

Kennis over goed datamanagement gebundeld in één tool

CESSDA biedt nieuwe online tool

CESSDA ERIC, het consortium van Europese sociaalwetenschappelijke onderzoeksdata-archieven, stelt met de Expert Tour Guide een schat aan informatie over goed datamanagement online beschikbaar. *Ricarda Braukmann*

Veel Europese sociaalwetenschappelijke data-archieven organiseren al jaren datamanagementworkshops voor onderzoekers. Elf archieven hebben hun bestaande kennis gebundeld en voor iedereen beschikbaar gesteld in één online training. Handig voor zowel wetenschappers, financiers en beleidsmakers die onderzoeksdata steeds meer FAIR (Findable, Accessible, Interoperable, Reusable) willen maken. In de training zijn alle stappen van het onderzoeksproces te volgen: de planning van een project, het orga-

niseren van de dataverzameling, het verwerken van gegevens en het archiveren en publiceren van onderzoeksdata. Elke stap is voorzien van uitleg en datamanagementtips, verrijkt met visualisaties, voorbeelden en extra informatie. Hoe organiseer ik mijn data op een logische manier? Hoe ga ik om met gevoelige data? Waar sla ik mijn data het beste op? Hoe kan ik mijn data delen met anderen? Al deze en andere vragen komen aan bod.

Input van onderzoekers

Ellen Leenarts, werkzaam bij DANS en projectleider van de Expert Tour Guide: "Een training moet aansluiten bij wat onderzoekers nodig hebben. We hebben via een workshop onderzoekers gevraagd mee te denken over de inhoud en vorm van de training, en ze hebben de eerste versie van de module geëvalueerd. De



Elf archieven hebben hun bestaande kennis gebundeld en voor iedereen beschikbaar gesteld in één online-training, de Expert Tour Guide. credits Verbeeldingskr8

CESSDA-datamanagementexperts hebben de input verwerkt in de module zoals deze nu is."

Leenarts is blij met het eindresultaat en ook de reacties van de onderzoekers zijn heel positief. Volgens Kas-

per Otten, één van de jonge onderzoekers die de training beoordeelde, is de module "een goede tool voor iedereen die nieuw is op het gebied van datamanagement". cessda.eu/DMGuide

E-DATA & RESEARCH

Jaargang 12 | nummer 2

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

E-data & Research verschijnt drie keer per jaar en wordt mogelijk gemaakt door: CentERdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

INHOUD

3 ODISSEI ontwikkelt nieuwe infrastructuur

3 CBS-data ontsloten voor onderzoek

4 Bruikbare WO2-data dankzij project TRIADO

5 De 'canoniciteit' van Vondel en De Ruyter



6 De taal van deze tijd naast de taal van toen

6 KNAW Humanities Cluster bouwt bruggen

7 DG Onderzoek en Innovatie Smits aan het woord

8 De KB bewaart websites voor de toekomst

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



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

Vier onderzoekers trekken met behulp van papieren en digitale bronnen na waar schilderijen en tekeningen uit de museumcollectie zich bevonden tussen 1933 en het einde van WOII.

Maarten Heerlien

Tussen 1933 en 1945 viel naar schatting een kwart van alle kunst in Europa door roof, confiscatie en gedwongen verkoop in handen van de nazi's. Een gedeelte daarvan kwam na de oorlog niet terug bij de rechtmatige eigenaren, veelal Joodse families. Sinds 2012 reconstrueert het team herkomstonderzoekers van het Rijksmuseum de herkomstgeschiedenis van ruim 20.000 werken, 1.270 schilderijen en 19.000 tekeningen; allemaal vervaardigd vóór 1946 en tussen 1933 en 2012 aan de collectie van het Rijksmuseum toegevoegd.

Standaardisatie

Herkomstonderzoek is niet uniek voor het Rijksmuseum. De Nederlandse Museumvereniging initieerde vanaf 1998 verschillende onderzoeken onder haar leden om mogelijke roofkunst te identificeren. Het grote aantal objecten waarmee het Rijksmuseum te maken heeft, vraagt echter om een gefaseerde aanpak. Standaardisatie van de werkwijze is daarin essentieel, aldus Merel van Erp, één van de vier herkomstonderzoekers: "Elk werk dat in aanmerking komt, onderzoeken we volgens dezelfde processtappen en bronnen. Dat maakt het onderzoek transparant en de betrouwbaarheid van de uitkomst aannemelijk." Objecten worden eerst onderworpen

Project Herkomstonderzoek 1933-1945

Rijksmuseum speurt naar sporen van onrecht



Dit stel zilveren zoutvaten van Johannes Lutma (1584-1669) uit 1639 zijn hoogstwaarschijnlijk afkomstig uit de collectie van Emma Budge, geboren Emma Ranette Lazarus (1852-1937), dochter van een joodse handelaar uit Hamburg. Na haar dood in 1937 werd haar kunstcollectie geconfisqueerd en verkocht op veiling bij het veilinghuis, nog onder de naam, Paul Graupe. Op het moment van de veiling was Graupe zelf al gevlucht en het veilinghuis geariseerd door Hans W. Lange. In maart 1960 werden de zoutvaten samen met een ander paar op een veiling bij veilinghuis Frederik Muller aangekocht door de Gemeente Amsterdam. Dit paar heeft de Gemeente Amsterdam aan het Rijksmuseum verkocht (inv. nr. BK-1960-13-B). Het andere paar bevindt zich in het Amsterdam Museum (inv. nrs. KA 8050 en KA 8051). credits Johannes Lutma (1584-1669), 1639 hdl.handle.net/10934/RM0001.COLLECT.18422

aan interne bronnen, zoals jaarverslagen, archieven en documentatie. Vervolgens wordt, indien nodig, een vaste verzameling externe publicaties, databases en aanvullende bron-

nen onderzocht. Doel is om een aansluitende reeks transacties vast te leggen: overdrachten van het werk vanaf 1933 tot aan opname in de collectie. Deze transactiegeschiedenis

wordt opgenomen in het Collectie Management Systeem en gepubliceerd in het online collectieportaal Rijksstudio, vergezeld van een bibliografie herkomstonderzoek.

Lacunes opvullen

Ook lacunes, missende schakels in de transactiereeks, worden gepubliceerd. Van Erp: "Van de helft van de nu onderzochte schilderijen bevat de transactiereeks één of meerdere lacunes in de periode 1933-1945. Die werken interpreteren we niet als roofkunst, we constateren enkel dat er na het doorlopen van het onderzoeksproces nog vragen openstaan en we maken die vragen inzichtelijk." Dankzij digitalisering komen steeds meer bronnen beschikbaar die kunnen helpen lacunes op te vullen, zoals de 15.000 aangifteformulieren van in de oorlog verloren kunst die het Nationaal Archief momenteel in het project Vermiste Werken online brengt. Van Erp: "Helaas kunnen we reeds onderzochte werken met lacunes niet direct tegen elke relevante nieuwe bron aanhouden. Wellicht bieden nieuwe vormen van geautomatiseerde bronnenanalyse in de toekomst een oplossing." Het project Herkomstonderzoek loopt nog zeker tot eind 2020.

rijksmuseum.nl/nl/onderzoek/herkomstonderzoek

GEHOORD & BIJGEWOOND

British Library Labs Symposium 2017

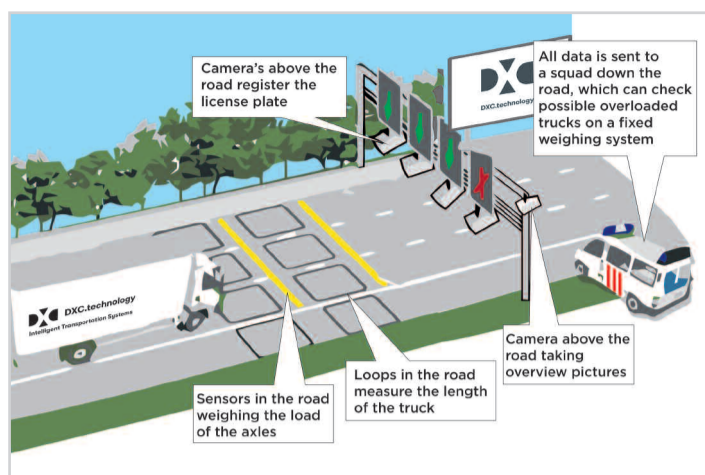
Lotte Wilms

Op 30 oktober vond bij de British Library (BL) in Londen het vijfde BL Labs Symposium plaats. Een jaarlijks initiatief om innovatief hergebruik van de digitale collectie van de British Library aan te moedigen.

Tijdens het symposium – dat geheel in het teken stond van hergebruik in alle mogelijke vormen – lichtte BL Labs de projecten toe die zij het afgelopen jaar heeft uitgevoerd. Ook reikte ze prijzen uit aan personen, teams en bedrijven die hun digitale collectie op een innovatieve manier hebben ingezet, in vijf categorieën: commercieel, wetenschap, educatie, artistiek én een prijs voor medewerkers.

Het BL Labs & Digital Scholarship-team houdt zich bezig met het promoten van de enorme collectie teksten, kaarten, muziek, afbeeldingen en handschriften van de British Library voor nieuwe vormen van hergebruik. Dit doet zij voornamelijk door partnerships aan te gaan en samen te werken in verschillende projecten. Een aantal van de tijdens het symposium gepresenteerde projecten viel op door de samenwerking met de creatieve industrie. De artistic award ging naar Michael Takeo Magruder. In Imaginary Cities heeft hij met behulp van gedigitaliseerde stadsplattegronden nieuwe steden gevisualiseerd. Dankzij een algoritme nemen deze plattegronden dagelijks in omvang toe. Leuk detail: ze zijn ook in 3D te renderen. Omdat deze kunstenaar graag de koppeling maakt tussen digitaal en fysiek, heeft hij ook enkele van deze nieuwe kaarten met behulp van oude verguld- en verzilver-technieken 'afgedrukt'.

Een ander mooi voorbeeld was het project dat BL Labs deed met Rob Sherman, een interactieve fiction writer-in-residence. Dat was alweer in 2014, maar daarom niet minder interessant. Sherman heeft bij de BL namelijk gewerkt aan een interactief boek, waarbij hij de collectie op een innovatieve manier heeft ingezet. Voor de tentoonstelling Lines in the Ice schreef hij een 19e-eeuwse karakter dat op zoek gaat naar de verloren schepen in de poolcirkel bij Canada. Sherman schreef niet alleen het verhaal, ook bakte hij scheepsbiscuit met recepten uit de BL-collectie en maakte hij – dank-



Gegevens over vervoer op de Nederlandse wegen zijn afkomstig uit online en papieren vragenlijsten en bijvoorbeeld uit zogenoemde Weigh in Motion (WIM)-systemen. credits DXC Intelligent Transportation Systems dxc.com/its

zij de hulp van restauratoren – een verloren gewaand scheepsjournaal. Op die manier gaf hij de tentoonstelling een extra dimensie. labs.bl.uk/British+Library+Labs+Awards

Wegvervoer onderzoeken door koppelen bronnen

Joris Mulder

In november organiseerde het Nederlandstalig Platform voor Survey Onderzoek (NPSO) een innovatiedag voor jonge onderzoekers. Eén van de sprekers was Jonas Klingwort, PhD-student aan de Universiteit van Duisburg-Essen. Hij vertelde over zijn promotieonderzoek waarin hij het wegvervoer in Nederland bestudeert door bestaande databronnen te combineren.

“In officiële statistieken zijn onderzoeken gebaseerd op steekproeftrekking en vragenlijsten gangbaar, maar alternatieve databronnen komen steeds meer beschikbaar”, zegt Klingwort. “Zo zijn de gegevens over vervoer op de Nederlandse wegen afkomstig uit online en papieren vragenlijsten (afgenomen onder vervoerders), maar worden ook sensordata verzameld met camera’s en sensoren door Weigh in Motion (WIM)-systemen op negen locaties in Nederland.

“We koppelen sensor- en vragenlijstdata die onafhankelijk van elkaar dezelfde doelvariabelen meten. Zo kunnen de antwoorden van de vragenlijst gevalideerd worden met behulp van Capture-recapture-technieken (meerdere keren hetzelfde meten, red.)”, licht Klingwort toe. In zijn onderzoek gebruikte hij alleen de records uit de WIM-database die overeenkwamen met records uit de vragenlijst. De nummerplaat diende hier als matchingvariabele.

“De resultaten van mijn onderzoek laten een aanzienlijke onder-rapportage zien: ritten die niet in de vragenlijst zijn opgegeven, maar wel zijn geregistreerd door de sensoren van de WIM-systemen. Dit toont dat het combineren van surveydata en sensordata zeer waardevol kan zijn”, concludeert Klingwort. “Bijvoorbeeld voor mensen die onderzoek doen naar en statistieken produceren over wegtransport, maar ook voor mensen die de CO2-uitstoot van het wegtransport onderzoeken en de invloed daarvan op het klimaat.”

De presentaties van Klingwort en andere jonge onderzoekers zijn beschikbaar op

npsso.net/evenementen/npsso-innovatiedag-22-november-den-haag

Huygens ING op Gala van de Wetenschap

Machteld Maris

Irene van Renswoude, onderzoeker bij de afdeling Wetenschapsgeschiedenis van het Huygens ING, sprak tijdens het Gala van de Wetenschap een column uit. Het was een sprankelende avond waarin topwetenschappers in een uitverkochte Stadsschouwburg vertelden over de nieuwste ontwikkelingen in hun vakgebied. Hier een ingekorte versie van de column.

‘Technologische vernieuwing’: Bent u er klaar voor? Of laat ik het anders vragen: wilt u eigenlijk wel innoveren? Heeft u de pictogrammen op uw smartphone weleens goed bekeken? Grote kans dat u gebruik maakt van Whatsapp. Dat groene vlakje met een telefoon, en niet zomaar een telefoon, zo’n ouderwetse, met een hoorn. Techniek mag zich dan in sneltreinvaart ontwikkelen, de samenleving is lang niet altijd klaar voor

innovaties. Elke ingrijpende vernieuwing gaat gepaard met angst om oude vaardigheden kwijt te raken en de controle te verliezen. En dat is precies de reden waarom vormgevers nieuwe media eruit laten zien als de oude. Om ons gebruikers gerust te stellen dat we de nieuwe informatiedrager kunnen vertrouwen. Vertrouwen is het smeermiddel van de technologische innovatie. En is dat erg? Is het nadelig voor ons als gebruikers dat we het hebben over *webpagina’s*, over *bladeren*, over *knippen en plakken*, alsof we nog steeds in een papieren wereld leven? Nee. Maar zolang we vasthouden aan oude vormen en oude taal, maken we onvoldoende gebruik van alle mogelijkheden van een inmiddels-al-niet-meer-zo-heel-erg-nieuwe technologie. En misschien kan dit ook niet anders. Als we terug in de tijd kijken, dan zien we dat de cognitieve patronen, en de sociale inbedding van nieuwe technieken, maar ook wetgeving en regulering, altijd hopeloos achterlopen. Maar wij onderzoekers moeten nu leren om op een nieuwe manier te kijken en te denken, om nog meer te profiteren van de digitale revolutie.” Lees de volledige column op huygens.knaw.nl/irene-van-rens-woude-op-gala-van-de-wetenschap/



Volgens Irene van Renswoude is vertrouwen het smeermiddel van de technologische innovatie. In een uitverkochte Stadsschouwburg sprak zij tijdens het Gala van de wetenschap haar column ‘technologische vernieuwing’ uit.

Het werd een sprankelende avond waarin topwetenschappers vertelden over de nieuwste ontwikkelingen in hun vakgebied.

credits Machteld Maris

AGENDA

19 - 22 februari • Barcelona

IDCC

Het thema van de 13e International Digital Curation Conference is ‘Beyond FAIR, from principles to practice to global join up’.

dcc.ac.uk/events/idcc18

13 - 14 maart • Berlijn

Open Science Conference 2018

Conferentie over onderzoeksdatamanagement met als thema de FAIR-principes en onderzoeksdata-infrastructuren.

open-science-conference.eu

16 maart • Den Haag

ESS Workshop

Ontmoetingsplaats voor gebruikers van de ESS-gegevens.

dans.knaw.nl/nl/actueel/ess-workshop-2018

19 - 20 maart • Amersfoort

ICT.Open 2018

Jaarlijkse conferentie waar onderzoekers op het gebied van ICT en micro-elektronica elkaar kunnen ontmoeten.

ictopen.nl

21 maart • Den Haag

Open dag DANS

DANS opent haar deuren en verzorgt samen met de Jonge Akademie en Promovendi Netwerk Nederland een middag over duurzame toegang tot en hergebruik van wetenschapsdata.

dans.knaw.nl/nl/actueel/agenda/open-day

26 maart • online

ICPSR summer program

Dien voor 26 maart een subsidie in voor het ICPSR summer program.

dans.knaw.nl/nl/over/diensten/easy/dans-subsidie-voor-icpsr-summer-program

10 april • Den Haag

MPN symposium

Symposium over het Mobiliteitspanel Nederland.

[Kimnet.nl](http://kimnet.nl)

22 - 25 mei • Chania

QQML 2018

De 10e Qualitative and Quantitative Methods in Libraries Conference is gericht op (data-)librarians.

qqml.org/event/qqml2018/

29 mei - 1 juni • Montreal

IASSIST & CARTO 2018

Jaarlijkse IASSIST-conferentie, met dit keer als thema ‘Once Upon a Data Point: Sustaining Our Data Storytellers’.

<https://www.library.mcgill.ca/iassist-carto2018/>

6 - 8 juni • Amsterdam

DHBenelux 2018

Jaarlijkse conferentie over digital humanities, met dit jaar als thema ‘Integrating Digital Humanities’.

2018.dhbenelux.org

22 - 24 juni • Toronto

ELPUB 2018

De 22e International Conference on Electronic Publishing, met als thema ‘Connecting the Knowledge Commons: From Projects to Sustainable Infrastructure’.

epress.uts.utoronto.ca/elpub2018/

COLOFON Uitgever: Stichting Uitgeverij E-data & Research Den Haag, 070-3494450, edata@dans.knaw.nl, edata.nl. Hoofd-/eindredacteur: Heidi Berkhout.

Redactie: Marika de Bruijne, Steven Claeysens, Machteld Maris, Erica Renckens, Saskia Scheltjens, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning.

Aan dit nummer werkten mee: Marnix van Berchum, Ricarda Braukmann, Arthur Charleston, Tessa Free, Laurie Hasselt, Maarten Heerliën, Cathelijne Hoorn, Harmen van der Meulen, Joris Mulder, Bas Nederveen, Kees Teszelszky, Thijs van der Veen, Lotte Wilms. Opmaak: Colette Sloots, Haarlem. Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Opplage: 6500 papier, 4500 digitaal. ISSN: 1872-0374. We hebben getracht alle belanghebbenden met betrekking tot het gebruikte beeldmateriaal te benaderen. Degenen die menen rechten te kunnen doen gelden, kunnen zich tot ons wenden. Toezending papieren en/of digitale versie is kosteloos aan relaties van de stakeholders en studenten in de alfa- en gamma-richtingen.

OVERNEMEN ARTIKELEN

Wilt u een artikel uit dit blad overnemen?

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

Nieuwe onderzoeksinfrastructuur voor sociale wetenschappen

ODISSEI komt op stoom

ODISSEI ontwikkelt een onderzoeksinfrastructuur voor de sociale wetenschappen. Wat gaat er precies gebeuren?

Erica Renckens

“Nederland behoort tot de wereldtop in de sociale wetenschappen, maar er zijn allerlei factoren die dat bedreigen”, vertelt Pearl Dykstra, hoogleraar Empirische Sociologie aan de Erasmus Universiteit. “Denk aan de beperkte middelen om bestaande datacollecties te consolideren en de beperkte computer-capaciteit om verrijkte databestanden te analyseren.”

ODISSEI moet ervoor zorgen dat Nederland zijn koppositie in het vakgebied behoudt. Afgelopen jaar is een start gemaakt met de ontwikkeling van deze onderzoeksinfrastructuur voor de sociale wetenschappen. Dykstra is een van de initiatiefnemers die ervoor hebben gezorgd dat het project inmiddels op NWO's roadmap voor grootschalige onderzoeksfaciliteiten is opgenomen. Momenteel is ze voorzitter van de Management Board van ODISSEI.

Data verbinden

Met name de samenwerking met het Centraal Bureau voor de Statistiek (CBS) maakt ODISSEI zo bijzonder, vertelt Dykstra. “Onderzoekers kunnen via remote access nu al gebruikmaken van de gegevens van het CBS, maar die gekoppelde register- en surveydata willen we binnen ODISSEI op een veilige en meer gebruiksvriendelijke manier beschikbaar maken. Omdat we ook focussen op interoperabiliteit, kunnen we via een datahub de verschillende dataverzamelingen van onze partners onderling verbinden. Voor de benodigde veiligheid en rekenkracht zijn we een samenwerking aangegaan met SURFsara.”

Tools voor analyse

ODISSEI maakt het mogelijk nieuwe actuele onderzoeksvragen te beantwoorden. Dykstra: “Wat zijn bijvoorbeeld de gevolgen van de decentralisatie van de zorg op lokaal niveau? En welk effect heeft het verhogen van de AOW-leeftijd op het grootouderschap?” Ook zal ODISSEI zich richten op het bijbrengen en ontwikkelen van tools om de gekoppelde data mee te analyseren. “We zullen deels werken met bestaande statistische technieken. Ook in andere vakgebieden zie je software die wij zouden kunnen gebruiken. Zo heeft CLARIAH een tool ontwikkeld waarmee je web-



ODISSEI, de Open Data Infrastructure for Social Science and Economic Innovation, is een belangrijke partner voor het CBS. ODISSEI ontwikkelt een duurzame wetenschappelijke infrastructuur voor de sociale wetenschappen in Nederland, en stimuleert het gebruik van de RA-faciliteit van het CBS bij haar deelnemers. Door de combinatie van de datarijksdom van het CBS, de rekenkracht en opslagcapaciteit van SURFsara en de infrastructuur van ODISSEI kunnen nieuwe onderzoeksvragen worden beantwoord, zoals bijvoorbeeld: “Welk effect heeft het verhogen van de AOW-leeftijd op het grootouderschap?” credits brfcs, pixabay

sites kunt ‘scrapen’. Of kijk naar de methode die we gebruiken om data uit accelerometers te analyseren. Oorspronkelijk is die tool ontwikkeld door vogelonderzoekers op Texel die wilden weten wanneer vogels vliegen of rusten. Zo kunnen

we elkaar allemaal verder helpen.” ODISSEI heeft inmiddels 27 deelnemers, waaronder alle universiteiten en verschillende kennisinstellingen. “Zij krijgen middels een soort abonnement allerlei voordelen, zoals reductie op het gebruik

van CBS-data en mogelijkheden voor dataverzameling. Ook willen we het prototype van de infrastructuur die er nu ligt, doorontwikkelen tot een pilot die we steeds verder kunnen uitbreiden.” odissei-data.nl

CBS-data beter toegankelijk voor onderzoekers

‘Net als internetbankieren’

Met de vernieuwde Remote Access-faciliteit kunnen onderzoekers makkelijker hun eigen data koppelen met die van het CBS. Ruurd Schoonhoven, senior relatie-beheerder bij CBS Microdata Services, vertelt.
Marika de Bruijne

Zelf analyses doen op brongegevens van het CBS en een eigen dataset koppelen aan microdata uit de CBS-bestandencatalogus, dat is de kern van de recent vernieuwde Remote Access(RA)-faciliteit van het CBS. Schoonhoven: “Voorheen moesten onderzoekers inloggen via een speciale remote-accesscomputer met vingerafdruklezer. Tegenwoordig werkt het systeem zoals internetbankieren: inloggen op de beveiligde omgeving van het CBS kan vanaf iedere computer. We stellen uiteraard wel eisen, zo moet de computer voorzien zijn van anti-virus-

software en moet het wifi-netwerk wachtwoordbeveiliging hebben. Ook voor de privacychecks die CBS uitvoert op te exporteren bestanden zijn meer flexibele varianten geïntroduceerd.”

Bestandencatalogus

Door de vernieuwing kunnen geautoriseerde onderzoekers makkelijker gebruikmaken van de CBS-bestandencatalogus met persoonsgebonden data over onderwijs en opleiding, wonen en werken, inkomen, gezondheid en uiteraard demografische gegevens. Maar ook bedrijfsgebonden gegevens over productie, innovatie en handel staan in de catalogus. “We hebben datasets van ongeveer driehonderd onderwerpen en over vijf tot twintig jaar beschikbaar”, legt Schoonhoven uit. “Omdat alle data onderling koppelbaar zijn, biedt dat uitgebreide mogelijkheden voor thema-overstijgend onderzoek.”

Jaarlijks maken ongeveer 700 onderzoekers hiervan gebruik. “Deelnemers aan ODISSEI krijgen een korting op de kosten”, aldus Schoonhoven.

En er is meer in beweging gezet via ODISSEI. Schoonhoven: “Mede via ODISSEI verkent het CBS met SURFsara of koppelingen op microniveau mogelijk zijn. Het combineren van de datarijksdom van het CBS met de rekenkracht en opslagcapaciteit van SURFsara kan grootschaligere kwantitatieve analyses mogelijk maken, bijvoorbeeld door klassieke sociaal-wetenschappelijke gegevens te combineren met genetische data.”

Meer weten?

Benieuwd naar de onderzoeksmogelijkheden? Neem dan contact op met CBS Microdata Services via microdata@cbs.nl of kijk op cbs.nl/microdata

KORT

CLARIAH-financiering voor Amsterdam Time Machine

Het CLARIAH-bestuur heeft besloten om het project Amsterdam Time Machine (ATM) een financiering van € 251.000,- toe te kennen. Dankzij deze financiering komt de geo-infrastructuur HisGIS van de Fryske Akademy beschikbaar voor ATM en maakt het voortaan onderdeel uit van de generieke CLARIAH-infrastructuur. Zo komen data en tools niet alleen ter beschikking voor onderzoek naar Amsterdam, maar ook voor geesteswetenschappelijk onderzoek in het algemeen. Daarnaast levert dit deelproject aan CLARIAH drie onderzoeks-use cases (in de taalkunde, sociale & economische geschiedenis en mediastudies), die elkaar versterken en samen de mogelijkheden van de CLARIAH-infrastructuur demonstreren. (TvdV) clariah.nl

Online survey digital humanities

In het kader van zijn promotieonderzoek naar de praktijk van digital history, doet Max Kemman (Universiteit van Luxemburg) onderzoek naar de samenwerking tussen geesteswetenschappers en informatici. Meer informatie hierover staat online: maxkemman.nl/about-survey. Graag zou Max iedereen die deelneemt aan een samenwerking tussen geesteswetenschappers en informatici (bijvoorbeeld in een centrum, lab, project, etc.) willen vragen om een online survey in te vullen. Dit duurt ongeveer tien minuten, met vragen in het Engels. (SC) maxkemman.nl/survey

Save the date: 21 maart open dag

Op 21 maart organiseert DANS samen met de Jonge Akademie en Promovendi Netwerk Nederland een open dag over onderzoeksdatamanagement, data-archivering en datahergebruik. Naast een debat met de Jonge Akademie, het Promovendi Netwerk Nederland en de andere aanwezigen over de mogelijke obstakels bij het delen en hergebruiken van (open) data en de oplossingen hiervoor passeren een aantal succesverhalen – in pitches van tien minuten – op het gebied van open data de revue. Ook is er een informatiemarkt over de diensten van DANS. De dag wordt afgesloten met een borrel. Kijk voor meer informatie en online aanmelding op de website van DANS. (HB) dans.knaw.nl/nl/actueel/agenda/open-day

Van analog naar digitaal bruikbaar WO2-onderzoeksarchief

TRIADO maakt archieven doorzoekbaar

Hoe maak je van kilometers ongestructureerd archiefmateriaal bruikbare digitale onderzoeksdata? Naar die methodiek is het project Tribunaalarchieven als Digitale Onderzoeksfaciliteit (TRIADO) op zoek.
Netwerk Oorlogsbronnen

Het meest geraadpleegde Tweede Wereldoorlogarchief van Nederland, het Centraal Archief Bijzondere Rechtspleging (CABR), dient als casus voor TRIADO. Op dit moment is het CABR alleen doorzoekbaar via een database, op naam van verdachten, en zijn de stukken niet digitaal raadpleegbaar. TRIADO, dat mogelijk is gemaakt dankzij het KNAW-Onderzoeksfonds, ontsluit en verkrijgt voor het eerst op een geautomatiseerde wijze dit rijke WO2-archief digitaal en op documentniveau. Hierdoor wordt het mogelijk om gegevens over tientallen personen, gebeurtenissen en plaatsen in dossiers te doorzoeken. En dus om grootschalig vergelijkend onderzoek te doen.

Digitale technieken

Het Netwerk Oorlogsbronnen, Nationaal Archief, Huygens ING en NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies zijn in 2017 van start gegaan met het generieke deel van het project. Hierin wordt



Een fysiek CABR-dossier wordt met TRIADO digitaal doorzoekbaar gemaakt.
foto Nationaal Archief

onderzocht welke digitale technieken gebruikt kunnen worden om hybride archiefmateriaal digitaal te ontsluiten. Begin 2018 is een steekproef uit het CABR gedigitaliseerd en doorzoekbaar gemaakt (ge-OCR'd). Met speciale software voor Named Entity Recognition en automatische datumextractie wordt vervolgens informatie over onder meer personen, locaties en datumaanduidingen vastgesteld. De uit de documenten gehaalde gegevens over wie, wat, waar en wanneer worden gekoppeld aan referentiedata uit onder meer de

WO2-thesaurus en de Basisregistraties Adres- en Gebouwen. Marielle Scherer, data-manager Huygens ING: "Deze verrijking-fase zorgt voor structurering van de data en een betere doorzoekbaarheid van de dossiers uit de steekproef".

Nieuwe vragen

Het projectteam is enthousiast. Anne Gorter, bij het Nationaal Archief projectleider nadere toegankelijkheid CABR: "Door deel te nemen aan dit project bereiden we ons voor op

de toekomst en op klantvragen die gaan komen". Edwin Klijn, projectleider TRIADO en programmamanager Netwerk Oorlogsbronnen: "Ik kan haast niet wachten om te zien welke nieuwe vragen we kunnen stellen aan dit archief en hoe digitale methodieken ons daarbij kunnen helpen".

Data laten spreken

Voor het tweede, onderzoeksdeel van TRIADO formuleert onderzoeker Ismee Tames van het NIOD met een wetenschappelijke adviescommissie verschillende kwantitatief-statistische onderzoeksvragen. Tames: "Denk aan vragen als: 'Wie werd wanneer door Nederlandse collaborateurs gearresteerd? Hoe verhoudt dit zich tot de ontwikkeling van het Duitse vervolgingsbeleid? Wie waren de collaborateurs en verzetslieden eigenlijk: zien we veranderingen in bijvoorbeeld leeftijd, gender of regio?' Dat we nu de data tot ons laten spreken in plaats van een van tevoren opgestelde lijst met namen terug te zoeken, is een veelbelovende stap voorwaarts".

bit.ly/TRIADOproject

Het Netwerk Oorlogsbronnen wordt gefaciliteerd door het NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies en gefinancierd door het Ministerie van VWS, Vfonds en VSBfonds.

SINDS KORT BESCHIKBAAR

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

CentERdata

• God in Nederland

Verschuivingen in de christelijke godsdienst, de godsdienstigheid van Nederlanders en de rol van de kerken in het publieke debat, daar wil het onderzoek 'God in Nederland' zicht op krijgen. De meest recente meting is in juni 2015 afgenomen in het LISS panel door Joris Kregting (KASKI, Radboud Universiteit). Voor de eerste keer via een websurvey werd de respondenten gevraagd hoe men in het eigen leven tegen religie en spiritualiteit aankijkt en op welke wijze men daar mee bezig is. Het onderzoek wordt sinds 1966 iedere tien jaar gehouden en is in boekvorm gepubliceerd. Het databestand God in the Netherlands is beschikbaar via LISS Data Archive. lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- CentERdata, april 2017 - mei 2017, Work and Schooling - Wave 10
- CentERdata, juni 2017 - juli 2017, Economic Situation: Housing - Wave 10
- Hubner, S.; Soest, A. van, juli 2017, Time Use and Consumption - Wave 5
- Scheele, D.; Fouarge, D., juni 2017 - juli 2017, Dutch Skill Survey (NSS) 2017
- Voet, J. van der, mei 2017, ABW Waste disposal



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

DANS

• Data nieuw aardewerkonderzoek beschikbaar

Het project Data Atlas van Byzantijnse en Osmaanse materiële cultuur: het archiveren van Middeleeuwse en Post-Middeleeuwse archeologische veldwerkgegevens uit het oostelijk Middellandse Zeegebied (600-2000), Phase 1 (Vroom, Dr. J.A.C., Faculty of Archaeology, Leiden University, 2017) toont de resultaten van aardewerkonderzoek vanuit opgravingen en surveys in het oostelijke Middellandse Zeegebied. Mede hierdoor is nu meer bekend over de typochronologie van aardewerk uit de periode 600 na Christus tot heden binnen het onderzoeksgebied en is meer kennis verzameld over onder meer de relatie tussen productie, distributie en gebruik van aardewerk in de vier stedelijke centra Athene, Butrint, Efeze en Tarsus. Het onderzoek is onderdeel van het NWO-VIDI Project (276-61-003) 'Material Culture, Consumption and Social Change: New Perspectives for Understanding the Eastern Mediterranean during Byzantine and Ottoman Times' (2010-2015) en ontving een KDP-subsidie vanuit DANS voor het ontsluiten van de data.

DOI: 10.17026/dans-zuq-ewst

Ook sinds kort beschikbaar:

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

- Huijgen, T.G. (Research Centre for Education and the Labour Market – ROA) (2017): HBO-Monitor 2016. DANS. DOI: 10.17026/dans-z5w-4yhk
- Chamberlain, B.P. (Imperial College London) (2017): Real-Time Community Detection in Full Social Networks on a Laptop. DANS. DOI: 10.17026/dans-2bc-4qgc
- Stoof, R (2017): Studentenonderzoek 2017. DANS. DOI: 10.17026/dans-xph-jdnu
- Kohnstamm Instituut, Universiteit van Amsterdam (2016): Pre-COOL tweejarigen-cohort vijfde meting 2014-2015. DANS. DOI: 10.17026/dans-xjb-8wvj



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



Keramische vondsten van de opgravingen.
credits N. Gail. oeaw.ac.at/oeai/forschung/tauforschung/ephesos-ayasuluk-forschungen

Wie wint de Nederlandse Dataprijs 2018?

Komend najaar wordt weer de Nederlandse Dataprijs uitgereikt. Een prijs voor een onderzoeker of onderzoeksgroep die extra bijdraagt aan de wetenschap door onderzoeksdata beschikbaar te maken voor aanvullend of nieuw onderzoek. Bij de vorige editie in 2016 sprak de jury over 'allemaal mooie voorbeelden van het toegankelijk maken en delen van onderzoeksdata'. De organisatie van de Nederlandse Dataprijs is in handen van Research Data Netherlands, een samenwerkingsverband tussen 4TU.Centre for Research Data, DANS en SURFsara. Binnenkort staat meer informatie over de Dataprijzen 2018 op de website van RDNL. (HB)

researchdata.nl

Update voor OpenSoNaR+

De webinterface van OpenSoNaR+, het referentiecensus voor geschreven en gesproken Nederlands, heeft begin dit jaar een update ondergaan. De 540 miljoen woorden aan allerhande teksten en ruim 900 uur aan spraakopnames in het corpus zijn nu toegankelijk voor iedereen met een CLARIN-account. Bovendien is de zoekfunctie flink verbeterd, met uitgebreidere zoekmogelijkheden en snellere resultaten tot gevolg. Zo kunnen gebruikers de zoekresultaten nu groeperen op contextwoorden of gebruikmaken van een random sample als steekproef. (ER)

opensonar.clarin.inl.nl/

Letterkundige Joep Leerssen over canonvorming en cultureel zelfbesef

‘Wat is de impactfactor van Vondel of Michiel de Ruyter?’

KB-fellow Joep Leerssen visualiseert het verloop van de ‘canoniciteit’ van historische figuren en teksten met behulp van data uit de KB-thesaurus en de DBNL. Erica Renckens

“Als wetenschappers worden we afgerekend op onze impactfactor. Hoeveel publicaties heb je in welke bladen? Hoe vaak worden die papers vervolgens weer geciteerd?” Joep Leerssen, hoogleraar Moderne Europese Letterkunde aan de Universiteit van Amsterdam, bedacht dat je zo’n impactfactor ook zou kunnen berekenen voor schrijvers, filosofen en andere belangrijke personen uit de Nederlandse geschiedenis. “Wat is bijvoorbeeld de impactfactor van Vondel of van De Ruyter? En hoe is die veranderd in de loop der tijd?”

Veertiende fellow

Tijdens zijn fellowship bij de Koninklijke Bibliotheek (KB) en het NIAS kreeg Leerssen de kans deze vragen te beantwoorden met behulp van de collecties gedigitaliseerde tekstbestanden van de KB. Vanaf september 2017 tot en met afgelopen januari was hij de veertiende fellow – de vijfde op rij die gebruikmaakte van digitaal, in plaats van analoog, materiaal. Eind januari sloot hij zijn fellowship af met een publiekslezing waarin hij zijn bevindingen deelde. E-data & Research sprak de onderzoeker ongeveer een maand daarvoor.

“Voordat ik begon met dit fellowship hield ik me al bezig met het analyseren van netwerken”, vertelt Leerssen. Zo stond hij aan de wieg van de Encyclopedia of Romantic Nationalism in Europe (ERNiE), een digitale encyclopedie met artikelen over belangrijke thema’s en personen uit negentiende-eeuws Europa. “De encyclopedie is opgebouwd als SQL-database. Die bevat onder andere uitgebreide briefwisselingen waar je hele mooie netwerkvisualisaties van kunt maken.”

Sneller onderzoek

Angst voor het gebruik van digitale methoden voor zijn onderzoek heeft Leerssen nooit gekend. “Eigenlijk werkt elke geesteswetenschapper sowieso met databases: een bibliografische kaartenbak is in feite al een database. Een computer kan die, nadat hij gedigitaliseerd is natuurlijk, alleen veel sneller doorzoeken dan je als onderzoeker kunt.”

Aan de hand van de impactfactor van culturele entiteiten, onderzocht Leerssen tijdens zijn fellowship de canonvorming in het Nederlandse culturele zelfbesef. “Op dat gebied zijn er grofweg twee scholen te onderscheiden. De ene stelt dat de Nederlandse identiteit oud en bestendig is, de andere dat dit tijdens het nationalisme van de 19e eeuw een flinke oppepper heeft gekregen. Dat wilde ik graag empirisch meten. Werden Erasmus en Michiel de Ruyter altijd al als nationale iconen gezien of zijn zij dat later pas geworden?” Om de ‘canoniciteit’, de impactfactor van een



“Het werken met de complexiteit van grote datasets is vaak moeilijker dan gedacht, maar het levert wel veel op.” foto Bart van Vliet

INTERVIEW

‘Een bibliografische kaartenbak is in feite al een database’

culturele entiteit, te bepalen, kon Leerssen niet alleen kijken naar het aantal vermeldingen in de KB-thesaurus (de digitale database waarop de catalogus draait) en in de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL). “Namen leven voort in standbeelden, naamgeving van instituten, straatnamen, maar dat zegt niks over hoe levendig het bijbehorende gedachtegoed is. Er is een verschil tussen naamsbekendheid en canoniciteit”, aldus Leerssen.

3D-film als spin-off

“Je kunt grofweg drie typen canoniciteit onderscheiden”, legt hij uit. “Anne Frank is een voorbeeld van actieve canoniciteit. Haar boek wordt nog steeds gelezen en belangrijk gevonden. Dan heb je academische canoniciteit, waarbij een naam belangrijk is in bepaalde vakkringen, maar bij het grote publiek verder inhoudelijk niet bekend is, zoals de

romans van Couperus. En ten slotte heb je nog banale canoniciteit, waarbij het louter om naamsbekendheid gaat. Een naam als Boerhaave kennen de meeste mensen alleen van de Boerhaavekliniek, maar slechts weinigen weten ook wie het was.”

Om dit aspect te registreren, keek Leerssen onder andere naar de mate van ‘procreativiteit’, ‘het vermogen om spin-offs op te leveren’. “Een boek dat alleen op een plank staat is eigenlijk dood; er moet wel iets mee gebeuren. Commentaren, herdrukken, bewerkingen. Kijk naar *Les Misérables* van Victor Hugo, dat boek is verfilmd, verstript en er is een musical van. En *Vondels Gijsbrecht van Aemstel* wordt elk jaar opnieuw gespeeld in de Amsterdamse Schouwburg – dat laat zijn sporen na.” Ook de vroege aanwezigheid in nieuwe media is een indicatie volgens Leerssen: “De overwintering op Nova Zembla, een canoniek thema van dichter Tollens tot de schoolplaat van Isings, was de eerste Nederlandse 3D-filmproductie.”

Zoekresultaten wegen

Vijf maanden is niet veel om zo’n onderzoeksvraag te beantwoorden, maar gelukkig stond Leerssen er niet alleen voor. “Ik heb hele goede hulp gehad van de informatici van de KB”, vertelt hij. “Zij schreven de scripts waarmee ik de enorme catalogi kon doorzoeken en verfijnden deze als dat nodig was. Dat bood de technische mogelijkheden.” De

scripts werden vervolgens geïmplementeerd in NodeGoat, de online infrastructuur waarmee onder andere ERNiE opereert, ontwikkeld door ex-studenten van Leerssen. “Na een week of acht kon ik al beginnen met datamining. Sinds november kwam de boel echt in een stroomversnelling. Inmiddels zijn er veertig- tot vijftigduizend entiteiten geïmporteerd. Dat gaat dan van personen als de Vlaamse dichter Maerlant, of Bilderdijk, tot werken als Van den vos Reynaerde.”

Leerssen moest alle zoekresultaten van deze entiteiten in Delpher en de DBNL zorgvuldig wegen. “Honderd hits in de 17e eeuw is iets anders dan honderd hits nu; daar moet een inflatiecorrectie overheen. En een vermelding door historicus Johan Huizinga legt meer gewicht in de schaal dan een anonieme vermelding in een obscuur dorpsblaadje. De absolute aantallen moesten gewogen aantallen worden.”

En welke school had dan gelijk wat betreft de Nederlandse identiteit? “Nou, je ziet een dip in de 18e eeuw en pieken rond 1870 en 1920-30”, zegt Leerssen. “Maar die pieken lijken vooral het nationalistische klimaat van de tijd aan te geven. De oorzaak-gevolgrelatie lijkt dus andersom te zijn. Dat vind ik het mooie van zo’n fellowship: je begint met de ontwikkeling van een instrumentarium en gaandeweg verandert je vraagstelling en de theorie.” De komende tijd zal Leerssen zijn bevindingen publiceren.

Nieuwe perspectieven

De digitale geesteswetenschappen ziet Leerssen als een uitdaging. “Het werken met de complexiteit van grote datasets is vaak moeilijker dan gedacht, maar het levert wel veel op. Het biedt nieuwe perspectieven.” De komende tijd zal hij zijn onderzoek nog voortzetten. “Het cultuurnationalistische gedachtegoed is nu weer heel actueel, met het populisme en de alt-rightbeweging. Dat is interessant om te onderzoeken.”

kb.nl/organisatie/kb-fellowship

Joseph Theodoor (Joep) Leerssen

Joep Leerssen is een Nederlands literatuurwetenschapper en historicus. Hij studeerde vergelijkende literatuurwetenschap en Engels aan de Universiteit van Aken en vervolgde met Anglo-Irish Studies aan het University College te Dublin. Daar voltooide hij ook zijn conservatoriumstudie. In 1986 promoveerde Leerssen cum laude aan de Universiteit van Utrecht op het proefschrift *Mere Irish and Fior-Ghael*. Sinds 1991 is hij hoogleraar Moderne Europese Letterkunde aan de Universiteit van Amsterdam. Van 1995 tot 2006 was Leerssen directeur van het Huizinga Instituut. In 2003 was hij gasthoogleraar aan Harvard University. In 2008 ontving hij de Spinozapremie. In 2009 werd hij erelid van de Royal Irish Academy. In 2010 werd hij door de Koninklijke Nederlandse Akademie van Wetenschappen benoemd tot Akademiëhoogleraar. Leerssen kreeg in 2015 de Dr. Wijngaendts Francken-prijs toegekend voor zijn boek *Spiegelpaleis Europa*.

Het Rijksmuseum evalueert eigen objectbeschrijvingen

Frisse blik op bestaande terminologie

Hoe gebruik je de taal van deze tijd, zonder de geschiedenis te kort te doen? Het Rijksmuseum evalueert sinds 2015 kritisch de eigen objectbeschrijvingen en past deze waar nodig aan, naast de bestaande invoer.

Heidi Berkhout

Om een breed samengesteld, hedendaags (inter)nationaal publiek enthousiast kennis te laten maken met kunst en geschiedenis, moet je als museum de juiste taal spreken. Bijvoorbeeld tijdens tentoonstellingen, in publicaties, op websites en in audiotours. Een belangrijke rol hierbij speelt de werkgroep Terminologie. De werkgroepleden beoordelen en wijzigen objectbeschrijvingen in Adlib, het collectie-managementsysteem (CMS) van het Rijksmuseum. Data uit dit CMS worden namelijk op de Rijksmuseumwebsite getoond en via API's uitgeleverd aan derden, zoals Europeana, Wikidata en Wikimedia.

Kritisch oordeel

“Sommige objectbeschrijvingen bevatten taalgebruik dat tegenwoordig ook als kwetsend kan worden ervaren. In andere beschrijvingen klinkt een (te) Eurocentrisch perspectief door. Om die reden houden we ons bezig met het kritisch beoordelen van de tot nu toe gebruikte terminologie,” aldus Bas Nederveen, informatiespecialist Collectie Informatie van de afdeling Research Services.

Iedere objectbeschrijving wordt afzonderlijk bekeken. Inmiddels zijn termen uit ongeveer achthonderd beschrijvingen tegen het licht gehouden, waarbij in ongeveer driehonderd gevallen is besloten om een nieuwe titel of beschrijving op te nemen in de database naast



de bestaande invoer. Het Rijksmuseum gooit dus geen ‘oude’ objectbeschrijvingen weg. Nederveen: “Alle informatie in de database blijft bewaard en wij documenteren onze afwegingen, zodat deze voor geïnteresseerden beschikbaar zijn en blijven. Wij onderzoeken de mogelijkheid om ook de oude beschrijvingen op de website te tonen, zodat het hele proces online te volgen is.”

Maatwerk

“Het maken van een nieuwe beschrijving is maatwerk. Dat doen we samen met vele andere experts. In de werkgroep zijn verschillende afdelingen van het Rijksmuseum vertegenwoordigd en wordt regelmatig overlegd

met externe conservatoren en informatie-specialisten. Ook wordt er samengewerkt met andere musea en advies ingewonnen bij belangvertegenwoordigers. Geïnteresseerden in deze terminologieproblematiek ontmoeten we graag tijdens een symposium later dit jaar, georganiseerd door het Museum voor Wereldculturen (als overkoepelende organisatie van de volkenkundige musea in Nederland), de Rijksdienst voor het Cultureel Erfgoed, de Nederlandse Museum Vereniging en het Rijksmuseum.

Meer informatie over dit symposium volgt via de websites van de organisatoren.”

rijksmuseum.nl/nl/onderzoek/terminologie

Nadere bestudering van een object of de documentatie levert vaak informatie op voor een betere beschrijving. Zo gaat het meisje op deze tekening niet langer door het leven als ‘Negerkind in locale kleding’, maar als ‘Portret van Affie’. Haar naam bleek gewoon op de tekening te staan.

credits Portret van Affie, vroeger bekend als ‘Negerkind in locale kleding’, Jacob Marius Adriaan Martini van Geffen (1821-’66), aquarel, gewassen inkt, potlood, ca. 1859

Stagiaire gezocht De werkgroep Terminologie zoekt een stagiaire voor 4 maanden (maart-juni), 2 dagen per week. Interesse? Mail naar B.Nederveen@rijksmuseum.nl.

Sinds oktober 2017 timmeren ze aan de weg, de drie musketiers van het Digital Humanities Lab. Teamleider Van Erp: “Laatst werden we de ‘incubator-groep’ genoemd, dat vind ik mooi.”

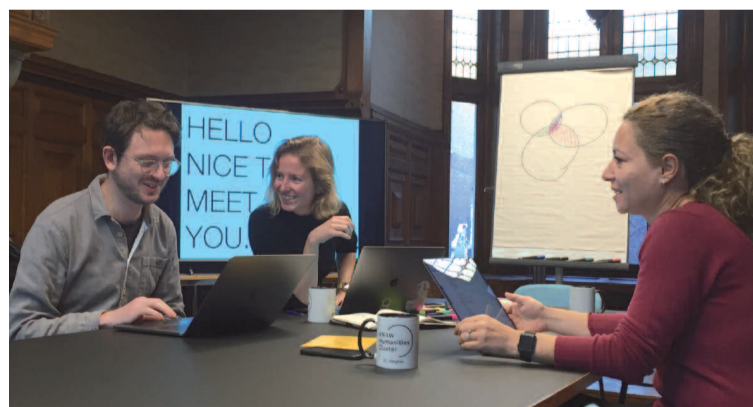
Machteld Maris

Het team heeft de opdracht om Digital Humanities in het KNAW Humanities Cluster te versterken en uit te breiden. Enerzijds door met het DHLab een eigen onderzoeksprogramma vorm te geven, anderzijds om onderzoeksprogramma's binnen en buiten het Cluster op te zetten. De verwachtingen zijn hoog gespannen.

Kruisbestuiving

Marieke van Erp: “We doen zelf onderzoek, hebben verschillende achtergronden en maken deel uit van de drie instituten in het Cluster. We vullen elkaar aan. Dat is een voordeel. Bovendien heeft het Cluster een unieke mix van onderzoekers in verschillende disciplines, en

DHLab: bruggenbouwers van het KNAW Humanities Cluster



Van links naar rechts: Melvin Wevers, historicus en digital humanities-specialist, teamleider Marieke van Erp, taaltechnoloog en specialist op het gebied van semantische analyse van teksten op entiteiten en gebeurtenissen en Adina Nerghes, communicatiewetenschapper en text analytics-specialist. foto Thijs van der Veen

onmetelijke datasets.” Melvin Wevers: “Het is zo goed om na te denken over andere, digitale manieren om een onderzoek aan te pakken.

Kunnen we nieuwe methoden aan dragen waarmee onderzoekers nog beter hun onderzoek kunnen uitvoeren? Computers kunnen onder-

zoeker op een nieuwe manier laten kijken naar hun eigen wetenschappelijk onderzoek.” Adina Nerghes: “Onze volgende stap is het stimuleren van collega's om nieuwe projecten op te zetten en nieuwe samenwerkingen aan te gaan.”

Digitaal én traditioneel

Alle drie benadrukken ze dat Digital Humanities een aanvulling op het traditioneel onderzoek is, niet een vervanging. Adina: “In de sociale wetenschappen is een vergelijkbare ontwikkeling zichtbaar. Digitale onderzoeksmethoden worden al jaren gebruikt, maar nu heeft iedereen het over ‘computational social science’. Er wordt onderscheid gemaakt tussen kwalitatief en kwantitatief onderzoek. Maar in mijn ogen staan ze niet tegenover elkaar, ze

vullen elkaar aan. Dat wij ons richten op Digital Humanities wil dus niet zeggen dat we daarmee de traditionele onderzoeksmethodes achter ons laten.”

Drie thema's

Marieke: “Na vele verkennende gesprekken hebben we nu drie thema's gekozen: Connections, Change en Entities and Events. Hiermee gaan we verder. We hebben veel in huis, maar we zoeken ook steeds verder naar samenwerkingen met collega's die ons hierin kunnen aanvullen. En andersom geldt dat natuurlijk ook, als er onderzoekers zijn van andere instellingen of organisaties die met onze data aan het werk willen, dan staan we daar open voor.

huc.knaw.nl/digital-humanities-lab

Robert-Jan Smits, directeur-generaal Onderzoek en Innovatie van de Europese Commissie:

‘Kenniss delen, daar gaat het om’

“We zien dat steeds meer onderzoekers bereid zijn hun data en kennis te delen in plaats van geobsedeerd te zijn door het ‘zo snel mogelijk publiceren’. Ik vind dit een positieve ontwikkeling.’ Aan het woord is Robert-Jan Smits, directeur-generaal Onderzoek en Innovatie van de Europese Commissie.

Heidi Berkhout

INTERVIEW

“Onderzoekers delen steeds vaker kennis, data en ervaring al tijdens het onderzoeksproces, met hulp van nieuwe (digitale) technologieën en toegenomen computercapaciteit. Zo wordt de wetenschap betrouwbaarder (we kunnen beter data verifiëren), efficiënter (we delen resources en werken beter samen) en transparanter (de werkwijze is toegankelijker). Het is van belang dat onderzoekers hierbij maximaal worden ondersteund, door universiteiten en door subsidieverstrekters zoals NWO. Ook de Europese Commissie draagt hieraan bij met de European Open Science Cloud. Dit initiatief beslaat een scala van maatregelen. Zo hebben wij onder Horizon2020, het EU-programma voor Onderzoek en Innovatie een krachtig beleid opgezet op Open Access Publicaties en op Open Access Data.”

DMP als verplichting

Smits vervolgd: “Data Management Plans (DMP’s) zijn bijvoorbeeld verplicht gesteld voor aanvragers van Europese onderzoeksgelden. DMP’s dragen bij aan het creëren van een cultuur van data stewardship in de onderzoeksweld en leidt ertoe dat de FAIR principes (data ‘vindbaar, toegankelijk, uitwisselbaar en herbruikbaar’ maken) realiteit worden. Onderzoekers worden betrokken bij onze stappen. Zo is in een survey vanuit OpenAIRE en



“Samen bouwen aan de European Open Science Cloud van morgen”.
credits EC DG Onderzoek & Innovatie

de FAIR-data-expertgroep onze benadering op het gebied van DMP’s en de FAIR-richtlijnen beoordeeld door onderzoekers. In een webinar begin januari werden vervolgens de resultaten van dit onderzoek gedeeld. Een mooi voorbeeld van co-design, het samen bouwen aan de

EUScienceInnov

Volg @EUScienceInnov, het twitteraccount van het Directoraat Generaal Onderzoek en Innovatie van de @EU_Commission. Ook op youtube.com/user/InnovationUnion staat informatie over bijvoorbeeld de effecten van Europese onderzoeksprojecten vanuit Horizon2020 en andere kaderprogramma’s op de samenleving.

EOSC, met stakeholders in het veld. Eigenlijk zouden alle onderzoeksfinanciers in Europa DMP’s verplicht moeten stellen. NWO doet dit al. Een andere positieve ontwikkeling is dat veel onderzoekers al uit eigen initiatief DMP’s maken, onafhankelijk van de eisen van onderzoeksfinanciers. Simpelweg omdat ze dit beschouwen als onderdeel van goed onderzoek in de globaal vernetwerkte onderzoeksweld. Nog een mooi voorbeeld is dat wetenschappelijke communities onderzoekers niet meer alleen willen beoordelen op ‘aantal publicaties’, maar ook op ‘goed datamanagement’. De GO FAIR-beweging draagt dit uit, en ook CESSDA heeft bijvoorbeeld een ‘Expert Tour Guide’ als ondersteuning bij datamanagement beschikbaar gesteld.

Robert-Jan Smits

Robert-Jan Smits studeerde aan de Universiteit Utrecht, aan het Institut Universitaire d’Hautes Études Internationales in Zwitserland en aan de Fletcher School of Law & Diplomacy in de Verenigde Staten. Na zijn studies werkte hij enkele jaren voor het Nederlandse Ministerie van Economische Zaken. In 1989 vertrok Smits naar de Europese Commissie in Brussel. Sinds 2010 is hij directeur-generaal van het Directoraat Generaal Onderzoek en Innovatie.

Infrastructuren koppelen

De European Open Science Cloud wil bestaande e-onderzoeksinfrastructuren aan elkaar koppelen. Hierdoor kunnen onderzoekers kennis en gegevens opslaan en delen over nationale grenzen en wetenschappelijke disciplines heen. DANS is hierbij een actieve partij. Dit alles vergt dat we in toekomst aanvullende eisen zullen moeten stellen. Zoals bijvoorbeeld aan het gebruik van unieke identificatie van documenten en auteurs (DOI’s, auteur-identificatie systemen zoals ORCID). Dit als een voorwaarde voor de vindbaarheid en de goede uitwisseling van documenten. Waar we ook voor willen gaan zorgen, is dat er beheersregels voor de European Open Science Cloud komen die gegevens uit publiek gefinancierd onderzoek, tot onderzoeks-‘commons’ verheffen, zodat de uitwisseling, het hergebruik en analyse van dit publiek gefinancierd onderzoek altijd mogelijk blijft. Wederom zijn hierbij de introductie van unieke identificatiesystemen essentieel.

Doe mee!

De European Open Science Cloud is een prachtig instrument voor onderzoekers. Mijn oproep is: doe met ons mee. Wij hebben een ‘coalition of Do-ers’ in het leven geroepen, bestaande uit organisaties die er samen met ons voor willen gaan. Word een ‘active player’, geef de Cloud gestalte. Samen bouwen we nu aan de European Open Science Cloud van morgen.”

https://ec.europa.eu/info/departments/research-and-innovation_nl

Koppeling van data in CLARIAH vergroot onderzoeksmogelijkheden

Centrale data-hub ANANSI verbindt data

Onderzoekers kunnen bij CLARIAH terecht voor onderzoeksdata uit drie domeinen: taalkunde, sociaal-economische geschiedenis en mediastudies. Een koppeling tussen de data van deze domeinen zorgt voor nog meer onderzoeksmogelijkheden.

Marnix van Berchum

De verschillende domeinen binnen CLARIAH beschikken over data en onderzoeksmethoden die ook potentieel interessant zijn voor andere vakgebieden. Data over het gebruik

van sociale media door jongeren kan bijvoorbeeld ook relevant zijn voor onderzoek over de ontwikkeling van taal. De mogelijkheden om zulke interdisciplinaire onderzoeksvragen te beantwoorden, waren tot dusver technisch beperkt.

ANANSI biedt de oplossing voor dit probleem: deze centrale data-hub synchroniseert de linked data uit de drie domeinen en verbindt ze. De data worden gegroepeerd in zoekmogelijkheden voor ‘personen’ (wie), ‘locaties’ (waar) en ‘concepten’ (wat), die via de webinterface beschikbaar zijn. Informatie over de

herkomst van de data, zoals welke bronnen zijn gebruikt en wie subsidie heeft verleend, is ook beschikbaar. De API van ANANSI maakt het mogelijk om de data (computationeel) te bevragen en op maat gemaakte websites te bouwen.

Beschikbaar en actueel

Met het ResourceSync-protocol (een state of the art protocol voor het synchroniseren van bronnen op het web) kunnen de CLARIAH-partners hun data beschikbaar stellen en actueel houden. Ook andere content providers, zoals archieven

en erfgoedinstellingen, kunnen hiervan gebruikmaken, aangezien alle softwarecode openbaar beschikbaar is. Zo neemt zowel de hoeveelheid data als de mogelijke onderlinge verbindingen toe.

Plannen voor 2018

Huygens ING, de bouwer van ANANSI, levert in de loop van 2018 meer functionaliteit voor onderzoekers, waaronder de mogelijkheid om zelf tabulaire data te uploaden en deze te converteren naar linked data. Ook zal de gebruiker in de toekomst zelf de zoekinterface van een

dataset kunnen configureren en data kunnen uitvoeren naar bijvoorbeeld CSV- of GraphML-bestanden.

ANANSI is gebouwd op Timbuctoo, door Huygens ING ontwikkelde software specifiek voor geesteswetenschappelijk onderzoek en de complexe, heterogene data die hierin gebruikelijk zijn. Timbuctoo ondersteunt verschillende interpretaties van dezelfde gegevens en houdt van alle data de herkomst bij. Voor meer informatie over Timbuctoo, zie timbuctoo.huygens.knaw.nl.

anansi.clariah.nl

De KB bewaart websites voor toekomstig onderzoek

Het post-truth-tijdperk gearchiveerd op het internet

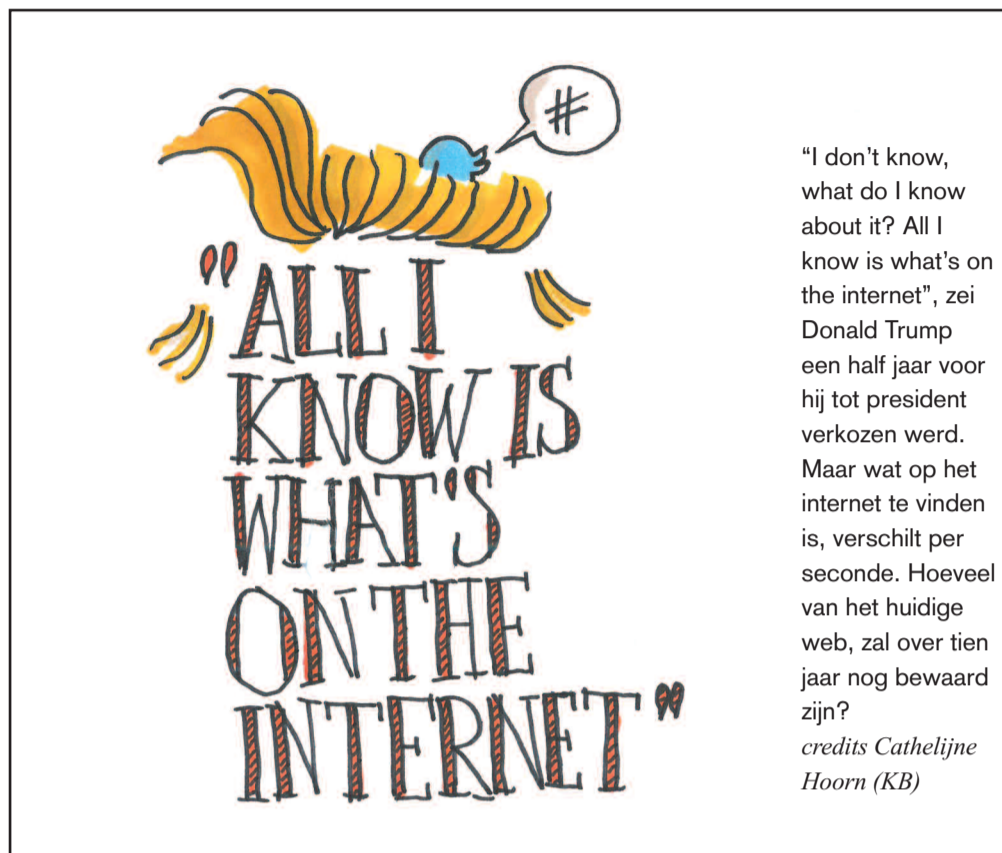
De gemiddelde webpagina is slechts negentig dagen online. Aangezien de geschiedenis van vandaag vooral online wordt opgetekend, is de kans op gaten in ons collectief geheugen groot.

Laurie Hasselt en Kees Teszelszky

Sinds 2007 streeft de Koninklijke Bibliotheek (KB) ernaar een zo representatief mogelijk deel van het Nederlandse web te bewaren. Tot nu toe bewaarde de KB vooral websites over geschiedenis, zoals websites van musea, historische verenigingen en erfgoedinstellingen. Veel minder aandacht ging naar sites die in de toekomst als bron kunnen worden gebruikt van onze geschiedenis. De belangrijkste oorzaak was een gebrek aan tijd en middelen: het totale Nederlandse nationale webdomein omvat meer dan tien miljoen sites.

Speciale webcollectie

Om aan de behoeftes van huidige en toekomstige onderzoekers tegemoet te komen, bouwt de KB nu een speciale webcollectie op van actuele, Nederlandse post-truth-websites. De collectie omvat circa 500 sites die in de toekomst een bron of spiegel van een, misschien zelfs overstreden, maatschappelijk fenomeen zijn. Volledigheid wordt niet nagestreefd, wel representativiteit. Het levert een bonte stoet aan websites op van anarchisten tot nationalist, van eindtijdpreppers tot aanhangers van de platte-aarde-theorie. Achter die ogenschijnlijke veelzijdigheid schuilt echter ook een gemene deler: het gros van deze sites is kritisch ten aanzien van de overheid en main-



"I don't know, what do I know about it? All I know is what's on the internet", zei Donald Trump een half jaar voor hij tot president verkozen werd. Maar wat op het internet te vinden is, verschilt per seconde. Hoeveel van het huidige web, zal over tien jaar nog bewaard zijn?

credits Cathelijne Hoorn (KB)

stream. Veel websites roepen op niet alles te geloven en vooral zelf onderzoek te doen.

Gearchiveerd én beschreven

Om de gebruiker van de KB-webcollectie te helpen begrijpen wat er anno 2017 online te vinden was, worden de websites gearchiveerd en wordt een uitgebreide collectiebeschrijving gemaakt. Bij deze beschrijving is veel aandacht voor de gebruikte selectie- en de uitsluitingscriteria, de documentatie van de werkwijze en contextinformatie.

Ook de onderlinge linkstructuur tussen de sites in de collectie wordt bewaard, waardoor ook infor-

matie kan worden vastgelegd over sites die niet gearchiveerd worden.

Het is de eerste keer dat een innovatieve techniek als online linkanalyse wordt gebruikt door de KB om relevante en belangrijke websites op te sporen en contextinformatie vast te leggen. Door deze techniek kan voor het eerst de websfeer rondom websites worden vastgelegd voor de onderzoekers van de toekomst. Dit project is voor zover bekend uniek in de wereld: nergens anders is een dergelijk scherp tijdsbeeld in een webcollectie vastgelegd.

kb.nl/organisatie/onderzoek-expertise/e-depot-duurzame-opslag/webarchivering

GELEZEN

Doekonderzoek oeuvre Vermeer

Counting Vermeer: Using Weave Maps to Study Vermeer's Canvases, RKD Studies, The Hague (RKD) 2017
Harmen van der Meulen

In het online boek *Counting Vermeer* worden de resultaten van innovatief en multidisciplinair doekonderzoek naar het complete schilders-oeuvre van Johannes Vermeer (1632-1675) gepubliceerd. Er wordt onder meer aangetoond hoe computer-gegenerateerd doekonderzoek – door draden in schildersdoeken te analyseren en zo 'vingerafdrukken' in het schilderij te herkennen – de toeschrijving van het eerder betwijfelde werk *Jonge vrouw aan een virginaal* (ca. 1670-1672) aan Vermeer bevestigt. Tevens ondersteunt de doekanalyse de theorie dat Vermeer tenminste twee paren schilderijen oorspronkelijk als bij elkaar horend heeft geschilderd, een feit waarover kunsthistorici in het verleden heftig van mening verschilden. De software die Johnson en zijn team gebruikten, wordt samen met het boek gratis beschikbaar gesteld.

countingvermeer.rkdmonographs.nl



COLUMN

To-do apps

Het heeft voor mij even geduurd voordat ik mijn gewone mobiele telefoon verving door een smartphone, maar toen ik die eenmaal had ging er een wereld voor me open. Wekelijks probeerde ik wel een nieuwe app uit, tot mijn telefoon er bijna mee vol stond.

Inmiddels bestaan er voor Androidtelefoons ruim 280.000 apps en voor iPhone's ruim 220.000. Naarmate dit aantal steeg, daalde de hoeveelheid apps op mijn smartphone: om de zoveel tijd kijk ik mijn mapjes door en alle apps die ik niet geregeld gebruik, verwijder ik. Keek ik vroeger nog weleens in de App Store of er interessante nieuwe dingen waren verschenen, nu doe ik dat nooit meer. Ik ben digitaal verzadigd.

Dacht ik. Tot ik onlangs, bij toeval, een app tegenkwam die uitkomst biedt op een terrein waar ik al jaren mee worstel. Ik heb het altijd nét te druk. Om alles gedaan te krijgen ben ik in to-do-lijstjes gaan geloven. Ooit maakte ik die op papier, maar in-

middels bestaan er natuurlijk allerlei apps en tools waarmee je to-do-lijstjes kunt samenstellen, want ik ben niet de enige die het altijd net te druk heeft.

Ik heb er, in de loop van de jaren, een hoop uitgeprobeerd. Wunderlist, Todoist, Any.do, Clear – van andere ben ik de namen alweer vergeten. Uiteindelijk ben ik een Evernote-fan geworden. Met Evernote kun je veel meer dan alleen lijstjes maken, maar de premiumversie is relatief duur en ook deze tool heeft, net als alle andere, z'n beperkingen.

Bij toeval stuitte ik onlangs op Google Keep. In eerste instantie was ik niet zo onder de indruk,

maar inmiddels gebruik ik deze gratis app een paar weken en ben ik er juist erg enthousiast over. In principe werkt deze app, die ook beschikbaar is voor je desktop of tablet, zoals de meeste tools waarmee je aantekeningen kunt vastleggen. Je kunt er geluid, beeld en hyperlinks in opslaan, je kunt notities voorzien van meta-



foto Leo van Velzen

data (zogenoemde tags) en je kunt er 'herinneringen' in aanmaken, die automatisch worden gesynchroniseerd met Google Agenda.

Met drie functies heeft Google Keep mijn hart gestolen. Eén: als je een notitie insprekt, wordt niet alleen het geluid bewaard, maar wordt de tekst ook onmiddellijk voor je uitgetikt. De kwaliteit is goed: een groot deel van deze column is foutloos door Google Keep uitgetikt.

Twee: als je een foto maakt van bijvoorbeeld een krantenartikel (wat ik vaak doe), wordt de tekst op de achtergrond automatisch uitgetikt en dus doorzoekbaar gemaakt door een programma dat

gebruikt van optische tekenherkenning (ocr). Die zogenoemde ocr-laag kun je zichtbaar maken door bij de notitieopties te klikken op 'Afbeeldingstekst ophalen'. Waarna je de tekst kunt knippen en plakken. Google Keep bevat dus feitelijk een gratis ocr-programma, wat je een hoop tikwerk kan schelen.

Een derde trucje gaf voor mij de doorslag. Als je een herinnering maakt, kun je niet alleen datum en tijd maar ook een locatie opgeven. Stel: je moet iets niet vergeten bij een bezoek aan een winkel. Als je de locatie van die winkel toevoegt aan een herinnering krijg je automatisch een seintje van je telefoon op het moment dat je er binnenstapt. Mits je natuurlijk je smartphone bij je hebt, maar ik ken bijna niemand die tegenwoordig nog zonder mobiele telefoon de deur uitgaat.

Ewoud Sanders

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