



E-DATA & RESEARCH

Jaargang 12 | 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

2 Hoe open access is Nederland anno 2018?

3 Martijn Kleppe, hoofd Onderzoek KB



4/5 spread 'Privacy'

- Gevolgen AVG voor data-archieven
- Gouden tips van lopend onderzoek
- ODISSEI biedt strikt beveiligde omgeving
- DataTag-tool bepaalt toegangsniveau data

7 Stand van zaken CLARIAH-CORE

8 Ewoud heeft een mooie taak voor bibliotheken

E-data kunt u gratis per post en/of digitaal ontvangen. Mail uw verzoek naar edata@dans.knaw.nl.



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

Nederlab geeft flinke impuls aan diachroon onderzoek

Miljoenen teksten digitaal doorzoekbaar

Op 1 juli wordt Nederlab opgeleverd: met miljoenen Nederlandse teksten biedt het onderzoekers een schat aan informatie over de Nederlandse taal en cultuur.

Mathilde Jansen

In 2012 kreeg Nederlab 3,2 miljoen euro toegekend door NWO. Nu, 6 jaar later, geeft Nederlab met zo'n 25 collecties aan teksten, lopend van de 13^e tot 21^e eeuw, een flinke impuls aan diachroon geesteswetenschappelijk onderzoek. Van kranten teksten tot dagboeken: in Nederlab vind je al-

'Na inlog kunnen onderzoekers en studenten een eigen corpus samenstellen'

lerlei soorten teksten. Het oudst zijn de dertiende-eeuwse teksten uit het Corpus Gijsseling, gevolgd door de veertiende-eeuwse teksten uit het Corpus Van Reenen-Mulder. Meer van deze tijd zijn teksten uit grote krantencorpora of hele romans afkomstig uit de Digitale Bibliotheek voor de Nederlandse Letteren (DBNL). Voor onderzoekers biedt Nederlab een schat aan informatie over de Nederlandse taal en cultuur.

Het Nederlab-consortium bestond uit het Meer-

tens Instituut, het Instituut voor Nederlandse taal, het Huygens ING, de Universiteit Nijmegen, de Koninklijke Bibliotheek, de DBNL en diverse dataleveranciers. Niet alleen veel data werden bijeengebracht, ook werd gezorgd voor een onderliggende infrastructuur en een groot aantal tools. Doordat de data taalkundig verrijkt zijn, is het mogelijk om bijzonder complexe zoekopdrachten uit te voeren over een heel breed corpus.

40 miljard annotaties

Hennie Brugman is IT-ontwikkelaar aan het Meertens Instituut en coördinator van Nederlab. Hij vertelt hoe de zoekmachine is opgebouwd: "Bij de oplevering bevat Nederlab dertig miljard woorden. Die hebben we zo veel mogelijk voorzien van taalkundige annotaties via automatische processen. Ieder woord is bijvoorbeeld gekoppeld aan een lemma en een woordsoort. Zo ontstaat een structuur van aan elkaar gekoppelde lagen van tekstwaarden. Het zoeken kun je doen door alle lagen heen; dat is de basis van de hele machine. Bij tien miljard woorden praat je dus al gauw over 40 miljard annotaties."

Ontwikkeling door tijd

Iedereen kan Nederlab gebruiken, maar onderzoekers en studenten hebben meer onderzoeksmogelijkheden na inlog. Zo kunnen ze hun eigen corpus samenstellen. Nicoline van der Sijs, taalkundige aan het Meertens Instituut, schreef destijds het projectvoorstel en is ook daarna nauw betrokken geweest bij Nederlab. Voor haar onderzoek maakt ze al regelmatig gebruik van het nieuwe onderzoeksportal: "Wat ik interessant vind, is dat je een ontwikkeling kunt volgen over een periode van eeuwen. Dat kon voorheen niet omdat alle corpora met Nederlandse teksten versnipperd waren. Ik denk dat we hierdoor nieuwe inzichten krijgen in de ontwikkeling van het Nederlands en de oorzaken van taalverandering."

www.nederlab.nl/onderzoeksportal

🛒 Het evangelie volgens Matteüs

gerne: non-fictie, bijbel/bijbelteksten

collectie: SoNaR

aantal hits: 2

woord:	God	zal	ook	met	hen	medelijden	hebben
lemma:	God	zullen	ook	met	hen	medelijden	hebben
pos:	SPEC	WW	BW	VZ	VNW	N	WW
kenmerken:	deeleigen	pv		init	pers	soort	inf
		tgw			pron	ev	vrij
		ev			obl	basis	zonder
					vol	onz	
					3p	stan	
					mv		

De zoekterm 'medelijden hebben' binnen de Nederlabportal 'geavanceerd zoeken' in 'zoeken in tekst' toont binnen de dataset 'Het evangelie volgens Matteüs' 2 hits met de volgende taalkundige tags: lemma (woord), pos (woordsoort, bijvoorbeeld spec = speciaal, ww = werkwoord, vz = voorzetsel), kenmerken verwijst naar de functie in de zin (pv = persoonsvorm, tgw = tegenwoordige tijd, ev = enkelvoud). Met dit resultaat kan een onderzoeker nagaan hoe vaak de combinatie 'medelijden hebben' in deze tekst voorkomt. credits Nederlab

OPROEP



Wint u de Nederlandse Dataprijs '18?

Op 28 november wordt 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.

Nomineer voor 1 juli

Bent of kent u zo'n onderzoeker met een goed voorbeeld van gedeelde data? U kunt tot 1 juli uzelf of een ander via researchdata.nl nomineren.

De organisatie van de Nederlandse Dataprijs is in handen van Research Data Netherlands. Kijk voor meer informatie over de prijs, zoals de categorieën, de prijzenpot en de criteria, op de website van RDNL. (HB)

researchdata.nl/diensten/dataprijs

Amsterdam Time Machine: reizen door tijd en ruimte

Het consortium Amsterdam Time Machine (ATM) wil op termijn een 'Google Earth' maken van het verleden. Gebruikers kunnen zich hiermee door Amsterdam verplaatsen in ruimte en tijd, op wijk-, straat- en huishoudniveau. ATM verbindt de Amsterdamse erfgoedinstellingen in AdamNet, de creatieve industrie, de gemeente Amsterdam en sociaal- en geesteswetenschappelijk onderzoekers aan zowel de universiteiten als bij de KNAW. Door een financiële bijdrage uit CLARIAH komt de geo-infrastructuur HisGIS van de Fryske Akademy beschikbaar voor ATM. Zo komen data en tools ter beschikking voor onderzoek naar Amsterdam, en voor geesteswetenschappelijk onderzoek in het algemeen. (MM) www.create.humanities.uva.nl/amsterdam-time-machine

Derde symposium Mobiliteitspanel Nederland

Marion Wittenberg

Verandert je reispatroon als je van de stad naar het platteland verhuist? Waarom kiezen sommige mensen ervoor om te gaan lopen of fietsen? Speelt je gezondheid een rol bij het kiezen van je vervoermiddel? Het Kennisinstituut voor Mobiliteitsbeleid (KiM) organiseerde samen met DANS op 10 april jongleden een symposium waar deze en andere onderwerpen aan bod kwamen. De meeste presentaties op dit goed bezochte symposium hadden betrekking op het Mobiliteitspanel Nederland (MPN), een online panel survey onder 4.000 respondenten van 2.000 Nederlandse huishoudens. Sinds de eerste opzet van het MPN in 2011, is de manier waarop respondenten een online vragenlijst beantwoorden veranderd, door de sterke toename van het gebruik van smartphones. Toon Zijlstra (KiM) ging in op dit fenomeen. 33% van de respondenten beantwoordde de vragenlijst in 2016 met een smartphone. Dit is een specifieke groep; stedselijk, jong, lager opgeleid, die op een andere manier moeilijk te bereiken is. Vandaar dat het belangrijk is om de vragenlijsten aan te passen aan het apparaat die de respondent gebruikt. Kortere en eenvoudiger vraagstelling en andere layout van de vragen zijn noodzakelijk om ook

GEHOORD & BIJGEWOOND



De data van het MPN zijn belangrijk voor het analyseren van keuzegedrag; wanneer neemt iemand welk vervoermiddel, hoe groot is de variatie hierin en waarvan is dit afhankelijk? Onderzoek hiernaar is belangrijk om effectieve beleidsmaatregelen te kunnen ontwerpen die het gebruik van meer duurzame vervoerswijzen stimuleren. foto Schutterstock

deze groep respondenten te bereiken.

Het MPN verzamelt ook data door middel van een dagboekje dat de respondenten drie dagen moeten bijhouden. Marie-José Olde Kalter (Universiteit Twente / Goudappel Coffeng) liet zien hoe belangrijk deze data zijn voor het analyseren van keuzegedrag; wanneer neemt iemand welk vervoermiddel, hoe groot is de variatie hierin en waarvan is dit afhankelijk. Onderzoek

hiernaar is belangrijk om effectieve beleidsmaatregelen te kunnen ontwerpen die het gebruik van meer duurzame vervoerswijzen stimuleren.

Naast presentaties over het MPN waren er ook bijdragen over het German Mobility Panel en de British Household Panel Survey.

Alle presentaties zijn beschikbaar op de DANS-website:

<https://dans.knaw.nl/nl/actueel/agenda/3rd-netherlands-mobility-pa->

nel-mpn-symposium

Meer informatie over de data van het MPN is te vinden op de website van het KiM:

<https://www.kimnet.nl/mobiliteits-panel-nederland/over-het-mpn>

Hernieuwd zelfvertrouwen op EuropeanaTech

Steven Claeysens

Half mei kwamen in Rotterdam op het gelijknamige stoomschip drie werelden bijeen: de wetenschap, het erfgoed en de creatieve industrie. EuropeanaTech-deelnemers uit alle windstreken monsterten voor twee dagen aan. Op de agenda stond de toekomst van Europa's digitale erfgoed op het web.

Tijdens het slotdebat concludeerde Jill Cousins, oud-directeur van Europeana, dat er in vergelijking met de vorige editie in 2015 te Parijs hernieuwd (zelf)vertrouwen lijkt te zijn. Het spook genaamd Google waart iets minder nadrukkelijk door erfgoedland en een aantal technieken om erfgoed en informatie bruikbaar online te plaatsen en tegelijk in het publieke domein te houden, bereiken de volwassen leeftijd. Bovendien maken steeds meer musea, bibliotheken en archieven ook gebruik van die mogelijkheid en

stellen ze hun digitale collecties vrijelijk ter beschikking, voor de wetenschap, de industrie en het grote publiek.

Veel aandacht ging uit naar IIF, wikidata en Linked Data, met een opgemerkt pleidooi door Robert Anderson (J. Paul Getty Trust) voor LOUD, Linked Open Usable Data. Ook toepassingen op basis van recente doorbraken in machine learning kwamen aan de orde, bijvoorbeeld voor automatische categorisering van vinders bij Naturalis of beeldherkenning in oude kranten bij de Franse nationale bibliotheek. Natuurlijk zijn nog lang niet alle zorgen van de baan of alle discussies beslecht. Zo bestaat er nog altijd onenigheid over de gewenste mate van centralisatie dan wel decentralisatie op het web en is de zoektocht naar een gebruikersinterface die meer biedt dan een klassieke zoekfunctie met zoektermen nog lang niet ten einde. De meest verontrustende boodschap bracht Herbert van de Sompel (Los Alamos National Laboratory). Hij herhaalde zijn bevinding dat een groot deel van de verwijzende hyperlinks in wetenschappelijke publicaties al na enkele jaren niet meer terug te vinden zijn. Gelukkig werkt hij, samen met anderen, ook aan oplossingen voor dit probleem.

pro.europeana.eu/event/europeana-tech-conference-2018

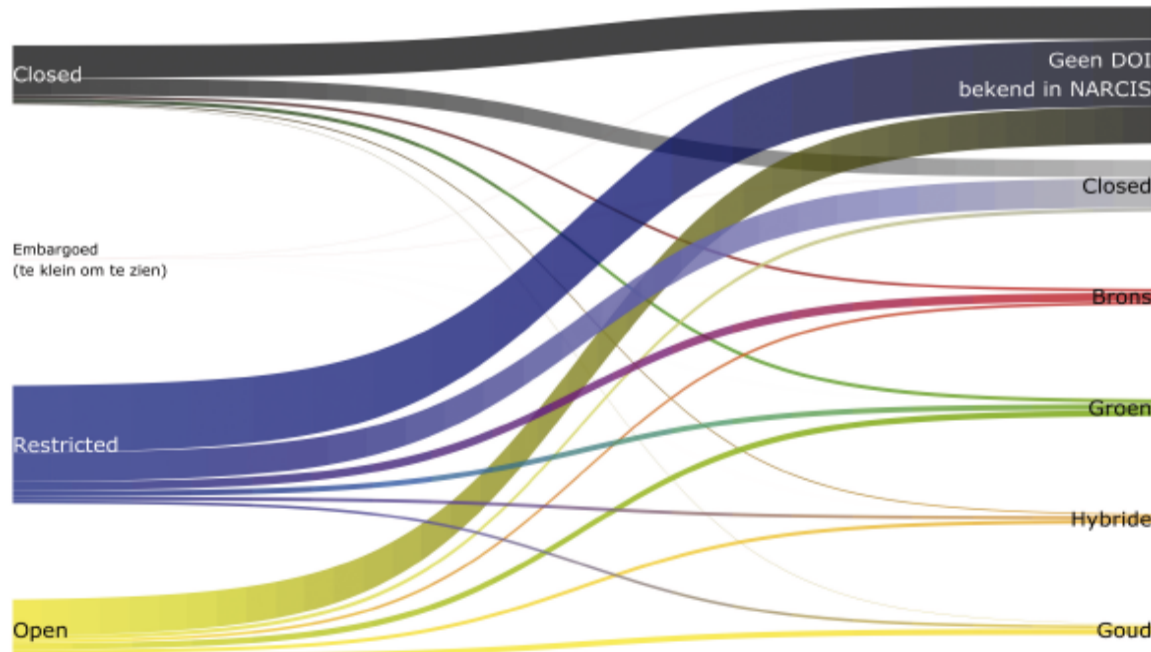
In 2020 moeten alle publicaties open access zijn

Hoe open access is Nederland anno '18?

In de 'Amsterdam Call for Action on Open Science' is afgesproken dat in 2020 alle wetenschappelijke publicaties open access moeten zijn. Hoe is de stand van zaken in Nederland? En hoe betrouwbaar is de beschikbare informatie?

Emil Bode

NARCIS, dé nationale portal voor wie informatie zoekt over wetenschappers en hun werk, toont een groot deel van de wetenschappelijke publicaties in Nederland, met informatie over de toegankelijkheid: open, restricted, embargoed of closed access. Om deze informatie te controleren en aan te vullen zijn de gegevens uit NARCIS vergeleken met die van Unpaywall, een vrij toegankelijke dienst om publicaties te vinden op basis van hun Digital Object Identifier (DOI), met links naar open access (OA)-versies. Unpaywall werkt met een ander classificatiesysteem dan NARCIS, namelijk op basis van de vindplaats van de



Een vergelijking: links de toegankelijkheid van artikelen beschikbaar via NARCIS, rechts de toegankelijkheid van dezelfde artikelen volgens Unpaywall. Credits Emil Bode

publicatie: goud (via een OA-journal), hybride (OA via website van de uitgever), groen (via een (betrouwbare) repository), brons (via het internet, zonder licentie of garantie dat de publicatie toegankelijk blijft), of closed (niet OA). De boven-

staande afbeelding toont de vergelijking tussen artikelen in NARCIS en Unpaywall.

Toegankelijkheid

De vergelijking geeft een beeld van de verhoudingen in toegankelijk-

heid van wetenschappelijke publicaties in Nederland volgens NARCIS en volgens Unpaywall: een groot deel van de artikelen is in bepaalde vorm toegankelijk, maar de situatie is vaak nog onduidelijk, er zijn beperking of de artikelen zijn

niet goed vindbaar. En in de gemaakte vergelijking komen alle combinaties voor, ook al zou dat niet moeten kunnen (een publicatie kan nooit open en closed zijn). Deze informatie kan als basis dienen om de systemen te verbeteren, bijvoorbeeld door onderzoeksinstituten hun publicaties op dezelfde wijze te laten classificeren en van alle publicaties een DOI door te geven aan NARCIS. Ook Unpaywall is niet perfect: in hun classificatie betekent closed dat ze geen open versies hebben gevonden. Dat kan omdat die er niet zijn, maar het kan ook zijn dat ze de open versie niet hebben gevonden. NARCIS krijgt informatie direct van de bron, en kan Unpaywall soms dus aanvullen. Uiteindelijk willen we naar een situatie waarbij duidelijk is welke artikelen via welke toegangscategorie beschikbaar zijn voor hergebruik, dat ze makkelijk vindbaar zijn, en dat de informatie betrouwbaar is.

narcis.nl
unpaywall.org

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.

Martijn Kleppe, kersvers hoofd Onderzoek van de KB:

‘Het geschreven woord is het uitgangspunt voor bewaren’

Historicus Martijn Kleppe is sinds kort het nieuwe hoofd Onderzoek van de Koninklijke Bibliotheek. E-data legt hem vijf uitspraken voor.

De vijf uitspraken komen van experts die eerder dit jaar de thema's uit de nieuwe onderzoeksagenda van de KB introduceerden. Steven Claeysens

1

De informatiemaatschappij

Valerie Frissen (directeur Stichting Internet Domeinregistratie Nederland en bijzonder hoogleraar ICT & Social Change, Erasmus Universiteit Rotterdam): ‘De rol van de bibliotheek in de informatiemaatschappij zou heel groot kunnen zijn, maar het is nog een hele weg om daar te komen.’

Kleppe: ‘In hun zoektocht naar informatie schakelen mensen online moeiteloos van Google naar Wikipedia naar Facebook, maar niet of heel beperkt naar bibliotheken. Terwijl het wel onze taak is om een gids te zijn in het informatielandschap. We moeten dus goed begrijpen hoe en waar mensen zoeken om ook op die plekken informatie aan te bieden waar ze van op aan kunnen. Daarnaast zien we dat mensen ook nog behoefte houden aan menselijk contact. Grote webwinkels als Amazon en Coolblue openen winkels in grote steden en cafés worden steeds meer gebruikt als ontmoetings- en werkplek. Een vraag die we ons dus stellen is: welke rol speelt een nationale bibliotheek, zowel digitaal als fysiek, in een veranderende informatiemaatschappij?’

2

Publicaties bewaren

Barbara Sierman (Digital Preservation Officer, KB): ‘Algoritmes bepalen wat we zien, weten we dan nog welke digitale publicaties we willen bewaren?’

Kleppe: ‘Het geschreven woord is het uitgangspunt van wat de KB bewaart. De vraag is alleen wat we nu nog verstaan onder het geschreven woord en welke typen publicaties we willen en kunnen opslaan. Digitale publicaties worden online geplaatst, herplaatst en gekoppeld aan data. Denk aan een multimediaal ebook, een blogpost, een tweet of een verrijkte wetenschappelijke publicatie. Wat kun je nog beschouwen als publicatie en hoe bewaar je die? Daarnaast zien we steeds meer gepersonaliseerde publicaties ontstaan, die op maat gemaakt worden voor consumenten. Hoe kunnen we dat soort uitingen bewaren en hoe zorgen we dat de data die we opslaan FAIR (Findable, Accessable, Interoperable en Reusable) blijven – ook als de technologie van vandaag op een dag verouderd is?’



“Het is onze taak om een gids te zijn in het informatielandschap”
foto Jos Uljee, Koninklijke Bibliotheek

INTERVIEW

‘Taaltechnologie en beeldherkenning helpt om de teksten steeds beter te interpreteren en te beschrijven’

3

Collecties ontsluiten

Johan Oomen (manager Kennis en Innovatie, Beeld en Geluid): ‘Handmatig kan niet meer, dus we hebben nieuwe technieken nodig, bijvoorbeeld gezichtsherkenning en sprekers-

herkenning. Het spannende is dat we zo heel andere vragen kunnen stellen aan die collecties.’

Kleppe: ‘Naast het duurzaam bewaren van onze collecties willen we deze ook optimaal beschikbaar stellen voor onze gebruikers. Het gaat dan over het verbeteren van de kwaliteit van digitale content zodat de computer het kan interpreteren, over het creëren van metadata, over het verrijken van de content én over het verspreiden ervan. Zeker bij het creëren van metadata heeft Beeld en Geluid grote stappen gemaakt waar de KB veel van kan leren. Omdat we steeds meer publicaties volledig digitaal tot onze beschikking hebben, kunnen we met behulp van taaltechnologie en beeldherkenning de teksten steeds beter interpreteren en automatisch laten beschrijven. Zo experimenteren we met het extraheren van personages en genres uit publicaties die we vervolgens kunnen toevoegen aan de metadata. Daarnaast kunnen we ook met behulp van beeldherkenning personen, voorwerpen en onderwerpen op bijvoorbeeld foto's in digitale kranten identificeren. We werken hard om dit soort toepassingen verder te ontwik-

Martijn Kleppe

Martijn Kleppe is historicus. Na de afronding van zijn proefschrift *Canonieke Icoonfoto's* (2013) werkte hij aan de Erasmus Universiteit Rotterdam en de Vrije Universiteit als onderzoeker in verschillende Digital Humanities-projecten gericht op het openen en koppelen van (audio)visuele collecties. Sinds februari 2016 werkt hij bij de KB, sinds maart dit jaar als hoofd van de afdeling Onderzoek.

kelen zodat we ze op termijn beschikbaar kunnen stellen via bijvoorbeeld Delpher. Hierdoor zou je in de toekomst ook kunnen zoeken naar een gebouw op een foto terwijl de naam van het gebouw of het woord gebouw helemaal niet in het bijschrift van de foto staat.’

4

Gebruik van collecties

Maarten de Rijke (hoogleraar Information Retrieval, Universiteit van Amsterdam): ‘De KB kan extreem open zijn. Vertel maar gewoon wat je met de data, de logs doet.’

Kleppe: ‘Naast onze gedigitaliseerde collecties hebben we ook veel data over het gebruik van onze collecties en diensten. Hiermee beschikken we over Big Data-achtige gegevens waarmee we het gedrag van onze gebruikers kunnen analyseren en hen eventueel kunnen helpen in hun zoektocht naar (online) informatie. Dat doen we op een ethisch verantwoorde manier in een veilige en betrouwbare omgeving. Zo kijken we bijvoorbeeld graag naar het Responsible Data Science-consortium dat de nadruk legt op de FACT-principes van Fairness, Accuracy, Confidentiality en Transparency.’

5

Impact op de samenleving

Harry Verwayen (directeur Europeana): ‘Begin met kijken of we wel de juiste dingen meten.’

Kleppe: ‘We willen steeds meer de lange termijn-impact van onze activiteiten in kaart brengen. Dat is geen eenvoudige opdracht, want hoe toon je je toegevoegde waarde? Hoe meet je maatschappelijke opbrengst? We weten wat ons bereik is onder de Nederlandse bevolking, hoeveel mensen onze diensten gebruiken en hoe tevreden ze zijn. Maar of hiermee op langere termijn de samenleving verandert, weten we niet. En dus willen we indicatoren ontwikkelen die de impact meten van de bijdrage van de KB aan een duurzame samenleving.’

Meer informatie over de KB en de toekomstplannen, zoals de vijf thema's: informatiemaatschappij, publicaties, ontsluiten en delen, klant en impact staat op de website van de KB.

kb.nl/organisatie/onderzoek-expertise/onderzoeksagenda-2018-2022

Nieuwe Europese privacywetgeving van kracht

Bent u al goed voorbereid op de nieuwe Europese privacywetgeving? Op 25 mei is de Algemene Verordening Gegevensbescherming (AVG) in werking getreden.

Ook als onderzoeker en onderzoeksondersteuner heeft u met deze nieuwe wet te maken. Zo eist de AVG dat data niet langer dan noodzakelijk bewaard worden. Ook is het verplicht om aan betrokkenen bij onderzoek toestemming te vragen voor het bewaren van hun persoonsdata.

Kortom: de AVG heeft gevolgen voor de wijze waarop met persoonsgegevens wordt omgegaan en hoe onderzoek kan worden gedaan. Lees enkele ervaringsverhalen in deze spread.

Voorbeelden van informed consent-verklaringen online beschikbaar

Heeft u van de betrokkenen bij uw onderzoek al expliciet toestemming gekregen om de gegevens uit interviews te archiveren en/of openbaar te mogen maken? Bij het deponeren van onderzoeksgegevens uit oral history of audiovisuele bronnen is

het aanleveren van toestemmingsverklaringen van de geïnterviewden en interviewer verplicht. DANS heeft voorbeelden van zogenoemde informed consentverklaringen opgesteld om onderzoekers te helpen bij het verkrijgen van toestemmingen. Download deze templates via de DANS-website. U kunt ook contact met ons opnemen voor meer informatie.

<https://dans.knaw.nl/nl/actueel/nieuws/voorbeelden-informed-consent-verklaringen-beschikbaar>

Gevolgen nieuwe AVG voor data-archieven

Wat mogen DANS en andere archieven nog met databestanden die persoonsdata bevatten?

Heiko Tjalsma

Deze vraag is op dit moment niet makkelijk te beantwoorden. Naast de AVG zijn namelijk een reeks van aanvullende wetten, gedragscodes en richtlijnen relevant voor de uitleg van de AVG. Deels zijn deze echter nog in de maak.

Nationale uitvoeringswet

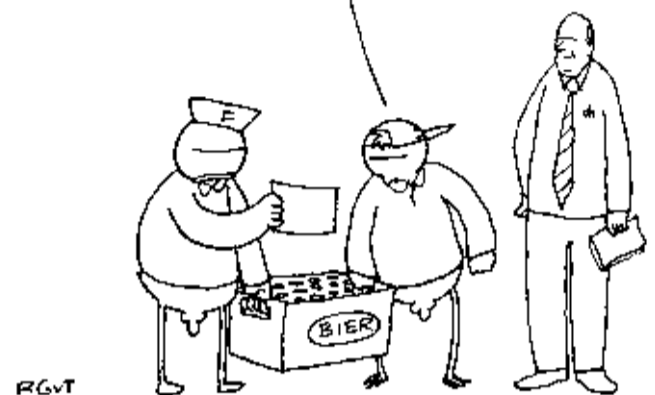
De Europese AVG opent expliciet de mogelijkheid om door middel van nationale wetgeving uitzonderingen voor het gebruik van persoonsdata ten behoeve van wetenschappelijk, historisch (archief) of statistisch onderzoek nader te bepalen. In Nederland is dat de UAVG – de Uitvoeringswet AVG. Deze wet volgt in grote lijnen de artikelen van de (oude) WBP. Daarmee heeft het onderzoeksveld al ervaring.

VSNU-gedragscode

Daarnaast komen er ook gedragscodes, op nationaal én op Europees niveau. In Nederland is, op initiatief van de VSNU, een gedragscode voor de omgang met persoonsdata in wetenschappelijk onderzoek in de maak. Ook wordt aan Europese gedragscodes per discipline gewerkt. Om deze codes door de Europese Commissie goedgekeurd te krijgen, moet representativiteit van het be-

FOKKE & SUKKE
VOELEN ZICH AANGETAST IN HUN PRIVACY

"ANDEREN DIE DIT PRODUCT KOZEN, KOCHTEN DAAR VAAK EEN ZAK CHIPS BIJ."



trokken onderzoeksveld aangetoond worden. Een Europese gedragscode zal uiteraard als gezaghebbend beschouwd worden, ook al is het 'soft law'. Tenslotte heeft de Europese privacy-waakhond (nu nog Article 29 Working Party) zeer strikt geformuleerde richtlijnen gepubliceerd, onder andere over hoe om te gaan met informed consent van (proef)personen die aan onderzoeksprojecten deelnemen.

Veiligheid data

Hoe het ook precies gaat uitpakken, een belangrijk aspect zal de veiligheid van de data worden, aangezien de AVG voorschrijft dat er voldoende technische en organisatorische maatregelen moeten zijn getroffen om de veiligheid van

persoonsdata te waarborgen. Het indelen van persoonsdata op grond van AVG-criteria in zogenaamde Datatags kan hier een grote rol gaan spelen (zie artikel op pagina 5). Voor zover dat mogelijk was, heeft de onderzoeksweld zich serieus voorbereid op de komst van de AVG en de daaraan gerelateerde harde en zachte wetgeving.

Drs. Heiko Tjalsma is beleidsadviseur/juridisch adviseur bij DANS www.rijksoverheid.nl/onderwerpen/privacy-en-persoonsgegevens/hulpbijprivacy.nl/ ec.europa.eu/info/law/law-topic/data-protection_nl/autoriteitpersoonsgegevens.nl/nl/onderwerpen/avg-nieuwe-europese-privacywetgeving

Zorgen dat zorgprofessionals ook vooraf kunnen handelen

Veilig opgroeien door machine learning

Het project 'Datagestuurd preventief werken' ontwikkelt een model om problemen bij gezinnen op tijd te identificeren of zelfs te voorkomen.

Marika de Bruijne

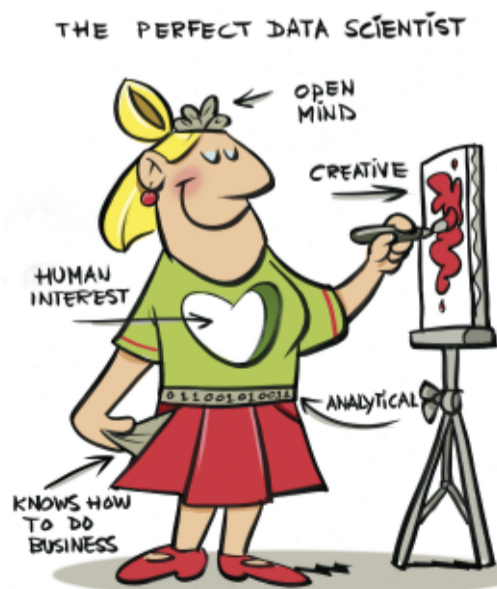
"Een zorgprofessional of leraar heeft vaak een onderbuikgevoel, maar weet dat gevoel niet precies te onderbouwen", vertelt Patricia Prüfer, projectleider bij CentERdata. "Samen met Sterk Huis, gemeente Tilburg en Data Science Centre van Tilburg University zoeken wij naar de onderbouwing van dat gevoel."

Diverse bronnen

"We hebben een lijst van indicatoren die een potentieel onveilige situatie voor kinderen voorspellen. De achterliggende data van deze indicatoren verzamelen we bij verschillende bronnen: gemeente, CBS, GGD, GGZ, leerplichtambtenaar, politie,

scholen en eerdere onderzoeken. Door de data te onderzoeken met hulp van machine learning, willen we de giftige cocktail van risicofactoren opsporen."

De grootste uitdagingen van het onderzoek blijken praktisch van aard. "Iedereen is enthousiast, iedereen steunt het onderzoek. Maar als je om de gegevens vraagt, vindt men



Patricia Prüfer, projectleider bij CentERdata: "Wees transparant, open minded en deel je kennis. Op weg naar de perfecte data scientist!" credit Marion van de Wiel

ze toch te gevoelig om beschikbaar te stellen, ook al is de aanpak niet op persoonsniveau maar op wijkniveau", vertelt Prüfer. Wat ook blijkt: de datakwaliteit is lang niet altijd optimaal. Prüfer adviseert: "Zoek de juiste contacten met bevoegdheden om data te delen, reserveer veel tijd voor gesprekken en voorlichting, zorg voor een hoger aggregatieniveau als data anders niet gedeeld mogen worden. Bovendien, om data steeds bruikbaar en het model steeds beter te maken, moet je de professionals die met een registratiesysteem werken, trainen. En wees transparant en deel je kennis, zo kweek je vertrouwen."

Lerende keten

Eind dit schooljaar moet de pilot het eerste voorspellingsmodel opleveren. "We willen interactieve sessies organiseren met beleidsmedewerkers, leerkrachten en ouders.

Door de vraag te stellen: "Herkennen jullie wat de analyses opleveren?", kan het model waar nodig bijgesteld en verbeterd worden. We zien het onderzoek als één levend systeem van data producenten, analisten en eindgebruikers, de hele keten is lerend", aldus Prüfer.

Kansen data science

Uiteindelijk is het doel om het model op landelijk niveau uit te rollen. Prüfer: "Vaak zien onderzoekers maar ook zorgprofessionals alleen wat ze al kennen. Zoals een huisarts die ontkent dat er sprake is van huiselijk geweld onder patiënten van zijn praktijk. Data science combineert bestaande informatie en biedt dus een kans om een tunnelvisie te doorbreken en open te kijken naar wat er echt aan de hand is."

www.centerdata.nl/nl/projecten-van-centerdata/pilot-datagedreven-preventief-werken

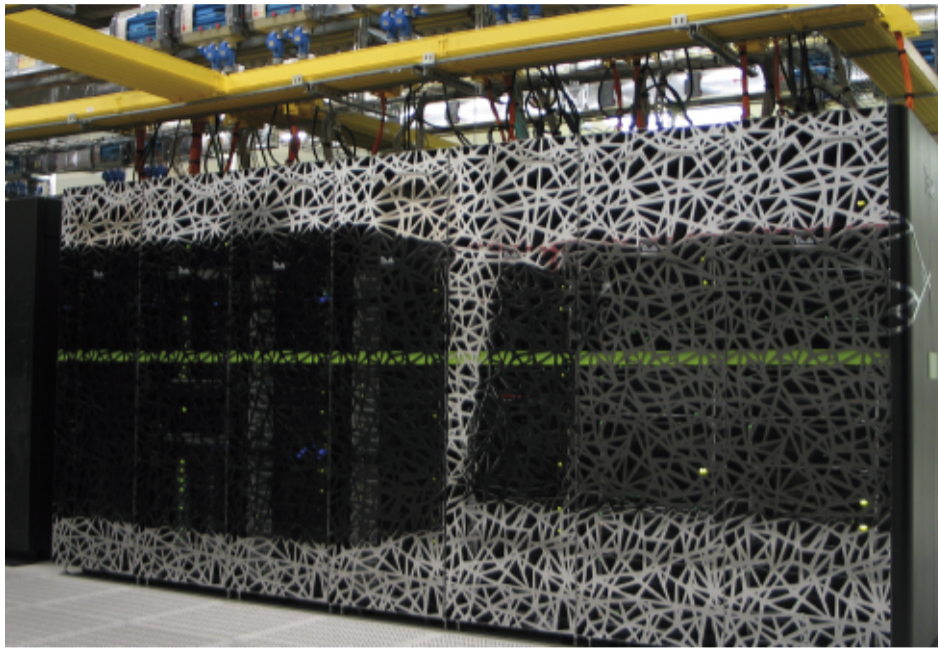
ODISSEI biedt strikt beveiligde omgeving

Veilig hoog-dimensionale en CBS-data combineren

Steeds meer grote cohort-onderzoeken verzamelen gegevens die hoog-dimensionaal zijn, zoals MRI-data of hartslag. ODISSEI Data Facility maakt het mogelijk om dergelijke data veilig te analyseren en te koppelen met gegevens van het CBS.

Marika de Bruijne

Aan het woord is VU-hoogleraar en oprichter van het Nederlands Tweelingen Register, Dorret Boomsma. “Er zijn werkelijk talloze onderzoeksvragen die met de nieuwe ODISSEI Data Facility beantwoord kunnen worden. Bijvoorbeeld: er is een oude discussie of het wonen in een stedelijke omgeving een risico is om schizofrenie te ontwikkelen. Er is een duidelijk verband tussen deze twee variabelen maar de oorzaak voor het verband is niet eenduidig. We hebben gegevens over woonplaats en verhuisgeschiedenis kunnen combineren met polygenetische scores en vonden een relatie tussen genetische kwetsbaarheid voor



De Nationale supercomputer Cartesius van SURFsara is het grootste systeem in Nederland op het gebied van high-performance computing en is vooral gewild vanwege de combinatie van zeer snelle processors, een zeer snel intern netwerk, veel geheugenruimte en de mogelijkheid grote datasets te verwerken. credits SURFsara

‘Dit jaar gaan een aantal projecten ervaring opdoen met deze dienst’

schizofrenie en stedelijkheid. Echter, we zien ook aanwijzingen dat niet zozeer de stedelijke omgeving risicoverhogend is, maar dat een kwetsbaarheid voor schizofrenie de kans op verhuizen naar een stad verhoogt.”

Onderzoekers konden gekoppelde data al analyseren in de omgeving van het CBS, maar dat gold niet voor hoog-dimensionale data. Hoe is dit nu mogelijk gemaakt?

“Er is een strikt beveiligde omgeving gecreeerd bij SURFsara die een extensie is van de beveiligde CBS-omgeving. Hoog-dimensionale data en CBS-gegevens, zoals verhuisgeschiedenis of opleidingsniveau, kunnen hier worden gecombineerd en geanalyseerd met geavanceerde software. De twee datasets ver-

laten als het ware hun eigen omgeving niet, ze worden alleen tijdelijk samengebracht in deze high performance computing omgeving, die op dat moment afgesloten is van andere systemen en gebruikers.”

Zijn de veiligheid en vertrouwelijkheid van de data gegarandeerd?

Zowel voor de CBS-gegevens als voor de privacygevoelige biologische gegevens staat veiligheid voorop. Boomsma: “Bij ons is niet bekend dat er elders een omgeving is gecreeerd waarin dergelijke koppeling mogelijk is, in de vorm waarin de veiligheid en vertrouwelijkheid van de data tijdens het hele proces volledig gegarandeerd blijven.”

Er is onlangs een geslaagde technische

pilot geweest. Welke uitdagingen kwamen jullie tegen?

“Dat varieerde van het opzetten van de beveiligde verbinding tussen CBS en SURFsara tot het juridisch vastleggen van alle verantwoordelijkheden in een samenwerkings- en verwerkersovereenkomst en de procedures rond het veilig koppelbaar maken van de databestanden.”

Voor wie is de faciliteit beschikbaar?

“Dit jaar laten we een handvol onderzoeksprojecten ervaring opdoen met de gerealiseerde faciliteit; na eventuele aanpassingen na deze pilot kunnen ook onderzoekers binnen het ODISSEI-consortium met de faciliteit werken.”

odissei-data.nl

ODISSEI

Het ODISSEI-consortium is een unieke samenwerking tussen ruim 20 cohorten en bijna 30 onderzoeksinstituten uit de sociale en economische wetenschappen.

DataTag-tool controleert privacygevoelige informatie

Beslisboom bepaalt privacyniveau dataset

De gevolgen van de nieuwe privacywetgeving zijn enorm als het gaat om het delen van persoonsgebonden data. Een data tag-benadering kan uitkomst bieden.

Peter Doorn

DANS heeft met het project GDPR DataTags een tool ontwikkeld om gevoelige persoonlijke data te kunnen classificeren volgens de nieuwe Algemene Verordening Geveensbescherming (in het Engels GDPR: General Data Protection Regulation). Hiermee is het mogelijk om data te controleren op privacygevoelige informatie.

Zes toegangsniveaus

Het project GDPR DataTags is geïnspireerd op de aanpak van data-

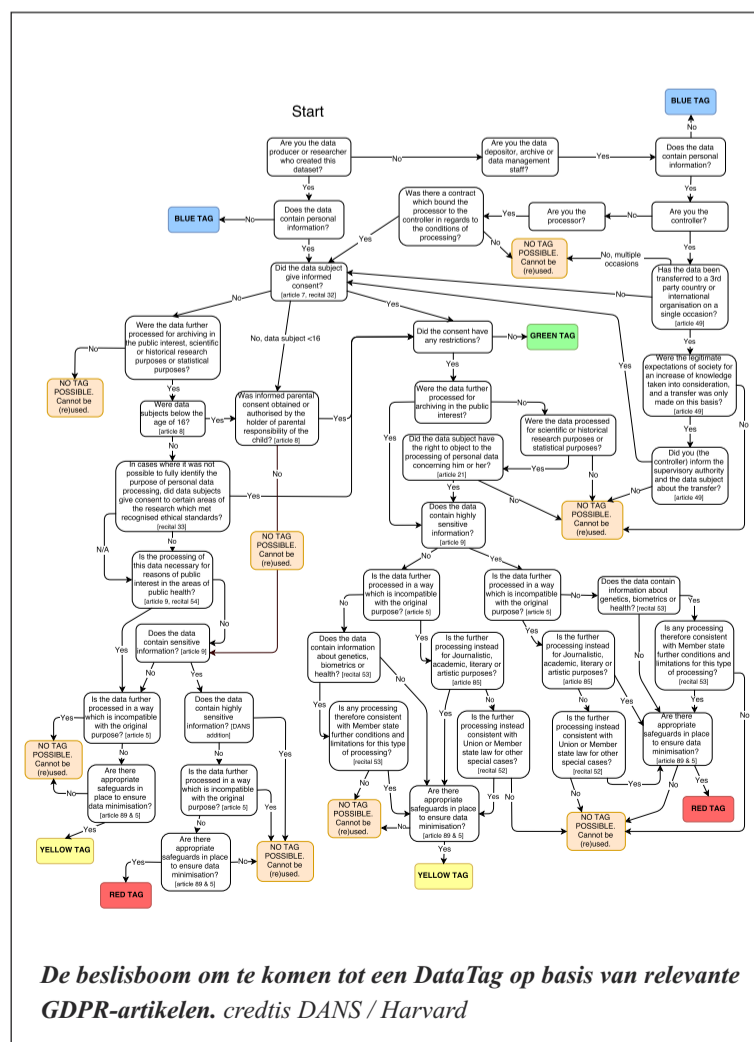
tags.org, ontwikkeld door Harvard University. Deze aanpak gaat uit van zes toegangsniveaus of tags, waarbij de deponeerder via het invullen van een vragenlijst aangeeft hoe anderen de data mogen gebruiken en hoe de gegevensbescherming moet worden geregeld. DANS heeft, in het kader van de ontwikkeling van de EUDAT Collaborative Data Infrastructure, de applicatie aangepast aan de Europese situatie.

Prototype

De eerste stap in het project was het analyseren en identificeren van de meest relevante artikelen van de GPDR voor Europa. Vervolgens werd dit vertaald in vragen voor een online vragenlijst-tool. Er is een prototype van de tool gemaakt, die

is besproken met juridische experts op het gebied van onderzoeksdata. De vragenlijst volgt een beslisboom, waarmee wordt vastgesteld welk niveau van bescherming voor een dataset is vereist, indien de dataset informatie over personen bevat. De tag die uit de beslisboom rolt, kan aan de dataset gekoppeld worden. Hiermee wordt in één oogopslag duidelijk, wat het privacyniveau van de gedeponeerde dataset is.

Het prototype van de vragenlijst-tool (aan de beslisboom en daaraan verbonden tags wordt nog verder gewerkt) is online beschikbaar via <https://goo.gl/cBgcMj>. De beslisboom is als .pdf beschikbaar via <https://goo.gl/fveAZ7>, en het volledige artikel via <https://goo.gl/HPn9Fq>.



SINDS KORT BESCHIKBAAR

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

CentERdata

• TIU Verkiezingsonderzoek

Voor de Tweede Kamerverkiezingen van 15 maart 2017 heeft CentERdata, met financiering van Data Science Center Tilburg, het stemgedrag van de Nederlandse bevolking onderzocht. Het ging om een wetenschappelijk experiment waarbij een alternatieve methode van peilen werd onderzocht. In totaal kregen 3.500 mensen 8 weken lang elke week een vragenlijst voorgelegd. Panelleden werd gevraagd om per partij aan te geven hoe groot de kans was dat ze op die partij zouden gaan stemmen. Bij de berekening van de voorspelde zetelverdeling werd ook rekening gehouden met kiezers die nog niet zeker wisten of ze zouden gaan stemmen. Het databestand Tilburg Election Study 2017 is beschikbaar via LISS Data Archive.

lissdata.nl

Ook sinds kort beschikbaar:

Studies LISS panel

- Bellemare, C.; Kroger, S.; Sarr, I., november 2016, Decision 2016
- Bresser, J. de; Soest, A. van, januari 2017 t/m april 2017, Tilburg Election Study 2017
- Buser, T.; Oosterbeek, H., maart 2017, Willingness to compete
- CentERdata, november 2017 – december 2017, Health - Wave 10
- CentERdata, oktober 2017 – november 2017, Social Integration and Leisure - Wave 10
- CentERdata, augustus 2017 – september 2017, Religion and Ethnicity - Wave 10
- CentERdata, september 2016 - oktober 2016, Family and Household - Wave 9
- CentERdata, januari 2016 t/m december 2016, Initial Questionnaire - 2016
- Elshout, M.; Giesen, R. van, april 2017, VWS Behavioral Choices of the Insured
- Fouarge, D., februari 2017, On lifelong learning 2017
- Knoef, M., december 2014, Expectations about Old-Age Income
- Regt, S. de, maart 2017, Commemorating in heterogeneous societies



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



Verkiezingsbord Utrecht foto ANP

DANS

• Data-publicatie interlinking

NARCIS, dé nationale portal voor wie informatie zoekt over wetenschappers en hun werk, geeft nu ook een overzicht van de onderzoeksdata in de data repositories van de Radboud Universiteit en de Universiteit Utrecht. Zo is samen met de Radboud Universiteit gewerkt aan het linken van de onderzoeksdata met relevante publicaties, om de data veel meer in context te tonen. Een goed voorbeeld hiervan is de dataset Migranten als Mediators (narcis.nl/dataset/RecordID/oai%3ARepository.ubn.ru.nl%3A2066%2F173004) waar vanuit de data wordt verwezen naar een proefschrift en andere boekdelen. NARCIS legt vervolgens zelf de link vanaf de publicatie naar de data. Van de Universiteit Utrecht zijn nu veertien open datasets in NARCIS terug te vinden. Datasets zijn voorzien van ORCID's en ISNI's (oftewel auteurs-identifiers) en ook informatie over de financiering van het onderzoek is aanwezig. Heeft u vragen over de datasets in NARCIS of wilt u ook dat uw repository met datasets (of publicaties) wordt geharvest, dan kunt u terecht bij Chris Baars, functioneel beheerder/informatiekundige bij DANS.

narcis.nl

Ook sinds kort beschikbaar:

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

- Brinkman, dr. A.B. (Radboud University); Stunnenberg, prof. dr. ir. H.G. (Radboud

University) (2018): Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation. DANS.

<https://doi.org/10.17026/dans-276-sda6>

- Centraal Bureau voor de Statistiek - CBS (2016): Enquête Beroepsbevolking - EBB - 2016. DANS. <https://doi.org/10.17026/dans-26j-x8wp>

- Centraal Bureau voor de Statistiek (CBS) (2016): Gezondheidsenquête 2016 - GECON 2016. DANS. <https://doi.org/10.17026/dans-xxa-e3m7>

- Centraal Bureau voor de Statistiek (CBS) (2015): Consumentenconjunctuuronderzoek - CCO 2015. DANS.

<https://doi.org/10.17026/dans-26s-v6xd>

- Centraal Bureau voor de Statistiek (CBS) (2016): Consumentenconjunctuuronderzoek - CCO 2016. DANS.

<https://doi.org/10.17026/dans-24f-8a8s>

- Frankema, Prof. dr. E. (Wageningen University); Woltjer, P. (Wageningen University); Dalrymple-Smith, A. (Wageningen University); Bulambo, L. (Wageningen University) (2017): An Introduction to the African Commodity Trade Database, 1730-2010. DANS.

<https://doi.org/10.17026/dans-xt9-fzkw>

- Gerritsen, S (Archeologie West-Friesland) (2013): Op de grens van de Stede. DANS.

<https://doi.org/10.17026/dans-2b8-4kf3>

- Kooi, drs. M. (BAAC bv) (2012): Gilze en Rijen (NB), Kerkstraat. DANS.

<https://doi.org/10.17026/dans-x9m-vey7>

- Weiss, Dr. D. (University of Oxford) (2018): A global map of travel time to cities. DANS.

<https://doi.org/10.17026/dans-ztx-2sd2>



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

AGENDA

6 – 8 juni • Amsterdam

DH Benelux Conferentie

Deze jaarlijkse conferentie biedt een platform voor samenwerking tussen betrokkenen bij Digital Humanities-projecten in de Benelux. 2018.dhbenelux.org

13 - 15 juni • Cambridge (VS)

Dataverse Community Meeting 2018

Een netwerkevent voor de Dataverse-community met als thema 'Dataverse in the Age of Data Science'. projects.iq.harvard.edu/dcm2018

13 - 15 juni • Edinburgh

IWSG 2018

Tiende internationale workshop over Science Gateways. sites.google.com/and.edu/iwsg2018

26 - 29 juni • Mexico-Stad

Digital Humanities

Het thema van deze Digital Humanities-conferentie is 'Bridges'. dh2018.adho.org

4 - 6 juli • Lille

LIBER Annual Conference 2018

Dit jaar is het thema 'Research libraries as an Open Science hub: from strategy to action'. libereurope.eu/events/liber-2018-lille-annual-conference

9 - 14 juli • Toulouse

EuroScience Open Forum 2018

Het thema van ESOF 2018 is 'Sharing Science: Towards New Horizons'. esof.eu

5 - 6 september • Brussel

OpenUP Conference

Het thema is 'Opening Up the Research Life Cycle: Innovative Methods for Open Science'. <http://openup-h2020.eu/openup-final-conference>

12 - 14 september • Leiden

STI 2018

Het thema van de conferentie is 'Science, Technology and Innovation indicators in transition'. <http://sti2018.cwts.nl>

17 - 19 september • Wenen

COASP 2018

Jaarlijkse conferentie over Open Access Scholarly Publishing. oaspa.org/conference

24 - 27 sept • Boston & Cambridge

iPRES 2018

Conferentie over digitale preservatie. ipres2018.org

13,8 miljoen voor CLARIAH PLUS

NWO heeft in het kader van de Nationale Roadmap voor Grootschalige Wetenschappelijke Infrastructuur €13,8 miljoen toegekend aan het project CLARIAH PLUS. Dit project is een voortzetting van het CLARIAH CORE-project, waarin de afgelopen vier jaar de basis is gelegd voor een digitale infrastructuur voor de geesteswetenschappen. CLARIAH-CORE focuste speciaal op taalkunde, sociaal-economische geschiedenis en mediastudies. In CLARIAH-PLUS worden daar disciplines aan toegevoegd die zich bezighouden met tekstinhoudelijke analyses, zoals letterkunde, geschiedenis, filosofie en theologie. Namens een nationaal consortium van universiteiten en geesteswetenschappelijke instituten nam



€13,8 miljoen voor CLARIAH PLUS. Credits NWO

de Principal Investigator Lex Heerma van Voss, directeur van het Huygens ING, de toekenning afgelopen april in ontvangst uit handen van minister Ingrid van Engelshoven van OC&W. (ER) clariah.nl

Overzichtsboek CLARIN verschenen

Onlangs verscheen bij uitgeverij Ubiquity Press (Londen) 'CLARIN in the low countries'.

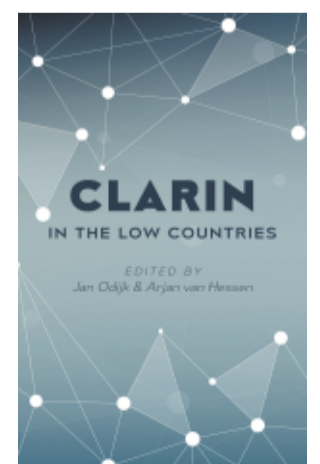
Dit boek geeft een overzicht van de resultaten van de activiteiten van CLARIN in Nederland en Vlaanderen. In Nederland werkten taalkundigen en technici tussen 2009 en 2015 aan de onderzoeksinfrastructuur voor onderzoekers in de geestes- en sociale wetenschappen die werken met talige data.

Vlaanderen droeg gedu-

rende deze periode ook aan verschillende subprojecten bij.

In meer dan dertig hoofdstukken gaan de betrokkenen uitgebreid in op de technische implementatie van de infrastructuur en de verschillende data, tools en services die via de infrastructuur beschikbaar zijn voor verschillende disciplines. Interessant voor zowel studenten, onderzoekers als ontwikkelaars. (ER)

<https://doi.org/10.5334/bbi>



CLARIN in the low countries. Credits Ubiquity Press

Voordat CLARIAH PLUS van start gaat, zal begin 2019 eerst het CLARIAH CORE-project worden afgerond. Hoe staat het ervoor? En wat vinden potentiële gebruikers ervan?

Erica Renckens

Sinds 2015 werkt een consortium van Nederlandse universiteiten en geesteswetenschappelijke instituten in CLARIAH CORE aan een digitale infrastructuur voor de geesteswetenschappen. Het doel is om geesteswetenschappers duurzaam toegang te verlenen tot grote dataverzamelingen en applicaties om deze data mee te verwerken. E-Data & Research vroeg drie potentiële gebruikers uit verschillende vakgebieden naar hun ervaringen en verwachtingen.

Onlangs heeft NWO het consortium €13,8 miljoen toegekend voor het project CLARIAH PLUS, waarin de infrastructuur verder zal worden uitgebreid.

Jack Hoeksema

“Het gaat steeds beter”

Samen met prof. Kees de Glopper onderzoekt Jack Hoeksema aan de Rijksuniversiteit Groningen de syntactische schrijfstijl van jongeren op basis van een verzameling opstellen. “We kijken naar allerlei maten van complexiteit: welke aspecten hebben scholieren al onder de knie en welke niet?” Daarbij maakt hij gebruik van de automatische ontleder Alpino op de site PaQu. “Omdat die site voor leken weinig uitdagend is, ontwikkel ik met de PaQumensen in SPOD (Syntactic Profiler of Dutch) een aantal voorgeprogrammeerde zoekopdrachten die CLARIAH-gebruikers kunnen



Volgens Hoeksema zijn de digital humanities niet meer weg te denken. Er zijn haken en ogen maar het gaat steeds beter.



Volgens Zandhuis besparen de tools om data om te zetten naar Linked Data enorm veel tijd, ook voor de volgende gebruiker. foto Martijn Smeets



Van der Molen voorziet dat er met de CLARIAH-infrastructuur veelzeggend inhoudelijk historisch onderzoek gedaan zal kunnen worden. foto Gerda van der Molen

CLARIAH bouwt verder aan duurzame toegang

Gebruikers infrastructuur delen hun ervaringen

inzetten voor heel veel doeleinden.” Volgens Hoeksema zijn de digital humanities niet meer weg te denken. “Natuurlijk zijn er haken en ogen, maar het gaat steeds beter.”

Ivo Zandhuis

“Tijdwinst door tools”

Onafhankelijk onderzoeker Ivo Zandhuis vergelijkt de introductie van digital humanities met energietransitie: “Ook daar zijn nog technische bezwaren, maar door eraan te beginnen, ontstaan er middelen en motivatie om oplossingen te ontwikkelen.” Zandhuis ziet vooral de mogelijkheden die de CLARIAH-infra-

structuur biedt. Samen met Richard Zijdeman van het IISG blaast hij momenteel een dataset uit EASY nieuw leven in. “De tools om data om te zetten naar Linked Data besparen enorm veel tijd, ook voor de volgende gebruiker. De mogelijkheid om zoekopdrachten uit te voeren op gecombineerde data levert natuurlijk ook veel tijdswinst op.”

Berrie van der Molen

“Combinatie van technieken”

Aan de Universiteit Utrecht doet Berrie van der Molen promotieonderzoek naar historische drugsdebatten in de Nederlandse media.

“Zodra de datavoorziening in de CLARIAH-infrastructuur op orde is, zal ik gebruikmaken van de Comparative Search-tool”, vertelt hij. “Dan kan ik zowel het krantenarchief van de KB als het radio- en televisie-archief van Beeld en Geluid onderzoeken met een combinatie van technieken voor distant en close reading.” Van der Molen voorziet dat er met de CLARIAH-infrastructuur veelzeggend inhoudelijk historisch onderzoek gedaan zal kunnen worden. “De ontwikkeling werpt voortdurend methodologische vragen op; de infrastructuur kan zich op basis daarvan doorontwikkelen.” clariah.nl

Eerste researcher-in-residence van 2018 aan de slag in de KB

Historische ontwikkeling van steden in kaart

Om de historische ontwikkeling van Nederlandse steden beter te begrijpen, verruimt stadsgeograaf Antoine Peris een paar maanden de technische universiteit voor de nationale bibliotheek.

Steven Claeysens

Jaarlijks nodigt de KB jonge, beloftevolle onderzoekers uit om in de bibliotheek met behulp van digitale technieken een vraagstuk naar keuze te lijf te gaan. Van de samenwerking profiteert zowel de onderzoeker als de bibliotheek: de onderzoeker krijgt hulp van programmeurs en collectiespecialisten, de KB leert van de wensen en methoden van de onderzoeker om zo collecties en dienstverlening te kunnen verbeteren. Bestaande of speciaal ontwikkelde tools die daarbij hun diensten bewijzen, worden opgenomen in het KB Research Lab zodat ook andere onderzoekers ze kunnen gebruiken en eventueel uitbreiden. Peris studeerde in Parijs en promo-

JONG TALENT

veert aan de TU Delft bij de sectie Urban en Regional Research in the Architecture and the Built Environment van de faculteit Bouwkunde. Hij is de eerste geograaf die bij de KB als researcher-in-residence aan de slag gaat:

Netwerken

“Ik ben stadsgeograaf en doe onderzoek naar stedelijke systemen. Ik ben geïnteresseerd in de netwerken die steden met elkaar verbinden en tot systemen organiseren. In mijn vakgebied wordt de centrale rol van informatie voor het proces van verstedelijking vaak benadrukt. Om te kunnen beslissen of je naar een andere plaats zal reizen of niet is kennis over de risico's en mogelijkheden vooraf immers noodzakelijk. Informatie is echter immaterieel en daarmee moeilijk te traceren.” Peris meent daar iets op gevonden te heb-

‘Voorheen was dit type onderzoek ondoenbaar’

ben. “Kranten vormen een interessante proxy. Vóór de opkomst van digitale media vormden ze de ruggraat van de informatieverstoring.”

Niet de eerste

“Mijn twee belangrijkste onderzoeksvragen zijn: welke steden werden vroeger het vaakst genoemd in de verschillende kranten in Nederland en welk beeld had men van die steden? Ik ben niet de eerste die



Peris: “De huidige technieken en online krantenarchieven zoals Delpher zorgen voor onderzoek op grotere schaal en langere periode.” foto Christel Swartouw-Hofmeijer

zulke vragen stelt of naar die data kijkt, maar voorheen was het ondoenbaar om dit type onderzoek op grote schaal uit te voeren. De huidige computermogelijkheden en de grote historische krantenarchieven, zoals Delpher, brengen daar verandering in. We kunnen nu onderzoek

doen op een veel grotere schaal en hebben de beschikking over data over een lange periode. Voor onderzoek naar steden is dat cruciaal, want steden worden over het algemeen niet op één dag gebouwd.” www.kb.nl/nieuws/2017/kb-researchers-in-residence-2018-bekend

Revolutionaire tools op DH Benelux

Complexe tekst? Eitje voor TAG!

Het Research & Development-team van het KNAW Humanities Cluster ontwikkelt digitale tools voor geesteswetenschappelijk onderzoek en presenteert dit op de DH Benelux van 6-8 juni.

Thijs van der Veen

Tekst kan worden gezien als een netwerk van informatie over inhoud, tekstvariatie en structuur. Het is lastig, zo niet onmogelijk, om al deze elementen (en de relaties ertussen) in kaart te brengen met behulp van bestaande tools zoals TEI-XML. Het R&D-team van het Humanities Cluster heeft hiervoor een nieuw model ontwikkeld, Text as Graph (TAG), dat gebruik maakt van een hypergraafstructuur. Teamleider en software ingenieur Ronald Haentjens Dekker: "Met TAG kunnen zeer complexe teksten gemakkelijk worden opgeslagen, bewerkt en bevestigd. TAG implementeert de tekstcollatielool HyperCollate, waarmee teksten met elkaar kunnen worden vergeleken op zeer gedetailleerd niveau (zowel variatie binnen één tekst als variatie tussen verschillende tekstversies). Deze manier van onderzoeken resulteert in een zeer rijke output.

DH Benelux 2018

De 5e Digital Humanities Benelux Conference vindt plaats van 6-8 juni 2018 op het Internationaal Instituut voor Sociale Geschiedenis (IISG) in Amsterdam.

Tijdens de conferentie komen interdisciplinaire DH-onderzoekers uit België, Nederland en Luxemburg samen om hun onderzoeksresultaten, tools en projecten met elkaar te delen.

Het KNAW Humanities Cluster is dit jaar de organisator.



Fysiek 3D-model van Text as Graph (TAG). Dit specifieke model van TAG laat zien hoe stukken tekst uit *The Hunting Of The Snark* van Lewis Carroll tegelijkertijd onderdeel uitmaken van elementen als bijvoorbeeld een excerpt, een page, een voice en een stanza.

Credits R&D-team KNAW Humanities Cluster

Het menselijk brein kan daardoor overweldigd worden, visualisatie van de output kan uitkomst bieden. Tijdens DH Benelux gaat het R&D-team in op de uitdagingen die dat met zich meebrengt in de lezing TMI? Visualisation as Research Instrument for Computational Philology.

Digitaal onderzoek

Een tweede lezing van het R&D-team tijdens de DH Benelux-conferentie is qua onderwerp minder technisch maar voor de toekomst van digitaal onderzoek minstens zo relevant. Marijn Koolen (R&D), Jasmijn van Gorp (Universiteit Utrecht) en Jacco van Ossenbruggen (CWI & VU) bespreken naar aanleiding van hun paper 'Lessons Learned from a Digital Tool Criticism Workshop' hoe wetenschappers in hun onderzoek het gebruik van digitale tools kunnen verbeteren.

Reflectie is het toverwoord en dan het liefst in een groep. Nu is samenwerking in digitaal geesteswetenschappelijk onderzoek sowieso aan te raden, omdat daar immers veel verschillende vaardigheden bij komen kijken. Traditionele geesteswetenschappers opereren meestal alleen, maar door in een groep te reflecteren op de denkstapen en keuzes die eenieder maakt bij het gebruik van digitale tools, kom je tot nieuwe inzichten en beter onderbouwde afwegingen in hoe, wanneer en waarom bij het inzetten van tools. Wellicht voor de hand liggende vragen, maar in de praktijk blijkt dat onderzoekers dit bewust of onbewust nog te weinig doen. Daarom organiseren Jacco, Jasmijn en Marijn tijdens DH Benelux 2018 hier een workshop over.

<http://2018.dhbenelux.org>

GELEZEN

The Open Science Training Handbook

Marjan Grootveld

Het "Open Science Training Handbook" is het resultaat van een week samen schrijven – en van de ervaring van veertien auteurs, die betrokken zijn bij ondermeer FOSTER, CESSDA en 4TU.ResearchData. Het handboek is bestemd voor trainers en frist in hoofdstuk twee de kern op van learning, teaching en training. Hoofdstuk één legt de inhoudelijke basis met informatie over onder andere open data, open software, citizen science en reproducible research. Het beschrijft systematisch de voornaamste kennis en vaardigheden, leerdoelen en eventuele obstakels en misvattingen. De laatste hoofdstukken zijn uitgesproken praktisch: na een checklist voor het organiseren van trainingsbijeenkomsten volgt een hoofdstuk met ruim twintig opdrachten. Hieronder zijn zowel ijsbrekers van enkele minuten als discussieopdrachten van een half uur, allemaal toegesneden op Open Science en met tips voor de uitvoering, waaronder "This exercise should be fun". Bij toepassing in een OpenAIRE-workshop bleek dit helemaal te kloppen. Handboek en literatuurlijst zijn online beschikbaar.

book.fosteropenscience.eu



COLUMN

Een mooie taak voor het netwerk van bibliotheken

Het gaat goed met de kranten. De lezersaantallen stijgen, vooral online. Die ontwikkeling is voor een belangrijk deel te danken aan nepnieuws. Door alle commotie rond nepnieuws en de schandalen rond Cambridge Analytica en Facebook zijn steeds meer mensen zich ervan bewust hoezeer zij gemanipuleerd worden. En dus keren zij terug naar oude, vertrouwde nieuwsbronnen: dagbladen en de website van bijvoorbeeld de NOS.

Voor mij geldt hetzelfde: ik vertrouw minder op de nieuwsfeeds van Google en begin en eindig mijn dag meestal op de NOS-site. Met Facebook heb ik nooit veel gedaan en vanwege Cambridge Analytica heb ik mijn account verwijderd – je moet ergens een grens trekken.

We zien dus een teruggaande beweging naar bronnen die hun betrouwbaarheid in het verleden hebben bewezen.

Ik denk dat we een vergelijkbare ontwikkeling zullen zien bij grote bibliotheken. Lang golden die als de kennisbewaarplaatsen bij uit-

stek. Een goede bibliotheek beschikte over de nieuwste handboeken, de beste en meest relevante tijdschriften en de bibliothecaris of balie medewerker (m/v) kon desgevraagd optreden als gids.

De afgelopen vijftien jaar zijn bibliotheken op grote schaal gaan digitaliseren, al dan niet in samenwerking met Google. Ik ga zelf alleen nog naar een bibliotheek als ik boeken moet inzien die nog niet online digitaal beschikbaar zijn.

Toch verwacht ik dat de komende jaren meer mensen lid zullen worden van een bibliotheek. Althans: als die zich, net als kranten, beter zullen positioneren als bronnen

van betrouwbare kennis. Wat ik nu zie is dit: een bibliotheek als de Koninklijke Bibliotheek (KB) zet zich onder meer in om grote datasets beter te kunnen analyseren met nieuwe technologieën en kunstmatige intelligentie. Dat is lovenswaardig, maar volgens mij zou het tegelijkertijd nuttig zijn om 'gewone' gebruikers – scholieren, studenten,



foto Leo van Velzen

ouderen – te leren hoe zij bijvoorbeeld nepnieuws kunnen onderscheiden van echt nieuws. En hoe je kunt vaststellen of informatie op een site betrouwbaar is of niet. Gewoon slimmer en creatiever zoeken op internet dus.

Dat is natuurlijk geen taak van de KB alleen. Er bestaat een netwerk van openbare bibliotheken. De hele infrastructuur ligt er dus om bijvoorbeeld op lokaal niveau workshops aan te bieden hoe je optimaal gebruik kunt maken van internet en van gedigitaliseerde bronnen van onder meer bibliotheken. Je zou hier ook filmpjes over kunnen maken (er zijn er al een paar), maar in mijn ervaring werken workshops beter, want die zijn interactiever. Overigens zouden dergelijke workshops volgens mij niet alleen zinvol zijn voor scholieren, studenten en ouderen. Ik kom ook geregeld docenten en zelfs bibliothecarissen tegen die baat zouden kunnen hebben bij zo'n workshop of bijscholing.

Internet is een uitdijend heelaal

waar je de prachtigste maar ook de onzinnigste informatie kunt vinden. Om niet te worden misleid door nepnieuws, keren mensen terug naar kranten en naar nieuwssites die zij vertrouwen. "De bibliotheek maakt je slimmer, vaardiger en creatiever", luidt de missie van sommige bibliotheken, waaronder de KB. Ik vind dat een mooie missie, die zou kunnen worden verwezenlijkt als het netwerk van bibliotheken cursussen zou gaan aanbieden over hoe wij optimaal gebruik kunnen maken van internet. Voor jong en oud, voor beginners en gevorderden.

Ik vermoed dat daar meer belangstelling voor zal bestaan dan voor slimme tools met de nieuwste technologieën, hoewel ook die moeten blijven worden ontwikkeld.

Ewoud Sanders

Taalhistoricus en journalist.

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