteit en de Landelijke Eenheid Politie. De data in WooGle worden continu aangevuld met nieuwe documenten.

Erfgoed voor de toekomst
Hoewel Marx zijn best doet om ingescande Woo-dossiers doorzoekbaar te maken, ziet hij liever dat het direct goed gaat bij de bron.

"Dit is geen 18e-eeuws materiaal, maar gewoon onlangs op de computer gemaakt. We roepen Woo-coördinatoren van bestuursorganen daarom ook op om dossiers direct naar WooGle te uploaden. Als het aan het begin al goed gaat, zou dit allemaal niet nodig zijn. Ik zie dat vooral Woo-coördinatoren bij gemeenten heel welwillend en enthousiast zijn om hieraan mee te werken." Onlangs heeft Marx de data uit WooGle gedeponeerd bij DANS, zodat ze ook toegankelijk zijn voor de onderzoeksgemeenschap. "Vooral vanuit Rechten is hier veel interesse voor. Maar ik zie het ook als een soort erfgoed voor de toekomst, zodat een historicus over vijftig jaar kan kijken hoe dat nou ging met die openbaarmakingen."

<https://woogle.wooverheid.nl/>
<https://doi.org/10.17026/dans-zau-e3rk>

SINDS KORT BESCHIKBAAR

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

CBS

Sinds februari 2023 is een bestand beschikbaar over Watergebruik van huishoudens; WGT. In de vragenlijst die voor dit onderzoek ontwikkeld is, wordt de respondent voor verschillende componenten gevraagd aan te geven wat hun situatie was op de invuldag. Ook nieuw zijn twee bestanden met dementie gegevens; indicatoren en populatie, en twee bestanden met kinderopvangtoeslagen gegeven; kinderen met kindertoeslag en ontvangers van kindertoeslag.



Een volledig overzicht van beschikbare bestanden staat in de microcatalogus op de CBS-website: [cbs.nl/microdata catalogus](https://cbs.nl/microdata-catalogus) Ook een link naar de bestanden die de afgelopen periode beschikbaar zijn gekomen staat online: [cbs.nl/microdata overzicht](https://cbs.nl/microdata-overzicht)



Centerdata

Voorspellers van ostracisme

Ostracisme is een veelvoorkomende, maar uiterst pijnlijke ervaring waarbij iemand wordt genegeerd en buitengesloten door anderen. Hoewel de negatieve gevolgen van ostracisme bekend zijn, is er veel minder bekend over de oorzaken en voorspellers van deze vorm van sociale afwijzing. Er is echter nog veel te leren over de factoren die leiden tot ostracisme. Recent onderzoek in het LISS panel heeft zich gericht op de voorspellers van deze vorm van sociale afwijzing, waaronder persoonlijkheidstrekken, sociale status en de dynamiek van sociale groepen.

lissdata.nl

De data de zijn beschikbaar via LISS Data Archive.
DOI: <https://doi.org/10.17026/dans-xrg-2h89>

Ook sinds kort beschikbaar:

Studies LISS panel

- Bartova, A., oktober 2021, Family policy.
DOI: <https://doi.org/10.17026/dans-xse-qc3q>
- Jacobs, K., februari 2020, Populism barometer - Wave 1.
DOI: <https://doi.org/10.17026/dans-xm9-4wxp>
- Jacobs, K., mei 2020, Populism barometer - Wave 2.
DOI: <https://doi.org/10.17026/dans-xnz-6avu>
- Jacobs, K., augustus 2020, Populism barometer - Wave 3.
DOI: <https://doi.org/10.17026/dans-23j-5zyr>
- Jacobs, K., november 2020, Populism barometer - Wave 4.
DOI: <https://doi.org/10.17026/dans-xka-sx5u>
- Jacobs, K., februari 2021, Populism barometer - Wave 5.
DOI: <https://doi.org/10.17026/dans-znd-njdj>
- Jacobs, K., mei 2021, Populism barometer - Wave 6.
DOI: <https://doi.org/10.17026/dans-zza-rm2z>
- Jacobs, K., augustus 2021, Populism barometer - Wave 7.
DOI: <https://doi.org/10.17026/dans-xck-vpba>
- Jacobs, K., november 2021, Populism barometer - Wave 8.



Credits: Markus Spiske via Pixabay

- DOI: <https://doi.org/10.17026/dans-z35-jqjb>
- Centerdata, 2021, Initial Questionnaire - 2021.
DOI: <https://doi.org/10.17026/dans-xxk-j997>
 - Sipma, T., Meer, T. van der, februari 2021, Dutch Parliamentary Election Study 2021 - Part 1.
DOI: <https://doi.org/10.17026/dans-zdy-qdcn>
 - Sipma, T., Meer, T. van der, maart 2021, Dutch Parliamentary Election Study 2021 - Part 2.
DOI: <https://doi.org/10.17026/dans-x6h-dwyx>
 - Sipma, T., Meer, T. van der, april - mei 2021, Dutch Parliamentary Election Study 2021 - Part 3.
DOI: <https://doi.org/10.17026/dans-zn6-ma4y>
 - Centerdata, oktober – november 2022, Social Integration and Leisure - Wave 15.
DOI: <https://doi.org/10.17026/dans-z8g-v3se>
 - Centerdata, november - december 2022, Health - Wave 15.
DOI: <https://doi.org/10.17026/dans-x4e-gj5>
 - Rouvroye, L., Henkens, K., april - mei 2021, Young people in the labor market - Young people on the labor market.
DOI: <https://doi.org/10.17026/dans-2ch-ra3w>
 - Rouvroye, L., Henkens, K., april - mei 2021, Young people in the labor market - Vignette study young people & job characteristics.
DOI: <https://doi.org/10.17026/dans-zp5-rsj2>
 - Rouvroye, L., Henkens, K., april 2021, Young people in the labor market - Vignette study employers & job characteristics.
DOI: <https://doi.org/10.17026/dans-x2g-v3rz>
 - Noordewier, M., Dijk, W. van., Nijkamp, R., Kunkel, D., september 2022, National Monitor Financial Worries - Wave 1.
DOI: <https://doi.org/10.17026/dans-2b8-mdns>
 - Noordewier, M., Dijk, W. van., Nijkamp, R., Kunkel, D., december 2022, National Monitor Financial Worries - Wave 2.
DOI: <https://doi.org/10.17026/dans-2bw-rsuy>
 - Stulp, G., februari - maart 2018, Social networks and fertility.
DOI: <https://doi.org/10.17026/dans-z6p-2v79>
 - Frey, V., Bouman, T., Goedkoop, F., mei 2022, Greener than others? - Part 1.
DOI: <https://doi.org/10.17026/dans-zex-ncw6>
 - Frey, V., Bouman, T., Goedkoop, F., juni 2022, Greener than others? - Part 2.
DOI: <https://doi.org/10.17026/dans-zgq-ezva>
 - Vollaard, H., Jansen, G., maart - april 2022, Local Voters Survey - Local Voters Survey 2022.
DOI: <https://doi.org/10.17026/dans-ze6-yfg2>
 - Meer, T. van der, mei 2019, NKO panel questions - Wave 1.
DOI: <https://doi.org/10.17026/dans-x5v-bgdt>
 - Meer, T. van der, juni 2020, NKO panel questions - Wave 2.
DOI: <https://doi.org/10.17026/dans-zph-an3h>
 - Meer, T. van der, juni 2021, NKO panel questions - Wave 3.
DOI: <https://doi.org/10.17026/dans-z4d-jfja>
 - Centerdata, september - oktober 2022,

- Assets - Wave 8.
DOI: <https://doi.org/10.17026/dans-xjr-9ff5>
- Scheele, D., januari - februari 2020, Survey Technostress.
DOI: <https://doi.org/10.17026/dans-x5n-82vj>
 - Hartevelt, E., juli 2021, Affective polarization - wave 2.
DOI: <https://doi.org/10.17026/dans-x2g-p5qj>



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

DANS

- De volgende datasets zijn open access beschikbaar via de DANS Data Stations of het online archiveringssysteem EASY:
- Terhorst, T.; Veerkamp, J. (Gemeente Amsterdam), 2023, 'De stadswal van 1586 (AAR 133)', <https://doi.org/10.17026/AR/8Z211F>, DANS. Data Station Archaeology, V2
 - Bazelmans, dr J. (Cultural Heritage Agency) (2023): Het molensterven in Nederland. Verantwoordingsdocument vormgeving databestand.
DANS. <https://doi.org/10.17026/dans-xwj-89ap>
 - Braat, Dr. E.C. (Utrecht University); Jong, Dr. B.M. de (2019): Between a Rock and a Hard Place: Interviews With a Former Double Agent.
<https://doi.org/10.17026/dans-zbd-6h33>
 - Marx, Dr M. (Informatics Institute, University of Amsterdam) (2023): Woogles dump.
DANS. <https://doi.org/10.17026/dans-zau-e3rk>
 - Boer, T. de (VU Amsterdam) (2021): Formal independence index of European Union agencies.
DANS. <https://doi.org/10.17026/dans-x8u-m3dd>
 - Alencar, Dr. A.P. (Erasmus University Rotterdam) (2014): Cultural values in news narratives coding.
DANS. <https://doi.org/10.17026/dans-xau-rhre>
 - Mulier Instituut (2022): Nationaal Sportonderzoek 2022 mei.
DANS. <https://doi.org/10.17026/dans-xgh-x9fe>
 - Yang, Y (Leiden University); Emmen, R (Leiden University); Emmen, R (Leiden University) (2022): Chinese-Dutch child ethnic prejudice and COVID-19.
DANS. <https://doi.org/10.17026/dans-xaa-8m5u>
 - Wolfs, MSc Z.G. (Open University of the Netherlands) (2020): Musical Perception and Vocal Abilities of Elementary School Children.
DANS. <https://doi.org/10.17026/dans-xrh-ayu6>
 - Lips, Drs. S. (Amsterdam UMC) (2023): SWING studie: Samen op Weg naar Integrale Geboortezorg.
DANS. <https://doi.org/10.17026/dans-x8m-rm88>
 - Lubbe, Dr R.H.J. van der (University of Twente) (2021): External and Internal Spatial Attention.
DANS. <https://doi.org/10.17026/dans-x2m-ug9s>
 - Post, Dr. P.M. (Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente) (2022): Data underlying the paper 'effects of food value-chain connection on land-use change'.

AGENDA

- 12 juni • Online**
Showcase Data Stations
Online event over de mogelijkheden van de DANS Data Stations.
<https://dans.knaw.nl/nl/agenda/presentatie-dans-data-stations/>
- 15 juni en 14 september • Online**
Lunchlezingen CLARIAH
<https://www.clariah.nl/events>
- Juni en september • Online**
DANS Open Hour SSH: live Q&A
In juni en september elke maandag van 10 tot 11 uur
<https://dans.knaw.nl/nl/agenda/open-hour-ssh-live-qa-op-maandag/>
- 28 juni en 28 september**
Tech/Data-dag CLARIAH
<https://www.clariah.nl/events>
- 29 juni • Den Haag**
Shaping the world of 3D
Verbeteren van het beheer en hergebruik van 3D datasets.
<https://dans.knaw.nl/nl/agenda/shaping-the-world-of-3d/>
- 31 juli - 11 augustus • Triëst**
CODATA-RDA Summer School
De Summer School ontwikkelt competenties in data-analyse en -beveiliging voor deelnemers uit alle disciplines en/of achtergronden, van wetenschappen tot geesteswetenschappen.
<https://indico.ictp.it/event/10195>
- 31 augustus • Rotterdam**
National Open Science Festival
Ook deze 3e editie draait om het delen van kennis en peer learning: het is gericht op zowel actieve onderzoekers van universiteiten (van toegepaste wetenschappen), academische ziekenhuizen en onderzoeksinstituten in Nederland, als op beleidsmakers die het mogelijk maken dat Open Science zich verspreidt over de academische wereld.
<https://opensciencefestival.nl/>
- 19 - 22 september • Illinois**
iPRES Conferentie 2023
De 19e editie van dit jaarlijkse event vindt plaats in Illinois (VS).
<https://ipres2023.us/>
- 2 november • Utrecht**
ODISSEI Conference for Social Science in the Netherlands
Samen discussiëren over onderwerpen als data, methodes, infrastructuur, ethiek en theoretisch werk over digitale en computationele methodes in sociaal-wetenschappelijk onderzoek.
<https://odissei-data.nl/nl/conferentie-2022/>

- DANS. <https://doi.org/10.17026/dans-zce-669a>
- Broekman, M.J.E. (Radboud University) (2023): Data from: Modelling impacts of habitat loss and fragmentation on mammal species.
DANS. <https://doi.org/10.17026/dans-x35-fxfj>
 - Bus, MSc J.D. (Wageningen University & Research); Walderveen, A. van (De Klomp Dierenartsen); Bolhuis, Dr. J.E. (Wageningen University & Research); Boumans, Dr. I.J.M.M. (Wageningen University & Research); Bokkers, Dr. E.A.M. (Wageningen University & Research) (2023): Protocol for Health Observations in Growing-Finishing Pigs.
DANS. <https://doi.org/10.17026/dans-xt2-xpej>



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

Nieuwe database over slavenhandelsreizen in maritiem Azië

Slavenhandelsreizen van Kaapstad tot Tokyo

Het Internationaal Instituut voor Geschiedenis (KNAW Humanities Cluster) bracht de afgelopen jaren gegevens over meer dan 4000 slavenhandelsreizen in maritiem Azië samen in de ESTA-database.

Mathilde Jansen

De afgelopen decennia is er wereldwijd veel onderzoek gedaan naar het slavernijverleden. Informatie over de trans-Atlantische slavenhandel is al sinds de jaren 90 bijeengebracht in de database slavevoyages.org. Voor de slavenhandelsreizen in Azië bestond zo’n database nog niet. De afgelopen jaren heeft een team van internationale onderzoekers zich daarom ingezet om een database op te zetten in het project Exploring Slave Trade in Asia (ESTA). Inmiddels zijn vijf grote datasets ingevoerd, voornamelijk uit Europese archieven van bijvoorbeeld de VOC en de Franse Oost-Indische Compagnie. Nog dit jaar wordt de database toegankelijk gemaakt voor een breed publiek.

Korte reizen

“De slavenhandelsreizen in Azië zijn anders dan die we kennen uit het trans-Atlantische gebied”, vertelt Merve Tosun, coördinator van het ESTA-project. “Bij het grote publiek is hier minder over bekend. In het trans-Atlantische gebied was er

altijd sprake van een enkele lange éénrichtingsreis van de slaafge-maakten tussen Afrika en Amerika, waar ze op plantages gedwongen te werk werden gesteld. In Azië bestond een slavenhandelsreis meestal uit een aaneenschakeling van korte en lange reizen, meerdere richtingen op. Een reis kon bijvoorbeeld in Madagaskar en via Ceylon (het huidige Sri Lanka) eindigen in Batavia (het huidige Jakarta), maar ook vanuit de zuidelijke kust van India bij de Kaap aankomen. Voor het opzetten van onze database was dat een uitdaging.”

Lokale archieven

Ook bestonden er in Azië verschillende vormen van gedwongen transport van slaven, legt Tosun uit. “Naast de meer bekende vorm van gedwongen transport voor gecommmercialiseerde slavenhandel, was er bijvoorbeeld uitwisseling van slaafgemaakten tussen vorsten en Europese mogendheden als onderdeel van tribuutafspraken, maar het kwam ook voor dat mensen als oor-



Map of the Indian Ocean Region, 1600 - 1700. Bron: National Archives, CC0

logsbuit werden meegenomen of ‘illegaal’ gekidnapt.” Al dit soort gegevens hebben de onderzoekers vastgelegd in de ESTA-database. De komende jaren zal deze nog meer worden uitgebreid. “De data die we nu hebben verzameld en gecureerd zijn vooral afkomstig uit Europese compagnie-archieven, maar deze

archieven bevatten nog veel meer informatie. Ook in de lokale archieven moet nog informatie aanwezig zijn. We zijn nu bezig met nieuwe subsidieaanvragen om deze archieven in te kunnen duiken. Op die manier willen we nog meer inzicht geven in dit stuk slavernijgeschiedenis.”

Lezers die beschikken over data rondom slavenhandel in Azië en die breder beschikbaar willen maken voor onderzoek, kunnen contact opnemen via de website van ESTA.

<https://esta.iisg.nl/>
<https://datasets.iisg.amsterdam/data-verse/IOMASTD>

TDCC’s ondersteunen FAIR data in verschillende domeinen

Bruggen bouwen in het Nederlandse onderzoekslandschap

De nieuwe thematische Digital Competence Centers zijn opgericht om domeinspecifieke uitdagingen rondom FAIR data aan te pakken. E-data sprak met het team van het thematische Digital Competence Center voor de sociale en geesteswetenschappen over de plannen voor de komende jaren.

Ricarda Braukmann

“Het verbeteren van de domeinspecifieke onderzoekspraktijken om zo wetenschappelijke ontdekkingen en vooruitgang te stimuleren” – zo zouden Nicole Emmenegger en Nils Arlinghaus het doel van het thematische Digital Competence Center (TDCC) in een zin samenvatten. Zij zijn in maart 2023 gestart als netwerkmanager en community coördinator van het TDCC voor de sociale en geesteswetenschappen (SSH).

De drie TDCC’s – SSH, LSH (Life Sciences and Health) en NES (Natural & Engineering Sciences) – zijn met financiering van NWO opgericht om een domeingerichte aanpak te bieden. Zij werken als aanvulling op de lokale DCC’s die onderzoekers aan hun eigen instituut ondersteunen, en de nationale expertisecentra (zoals DANS, SURF of het Netherlands eScience Center) die algemene diensten en infrastructuur leveren. Het werk van de TDCC’s richt zich ook op het beleid en de wetenschapsstrategie in samenspraak met onder meer Open Science NL en andere nationale initiatieven. De focus van de TDCC’s ligt op domeinspecifieke uitdagingen rondom Open Science en FAIR data. Elk TDCC heeft in de afgelopen periode al een aantal knelpunten geïdentificeerd die de komende tijd aangepakt wordt. “Veel SSH-onderzoekers werken met kwalitatieve data, maar de richtlijnen en aanbevelingen rondom FAIR data en Open

Science zijn vaak niet gericht op dit type data. Dit willen we verbeteren”, legt Emmenegger uit. “Een ander belangrijk kenmerk van het SSH-domein is het werken met gevoelige data, zoals data met copyright of persoonsgegevens die verwerkt moeten worden volgens de AVG.”

Roadshows en visites

Momenteel zijn Arlinghaus en Emmenegger druk bezig om het hele land door te reizen en kennis te maken met het veld. “We gaan langs bij alle lokale DCC’s om te praten over de ervaringen van de SSH-gemeenschap”, zegt Arlinghaus. Hij geeft aan dat beiden zich ook richten op een nauwe samenwerking met de infrastructuren van het domein, zoals CLARIAH en ODISSEI. “Er zijn al veel goede werkwijzen en tools beschikbaar, maar het landschap is erg groot en divers waardoor we

niet altijd de juiste links leggen. De drie TDCC’s willen helpen om de verschillende stukken van het landschap met elkaar te verbinden.”

Non-competitive funding

Om de ontwikkeling van het landschap te ondersteunen en de knelpunten aan te pakken, stellen de drie TDCC’s financiering beschikbaar. In tegenstelling tot de meeste NWO-beurzen volgt dit een model van non-competitive funding. Arlinghaus licht dit toe: “Wij gaan projecten financieren op basis van de behoeften van de gemeenschap. En moedigen verschillende organisaties aan om samen te werken aan een gemeenschappelijk projectvoorstel.” Emmenegger en Arlinghaus horen graag over de ideeën die leven: “Neem vooral contact met ons op, of spreek ons aan op een community event.”

<https://tdcc.nl/>

Marieke van Erp ontvangt ERC Consolidator Grant van 2 miljoen euro

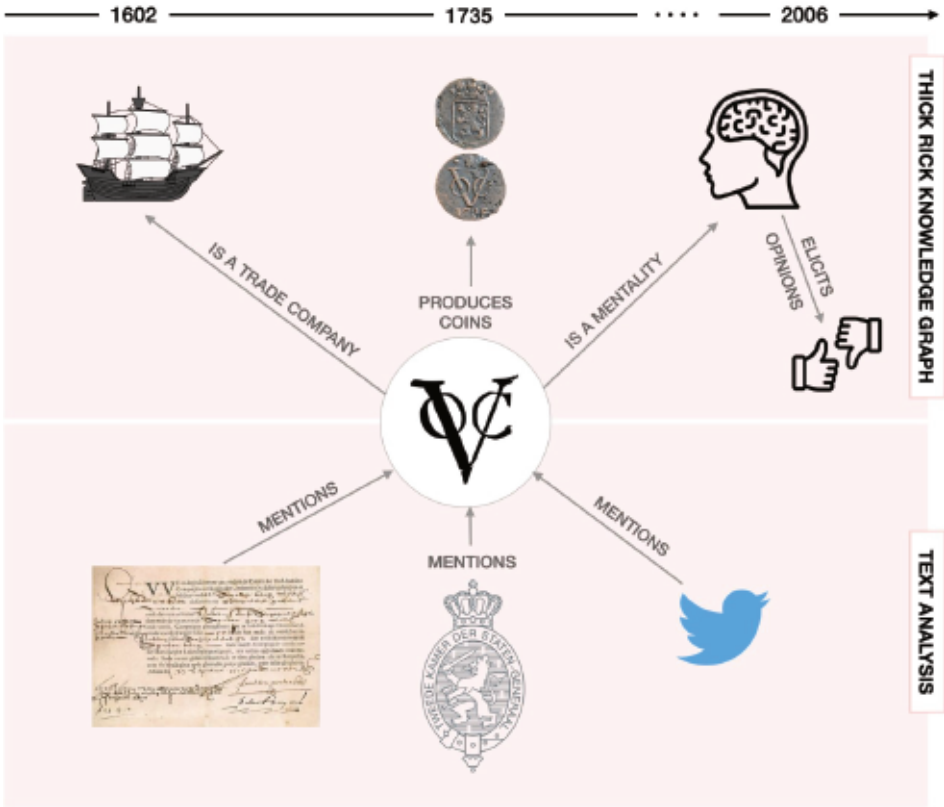
Taalmodellen verbeteren met semantisch web

Door taal- en semantische webtechnologie te combineren, wil Marieke van Erp (DHLab, KNAW Humanities Cluster) de mogelijkheden van het Digital Humanities-onderzoek vergroten. *Mathilde Jansen*

Met taaltechnologie kun je allerlei informatie uit teksten halen. Toch is Marieke van Erp, hoofd van het DHLab, vooral geïnteresseerd in de vragen die daarna komen: Hoe verhoudt die informatie zich tot elkaar en wat vertelt het ons over de wereld? Met de huidige technologie is die vraag vaak nog moeilijk te beantwoorden, maar Van Erp wil hier verandering in brengen. Dit jaar kreeg zij een ERC Consolidator Grant waarmee ze de komende vijf jaar fundamenteel onderzoek gaat doen naar het analyseren van historische teksten met computergestuurde technieken.

Semantisch web

In het geesteswetenschappelijk onderzoek staan concepten centraal die voortdurend veranderen, legt Van Erp uit. Daardoor kun je de informatie niet opslaan in een statische database. “Neem de VOC. Dat was in het oprichtingsjaar 1602 een heel andere organisatie dan in 1798, het jaar van opheffing. Of het concept Nederland: niet alleen de grenzen zijn in de



Semantisch webmodel van de VOC. De onderste laag laat zien dat er informatie uit teksten gehaald wordt, die wordt gekoppeld aan een entiteit of concept. De bovenste laag laat zien dat die informatie aan een jaartal gekoppeld is. Bron: <https://trifecta.dhlab.nl> Credits: Marieke van Erp/Ingrid Arcas

loop van de tijd veranderd, maar ook de geopolitieke entiteit. En ook personen hebben door de jaren heen verschillende rollen. Hoe ga je ermee om als je dit soort informatie wilt analyseren vanuit een historisch perspectief?”

Voor Van Erp ligt het antwoord in het combineren van taaltechnologie en het semantisch web. Ook binnen het project Odeuropa, waarbij gedigitaliseerde erfgoedcollecties worden doorzocht op geurbeschrijvingen, maakt de

onderzoeker gebruik van deze techniek. Daarbij wordt informatie uit een tekst opgeslagen via zogenaamde semantische webmodellen, of ‘knowledge graphs’. Het is een groot informatienetwerk, dat een beetje vergelijkbaar is met de organisatie van ons brein.

Oneindige verbindingen

Elk knooppunt in dit netwerk is met drie andere concepten verbonden. De onderzoeker legt uit hoe zo’n kenniseenheid eruit ziet: “Het zijn allerlei losse feitjes: Arnold Schwarzenegger (1) is geboren in (2) Oostenrijk (3). Oostenrijk (1) ligt in (2) Europa (3). Europa (1) is een (2) werelddeel (3). Aan elk knooppunt in het netwerk kun je informatie blijven toevoegen. Hele databases die op deze manier zijn georganiseerd kun je aan elkaar koppelen, en dan heb je opeens allerlei informatie die je kunt bevroegen. De tijdsdimensie kun je toevoegen door elk knooppunt een tijdslimiet mee te geven.”

Behalve goed gedocumenteerde feiten, wil Van Erp ook onderbelichte feiten uit de geschiedenis kunnen toevoegen aan zo’n informatienetwerk. “Op de Banda-eilanden heeft de VOC verschrikkelijke dingen gedaan, die nu niet goed zijn gedocumenteerd. Ook die informatie wil ik toevoegen, zodat we een completer beeld krijgen van de geschiedenis, vanuit meerdere perspectieven. Uiteindelijk moet het een blauwprint van stukjes software opleveren, zodat dit soort technologie gedeeld kan worden.”

<https://trifecta.dhlab.nl>

Vijf jaar na de invoering van de Europese privacy wet

AVG. Weg ermee?

Vijf jaar geleden, om precies te zijn op 25 mei 2018, trad de AVG (Algemene Verordening Gegevensbescherming) in werking. Hoofddoel van deze EU wet was om de privacy van de Europese burgers beter te beschermen door de verwerking van persoonsgegevens aan uniforme regels te binden. *Peter Doorn*

Ook voor de wetenschap golden de nieuwe regels. Universiteiten en andere wetenschapsorganisaties, waaronder ook DANS, troffen maatregelen om aan die regels te voldoen. Eerder dit voorjaar stuurde de Algemene Rekenkamer een brief aan de Tweede Kamer waarin zij stevige kritiek uitte op de manier waarop de AVG werd uitgevoerd. DANS waarschuwde al vóór de invoering dat de nieuwe wet onnodige en ongewenste belemmeringen voor wetenschappelijk datagebruik kon opwerpen. De Rekenkamer toont aan dat overheidstaken in brede zin worden gehinderd, niet in de eerste plaats door de wet zelf, maar vooral door de manier waarop de wet is geïnterpreteerd en toegepast. Deze kritiek ligt precies in lijn met waar DANS destijds

al voor vreesde. Zoals vaak gebeurt bij de invoering van nieuwe wetgeving was de voorbereiding in de wetenschap op de invoering van de AVG (die al enkele jaren eerder was aangekondigd) grotendeels afwezig. Er kwam een paniekgolf op gang, die doorgaans leidde tot veel strengere regels dan de wet zelf vergde. Juristen kwamen soms met de simpele oplossing om maar helemaal geen persoonsgegevens meer te delen, dan kon er ook niets fout gaan. Want: op overtreding van de wet stonden hoge boetes. Onder het motto: beter safe than sorry werd stevig aan de privacy-noodrem getroffen.

Wat is privacygevoelig?

Sommige privacy-opvattingen waren zo extreem, dat zelfs wetenschappelijke artikelen als privacygevoelige informatie werden gezien, louter omdat er een auteursnaam boven staat. Ook Data Management Plannen, die misschien om andere redenen als vertrouwelijk kunnen gelden, werden uit naam van de privacy beschermd geacht. Hierdoor werd het voor DANS onmogelijk om te zien welke projecten archiefwaardige data zouden opleveren. Toch worden dit soort kwesties expliciet in de

AVG ondervangen: als het voor de uitoefening van publieke taken noodzakelijk is, mag persoonsinformatie wel degelijk worden gedeeld, mits doel en wijze duidelijk en gemotiveerd zijn vastgelegd. Dit is precies één van de manco’s die door de Rekenkamer worden genoemd: zo wordt bijvoorbeeld de fraudebestrijding in de zorg gefrustreerd omdat de vele betrokken organisaties hun gegevens niet delen, met een onterecht beroep op de privacywet.

Beschermen van burgers

Natuurlijk was en is het doel van de privacywetgeving loffelijk. Vraag je het aan ChatGPT, dan antwoordt het systeem braaf dat de AVG “heeft geleid tot meer verantwoorde en zorgvuldige omgang met persoonsgegevens” en “tot grotere bewustwording van de privacyproblematiek”. Belangrijk doelwit van de wet waren (en zijn) de grote ICT-bedrijven, die steeds meer informatie over burgers vastleggen. Dergelijke bedrijven waren veel beter voorbereid op de AVG. Hoewel de wet bedrijven zoals Google heeft gedwongen om veranderingen aan te brengen in hun processen van gegevensverzameling en -verwerking

om te voldoen aan de privacyeisen, is de hoeveelheid informatie die zij vastleggen over personen zo goed als zeker niet verminderd. Immers, je kunt praktisch geen gebruik maken van de meeste Google-diensten zonder in te stemmen met de privacyverklaring, waarin je Google toestaat je persoonsgegevens te verzamelen en te gebruiken.

ChatGPT gebonden aan regels

ChatGPT antwoordt op de vraag wat hij over mij weet: “Als AI-model heb ik geen directe toegang tot persoonlijke informatie over jou, tenzij je mij specifieke informatie verstrekt tijdens onze interactie. Ik ben geprogrammeerd om te reageren op vragen en commando’s op basis van mijn trainingsdata en algoritmes, en ik houd geen gegevens bij over individuele gebruikers. Het is belangrijk om te benadrukken [...] dat ik als AI-model gebonden ben aan de privacyregels en -voorschriften van de bedrijven en organisaties die mij gebruiken.” Vijf jaar is geen lange tijd om de effectiviteit van een nieuwe wet te evalueren. Desalniettemin is de vraag van de Rekenkamer gerechtvaardigd of de uitvoering van de AVG niet aanzienlijk beter kan.

De ontwikkeling van een landelijk dashboard

Hoe kunnen data helpen de waterkwaliteit te verbeteren?

Overall in Nederland zijn sloten, beken, rivieren en meren te vinden. De kwaliteit van deze waterlichamen is echter lang niet altijd zoals we die graag zouden zien. Het ministerie van Infrastructuur en Waterstaat was daarom benieuwd of een datagedreven aanpak kan bijdragen aan beter beleid. *Mara Verheijen*

Samen met twee andere organisaties (adviesbureau SeederDeBoer en het Informatiehuis Water) ontwikkelde Centerdata een landelijk dashboard dat data over bijvoorbeeld schadelijke stoffen en de visstand inzichtelijk maakt. Met dit landelijke dashboard kan de kwaliteit van verschillende waterlichamen beter gemonitord worden en beleid gericht ingezet worden. De Europese Commissie is niet tevreden over de huidige waterkwaliteit in Europa. Daarom is de Kaderrichtlijn Water (KRW) opgesteld. Volgens deze Europese richtlijn moet al het grond- en oppervlaktewater in Europa in 2027 schoon en gezond zijn. Dat betekent onder andere dat waterlichamen als sloten, beken, rivieren en meren goede omstandigheden moeten bieden voor planten en dieren,



Natuurlijk waterlichaam. Bron: Pixabay

dat schadelijke stoffen zoveel mogelijk uit het water gefilterd moeten worden en grondwater geschikt moet zijn voor drinkwaterwinning. **Inzichten uit dashboards** Op dit moment zijn er verschillende dashboards in gebruik die een overzicht geven van de waterkwaliteit. Deze gaan echter alleen over lokale

waterlichamen en zijn kwalitatief van aard: ze laten zien of de kwaliteit goed, matig of slecht is. Om goed beleid te kunnen ontwikkelen is er meer kwantitatief inzicht nodig. Hoe goed of slecht is de kwaliteit precies? Hoe groot is het verschil tussen de huidige en de gewenste kwaliteit? En hoe ontwikkelt dit zich over de jaren heen?

Pilot Om te laten zien welke inzichten een datagedreven aanpak kan bieden en hoe dit kan bijdragen aan beter beleid, voerde Centerdata voor het ministerie van Infrastructuur en Water een pilot uit. Data scientists van Centerdata koppelden beschikbare data uit het hele land aan elkaar en brachten zo een grote hoeveel-

heid informatie van 17 waterlichamen samen in één overzichtelijk dashboard. Daarmee kan de ontwikkeling van specifieke kenmerken zoals de zuurgraad of de aanwezigheid van vissen over verschillende jaren per waterlichaam gemonitord worden. Er kan dan bijvoorbeeld gekeken worden of de afname van een specifieke schadelijke stof (zoals PFAS) in het water in lijn is met gestelde doelen. Ook kan er gekeken worden naar het effect van beleidsinterventies om te bepalen of deze ook daadwerkelijk doen wat ze zouden moeten doen. Zorgt de aanleg van natuurvriendelijke oevers bijvoorbeeld ook echt voor een betere waterkwaliteit? Als we dit beter begrijpen, kunnen beleidsmaatregelen gericht ingezet worden.

Vervolgstappen Op dit moment is informatie van 17 waterlichamen opgenomen in het dashboard. In de toekomst kan het dashboard mogelijk verder ontwikkeld worden om een landelijk overzicht te geven van alle waterlichamen in Nederland en zo de actuele stand van zaken rondom het KRW-programma te laten zien. www.centerdata.nl/dashboarding

125 jaar geleden begon het met 5 man personeel

Het CBS door de jaren heen

Met kantoren in Den Haag, Heerlen en Bonaire, 600 statistische onderzoeken per jaar, 2.000 medewerkers, 75.000 vermeldingen in de media en 14 miljard cijfers in StatLine, valt het CBS niet meer weg te denken uit de hedendaagse datawereld. Maar hoe en wanneer is het ooit begonnen? Een blik terug in de tijd. *Astrid de Kock*

Het CBS werd opgericht op 9 januari 1899 en startte met 5 man personeel op het Binnenhof, onderin het gebouw van de Eerste Kamer, waar de Centrale Commissie voor de Statistiek haar secretariaat had. Via enkele omzwervingen kwam het CBS terecht op de Oostduinlaan in Den Haag en van 1970 tot 2009 huisde het CBS in de locatie aan de Prinses Beatrixlaan nr. 428 in Voorburg. Een gebouw met een vloeroppervlak van 60 duizend vierkante meter, veertien verdiepingen – vier met laagbouw, tien met hoogbouw – en vier vleugels.

De start in Heerlen

Op 1 mei 1974 gingen de eerste CBS-medewerkers in Heerlen aan de slag in het voormalig directiegebouw van de Oranje Nassaumijnen. In november 1978 namen de CBS'ers in Heerlen hun intrek in een nieuw gebouw, een betonnen monoliet, een schakeling van 16 gelijkvormige eenheden in een rechthoekig stelsel met een centrale binnenplaats en omringd door groen. Zowel de vestigingen in Heerlen als Voorburg zijn inmiddels vervuild voor een nieuw gebouw. Waarbij 'Voorburg' terugging

naar Den Haag, naar de relatief nieuwe wijk Leidschenveen. Meer informatie en foto's kun je vinden op de website van het CBS.

Mechanisatie en automatisering

Dat er niet alleen wat huisvesting betreft veel gebeurd is, zal duidelijk zijn. De mechanisatie en automatisering startte in 1916. In dat jaar nam het Centraal Bureau voor de Statistiek (CBS) ponskaartenapparatuur in gebruik voor sorteer- en telwerk en was daarmee de eerste ponskaartengebruiker in Nederland. In 1960 werd de eerste computer geïntroduceerd en vanaf 1980 groeide het gebruik van terminalverbindingen met mainframes en minicomputers. Na 1985 werden de terminalnetwerken vervangen door netwerken van pc's. Al deze wetenswaardigheden zijn terug te vinden in het document 'Een eeuw statistieken' (Erwich en van Maarseveen). Een prachtig document dat is geschreven ter gelegenheid van het eeuwfeest, inmiddels ook alweer bijna 25 jaar geleden.

Het CBS nu en later

En hoe is het dan nu? Bijna 125 jaar na dato? Niet alleen verzamelt en publiceert het CBS statistisch materiaal. Tegenwoordig is het zelfs mogelijk dat 'derden' via Remote-Access zelf onderzoek kunnen doen op deze data. Wie had dat ooit kunnen denken, destijds, op 9 januari 1899? De vraag is dan ook: 'waar staan 'we' in 2099'? De 'glazen bol' houdt het geheim. Hoeveel CBS-vestigingen zullen er dan zijn en waar? Bestaat Statline nog tegen die tijd? Zitten 'we' dan



CBS in Heerlen. Credits: CBS



CBS in Voorburg. Credits: CBS

inmiddels op 6.000 statistische onderzoeken per jaar in plaats van 600? En hoeveel medewerkers zullen het CBS draaiend houden? Wie, o wie, van de lezers van dit artikel gaat het meemaken? Het geheim wordt onthuld, later, aan onze (klein)kinderen. <https://www.cbs.nl/over-ons/dit-zijn-wij/onze-historie>

Nieuwe erfgoed-data in CLARIAH

Via de Heritage Data Research-call financiert CLARIAH momenteel drie onderzoeksprojecten die data uit de erfgoedsector beschikbaar maken voor de onderzoeksgemeenschap. De projecten bieden erfgoedinstellingen de kans om hun data verder te ontwikkelen en te delen volgens de FAIR-standaarden. In het project PARAWARD (Thunnis van Oort, Radboud Universiteit) betreft het de registers van de inwoners van Paramaribo tussen 1828 en 1846, in FAIR Photos (Leon van Wissen, Universiteit van Amsterdam) gaat het om de collectie van Fotopersbureau De Boer van het Noord-Hollands Archief, en in TW-NL (Amaury de Vicq, Rijksuniversiteit Groningen) om de Memories Database, met de nalatenschappen van Nederlanders die in 1921 overleden. (ER)

<https://bit.ly/3MG7IHH>



Harry Mulisch op het dak van de Bavokerk in Haarlem, 11 maart 1957. Bron: Noord-Hollands Archief, collectie Fotopersbureau De Boer, NL-HlmNHA_1478_13925B00_01

Gebruikerservaringen helpen onderzoeksinfrastructuur te verbeteren

Beeldvorming vluchtelingen onderzocht in MediaSuite

De representatie van Oekraïense vluchtelingen op de Nederlandse televisie is in de loop van het jaar na de Russische inval veranderd. Dat blijkt uit onderzoek met CLARIAH's MediaSuite.

Erica Renckens

‘Het zijn onze burenen.’ ‘Ze zijn net als wij.’ De berichtgeving over Oekraïense vluchtelingen in Nederland op de Nederlandse televisie was vlak na de Russische inval gerust uitbundig te noemen. Iedereen bood humanitaire hulp aan, van inzamelingsacties voor geld en spullen tot plekken in gastgezinnen. Na ongeveer drie maanden kandelde het beeld geleidelijk, en raakte het onderwerp overschaduwd, mede door de vluchtelingencrisis in Ter Apel. Dat is het beeld dat naar voren komt uit de eerste analyse van mediaonderzoeker Andrea Meuzelaar (UvA) in de MediaSuite, CLARIAH's online onderzoeksomgeving voor audio en video.

“Met verschillende zoektermen heb ik in de MediaSuite gezocht naar verslaggeving over Oekraïense vluchtelingen in Nederland vanaf 24 februari 2022 tot één jaar later”, vertelt Meuzelaar. “Zo kwam ik uit op 358 uitzendingen van vooral NOS Journaals, actualiteitenprogramma's en talkshows.”

Niet meer naar Hilversum

Meuzelaar voert dit onderzoekproject binnen de Fellowship Call van CLARIAH. Aan de hand van de gebruikerservaringen van onderzoekers in deze projecten verbetert CLARIAH de onderzoeksinfrastructuur. “Natuurlijk loop ik soms nog tegen kleine bugs aan, maar al met al ben ik dolenthousiast over de MediaSuite”, vertelt Meuzelaar. “Ik ben vooral te spreken over alle verschillende functionaliteiten. In de zoekresultaten kan ik bijvoorbeeld eenvoudig uitzendingen bookmarken en zo een eigen corpus samenstellen. Daarbinnen kan ik ook weer bookmarkgroepen creëren en zo verschillende subcorpora rondom



Oekraïense moeders met kinderen in Nederland. Bron: MediaSuite

bepaalde thema's maken, waar je dan ook weer op kunt filteren. Verder maak ik veel annotaties bij de video's waarin ik beschrijf wat me opvalt, bijvoorbeeld aan het taalgebruik of het beeld.”

Een verademing vergeleken bij het promotieonderzoek dat ze tien jaar geleden uitvoerde. “Toen onderzocht ik de geschiedenis van de tv-representatie van islamitische immigranten in Nederland. Zeven jaar lang ging ik naar Hilversum om het archief van Beeld en Geluid in te duiken. Het ging om extreem veel materiaal, dus ik moest een ingewikkeld systeem bedenken om alles bij te houden. Had de MediaSuite destijds maar bestaan.”

Beeldgebruik makkelijker te onderzoeken

De komende tijd zal Meuzelaar de verschillende thema's die ze ziet terugkeren in haar corpus verder uitwerken. “Bijvoorbeeld over het beeldgebruik. Je ziet vooral veel moeders met kinderen, terwijl je als het gaat over andere vluchtelingen eerder mannen en hekken ziet. Soms wordt dat verschil in behandeling met andere vluchtelingen expliciet benoemd in de programma's, dat is ook interessant om naar te kijken. Of de Nederlander als goede Samaritaan, die zijn huis opent of spullen inzamelt – die berichtgeving overschaduwde soms de vluchtelingen zelf. Genoeg te onderzoeken!”

<https://www.clariah.nl/projects/reurtv>

COLUMN

Historische kansen

Als digitale geesteswetenschapper leef ik in twee werelden. In de ene werk ik te midden van mensen die voldoening halen uit hun vermogen virtuele werkelijkheden te scheppen uit eentjes en nulletjes. Van digital twin tot klimaatmodel, van linked open data tot de optimalisering van GPU's, ze beheersen het allemaal. De andere wereld bestaat uit archief- en bibliotheekbezoekers die de gave hebben het verleden te reconstrueren. Zij analyseren de geschiedenis, brengen haar in verhalen weer tot leven, of proberen lering te trekken uit historische situaties. Het valt niet altijd mee om juist in deze twee werelden te leven. Zij zouden in elkaars verlengde kunnen liggen, of flink wat overlap kunnen vertonen. Maar dat is niet zo. De eerste wereld is de vleesgeworden digitalisering. Daar hoeft je niemand uit te leggen wat data of software is. Men spreekt over versiebeheer, refactoring, C++ en libraries, en men

begrijpt elkaar nog ook. De tweede wereld is marginaal gedigitaliseerd. Men weet wat MS Word is, is in staat verstandig over het internet te surfen, en kan onderzoeksmateriaal uit online archieven downloaden. Daar is meestal alles mee gezegd.

Waarom zouden de humaniora zich ook interesseren voor computers? Geesteswetenschappers vertrouwen op het menselijk brein, en dat brein is nog altijd ongeëvenaard. Zelfs ChatGPT kan er qua creativiteit, interpretatievermogen en serendipische associatiekracht niet tegenop. En natuurlijk gaat de digitalisering niet geheel aan de geesteswetenschappen voorbij. Voor allerhande beschouwingen over de digitalisering – ethisch, filosofisch, theoretisch, historisch – ben je bij hen aan het juiste adres. Het probleem zit hem niet in het denken, maar in het doen. In vergelijking met elk ander domein, inclusief de sociale weten-

schappen, lopen de geesteswetenschappen wat de digitalisering betreft ernstig achter. Toegegeven, de vergelijking gaat soms mank. Het is voor een computer heel wat eenvoudiger de bacteriële biodiversiteit van veengronden in kaart te brengen dan de culturele betekenis van negentiende-eeuwse chicklit te duiden. Dat ontslaat de geesteswetenschappen echter



Joris van Eijnatten Credits: Netherlands eScience Center

niet van hun taak, nee, hun morele plicht, zich intensiever op de digitalisering te oriënteren.

Op de enorme potentiële opbrengst van elementaire algoritmische vaardigheden voor het onderzoek in het geesteswetenschappelijk domein ga ik hier niet in. Waar het mij vooral om gaat, is het wetenschappelijk onderwijs. Als wetenschappelijk onderzoek maatschappelijk relevant moet zijn, dan geldt dat ook voor wetenschappelijk onderwijs. Toch heeft geen hoogleraar, afdelingshoofd, vakgroepsvoorzitter, instituutsdirecteur of decaan mij tot nog toe kunnen uitleggen, waarom curricula die elementaire algoritmische vaardigheden negeren, in het huidige tijdsgewricht nog voldoende relevant zijn. Zeker, het is óók relevant om heel snel veel teksten te analyseren, en om leesbare rapporten te schrijven. Daar zijn geesteswetenschappers goed in – maar pas op, want dit is

een vaardigheid waarover ChatGPT intussen ook al beschikt.

De digitalisering is permanent en alomtegenwoordig en dus maatschappelijk relevant. Als de geesteswetenschappen de bedoeling hebben nieuwe generaties van politici, journalisten, beleidsmakers, ambtenaren en kritische burgers op te leiden, dan wordt het hoog tijd bestaande curricula op de schop te doen. Geen student zou zonder substantiële digitale vaardigheden de poorten van de academie mogen verlaten.

Joris van Eijnatten

Prof. dr. Joris van Eijnatten is algemeen directeur van het Netherlands eScience Center en sinds 2020 hoogleraar Digitale Geschiedenis aan de Universiteit Utrecht. In 2022 werd hem (samen met Pim Huijnen) de Richard Deswarte Prize in Digital History toegekend. Hij geeft de column door aan Menno Rasch.