

DANS biedt toegang tot schatten aan informatie

Klein én fijn: subsidie voor kleine dataprojecten

Met de oprichting van DANS in 2005 werden ook de Kleine Data-projecten-subsidies (KDP's) in het leven geroepen. Anno 2019 zijn al zo'n honderd projecten succesvol afgerond. *Nina Westzaan*

“Door gebrek aan tijd, kennis of middelen liggen veel waardevolle data nog bij onderzoekers ‘in de la’.” Aan het woord is Henk Harmsen, adjunct-directeur van DANS. “Iedereen zou altijd toegang moeten hebben tot deze informatie. DANS stimuleert onderzoekers om hun gegevens te delen, bijvoorbeeld door gebruik te maken van onze diensten of een training over researchdata-management te volgen. De KDP-subsidie is ook een manier om onderzoekers te stimuleren waardevolle data FAIR te maken.”

Open access

Een van de KDP-subsidies werd verstrekt aan ‘Tell Deir Alla’ (2018). Dit project, geleid door dr. Gerrit van der Kooij, ontsluit een schat aan gedigitaliseerde, open access veldwerkdokumentatie. Het gaat om opgravingswerkzaamheden van 1976 tot 2009, van een uniek inter-



De data van ‘Excavation archive of the Tell Deir Alla project, Jordan’ is open access beschikbaar via EASY. foto dr. G. van der Kooij. DOI: 10.17026/dans-xs5-kpqn

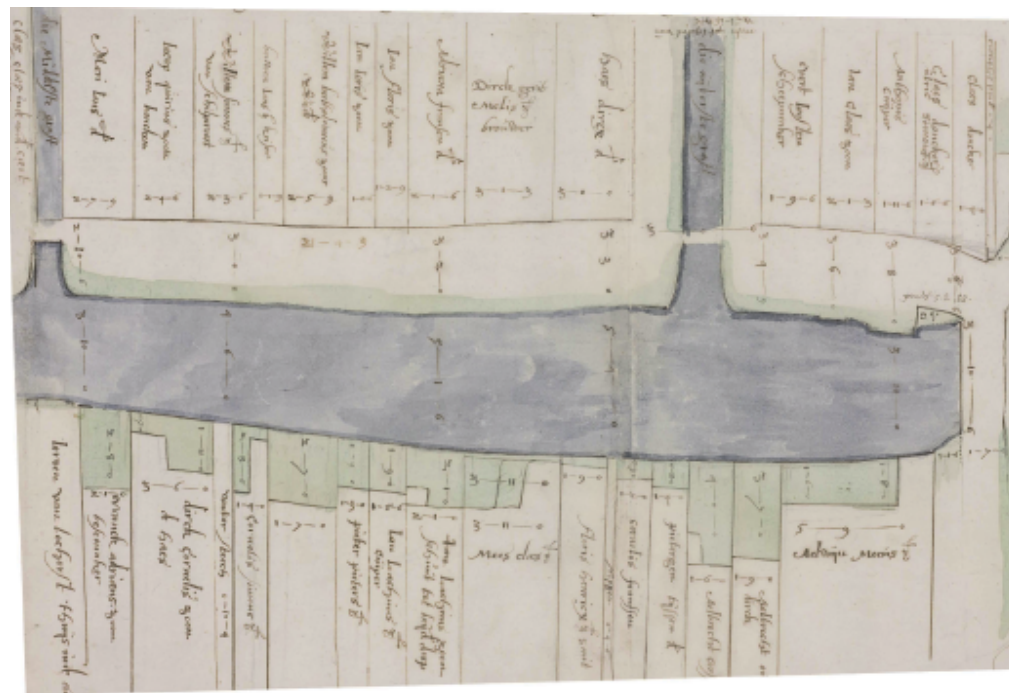
nationaal-regionaal samenwerkingsproject in de Jordaanvallei. Voorheen was het een analogoog archief, door de KDP-subsidie van DANS hebben Nederlandse en internationale onderzoekers nu direct toegang tot onder andere de verslagen, kleurenfoto's en veldtekeningen. “Open access biedt handvatten voor een voortdurende dialoog

over de data en een hernieuwde blik op opvattingen over het verleden,” stelt Van der Kooij. Een andere KDP-subsidie werd verstrekt aan dr. Roos van Oosten en dr. Arie van Steensel. Hun KDP ‘Ranking the Towns’ heeft bijgedragen aan de realisering van fase twee van het grotere project ‘Mapping Historical Leiden: A Dynamic and Digital Atlas’. “Voor een gezond wetenschappelijk veld zijn kleine projecten onmisbaar,” stelt Van Oosten. In haar KDP bracht ze van acht steden informatie uit verschillende bronnen bijeen in een GIS-kaart. “GIS is een machtig mooi middel om stadshistorische ontwikkelingen te analyseren en te visualiseren en bij uitstek geschikt voor interdisciplinaire samenwerking, maar een goede kaart en onderliggende databases maken, is ook complex en tijdsintensief. Over één lijn op de kaart kan soms een historische verhandeling geschreven worden. Waarom staat die lijn daar en niet een huizenblok verderop?” De KDP-subsidie en het deponeren van de data luiden het begin in van een volgende fase. “In de nabije toekomst bouwen de betrokken onderzoekers voort op de inzichten, kennis en vaardigheden die dit KDP-project opleverden.”

Maak kans op een KDP-subsidie

Meer voorbeelden van verstrekte KDP's en informatie over het aanvragen van een subsidie staat op de website van DANS.

dans.knaw.nl



Deze deelkaart van het middeleeuwse grachtenbeeld van Leiden is met ander kaartmateriaal uit de dataset te projecteren in GIS-applicaties. Alle deelkaarten samen vormen een totaalkaart met coördinaten. credits dr. A. van Steensel. DOI: 10.17026/dans-xng-asq2

Digitale Charterbank Nederland online

172.000 oorkonden doorzoekbaar

De Digitale Charterbank Nederland maakt historisch onderzoek in Nederlandse oorkonden mogelijk.

Jan Burgers en Rik Hoekstra

Oorkonden zijn vaak de oudste overgebleven documenten van een plaats. Alleen al op Nederlands grondgebied zijn er tienduizenden bewaard gebleven, vanaf de tiende

tot het begin van de negentiende eeuw. Als precies gedateerde juridische documenten vormen charters, zoals oorkonden ook genoemd worden, een exacte en betrouwbare kennisbron. De rechtshandelingen, plaatsen en personen weerspiegelen de toenmalige realiteit, vrijwel alle aspecten van de maatschappij passeren de revue. Op basis van charters kunnen politieke, sociale, economi-

sche en kerkelijke ontwikkelingen worden onderzocht. Ook kunnen familiegeschiedenissen of topografische omstandigheden worden gereconstrueerd, zowel lokaal als in een breder verband, zowel in een momentopname als over een langere periode.

De juiste informatie

Tot op heden werd dit rijke materi-

aal weinig gebruikt in historisch onderzoek. Oorkonden liggen namelijk verspreid over honderden archieffondsen in tientallen archiefbewaarplaatsen. Hoe kan een onderzoeker in dat gefragmenteerde veld de juiste informatie vinden? De Digitale Charterbank Nederland (DCN) biedt de oplossing.

Vervolg op pagina 2



E-DATA & RESEARCH

Jaargang 13 | 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: Centerdata, CLARIAH, DANS, Huygens ING, de Koninklijke Bibliotheek en het Rijksmuseum.

INHOUD

3 DH2019-spreker Johanna Drucker aan het woord

5 Hermens over digitale hulp voor technische kunstgeschiedenis



6 Bronnen volkstellingen checken met tabellen

6 Data verzamelen met ODISSEI en LISS panel

7 Flora en fauna online beter vindbaar maken

7 Onderzoek met rapbot: wat is echt en wat niet?

8 Gastcolumnist Dyksta: Samen kom je verder

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

Participatie en co-creatie bij DEN-event

Fricke Oosten en Saskia Scheltjens

Op 4 en 5 maart organiseerde Kennisinstituut cultuur en digitalisering DEN hun jaarlijkse event in Rotterdam. Voor de eerste keer werd ook de bijkomende doelgroep van cultuurproducerende instellingen bediend, hen werd getoond wat digitalisering biedt aan cultuurmakers zelf. De tweede dag, merkbaar drukker dan de eerste, stond in het teken van de oorspronkelijke doelgroep, de erfgoedsector en de mogelijkheden die digitale collecties bieden voor educatie en inclusiviteit. Participatie en (co-)creatie waren onderwerpen waar beide doelgroepen elkaar vonden. De keynote, Kati Price (hoofd Digital Victoria & Albert Museum), vertelde over het werk om hun online collecties opnieuw in te richten en digitaal aan te bieden. Hiervoor werd opnieuw kritisch gekeken wie de belangrijkste gebruikersgroepen zijn en of deze wel echt bediend worden door de huidige inrichting. Musea richten zich veelal op onderzoekers, terwijl bezoekers diverse redenen hebben om online collecties te gebruiken. Price benoemde het voor veel erfgoedinstellingen herkenbare “William Morris probleem” dat informatie over hun populairste

GEHOORD & BIJGEWOOND

ontwerper in verschillende silo’s opgeborgen zit en niet in één keer te vinden is. Natuurlijk zijn het niet alleen grote erfgoedinstellingen die mooie stappen maken. Verschillende sprekers deelden succesverhalen over hoe ze hun collecties beter onder de aandacht wisten te brengen door samenwerking met nieuwe partners of het publiek. den.nl

Volop discussie tijdens CLARIAH Toogdag

Erica Renckens

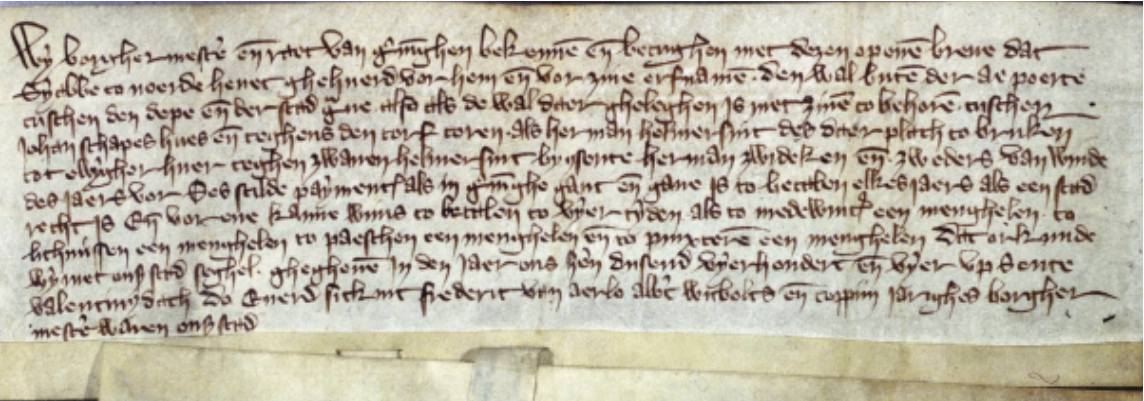
Sinds de start van CLARIAH is het aantal softwareprogramma’s dat geesteswetenschappers kunnen gebruiken bij hun onderzoek flink toegenomen. Hebben al deze tools een intuïtieve, gebruiksvriendelijke interface nodig of mogen we inmiddels van geesteswetenschappers verwachten dat ze op zijn minst een beetje kunnen programmeren? Deze vraag kwam verschillende keren naar boven tijdens CLARIAH’s Toogdag, vrijdag 5 april bij Beeld en Geluid in Hilversum. Uit de felle discussie bleek dat de revolutie in de geesteswetenschappen nog in volle gang is. Waar zo’n tien jaar



Jan Odijk (directeur CLARIAH) en Lex Heerma van Voss (Principal Investigator CLARIAH) in gesprek tijdens de CLARIAH Toogdag, begeleid door dagvoorzitter Malou van Hintum. foto Margriet van den Berg

geleden programmeerkennis onder geesteswetenschappers nog een unicum was, ligt deze competentie nu juist meer in de lijn der verwachting. ‘Geen natuurkunde-aio zal vragen om een interface voor dummies, waarom een geesteswetenschapper dan wel?’, klonk het in de zaal. Waarschijnlijk zullen er toch in bepaalde disciplines onderzoekers blijven die niet leren programmeren, zo moesten de aanwezigen toegeven. Voor deze digibeten moet een gebruikersschil beschikbaar zijn, al

zal dit altijd met beperkingen over de te stellen onderzoeksvraag komen. Verschillende geesteswetenschappers lieten gedurende de dag zien welke onderzoeksvragen zij met behulp van de CLARIAH onderzoeksinfrastructuur hebben kunnen beantwoorden. Ook deelden zij hun wensen voor aanpassingen, waarvan de meeste zijn meegenomen in de plannen voor CLARIAH-PLUS, het vervolgproject dat dit jaar van start is gegaan. Nieuw in deze voortzetting is werkpakket 6, waarvan werkpakketleider Karina van Dalen-Oskam de plannen presenteerde. Dit nieuwe focusgebied richt zich op de inhoudelijke analyse van tekst, wat in zo goed als alle geesteswetenschappelijke disciplines van pas komt. Over de plannen voor CLARIAH-PLUS is elders in dit blad meer te lezen. Na een lange dag vol presentaties en interactie werden de discussies tijdens de borrel onder het genot van een hapje en een drankje volop voortgezet. clariah.nl



Voorbeeld van een oorkonde: Burgemeesters en Raad van Groningen verklaren dat Syabbe to Noerde een stuk van de stadswal buiten de Aapoort huurt (r. 1-2: ‘...dat Syabbe to Noerde hevet ghehuert vor hem ende vor zine erfnamen den wal buten der Ae poerte’), tussen het huis van Johan Schape en de Turftoren, voor zes schild en een kan wijn per jaar, 14 februari 1404. credits Groninger Archieven, Stadsbestuur van Groningen (1), inv. nr. 190.1.

Vervolg van pagina 1

Digitale Charterbank Nederland

Via DCN zijn de beschrijvingen van meer dan 172.000 oorkonden beschikbaar, waarin full-text kan worden gezocht. De database bevat de geaggregeerde data uit de archiefinventarissen die oorkonden bevatten. Die inventarissen zijn overal op ongeveer dezelfde manier gestructureerd en online gezet. DCN is opgezet samen met het Groningse bedrijf De Ree, dat de digitale archiefsystemen van een groot aantal Nederlandse archieven beheert, en

met de medewerking van een aantal niet bij dat bedrijf aangesloten archiefdiensten. Op deze manier is een grote hoeveelheid informatie direct ontsloten voor alle geïnteresseerde onderzoekers.

Naar het archief

Voor alle duidelijkheid: de teksten van de oorkonden zelf zijn slechts in een klein aantal gevallen beschikbaar. Vaak moet de onderzoeker nog steeds naar het archief om stukken te

raadplegen, maar veel onderzoek kan ook al gebeuren op basis van de gegevens die in DCN te vinden zijn. Bovendien worden steeds meer archiefstukken, waaronder oorkonden, gescand en digitaal beschikbaar gesteld. Op een dergelijke manier zijn de afbeeldingen van zo’n 25.000 oorkonden nu al toegankelijk via DCN.

charterbank.huygens.knaw.nl

Beeldherkenning en het oude boek

Steven Claeyskens

Sommige toepassingen van automatische gezichtsherkenning in de publieke ruimte geven anno 2019 voldoende aanleiding voor prangende maatschappelijke vraagstukken en zelfs rechtszaken. Het geeft aan hoe verfijnd de technieken voor beeldherkenning of computer vision intussen zijn. De software is steeds beter in staat afbeeldingen en bewegend beeld te ontleden en wat er afgebeeld staat ook te identificeren. De mogelijke toepassingen zullen ons denkelijk nog verrassen. Op 27 maart organiseerden uitgeverij Brill, de Koninklijke Bibliotheek en CERL (Consortium of

Research Libraries) de workshop ‘Computer Vision in Book History’ om die mogelijkheden voor de meesttijds onschuldige wereld van de boekgeschiedenis te onderzoeken. Boekhistorici, uitgeverijen en bibliotheken vinden de weg naar beeldherkenning met succes, zo bleek. Giles Berger, Visual Geometry Group, Oxford University, trad aan als de ervaren leidsman. Jessie Wei-Hsuan Chen (UU) trakteerde het publiek op haar onderzoek met behulp van VGG Image Search Engine (VISE), een open source-project van Bergers groep, naar het gebruik van houtblokken in vroegmoderne botanische werken. Cristina Dondi (Oxford University) toonde de mogelijkheden van het 15eILLUSTRATION-project om in gedigitaliseerde 15de-eeuwse boeken te zoeken op illustraties. Thomas Smits (UU) gaf een voorproefje van het kwantitatief onderzoek dat hij samen met Melvin Wevers (DHLab KNAW) uitvoert naar de representatie van gender in advertenties in Nederlandse kranten uit de tweede helft van de twintigste eeuw. Heleen Wilbrink (Aincient) demonstreerde de kracht van beeldherkenning voor het zoeken op afbeeldingen van Egyptische artefacten met het CLEO-platform. Er waren nog meer tot de verbeelding sprekende voorbeelden, maar het laatste woord was aan keynote-spreker Cornelis van Lit (UU), onderzoeker van de laat-middeleeuwse islamitische filosofie en theologie en hoofdredacteur van de The Digital Orientalist. Van Lit bereidde de aanwezigen voor op de computer vision-revolutie met een kleurrijke rondgang langs de mogelijkheden en kansen voor beeldherkenning binnen de islamitische filologie en handschriftkunde.

cerl.org/services/seminars/cvinbook history



Tijdens de workshop ‘Computer Vision in Book History’ bereidde Cornelis van Lit de aanwezigen voor op de computer vision-revolutie met een kleurrijke rondgang langs de mogelijkheden en kansen voor beeldherkenning binnen de islamitische filologie en handschriftkunde.

foto Steven Claeyskens

Nood aan innovatie en verandering van ons gedrag

“Welke impact hebben de Digital Humanities?”

Op DH2019 in juli spreekt onder andere Johanna Drucker. Zij vraagt zich af of we Digital Humanities op een sociaal rechtvaardige manier kunnen verduurzamen. Hoe complex is sustainability? Steven Claeysens

Drucker, Breslauer Professor of Bibliographical Studies aan de University of California, Los Angeles (UCLA), gaat het tijdens haar keynote op Digital Humanities 2019 hebben over de toekomst van Digital Humanities (DH), onder de titel ‘The complexity of sustainability’. Ze stelt zichzelf, en haar publiek, daarbij grote vragen. “Is DH intellectueel levensvatbaar? Worden er werkelijk nieuwe ontdekkingen gedaan of beperkt het routinematig gebruik van tools en platforms wellicht onze verbeelding? Als DH institutioneel kan overleven, waar hoort het dan thuis en wat zijn de vereiste middelen? Zijn pedagogische initiatieven hernieuwbaar en welke vorm zouden ze moeten aannemen om kritische betrokkenheid te bevorderen in plaats van consumentengedrag te repliceren?” Terwijl Drucker deze vragen stelt, benoemt ze dat zulke kwesties leiden tot veel grotere, wereldomvattende vragen. “Welke impact hebben de digitale geesteswetenschappen op onze culturele systemen? Wat zijn de ecologische kosten? Wat is de politieke prijs? Daar waar een digitale infrastructuur een vanzelfsprekendheid is, worden exploitatie en misbruik van arbeid, maar ook schade aan de natuurlijke en sociale omgeving maar al te vaak genegeerd. Kunnen we dit complex van uitdagingen op een verantwoorde wijze behappen? Kunnen we Digital Humanities op een sociaal rechtvaardige manier verduurzamen?”

Zijn er leerzame voorbeelden om deze grote vragen samen te kunnen beantwoorden? Drucker noemt het UNESCO Science Report.



Johanna Drucker is boekkenstenaar en hoogleraar aan UCLA. Ze is internationaal vermaard voor haar werk over de geschiedenis van grafisch design en visualisaties, typografie, experimentele poëzie en Digital Humanities. Ze droeg bij aan Digital Humanities (MIT Press, 2012) en publiceerde diverse monografieën. Recent verscheen Graphesis: Visual Forms of Knowledge Production (Harvard University Press, 2014). foto Drucker

“Dit rapport committeert zich onomwonden aan duurzaamheid. Amerika, en met name de Verenigde Staten, zou veel kunnen leren van de Southern African Development Community, de European Energy Commission en Renewable Cities Australia. Het verschil tussen dit soort structurele programma’s en de dagelijkse realiteit in de Verenigde Staten is opmerkelijk. De retoriek van Amerikaans exceptionalisme en technologische vooruitgang staat nog steeds voorop, terwijl we te rade zouden moeten gaan bij onze collega’s over de hele wereld en veel actiever deel zouden moeten nemen.”

Het thema van de DH-conferentie is Complexities. Volgens de call for papers betekent dit ook ‘de volgende generatie betrekken, DH onderwijzen aan studenten – de mensen die met de complexiteit van de toekomst zullen moeten leven.’ Hebt u advies voor de volgende generatie

(Digital Humanities) wetenschappers?

“Het is echt belangrijk om te begrijpen dat elke keuze die we maken, deel uitmaakt van een mondiaal systeem van ecologische, sociale en politieke verantwoordelijkheid. Er is op wereldschaal nood aan innovatie en verandering van ons gedrag, en dat op zoveel terreinen, persoonlijk én professioneel. Heroverweeg onze omgang met energiebronnen, creëer wereldwijd eerlijke arbeidsomstandigheden, bouw de materiële technologische infrastructuur radicaal anders op en dring de patronen van consumentengedrag en absorptie die ons aan het scherm gekluisterd houden terug. De uitdagingen zijn nog nooit zo groot geweest. Dat zou meer dan voldoende inspiratie moeten geven.”

De call stelt ook dat ‘Complexities mensen willen inspireren’. Wie (of welk project) inspireerde u recentelijk als onderzoeker?

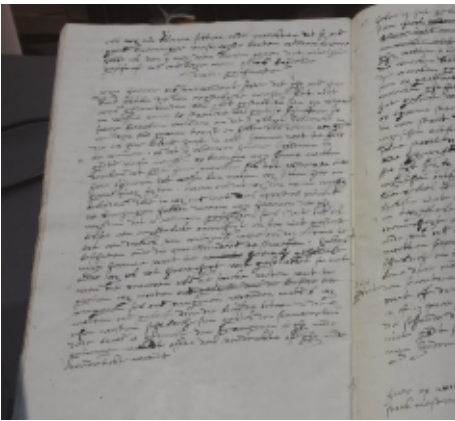
“Het werk van mijn collega’s Miriam Posner en Sarah Roberts vind ik inspirerend. Posner heeft onderzoek gedaan naar toevoerlijnen en interfaces die worden gebruikt door grote bedrijfssystemen en hoe ze de ecologische en de menselijke prijs bij grootschalige productie van goederen aan het zicht onttrekken. Roberts onderzoekt de arbeidsomstandigheden van werknemers die content modereren op sociale media. Ze kijkt daarbij ook naar de emotionele tol die deze mensen betalen. Allebei leggen ze hetzelfde bloot: iedere gebruiker van digitale systemen is medeplichtig aan de productieomstandigheden. Ook al houden we ze graag buiten beeld, alsof dat ons beschermt tegen de verantwoordelijkheid. Niet dus.”

Kom naar DH2019

Nieuwsgierig naar de lezing van Drucker en vele anderen? Kom ook naar de grootste internationale Digital Humanities-conferentie van dit jaar in TivoliVredenburg, Utrecht, van 9 tot 12 juli. Inschrijven kan tot 1 juli. dh2019.adho.org

Pamfletten bron letterkundig onderzoek

In het kader van haar onderzoek naar intra-auteur-variatie in de zeventiende eeuw (Language Dynamics in the Dutch Golden Age) bestudeerde promovenda vroegmoderne letterkunde Cora van de Poppe de rol van taalvariatie. Heidi Berkhout



Het egodocument van weduwe Van Wanray. foto Cora van de Poppe

“Mijn hypothese is dat taalkeuzes tot op zekere hoogte beïnvloed zijn door andere ‘verhaalen’. Rond 1622 schreef de remonstrantse weduwe Willemken van Wanray: Het verhael van gennge my, Wyllemken van Wanray [...] is overcomme van de Magestraet [...]. In dezelfde periode verschenen in de Noordelijke en Zuidelijke Nederlanden ‘verhael’pamfletten over conflicten tussen remonstranten en contraremonstranten. De door Stichting Vrijwilligersnetwerk Nederlandse Taal (SVNT)

getranscribeerde pamfletten heb ik geanalyseerd in relatie tot Van Wanray’s egodocument, met oog voor met name narratieve structuur en variatie in werkwoordtijden. Wat

blijkt? De ‘verhaalen’ volgen een vaste opbouw en beschrijven gebeurtenissen met vaste werkwoorden, -tijden en andere talige eigenschappen. Er is sprake van nauwe verwendheid tussen de teksten.” Dit onderzoek kon plaatsvinden omdat vrijwilligers van de SVNT onder leiding van prof. dr. Nicoline van der Sijs en drs. Hans Beelen oud-Nederlandse teksten (15e-18e eeuw) transcriberen en beschikbaar en toegankelijk maken. De bevindingen van Van de Poppe zijn gepubliceerd in Vooy, tijdschrift voor letteren. De pamfletten zijn beschikbaar via de Thematische collectie: Digitale tekstuutgaven, vervaardigd door de Stichting Vrijwilligersnetwerk Nederlandse Taal in het online archiveringssysteem EASY.

DOI: 10.17026/dans-xxv-frex

KORT

7.000 brieven Johan de Witt online

Onlangs stelde het Huygens Instituut zeventienduizend brieven van Johan de Witt online beschikbaar. Het zijn de eerste brieven afkomstig uit een compleet archief van de diplomatieke correspondentie van de raadpensionaris, beheerd door het Nationaal Archief. De overige 28.000 brieven zullen de komende jaren aan deze online catalogus worden toegevoegd. Het betreft correspondentie uit de periode tussen 1649 en 1672 uit heel Europa. De brieven zijn vooral afkomstig van diplomaten en geschreven in het Nederlands, Frans, Engels en Latijn. Ze worden ontsloten in samenwerking met het Nationaal Archief, Oxford University en de Bodleian Library. (MJ) huygens.knaw.nl



Portret door Jan de Baen van Johan de Witt, raadpensionaris van Holland.

credits <http://hdl.handle.net/10934/RM0001.COLLECT.5869>

Europeana lanceert Newspapers

In mei vond de officiële lancering plaats van Europeana Newspapers. Deze collectie is de voortzetting van het gelijknamige EU-project waarin onder leiding van de Berlijnse Staatsbibliothek 18 Europese instellingen vier eeuwen aan kranten digitaal beschikbaar stelden. De nieuwe collectie omvat ruim 4 miljoen gedigitaliseerde kranten uit Europa. Bijna een kwart daarvan is full-text doorzoekbaar. Om de vindbaarheid van relevante informatie te vergroten, heeft Europeana een aantal functionaliteiten aan de collectieportal toegevoegd, waaronder de mogelijkheid om te zoeken op een datumbereik. Ook zijn de krantendata te downloaden via de API op Europeana Pro. (MH) newspapers.europeana.eu

Parthenos geeft richtlijn voor FAIR-RDM

De Parthenos Guidelines to FAIRify datamanagement and make data reusable geeft in twintig richtlijnen weer hoe data uit de geesteswetenschappen en de sociale wetenschappen meer FAIR gemaakt kunnen worden. PARTHENOS staat voor Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies. Het project is interdisciplinair en verbindt bestaande Europese infrastructuren en diverse projecten om voor onderzoekers een meerwaarde te creëren op het gebied van taal, cultureel erfgoed, geschiedenis en archeologie. Later dit jaar verschijnt de gids ook in andere talen, zoals Italiaans, Duits en Grieks. (FvG) parthenos-project.eu

SINDS KORT BESCHIKBAAR

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

CentERdata

Huiselijk geweld in Nederland

In opdracht van het Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC) is in het LISS panel van juni t/m augustus 2017 een vragenlijst afgenomen over de prevalentie van huiselijk geweld in Nederland. Dezelfde vragenlijst werd ook afgenomen door het CBS onder een ‘verse’ aselecte steekproef. Het dubbel uitvoeren had primair als doel om te onderzoeken of één van beide manieren duidelijk andere uitkomsten zou opleveren. Begin dit jaar zijn de uitkomsten gepubliceerd in een rapport, beschikbaar op de WODC-website. Uit de resultaten blijkt dat de overeenkomsten tussen het LISS-panel en het CBS-onderzoek vrij groot zijn. De data van de afname in het LISS panel - Personal Safety 2017 - zijn beschikbaar via LISS Data Archive.

lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Pieters, R.; Giesen, R. van, december 2016 – maart 2017 – juni 2017 – september 2017, Tilburg Consumer Outlook Monitor
- CentERdata, december 2017 t/m maart 2018, Politics and Values - Wave 10
- CentERdata, oktober – november 2018, Economic Situation: Assets - Wave 6
- CentERdata, juli – augustus 2018, Economic Situation: Housing - Wave 11
- CentERdata, november – december 2018, Health - Wave 11



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

DANS

Stormvloed aan data vanuit PAN

Portable Antiquities of the Netherlands (PAN) is een databank waarop vondsten van metaaldetector hobbyisten online ontsloten worden. De collectie bestaat uit onder andere kledingspelden, munten, haar-



Huiselijk geweld in Nederland
foto Alexas / Pixabay

pinnen, kunstvoorwerpen en wapens. De vondsten worden professioneel gefotografeerd en gedetailleerd beschreven. DANS zorgt voor de duurzame archivering van de PAN-databank en voor ontsluiting via ARIADNE en Europeana, zodat deze Nederlandse vondsten ook in Europees verband getoond en onderzocht kunnen worden. Een groot aantal opmerkelijke vondsten zijn nu zichtbaar in EASY.

DOI: 10.17026/dans-zbt-nexy

Ook sinds kort beschikbaar:

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

- Balint, K.; Das, prof. dr. E. (Radboud University); Stel, G.; Hoppener, M. (2019): Can a Funny Story with an Ambiguous Role Model Promote Dental Hygiene in Children? DANS. DOI: 10.17026/dans-zsh-qthf
- Boon, Drs. P.J. (2018): Clay pipe heelmark photos for identifying finds. DANS. DOI: 10.17026/dans-xgk-2uek
- Kooij, Dr G. van der (University of Leiden); Franken, Prof H.J. (University of Leiden); Ibrahim, Prof M.M. (Yarmouk University, Jordan); Kafafi, Prof Z.A. (Yarmouk University,

Jordan) (2009): Excavation archive of the Tell Deir Alla project, Jordan - Paper Archive. DANS. DOI: 10.17026/dans-x4n-wkhu

- Oosten, Dr R.M.R. van (Universiteit Leiden) (2018): Ranking the towns: Nieuw licht op de bevolkingsdichtheid van middeleeuwse steden: Archeologische opgravingen Leiden centrum. DANS. DOI: 10.17026/dans-zwm-r834
- Peursen, prof. dr. W.T. van (Faculty of Theology, Vrije Universiteit Amsterdam); Veldman, MA J.G. (Faculty of Theology, Vrije Universiteit Amsterdam) (2018): Linking Syriac Liturgies. DANS. DOI: 10.17026/dans-26t-hhv7
- Pijls, Dr. B.G.C.W. (LUMC) (2018): Segmental induction heating of orthopaedic metal implants. DANS. DOI: 10.17026/dans-z5e-qerx
- Sociaal en Cultureel Planbureau (SCP) (2016): Continu Onderzoek Burgerperspectieven 2016 - COB2016. DANS. DOI: 10.17026/dans-2xr-3pa7
- Tump, drs M. (BAAC bv) (2018): Sporen onder Betonson. Onderzoek naar bewoningssporen uit de bronstijd, ijzertijd, Romeinse tijd, middeleeuwen en nieuwe tijd onder de voormalige betonfabriek te Son gemeente Son en Breugel. DANS. DOI: 10.17026/dans-x4v-92r9



PAN ontsluit vondsten van hobbyisten, zoals dit massief gegoten beeldje van een everzwijn credits www.portable-antiquities.nl/pan/#/object/public/PAN-00001571



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

AGENDA

10 - 13 juni • Hamburg
14th International conference on Open Repositories 2019
Het thema is 'All the user needs'.
or2019.blogs.uni-hamburg.de/cfp

12 - 14 juni • Ljubljana
IWSG 2019
Elfde internationale workshop over Science Gateways.
iwsig2019.eu

13 - 14 juni • Philadelphia
NASKO 2019: North American Symposium on Knowledge Organization
Dit jaar is het thema 'Community and Computation'.
iskocus.org/nasko2019-cfp.php

19 - 21 juni • Cambridge
Dataverse Community Meeting 2019
Jaarlijks netwerkevent voor de Data-verse-community.
projects.iq.harvard.edu/dcm2019/home

26 - 28 juni • Dublin
LIBER Annual Conference 2019
Dit jaar is het thema 'Research Libraries for Society'.
liberconference.eu/dublin2019

1 - 5 juli • Leiden
Satelite workshops DH2019
Diverse workshops voorafgaande aan DH2019.
dhafrika.blog

9 - 12 juli • Utrecht
DH2019
Workshops, lezingen en demonstraties over digitale analyse- en presentatietechnieken.
dh2019.adho.org

11 - 13 september • Luik
DH conferentie Benelux 2019
Voor betrokkenen bij Digital Humanities-projecten in de Benelux.
2019.dhbenelux.org

16 - 20 september • Amsterdam
iPRES 2019
Jaarlijkse conferentie over digitale preservatie.
ipres2019.org

24 - 26 september • Kopenhagen
COASP 2019
Conferentie over Open Access Scholarly Publishing.
oaspa.org/event/coasp-2019

CLARIAH-as-a-Service

CLARIAH PLUS duikt in de tekstinhoud

Begin dit jaar is CLARIAH PLUS van start gegaan, als vervolg op het CLARIAH CORE-project. Wat kunnen we verwachten?
Erica Renckens

“In CLARIAH CORE richtten we ons op drie vakgebieden: taalkunde, mediastudies en sociale en economische geschiedenis. Maar onze ambitie was altijd al groter”, vertelt Lex Heerma van Voss, Principal Investigator (PI) van CLARIAH. “Naast het uitbreiden van de digitale onderzoeksinfrastructuur voor deze drie vakgebieden, wil CLARIAH de komende jaren ook andere vakge-

bieden binnen de geesteswetenschappen bereiken.”

Taaldata centraal

Daarvoor is een nieuw werkpakket in het leven geroepen, waarin taaldata centraal staan. “In het werk-

Op de hoogte blijven?

Historici, literatuurwetenschappers, filosofen en andere geïnteresseerden kunnen zich aanmelden voor de nieuwsbrief van het nieuwe werkpakket door een mail te sturen naar karina.van.dalen@huygens.knaw.nl

pakket taalkunde is taal het doel, in het nieuwe werkpakket is taal het middel”, aldus Karina van Dalen-Oskam, hoogleraar Computationale Literatuurwetenschap en leider van het nieuwe werkpakket. “Dit nieuwe werkpakket is er voor onderzoekers die met teksten inzicht verkrijgen in culturele en historische ontwikkelingen. Het gaat dan meer om de inhoud dan om de taal.”

Van Dalen-Oskams team is nog bezig een keuze te maken welke tools ze willen verbeteren. “We denken bijvoorbeeld aan manieren om automatisch meningen in teksten te herkennen en aan het herkennen van

de beschrijving van gebeurtenissen. Ook willen we zorgen dat gegevens en tools voor onderzoek naar betwist auteurschap bij elkaar komen, en dat onderzocht kan worden hoe teksten uit bepaalde categorieën door de tijd heen zijn veranderd.”

Analyses CLaaS

Om dit alles technisch mogelijk te maken, bouwen de CLARIAH-ontwikkelaars CLaaS: CLARIAH-as-a-Service. “Dat is een op CLARIAH toegepaste versie van een PaaS, Platform-as-a-Service”, legt Heerma van Voss uit. “Veel data en tools in de geesteswetenschappen worden aangeboden via een inter-

face. De gebruiker logt in op een server van de aanbieder, zoekt daar iets op en bewerkt of downloadt de data. Voor veel gebruikers is dat voldoende, maar er zijn ook onderzoekers die ingewikkeldere combinaties van datasets en tools hebben, of bewerkingen willen uitvoeren die heel veel rekenkracht vergen. Dat maakt CLaaS mogelijk. Voor de komende tijd staan er tientallen punten op onze agenda. Persoonlijk vind ik het belangrijk dat we de interfaces voor de grote groep gebruikers nog gebruiksvriendelijker maken. De MediaSuite is daar al een heel mooi voorbeeld van.”
clariah.nl

Prof. dr. E. Hermens, Hoogleraar Rijksmuseumleerstoel Atelierpraktijken en Technische Kunstgeschiedenis:

‘Online op bezoek in een zeventiende-eeuws atelier’

Hoogleraar op de Rijksmuseumleerstoel Atelierpraktijken en Technische Kunstgeschiedenis prof. dr. Erma Hermens spreekt over de evolutie van het vakgebied en de rol van digitale middelen daarin.

Maarten Heerlien

“Technische Kunsthistorici onderzoeken het maakproces van een erfgoedobject. Ze ontleiden als het ware het object en stellen zich daarbij vragen over bijvoorbeeld de gebruikte materialen, technieken en gereedschappen, over degradatieprocessen. De onderzoeksvraag, doorgaans kunsthistorisch van aard, komt voort uit het object zelf. De methodiek om deze vraag te beantwoorden, is inherent interdisciplinair. Technische Kunsthistorici ontlelen aan steeds meer disciplines kennis, methoden en technieken, zowel binnen de Geesteswetenschappen als de Natuurwetenschappen.”

Sneeuwbaaleffect

Communicatie staat centraal binnen Technische Kunstgeschiedenis. Technisch Kunsthistorici moeten vooral weten welke wetenschappelijke analyses een specifiek onderzoek verder helpen en bij wie ze daarvoor kunnen aankloppen. Als voorbeeld beschrijft Hermens een actueel onderzoek. “In de collectie van het Rijksmuseum bevindt zich een schild van een reuzenschildpad met daarop een schildering van Prins Frederik Hendrik van rond 1630. Een curiosum want die dieren komen hier niet voor. Vanuit Technisch-Kunsthistorisch perspectief is het interessant om te weten waarom de schilder dat portret op die drager maakte. Daartoe doen we samen



In de collectie van het Rijksmuseum bevindt zich een schild van een reuzenschildpad met daarop een schildering van Prins Frederik Hendrik van rond 1630. Onderzoekers uit verschillende disciplines werken samen aan de integratie en interpretatie van de data van dit erfgoedobject. credits Rijksmuseum, objectnummer NG-NM-2970



“Er is grote behoefte aan digitale technieken en tools, ontworpen rondom de specifieke vereisten voor de analyse van objecten en de betekenisvolle integratie en duiding van de resulterende data. Daarin liggen ook kansen voor de datawetenschap.” foto Bart van Vliet

met een schildpaddenexpert DNA-analyse om de soort te achterhalen. Het leefgebied van de soort geeft informatie over 17de-eeuwse handelsroutes. Het is ook interessant om te weten of de schilder het schild heeft bewerkt voordat hij het als ‘canvas’ gebruikte. Om dit te kunnen zien, gebruiken we scanningstechnieken. We bestuderen ook de bredere context van het object, zoals de relatie met natuurhistorische verzamelingen van die tijd, de symbolische betekenis van schildpadden en andere portretten van Frederik Hendrik. Zo’n vraagstuk veroorzaakt dus een sneeuwbaaleffect waarbij steeds onderzoekers uit andere disciplines onderdeel van het team uitmaken. Dat levert een heterogene verzameling data op. De integratie en interpretatie van die data is de kern van dit vakgebied.”

Een eigen CT-scanner

“Samenwerking met datawetenschappers is cruciaal,” vervolgt Hermens. “Er is grote behoefte aan digitale technieken en tools, ontworpen rondom de specifieke vereisten voor de analyse van objecten en de betekenisvolle integratie en duiding van de resulterende data. Daarin liggen ook kansen voor de datawetenschap. Zo werken we in het door NWO gefinancierde NICAS-project CT for Art: from Images to Patterns (IMPACT4ART) samen met het CWI aan de doorontwikkeling van CT-scanningstechnologie en -datamodellering. Hiermee kunnen we de binnenkant van erfgoedobjecten bestuderen door ze virtueel uit elkaar te halen. We kunnen dan bijvoorbeeld sporen van gebruikt gereedschap opsporen en die vergelijken met sporen uit andere objecten. Zulk kwantitatief onderzoek legt nieuwe ken-

INTERVIEW

‘Het maakproces van erfgoedobjecten onder de loep’

nis bloot over ambachtelijke technieken. Om dat onderzoek in te bedden, krijgt het Rijksmuseum een eigen CT-scanner, gebaseerd op het Flex-ray lab van het CWI. Een primeur voor de Nederlandse kunstmusea.”

Het maakproces

Dat erfgoedinstellingen steeds meer objecten en kennisbronnen digitaal beschikbaar stellen, draagt bij aan de ontwikkeling van het vakgebied, aldus Hermens: “Mijn onderzoek richt zich onder andere op wat ik de ‘choreografie van het maakproces’ noem: de vaak complexe taakverdeling in het atelier en opeenvolging van beslissingen en handelingen die kunstenaars en ambachtslieden volgden bij de vervaardiging van hun product. Dit maakproces staat vaak deels vastgelegd in kunstenaars-handleidingen, het Rijksmuseum bezit zo’n 1500 van deze handleidingen. Veel stappen in het maakproces beruisten echter ook op zogenaamde tacit knowledge: technische kennis gebaseerd op jaren ervaring, leidend tot een directe hand-mind verbinding die zich moeilijk schriftelijk laat vastleggen. Die kennis proberen we te achterhalen met analyse en reconstructies. Het volledige maakproces

openbaart zich door gedigitaliseerde kunstenaarshandleidingen te combineren met het object en de verzamelde data over tacit knowledge.”

Scherpere definitie

De positionering van het vakgebied is een belangrijk vraagstuk voor de nabije toekomst, volgens Hermens. “Men beweert dat Technische Kunstgeschiedenis de Kunstgeschiedenis van de eenentwintigste eeuw is. Persoonlijk zie ik dat anders, er zijn immers andere takken binnen de Kunstgeschiedenis waar het maakproces een kleinere rol speelt. Maar er is veel discussie over de logische plaats van technisch kunsthistorisch onderzoek. De inbedding ervan binnen het brede onderzoeksveld is de uitdaging voor de komende tijd. Misschien moeten we niet spreken van een afzonderlijk vakgebied maar over een interdisciplinaire methode binnen het domein van de Kunstgeschiedenis. Met het erfgoedobject als vertrekpunt, zie ik de Technische Kunstgeschiedenis, opererend op het snijvlak van kunst en wetenschap, als primaire vragensteller over het maakproces, waarbij binnen én buiten de grenzen van dat gebied wordt gezocht naar de antwoorden. Interdisciplinaire samenwerking met bijvoorbeeld Wetenschapsgeschiedenis, Restauratie, Conservation Science, Antropologie, afhankelijk van de vraagstelling, is cruciaal. Een scherpere definitie van Technische Kunstgeschiedenis als een interdisciplinaire methodologie die helpt om het belang van het onderzoeksgebied en de verdere ontwikkeling daarvan te onderstrepen, is daarbij essentieel.”

“En”, besluit Hermens, “we willen de relevantie van het vakgebied meer onder de aandacht brengen bij het publiek. Daarvoor kijken we naar integratie van de onderzoeksresultaten in grotere initiatieven. Denk bijvoorbeeld aan Time Machine. Daarin kun je virtueel een vroegmodern woonhuis binnengaan. Op basis van de verrijkte kunstenaarshandleidingen zouden we mensen een zeventiende-eeuws atelier kunnen laten binnentreden. Ze zouden de choreografie in zo’n werkplaats kunnen ervaren en de gemaakte objecten virtueel ontleiden. Zou dat niet fantastisch zijn?”

lookingthroughartblog.wordpress.com

Erma Hermens

Erma Hermens is hoogleraar op de Rijksmuseumleerstoel Atelierpraktijken en Technische Kunstgeschiedenis aan de Faculteit der Geesteswetenschappen van de Universiteit van Amsterdam. Ze studeerde Kunstgeschiedenis en Italiaans en promoveerde in de Kunstgeschiedenis aan de Universiteit Leiden. Ze is tevens opgeleid tot en werkte als schilderijenrestaurator. Van 2006 tot 2016 was Hermens Associate Professor aan de Universiteit van Glasgow, waar zij het masterprogramma in Technische Kunstgeschiedenis opzette. Naast haar Hoogleraarschap is zij sinds 2016 senior onderzoeker Technische Kunstgeschiedenis bij het Rijksmuseum. Hermens is lid van de wetenschappelijke commissie van het Netherlands Institute for Conservation, Art and Science (NICAS).

Controle op bronmateriaal mogelijk

Volkstelling 1947 in gedetailleerde tabellen online

224 jaar na de eerste volkstelling zijn de data over bevolkingsgrootte, leeftijden, geslacht en beroepsgroepen steeds beter online te raadplegen. *Felix van Gelderen*

De eerste integrale telling van de Nederlandse bevolking vond plaats in 1795 onder Franse invloed, ter voorbereiding op de Nationale Vergadering in 1796. Pas 34 jaar later werd bij Koninklijk Besluit de tweede volkstelling gehouden, met elke tien jaar een nieuwe telling. Vanaf 1889 werd de beroepentelling toegevoegd en tien jaar later kwam hier de woningtelling bij. Het Centraal Bureau voor de Statistiek werd bij oprichting in 1899 verantwoordelijk voor de tellingen. Vanaf 1971 namen echter negatieve geluiden rondom privacy-schending toe. Publieke protesten en negatieve media-aandacht zorgden voor een uitstel van de geplande telling van 1981. En van dit uitstel kwam afstel.

Bron voor onderzoek

De tellingen zijn een fantastische bron voor onderzoek, maar de oorspronkelijke tabelstructuur kent ook beperkingen. Door de jaren heen veranderden bijvoorbeeld plaatsnamen, fuseerden gemeenten en werden beroepen anders ingedeeld. Handmatig zijn datasets uit tabellen nog wel te analyseren, maar dit kost veel tijd. Digitale technieken, zoals Linked

Data, bieden de uitkomst. Ook de digitale ontsluiting van de tabellen biedt meer mogelijkheden.

Digitale toevoeging

Jan Jonker, adviseur dataprojecten bij DANS: “Onlangs zijn we weer een stap verder gegaan. Meer dan 30.000 afbeeldingen met data-entry van de Volks- en beroepentelling 1947 zijn naar Excelbestanden overgezet en via EASY beschikbaar gesteld. Daar waar het bronmateriaal totalen of subtotalen bevat, is aan het Excelbestand een controletelling toegevoegd. Doordat in de onderliggende datasets een eenduidige relatie is tussen bronmateriaal en Excelbestand, is snel te controleren of het bronmateriaal inconsistenties bevat. Hierdoor weet een onderzoeker nu voor statistische analyse, in hoeverre inconsistentie in het bronmateriaal de analyse kunnen verstoren. Een mooie digitale toevoeging op het oorspronkelijke materiaal.”

Bij het (digitaal) beschikbaar stellen van de gegevens uit de Nederlandse Volkstellingen zijn diverse partijen betrokken (geweest), zoals het Nederlands Instituut voor Wetenschappelijke Informatiediensten (NIWI-KNAW), het Centraal Bureau voor de Statistiek (CBS), het Internationaal Instituut voor Sociale Geschiedenis (IISG) en de Stichting Historische Databank Nederlandse Gemeenten (HDNG). De gegevens zijn gepubliceerd in DANS Data Guide 16 en online beschikbaar via EASY, het online archiveringssysteem van DANS.

DOI: 10.17026/dans-zs3-cf4m

Tabel 9. Bevolking naar huidige werkkring en woonplaats in Augustus 1939 (deel van de tabel)									
gemeente (wijk) blad		Amsterdam Oud Oost 6							
Gemeente		Werkkring op 31 Mei 1947							
Woonplaats in Augustus 1939		In de nijverheid		In de landbouw		In visserij en jacht		>	
1		M	V	M	V	M	V	>	
		2	3	4	5	6	7	>	
Zelfde gemeente als 31-5-'47		15763	4388	343	1	7		>	
Andere gemeente in de prov. Groningen		68	14	2				>	
Andere gemeente in de prov. Friesland		74	12	5				>	
Andere gemeente in de prov. Drenthe		54	10	1		1		>	
Andere gemeente in de prov. Overijssel		73	15	2	1	1		>	
Andere gemeente in de prov. Gelderland		149	30	3				>	
Andere gemeente in de prov. Utrecht		116	26	3				>	
Andere gemeente in de prov. Noordholland		330	215	35		10		>	
Andere gemeente in de prov. Zuidholland		326	81	9		5		>	
Andere gemeente in de prov. Zeeland		41	17	5				>	
Andere gemeente in de prov. Noordbrabant		101	31	2				>	
Andere gemeente in de prov. Limburg		47	14					>	
Totaal Nederland		17642	4853	410	2	24		>	
Oost- of West-Indië		33	9	1				>	
Duitsland		30	13	3				>	
Belgie		28	7					>	
Groot-Brittannië			1					>	
Frankrijk		3	2	1				>	
Italië								>	
Polen			5					>	
Oostenrijk		1	1					>	
Overige landen		14	6			1		>	
Totaal buitenland		109	44	5		1		>	
Woonplaats Augustus 1939 onbekend of nadien geboren		32	3	2				>	
Algemeen totaal		17783	4900	417	2	25		>	
Controletellingen Totaal Nederland		-500	0	0	0	0	0	>	
Controletellingen Totaal buitenland		0	0	0	0	0	0	>	
Controletellingen Algemeen totaal		0	0	0	0	0	0	>	

Tabel 9 bevat informatie over de voormalige wijk Oud Oost in Amsterdam. In kolom 2 is de som van de aantallen voor gemeenten in Nederland niet gelijk aan het getal voor Totaal Nederland. De Excelbestanden met controletellingen zorgen ervoor dat zulke inconsistenties snel te lokaliseren zijn. credits Jan Jonker



Naast initiatieven om de volkstelling van 1971 te boycotten, was er een overheids-campagne tegen de negatieve media-aandacht. credits CBS.nl

Negen projecten gehonoreerd in eerste LISS-call van ODISSEI

Paneltijd voor onderzoekers

ODISSEI biedt onderzoekers de mogelijkheid om data te verzamelen in het LISS panel van CentERdata. Dr. Mariëlle Cloïn onderzoekt de veerkracht van ouders. *Evi de Cock*

ODISSEI staat voor ‘Open Data Infrastructure for Social Science and Economic Innovations’ en heeft als doel het ontwikkelen van een duurzame nationale data-infrastructuur voor de sociale wetenschappen. Onderzoekers verbonden aan een organisatie die deelneemt aan ODISSEI konden vorig jaar een voorstel indienen om data te verzamelen via het LISS panel. Dit panel bestaat uit zo’n 7.000 personen uit ongeveer 4.500 huishoudens en is een representatieve afspiegeling van het bevolkingsregister van het CBS. Negen van de 31 ingediende voorstellen werden gehonoreerd met gratis paneltijd. Eén daarvan is van dr. Mariëlle Cloïn, senior onderzoeker bij Tranzo (Tilburg University). Cloïn: “In veel onderzoek naar ouderschap gaat de aandacht uit naar zogenoemde risicofactoren en (mentale) problemen zoals depressies. In andere onderzoeksgebieden is er de laatste jaren juist veel aandacht voor het concept veerkracht,



illustratie Frank Lansbergen

datgene wat maakt dat mensen met hun uitdagingen om kunnen gaan. Voor een onderzoek naar de invloed van beschermende factoren op het welzijn van ouders wilde ik gebruik maken van het LISS panel.”

Koppelen van gegevens

Cloïn verzamelt bij ouders met minstens één kind onder de 13 jaar nieuwe data over beschermende factoren in relatie tot ouderschap. Deze gegevens worden vervolgens gekoppeld aan gegevens die al eerder verzameld zijn bij dezelfde ouders in het LISS panel, zoals informatie over hun gezinssamenstelling en hun

mentale gezondheid. Zo kan de relatie tussen de aanwezigheid van beschermende factoren en het welzijn van ouders onderzocht worden.

“De resultaten zijn interessant voor onderzoekers en professionals die werken met ouders met jonge kinderen. Verschillende GGD’en die samenwerken met Tranzo binnen de Academische Werkplaatsen Publieke Gezondheid Brabant en Jeugd hebben aangegeven interesse te hebben in de resultaten“, vertelt Cloïn. “Het LISS panel is laagdrempelig in gebruik en er zit ontzettend veel informatie in. Ook bevat het genoeg respondenten om er bepaalde specifieke groepen uit te lichten, in mijn geval ouders met jonge kinderen. Met dataverzameling in het LISS panel kan ik ouderlijke veerkracht onderzoeken.”

Mogelijkheden LISS panel

In het najaar van 2019 zet ODISSEI een nieuwe call uit om data te verzamelen in het LISS panel. Contact opnemen met CentERdata voor de mogelijkheden voor dataverzameling in het panel kan ook. lissdata.nl

Schrijf een data paper in Research Data Journal

Het Research Data Journal (RDJ), een uitgave van uitgeverij Brill en DANS, is een peer reviewed, online only en open access data journal. Wilt u ook een data paper schrijven?

Dit zijn korte publicaties waarin de dataset wordt beschreven: de context van het onderzoek, de probleemstelling, gebruikte methoden en een globaal profiel van de dataset.

Conclusies zoals in een gewoon wetenschappelijk artikel zijn niet vereist, wel is er ruimte voor afsluitende opmerkingen. De data dienen gedeponeerd te zijn in een gecertificeerd trusted repository, zoals het online archiveringssysteem EASY van DANS. (HB)

brill.com/RDJ



Taaltechnologie doorzoekt miljoenen online krantenartikelen

Samen dieren vinden in Delpher

In het SERPENS-project werken lexicologen, taaltechnologen en historisch ecologen samen om krantenartikelen over flora en fauna beter vindbaar te maken.

Marieke van Erp

Historisch ecologen bestuderen de impact van de natuur op het menselijk handelen door de tijd. Kranten zijn hiervoor een uitstekend medium omdat ze alledaagse gebeurtenissen beschrijven, frequent gepubliceerd worden en met dank aan de Koninklijke Bibliotheek gemakkelijk digitaal te doorzoeken zijn via Delpher.nl, de historische database van Nederland.

Zoekresultaten op ‘wolf’

Relevante artikelen over flora en fauna vinden in miljoenen online artikelen is echter geen kwestie van een zoekterm invullen en dan de resultaten bekijken. Je moet rekening houden met de historische taalontwikkelingen in spelling en woordenschat, en er is een efficiënte manier nodig om de gewenste data uit Delpher te halen en voor onderzoek geschikt te maken. Bovendien kom je bij het zoeken naar bijvoorbeeld een ‘wolf’ deze niet alleen als lid van de soort *Canis lupus* tegen, maar ook als achternaam (John de Wolf), als locatie (Wolf Rock, 15 kilometer van Kaap Land’s End) of



Het woord ‘wolf’ levert vele zoekresultaten op. Hiernaast drie voorbeelden:

- **Zoekplaatje Rood Kapje en de wolf**
credits Provinciale Geldersche en Nijmeegsche courant. Nijmegen, 21-09-1935.
<https://resolver.kb.nl/resolve?urn=MMRANM02:000035671:mpeg21:a0105>,
- **Het wondere wolfje**
credits Limburgsch dagblad. Heerlen, 03-12-1990. <https://resolver.kb.nl/resolve?urn=ddd:010624151:mpeg21:a0083>.
- **De wolf in den Haagschen Dierentuin**
credits Bredasche courant. Breda, 16-06-1932. <https://resolver.kb.nl/resolve?urn=MMSAB03:000065143:mpeg21:a0057>

bijvoorbeeld in de naam van een filmscript ‘The Sea Wolf’. Alleen al tussen 1930 en 1939 levert het trefwoord ‘wolf’, zonder synoniemen, 66.809 resultaten op. Het is ondoenlijk om deze resultaten handmatig te analyseren.

Supervised machine learning

Uit het DiaMaNT-lexicon van het Instituut voor de Nederlandse Taal werden historische spellingvarianten en synoniemen van onze zoekterm gehaald. Vervolgens werden de data via de API van Delpher binnengehaald en in een databaseomgeving gezet. Een historisch lexicon maakte een inschatting van de OCR-

kwaliteit. Om de relevante artikelen te selecteren, is een automatisch classificatiesysteem ingezet om het onderscheid dier versus geen dier te kunnen maken. Ook gaf het aanvullende informatie: bevat het artikel bijvoorbeeld een beschrijving van een jachtgebeurtenis of een verslag over materiële schade door een dier. Het systeem gebruikte een ‘supervised machine learning’-algoritme: het leerde van een door een domeinexpert geannoteerde trainingsdataset door in artikelen karakteristieken te herkennen die tot een bepaalde classificatie leiden. Getest op nieuwe artikelen, wist het systeem in 92% van de gevallen

correct te identificeren of een artikel al dan niet over een dier gaat. Uiteraard willen historisch ecologen nog veel meer weten. Door de computer de gigantische bak resultaten te laten filteren, kunnen zij zich richten op de diepere analyse van relevante artikelen. SERPENS is een samenwerking tussen de Radboud Universiteit Nijmegen, het Instituut voor de Nederlandse Taal en het KNAW Humanities Cluster. SERPENS is mogelijk gemaakt door het NWO-project CLARIAH-CORE.

clariah.nl/projecten/research-pilots/serpens

Lowlandsgangers testen eerste hiphopgenerator

Machine learning: wat is echt en wat is nep?

Ruim 700 bezoekers testten afgelopen zomer de rapbot die ontwikkeld werd binnen het project Deepflow. Taaltechnoloog Folgert Karsdorp praat ons bij over de resultaten. Mathilde Jansen

Karsdorp, postdoc aan het Meertens Instituut, ontwikkelde samen met onderzoekers van de Universiteit Antwerpen een hiphopgenerator, de rapbot. Deze genereert zelf hiphopteksten op basis van 64.000 (vooral Amerikaanse) originele raps. Tijdens Lowlands, een drie dagen durend festival boordevol muziek, theater, film, comedy, literatuur en wetenschap, lieten ze de rapbot testen in een zogenaamde MC Turing-test. Bezoekers aan Lowlands Science zagen een tekst op een beeldscherm. Van elke tekst moesten ze beoordelen of de tekst een echte rap was of nep. Karsdorp wilde daarbij vooral weten: waar letten mensen op? En: werkt het bij iedereen hetzelfde?

Het antwoord op de laatste vraag is ‘nee’. Uit de analyses blijkt dat er grote verschillen zijn tussen de antwoorden van zogenaamde amateurs

en experts. “De eerste groep gokt maar wat”, zegt Karsdorp, “terwijl de tweede groep let op talige eigenschappen van de tekst, zoals rijm, alliteratie en flow.”

Letter- of woordniveau

In totaal werden 6 taalmodellen getest. Het eerste model werkt op



letterniveau: het computermodel bepaalt per letter wat de meest waarschijnlijke letter is die volgt. Het tweede model werkt op woordniveau, het derde combineert beide modellen. “Een voordeel van het lettermodel is dat je te maken hebt met een klein vocabulaire”, legt Karsdorp uit, “26 letters en nog

‘Rapbot genereert zelf hiphopteksten’

allerlei interpunctie en diakritische tekens, zeg een stuk of 100 symbolen. Een nadeel van dit letter voor letter genereren, is dat je na 100 letters maar een paar woorden verder bent. Dan kun je beter overgaan op het woordniveau, maar daar is het probleem juist dat je vocabulaire heel groot wordt. In ons corpus had-

den we bijvoorbeeld te maken met 380.000 unieke woorden. Naarmate je vocabulaire groeit, heeft de computer meer trainingsmateriaal nodig om de juiste voorspellingen te doen. Daarom hebben we een hybride model toegevoegd.” Vervolgens werden alle modellen getest met en zonder bepaalde talige condities zoals de flow van de regels, rijm-elementen en ritmische aspecten.

Sterkste effect

Karsdorp: “Bij het lettermodel pikten de deelnemers de gegenereerde fragmenten er het makkelijkst uit. Bij het hybridemodel zat het slagingspercentage echter ongeveer op kansniveau. Als er talige condities



Folgert Karsdorp studeerde Nederlands en Taalkunde in Leiden en promoveerde in 2016 cum laude aan de Radboud Universiteit. Voor zijn proefschrift ontwikkelde hij een computermodel om de hervertellingen van verhalen te simuleren. foto Evy van Schelt

aan het model waren toegevoegd, presteerden alle modellen beter – en werd het dus moeilijker voor de deelnemers. Maar het sterkste effect hadden ze op het woordmodel. Ritmische aspecten zitten op het lettergreepniveau, en vallen daardoor het beste samen met het woordmodel.”

Karsdorp gebruikt de resultaten van dit onderzoek om nieuwe modellen voor culturele verandering te ontwikkelen. Niet alleen in taal, maar ook bijvoorbeeld in melodie. Voor zijn onderzoek put hij onder andere uit de liederenbank en volksverhalenbank van het Meertens Instituut. deep-flow.nl

Een schat aan informatie, klaar om te benutten

CBS Microdata in beeld

Universitair hoofddocent Anne Gielen onderzocht de langetermijneffecten van het aanpassen van de arbeidsongeschiktheidsvoorziening in de jaren negentig. Alle gegevens waren er al, bij CBS Microdata Services.
Marika de Bruijne



Anne Gielen foto Karoly Effenberger

Midden jaren negentig leidde een grote beleidshervorming in Nederland ertoe dat de arbeidsongeschiktheidsvoorziening minder generoos werd. Gielen (Erasmus Universiteit Rotterdam) vertelt: “We hebben bestudeerd of er een direct verband bestaat tussen uitkeringsafhankelijkheid bij ouders en de kans op economisch succes van hun kinderen op latere leeftijd. We keken in het bijzonder naar de individuen die een arbeidsongeschiktheidsuitkering ontvingen midden jaren negentig tijdens de beleidshervorming. We maakten gebruik van verschillende data die het CBS beschikbaar stelt via de remote access voorziening: informatie over economische uitkomsten zoals werk, loon of uitkeringsontvangst, maar ook informatie over onderwijsniveau, medicatiegebruik en criminele activiteiten.” Door CBS Microdata Services zijn

deze gegevens onderling koppelbaar op microniveau. “Dankzij de koppeling van de gegevens van de ouders met die van hun kinderen, kon heel specifiek gekeken worden naar intergenerationale effecten, hoe de situatie van een ouder de toekomstige situatie van een kind beïnvloedt. Dit, en het feit dat met de CBS-data gekeken kan worden naar effecten op lange termijn, maakt het werken met de CBS-data uniek”, aldus Gielen.

“De resultaten laten zien dat een lagere arbeidsongeschiktheidsuitkering, of zelfs helemaal geen uitkering meer, positieve gevolgen heeft op de kinderen”, vertelt Gielen. “Wanneer de kinderen zelf volwassen zijn, blijken zij minder vaak afhankelijk van een arbeidsongeschiktheidsuitkering, verdienen ze meer en zijn ze hoger opgeleid. Ook vertonen ze minder vaak crimineel gedrag en hebben ze minder vaak medicatie voor psychische

problemen.” De onderzoekers geven hier drie verklaringen voor: ouders zijn weer gaan werken, waardoor de kinderen meer leren over de arbeidsmarkt, de thuissituatie is verbeterd doordat ouders weer zijn gaan werken, of de kinderen verwachten minder van overheidssteun en anticiperen hierop door meer te investeren in hun onderwijs en latere kansen op de arbeidsmarkt.

Verloop datakoppeling

“De datakoppeling verliep doorgaans heel goed, mede dankzij de uitstekende documentatie en de support van het CBS. Het bewerken van de erg grote datasets in de beperkte computercapaciteit is wel een uitdaging. Advies voor onderzoekers die overwegen om de CBS Microdata te gebruiken? Vooral doen! Het CBS heeft een schat aan informatie liggen, klaar om benut te worden door onderzoekers.”

Het onderzoek “Intergenerational Spillovers in Disability Insurance” is beschikbaar op:
www.nber.org/papers/w24296
cbs.nl

Onderzoekers verbonden aan ODISSEI-deelnemers kunnen korting krijgen op het gebruik van CBS Microdata.

GELEZEN

*What to keep:
A JISC research data study*
Marjan Grootveld

Hoe bepaal je welke onderzoeksdata en documentatie het bewaren waard zijn? JISC deed onderzoek naar criteria, obstakels en mogelijke oplossingen. Het goed leesbare rapport is tweeledig: na een verslag van desk-research, dat ook vier Britse checklists vergelijkt, volgen tien aanbevelingen. Neil Beagrie interviewde hiervoor vertegenwoordigers van kennisinstellingen, repositories, uitgevers en onderzoeksfinanciers. Verfrissend is de aanbeveling om de relatieve kosten en baten te onderzoeken van verschillende niveaus van datacuratie voor de twee voornaamste use cases: reproduceerbaarheid van onderzoek en herbruikbaarheid van de data. De inspanning om data herbruikbaar te maken – en te houden – zou namelijk groter zijn dan die voor reproduceerbaarheid en wetenschappelijke integriteit.

Ook adviseert het rapport dat publicaties altijd met een ‘data availability statement’ naar onderliggende data moeten verwijzen. Hopelijk maakt JISC of Beagrie nog een handzame checklist voor onderzoekers en onderzoeksondersteuners om data te selecteren.

repository.jisc.ac.uk/7262

COLUMN

Samen kom je verder

Het universitaire beleid was de afgelopen dertig jaar gericht op het bevorderen van excellentie in onderzoek. Dit beleid toont nu zijn keerzijde, zo is te lezen in het recente Rathenau-rapport Excellent is niet gewoon (2018). Er is een academische cultuur ontstaan gericht op competitie. De wedijver speelt vooral tussen individuele onderzoekers, die zeer getraind zijn in het vinden van een specialistische onderzoeksniche in hun vakgebied waar zij belangrijke, nieuwe kennis aan kunnen toevoegen. De nadruk op individuele prestaties is echter niet bevorderlijk geweest voor het streven naar samenwer-

king – tussen onderzoeksgroepen, tussen disciplines en met partijen buiten de academie.

Grootse wetenschap kan juist door samenwerking ontstaan. Recent hebben we daar een prachtig voorbeeld van gezien. De eerste foto van een zwart gat kon alleen worden gemaakt door een wereldwijd consortium van sterrenkundigen en telescoop-eigenaren. Gelukkig begint ook in het brede domein van de sociale en geesteswetenschappen het inzicht te groeien dat gezamenlijk optrekken loont. In het rapport Samen sterker: Beeld van het SSH-domein (2018) roept kwartiermaker Mark Bovens op tot het maken van strategische plannen waarin gestructureerd wordt nagedacht over onderzoeksstrategieën en onderlinge afstemming. Het is belangrijk om aandacht te hebben voor afzonderlijk disciplines binnen de sociale en geesteswetenschappen, én te investeren in inhoudelijke samenwerking over de grenzen van disciplines. Veel maatschappelijke

vraagstukken vragen immers om een multidisciplinaire benadering. Een landelijke visie en onderzoeksagenda is gebaat bij een gemeenschappelijke infrastructuur, één die de toegang tot en hergebruik van data faciliteert, afstemming tussen dataverzamelingen bewaakt, interoperabiliteit van datasets bevordert, helpt bij het onderhouden van cruciale langlopende dataverzamelingen, en verbetering van methoden van dataverzameling mogelijk maakt. Met gepaste trots noem ik ODISSEI, de Open Data Infrastructure for Social Science and Economic Innovations, eind 2016 opgericht op initiatief van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO), het Centraal Bureau voor de Statistiek (CBS) en de Erasmus Universiteit Rotterdam (EUR). Inmiddels heeft ODISSEI 33 deelnemers: faculteiten (sociale wetenschappen, economie en bedrijfskunde, ruimtelijke wetenschappen, gezondheidsbeleid), CentERdata, planbureaus, Netherlands eScience



Center, Nivel, SURFsara en onderzoeksinstituten van NWO en de Koninklijke Akademie van Wetenschappen (KNAW). Alle deelnemers dragen financieel bij aan ODISSEI en onderschrijven daarmee dat data-infrastructuur een gedeeld goed en een gedeelde verantwoordelijkheid is.

ODISSEI heeft inmiddels verschillende faciliteiten voor het veld van de sociale wetenschappen gecreëerd. Een eerste betreft subsidies voor het gebruik van CBS-microgegevens over personen, huishoudens en bedrijven. De reikwijdte van de informatie die de CBS-registers bevatten over

gezondheid, wonen, inkomen, criminaliteit, voorzieningengebruik, enzovoorts, is ongekend. Een tweede faciliteit betreft subsidies voor eigen dataverzamelingen via het LISS panel (Longitudinal Internet studies for the Social Sciences). Uiteraard kunnen de nieuwe data gekoppeld worden aan reeds verzamelde informatie over de panelleden en aan CBS-microgegevens. Een derde betreft de zogenaamde ODISSEI Data Facility (ODF), een gesloten omgeving op de Cartesius supercomputer van SURFsara, waar sociaal wetenschappers hun data kunnen verrijken met die van het CBS en tegelijkertijd gebruik kunnen maken van een buitengewoon groot rekenvermogen. Een faciliteit als de ODF bestaat nergens anders ter wereld.

Door samen te werken, beoogt het consortium van ODISSEI-deelnemers een betere data-infrastructuur te verwezenlijken, die de basis zal zijn voor betere wetenschap, die vervolgens kennis zal inzetten voor een betere samenleving. Alleen gaat het misschien sneller, maar samen kom je verder.

Pearl A. Dykstra

Ik geef de column door aan Maarten van Ham.

Pearl Dykstra

Pearl Dykstra is hoogleraar Empirische Sociologie aan de Erasmus Universiteit Rotterdam en vicevoorzitter van de High Level Group van wetenschappers die de Europese Unie adviseren. Dykstra is een internationaal gewaardeerde specialist op het gebied van intergenerationale solidariteit, vergrijzende samenlevingen, gezinsverandering, ouder worden en de levensloop, en eenzaamheid. Naast wetenschappelijk directeur van ODISSEI is ze ook lid van de KNAW en diende ze van 2011-2016 als vicevoorzitter van de KNAW.