

DANS en uitgeverij Brill lanceren Research Data Journal

Nieuw online tijdschrift voor onderzoeksdata

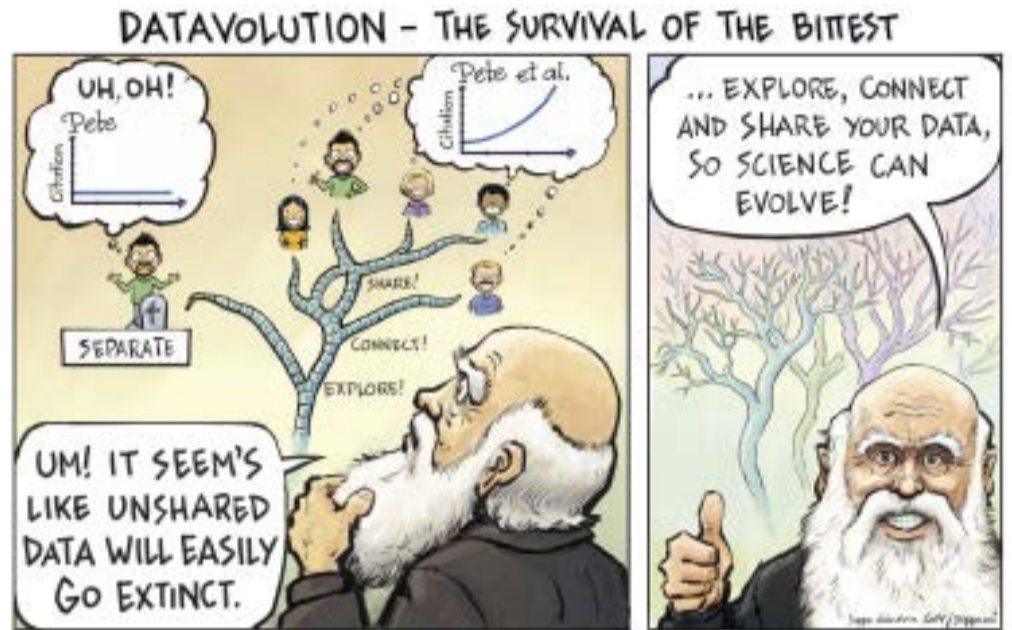
Onlangs is het eerste nummer van het *Research Data Journal for the Humanities and Social Sciences* gelanceerd: een nieuw, peer reviewed, online only en open access data journal.

Leen Breure

Het Research Data Journal bevat korte publicaties (data papers) waarin onderzoekers hun dataset beschrijven: de context van hun onderzoek, de probleemstelling en gebruikte methoden. Daarop volgt een globaal profiel van de dataset, bijvoorbeeld in termen van algemene karakteristieken of opmerkelijke resultaten. Conclusies zoals in een gewoon wetenschappelijk artikel zijn niet vereist, maar er is wel ruimte voor afsluitende opmerkingen. Lezers kunnen middels een commentaarveld reageren op de inhoud. De data dienen gedeponerd te zijn in een 'trusted repository', zoals het online archiveringsysteem EASY van DANS.

Voor en door onderzoekers

DANS heeft in het verleden samen met SURF actief bijgedragen aan de bekendheid van 'verrijkte publicaties': digitale artikelen over wetenschap-



illustratie www.seppo.net

pelijke onderzoek die nauw geïntegreerd zijn met de data van het onderzoek. Dit journal bouwt voort op deze traditie. Het tijdschrift biedt onderzoekers een online platform om eigen onderzoek onder de aandacht van geïnteresseerden te brengen. Bovendien geeft het lezers een snelle blik op onlangs afgerond onderzoek, bevat het (interactieve) datavisualisaties en zo veel mogelijk directe toegang (via hyperlinks) tot de data zelf (bijvoorbeeld in de vorm van een gelinkt on-

line spreadsheet dat een sample van de dataset bevat).

Het Research Data Journal verschijnt tweemaal per jaar. Onderzoekers kunnen zelf een publicatie in het Research Data Journal verzorgen. Neem voor de mogelijkheden en eventuele ondersteuning bij en uitbesteding van technische onderdelen contact op met DANS.

brill.com/RDJ

Reference rot en content drift: serieuze problemen

Eeuwige webcitatie?

Refereren naar webbronnen in wetenschappelijke publicaties is populair, maar een derde van de hyperlinks werkt na een paar jaar niet meer. Een serieus probleem.

René van Horik

Wetenschappelijk onderzoek zonder bronvermelding is ondenkbaar. Voor het citeren van bestaande wetenschappelijke literatuur bestaan al jaar en dag robuuste oplossingen. Verwijzingen naar informatie 'ergens' op het web hebben echter last van reference rot: de website waarnaar verwezen wordt, bestaat niet meer of van content drift: de hyperlink verwijst niet meer naar de oorspronkelijke informatie.

Minder dan 50%

Het Hiberlink-project bestudeerde een corpus van 6.400 elektronische dissertaties met 46.000 hyperlinks uit de periode 2003 - 2010.

Eén van de conclusies: een derde van de hyperlinks werkt na een paar jaar niet meer. Ook bleek de kans dat de geciteerde online content opgenomen wordt in een data archief minder dan 50%.

Het Hiberlink-project adviseert drie

```
<a href="http://www.newyorker.com/magazine/2015/01/26/cobweb" data-versionurl="http://web.archive.org/web/20150219094636/http://www.newyorker.com/magazine/2015/01/26/cobweb" data-versiondate="2015-02-19T09:46:36">Cobweb Article</a>
```

Een robuust link bevat het webadres (URI), datum en tijd waarop de archiefkopie gemaakt is en het originele webadres. Richtlijnen voor het citeren van referenties op het web dienen de onderdelen van deze robuust link te bevatten. Een aanzet voor de syntax staat op: robustlinks.mementoweb.org/tools. Credits insights.uksg.org/articles/10.1629/uksg.237

momenten waarop reference rot kan worden voorkomen.

Geschikte momenten

Het meest geschikte moment om reference rot te voorkomen, is tijdens de voorbereidingsfase. Tijdens het aanbrenge van webcitatie in de publicatie, kan referentiemanagementsoftware snapshots maken van de webpagina waarnaar gerefereerd wordt. Zo heeft het projectteam een plugin gemaakt voor het opensource programma Zotero. Deze plugin maakt een duurzame referentielink naar de bron waarvan door een webarchief een archiefkopie wordt gemaakt. Ingrijpen tijdens de indieningsfase, als de publicatie wordt uitgegeven of gepubliceerd, is ook mogelijk, hoewel er een risico is dat

geciteerde webbronnen al corrupt zijn geworden. Hiberlink heeft een prototype oplossing ontwikkeld voor uitgever op basis van het Open Journal System (OJS). De OJS-plugin maakt automatisch snapshots van de opgenomen referenties. Dan is het nog mogelijk om in te grijpen tijdens de post-publicatiefase, wanneer de publicatie wordt geraadpleegd en de citaties door de lezer worden opgevraagd.

De lezer van het artikel kan de geciteerde webbronnen vinden in webarchieven. Voor dit terugreizen in de tijd is de Memento-plugin beschikbaar. Voorwaarde is wel dat de geciteerde webbronnen opgenomen zijn in een webarchief, waarvan er al een aantal bestaan.

hiberlink.org

Data platform Mendeley duurzaam in archief DANS

Onlangs is Mendeley Data gelanceerd. De door onderzoekers op dit platform gepubliceerde onderzoeksdata worden bij DANS voor de lange termijn opgeslagen. Met deze samenwerking kan DANS nog meer digitale onderzoeksgegevens langdurig bewaren.

De dienst biedt mogelijkheden om data en publicaties beter aan elkaar te linken, onder meer door het gebruik van DOI's.

Duurzame toegang

Peter Doorn, directeur van DANS: "DANS bevordert duurzame toegang tot digitale onderzoeksgegevens. Om het delen en hergebruiken van data, ook op de lange termijn, te bevorderen, deponeren onderzoekers hun data in het DANS-archief. Daarnaast verzorgen we lange termijn opslagdiensten voor verschillende publieke en private partijen. Dit houdt in dat de data van deze organisaties in het DANS-archief terechtkomen. De metadata stellen we open access via EASY beschikbaar voor alle onderzoekers."

data.mendeley.com



E-DATA & RESEARCH

Jaargang 10 | nummer 1

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 RIVM.

INHOUD

2 Verslagen van events in Gehoord en bijgewoond

3 Wikipedia-gemeenschap wint Erasmusprijs 2015

4 Deze datasets zijn ook sinds kort beschikbaar



5 Charles van den Heuvel over Digital Humanities

6 Online virtuele collectie inspiratiebron voor velen

6 Over een onderzoek naar het big data-landschap

7 Veilig online datagebruik binnen ieders handbereik

8 Leer digitaal: Massive Open Online Courses

8 Ewoud over onderzoeken met één druk op de knop



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

Nieuwe dataverzameling over nationale mobiliteit

Marion Wittenberg

Op 15 september vond bij NWO het eerste MPN-symposium plaats. MPN staat voor Mobiliteitspanel Nederland, een gezamenlijk initiatief van het Kennisinstituut voor Mobiliteitsbeleid (KIM), onderzoeksbureau Goudappel Coffeng en de Universiteit Twente. Zowel beleidsambtenaren als onderzoekers op het gebied van mobiliteit vormden het publiek van dit druk bezochte symposium. Sascha Hoogendoorn-Lanser (projectleider van het MPN) trapte af met uitleg over de opzet van deze longitudinale dataverzameling. Behalve door middel van vragenlijsten, wordt ieder najaar met dagboekjes informatie verzameld over het reisgedrag van 4.000 respondenten uit 2.000 huishoudens. Tijdens de pauzes stond dit online dagboekje op twee laptops klaar, zodat de aanwezigen zelf konden ervaren hoe het invullen hiervan in z'n werk gaat. Karst Geurts (Universiteit Twente) vergeleek het MPN met ander mobiliteitsonderzoek. Het unieke van het MPN is dat het inzicht geeft in de dynamiek en variatie in het mobiliteitsgedrag van personen en huishoudens, denk aan het effect van veranderingen in de gezinssituatie (zoals kinderen krijgen, scheiden, van baan veranderen). Ook meet het MPN ICT-gebruik en attitudes ten opzichte van mobiliteit. Nina Schaap (KIM) ging in op de mogelijkheden om meer- en maatwerk te laten uitvoeren bij de dataverzameling. Verder werden resultaten van de eerste analyses op de MPN-data gepresenteerd.

Marie-José Olde Kalter (Goudappel Coffeng) presenteerde haar onder-



In veel gezinnen geldt dat als de ene partner de auto heeft, de ander uit moet wijken naar andere middelen van vervoer. Het MPN geeft inzicht in deze huishoudinteracties. foto Shutterstock

zoek naar huishoudinteracties. Hoe vindt er afstemming plaats over het gebruik van de auto en hoe beïnvloedt dit de vervoerskeuze van de verschillende gezinsleden? Eline Schepers (KIM) presenteerde haar onderzoek naar de impact van ICT-gebruik op vrijetijdsmobiliteit. Tot slot liet Eric Balster (CentERdata) zien hoe de data van het MPN ontsloten worden via Survey Data Nederland.

Tijdens de discussies was het interessant om te zien hoe beleids- en onderzoeksvragen elkaar afwisselden en aanvulden. De symposiumpresentaties en achtergrondinformatie over het MPN zijn te vinden op kimnet.nl/mobiliteitspanel-nederland. Voor wie zelf met de MPN-data aan de slag wil, de eerste wave is online beschikbaar.

mpndata.nl

Workshop: 'Beyond methods of mining'

Steven Claeysens

Het Utrechtse HERA-project Asymenc (Asymmetrical Encounters) kamt omvangrijke Europese gedigitaliseerde collecties, zoals het Nederlandse Delpher-corpus, uit op zoek naar culturele aspecten van Europese identiteit. Centraal staat de vraag in welke mate Europese culturen een referentiecultuur voor elkaar vormden in de jaren 1815-1992 – en of zulke transnationale referentieculturen bijdroegen aan de Europese integratie. De (bij voorkeur digitale) technieken die in het project worden toegepast, waren half september het onderwerp van een tweedaagse workshop. Buitenlandse historici presenteerden hun onderzoeken op basis van gedigitaliseerde historische bronnen en de

inzet van Text and Data Mining-technieken. Jane Winters van de University of London stipte een aantal kenmerken van hedendaagse geschiedschrijving aan met een keynote over 'digitale geschiedenis'. Ze wees o.a. op de mogelijkheden van 'big data', van visualisaties en op het groeiend belang van samenwerking, maar ook op het risico dat we de grote hoeveelheid audiovisueel materiaal uit het oog verliezen als we ons beperken tot tekstuele data. Digital Humanities zou bovendien meer mensen en hun verhalen voor het voetlicht moeten brengen, meer 'history from below'. Tijdens de tweede dag deelden enkele Utrechtse onderzoekers hun ervaringen uit de onderzoekspraktijk. Hermione Giffard trapte vervolgens een levendige slotdiscussie af. Ze hield de aanwezigen voor wat kan gebeuren met die historische onderzoekspraktijk, als onderzoekshandelingen aan de computer worden afgegeven: de kwaliteit van de primaire bronnen laat zich slecht wegen, de zoekinterfaces en andere tools ontwikkelen zich meer en meer tot 'black boxes', een binaire epistemologie dreigt de 'historische' epistemologie te verdrukken, etc. Een definitieve conclusie werd niet bereikt tijdens de brede discussie, maar het onverminderde belang van kritische omgang met bronnen werd benadrukt, terwijl een pleidooi voor uitbreiding naar software-, tool- en algoritme-critiek algemene instemming vond.

asymenc.wp.hum.uu.nl

DH2015: het politieke gezicht van DH

Steven Claeysens

Het jaarlijkse Global Digital Humanities-congres vormde met meer dan 250 presentaties en posters een indrukwekkende staalkaart van hedendaagse, digitale onderzoeksmethoden in de geesteswetenschappen. Het congres kende drie zeer verschillende keynotes.

De eerste dag was er een wat warrig verhaal van cultuurhistoricus Jeffrey T. Schnapp over 'little data' en 3D-technieken in de museale wereld. De tweede dag hield antropologe Genevieve Bell een zeer onderhoudende lezing over de menselijke fascinatie met én angst voor robots (via YouTube zijn optredens van haar te bekijken.) Op de laatste dag toonde hacker-historicus Tim Sherratt het politieke gezicht van de Digital Humanities in een beschouwing over openheid van informatie, het belang van herinneren en de macht van de hacker.

Toen tijdens de openingsceremonie een kleine stoet mannen het podium beklom, ontspon zich op Twitter wellicht de opvallendste discussie van het congres: heeft de Digital Humanities-gemeenschap een genderprobleem? Twitter meende in ieder geval van wel en dit standpunt werd de volgende dag ook krachtig uitgedragen door Deb Verhoeven, de voorzitter van de programma-commissie. Haar oproep om gedurende de komende dertig jaar de vermeende 80-20-genderverhouding in de Digital Humanities om te draaien, vond dan weer minder weerklank bij de aanwezigen. Die discussie zet zich volgend jaar, in Krakau, ongetwijfeld voort. dh2015.org

12 miljoen historische krantenpagina's op woordniveau doorzoekbaar

Europese erfgoeddata klaar voor gebruik

Europeana brengt gegevens over miljoenen gedigitaliseerde objecten bij elkaar, maar nog te weinig onderzoekers vinden de weg naar al die data. Daar wil Europeana Research verandering in brengen. Steven Claeysens

'Benadruk wel dat er nog veel werk aan de winkel is!' Het is de slotopmerking van Alastair Dunning, programmamanager bij Europeana en The European Library (TEL) en sinds dit jaar ook het gezicht van het Europeana Research-programma. De Schot vertelt enthousiast over zijn plan om de enorme hoeveelheid Europese erfgoeddata voor de wetenschap beter te ontsluiten. Zo wil Dunning data opdelen in han-



Alastair Dunning

foto Inge Angevaere

teerbare en logische datasets en zodanig beschrijven en documenteren dat onderzoekers er vrijer mee aan

de slag kunnen. Hij laat zich daarbij graag inspireren door Europeana Labs, dat data levert aan de creatieve industrie en scholen, maar vooral ook door de wensen en ideeën van onderzoekers zelf.

Aansluiten op wensen

'We onderzoeken nog hoe en waar we de data het best kunnen laten aansluiten op de reguliere datastroom van wetenschappers. Willen ze een API of downloads? Maken we gebruik van GitHub of Figshare? Of kunnen we de data beter toevoegen aan bestaande nationale of internationale onderzoeksinfrastructuren, waar dan verder op gebouwd kan worden?' Wie hier goede ideeën over heeft, mag altijd bij hem aankloppen.

De eerste dataset waar hij zijn pijlen op richt, is de omvangrijke collectie gedigitaliseerde historische kranten uit heel Europa die de afgelopen jaren tezamen gebracht is. 'Twintig miljoen pagina's, waarvan twaalf miljoen op woordniveau doorzoekbaar, horen als ruwe data de weg te vinden naar Europese universiteiten. Natuurlijk werpt het auteursrecht nog een lastige versperring op, maar een groot deel is Publiek Domein en dus redelijk eenvoudig te verspreiden. Daarnaast heeft Europeana ook een ongekende hoeveelheid afbeeldingen. Dat moet interessant zijn voor bepaalde onderzoekers.'

Vruchtbaar inzetten

Om beter voeling te hebben met het onderzoeksveld en te begrijpen hoe

al die data vruchtbaar ingezet kunnen worden, broedt Dunning ook op een plan om een klein onderzoeksprogramma te starten. Momenteel zet hij daartoe een 'editorial board' op. Het programma moet onderzoekers uit Europa de mogelijkheid bieden om voor korte of langere tijd onderzoek te verrichten op basis van al dat fraais.

GLAM-data

Maar wat hebben Nederlandse wetenschappers aan Europeana Research? 'Veel internationale GLAM-data, dat wil zeggen gedigitaliseerde objecten en bijbehorende metadata van Galleries, Libraries, Archives and Museums. Probeer maar.'

research.europeana.eu

COLOFON Uitgever: Stichting Uitgeverij E-data & Research Den Haag. Redactieadres: Postbus 93067, 2509 AB Den Haag, 070-3494450, edata@dans.knaw.nl, www.edata.nl.

Hoofd-/eindredacteur: Heidi Berkhout. Redactie: Marika de Bruijne, Steven Claeysens, René van Horik, Ilja Nieuwland, Rutger Nugteren, Erica Renckens, Marion Wittenberg. Redactiesecretariaat: Lucas Pasteuning.

Aan dit nummer werkten mee: Eric Balster, Leen Breure, Jos Damen, Robin Duinker, Ferry Gouwens, Alissa de Koning, Milo van de Pol, Eelco Rensink, Mariëtte Schipper-Heesters. Opmaak: Colette Sloots, Haarlem.

Productie: Amsterdam University Press. Druk: Ten Brink, Meppel. Webmaster: Sonja Duijkers. Oplage: 6.000 papier, 4.100 digitaal. ISSN: 1872-0374.

We hebben getracht alle belanghebbenden met betrekking tot het gebruikte beeldmateriaal te benaderen. Degenen die menen rechten te kunnen doen gelden, kunnen zich tot ons wenden. Toezending papieren en/of digitale versie is kosteloos aan relaties van de stakeholders en studenten in de alfa- en gamma-richtingen. Wilt u een artikel uit E-data overnemen? Neem dan contact op met de hoofdredacteur. Overname van artikelen geschiedt altijd onder vermelding van de bron E-data & Research en de naam van de auteur van het artikel.

Kruisbestuiving wetenschap en Wikipedia

Erasmusprijs 2015 voor Wikipedia-gemeenschap

De Erasmusprijs 2015 gaat naar Wikipedia als community, naar de tienduizenden vrijwilligers die helpen bij het verspreiden van kennis via de gelijknamige online encyclopedie.

Jos Damen (ASC Leiden)

Sinds 2010 werken Wikipedians bij kennisinstellingen in binnen- en buitenland. Bij archieven en bibliotheken, maar ook bij organisaties als de Royal Society of Chemistry, Cancer Research UK en het Smithsonian Institute liepen enkele maanden Wikipedians in Residence rond. Die probeerden een verbinding te leggen tussen de instellingen en Wikipedia. Daarvan hadden beide partijen nut. Wikipedia kreeg bijzondere afbeeldingen voor de online encyclopedie. De kennisinstellingen kregen een extra podium voor hun afbeeldingen, waardoor afbeeldingen soms honderden keren meer werden gebruikt dan in de eigen databases.

700.000 afbeeldingen

In Nederland waren organisaties als de Stichting Academisch Erfgoed, de Koninklijke Bibliotheek, het Nationaal Archief, het Tropenmuseum, Nederlands Instituut voor Beeld en Geluid, de Rijksdienst voor het Cultureel Erfgoed, het Rijksmuseum, universiteiten van Maastricht en Utrecht en instellingen als het NIOD, Vredespaleis, EYE, KITLV en het Afrika-Studiecentrum betrokken. Soms participeerden ze als zelfstandige instellingen, soms in groepsverband met wel 12 instellingen, zoals de Werkgroep Speciale Wetenschappelijke Bibliotheken. In totaal doneerden Nederlandse instellingen al ruim 700.000 afbeeldingen aan Wikimedia Commons en verbeterden ze de kwaliteit van de artikelen in de online encyclopedie.



Festival Kameroen 2014
foto Wouter van Beek / African Studies Centre Leiden

Voor wetenschappers kan Wikipedia op verschillende manieren van belang zijn. De afbeeldingendatabase Wikimedia Commons bevat 28 miljoen bestanden die vrij te gebruiken zijn. Door kennis of bestanden te delen, vergroot je het bereik enorm. Wikipedia hoort bij de vijf meest gebruikte websites ter wereld, en is de enige niet-commerciële website in de top-10. Een voorbeeld: Wikipedia-artikelen met door Nederlandse kennisinstellingen gedoneerde afbeeldingen werden in 2014 maar liefst 2 miljoen keer bekeken. Het hergebruik van de afbeeldingen is enorm groot.

Nieuwe initiatieven

Ook in de komende periode zullen Nederlandse instellingen structureel aan de encyclopedie bijdragen. Zo werkt bij Naturalis Wikipedian Hans Muller dit jaar aan een project waarmee nu al meer dan 30.000 afbeeldingen zijn ingebracht in Wikimedia Commons, en er volgen naar verwachting nog vele tienduizenden vogelafbeeldingen. Verder werken het Nationaal Archief en de Koninklijke Bibliotheek aan een project rond verzetskranten uit de Tweede Wereldoorlog en zijn er plannen voor een nieuwe donatie van het Tropenmuseum.

Prestigieuze prijs

Sinds 1958 wordt de Erasmusprijs uitgereikt. Deze prestigieuze prijs wordt toegekend aan een persoon of instelling die een belangrijke bijdrage heeft geleverd op cultureel, sociaal of sociaal-wetenschappelijk terrein. De prijs, en het bijbehorende bedrag van €150.000, wordt in november door koning Willem-Alexander aan vertegenwoordigers van Wikipedia uitgereikt. Op 23 november organiseert de KNAW een symposium 'Reading Wikipedia', en op 25 november houdt Wiki-oprichter Jimmy Wales een lezing in Paradiso.

commons.wikimedia.org
erasmusprijs.org

De hier getoonde afbeeldingen zijn allen gedeponeerde afbeeldingen aan Wikimedia Commons door betrokken organisaties.



Taj Mahal, India 1890
foto KITLV / Clifton & Co



Initiatierite, Sierra Leone 1935
foto Sjoerd Hofstra / African Studies Centre Leiden



Dochters van de radja van Boeileleng
foto KITLV / Isidore van Kinsbergen

Wat doet een socioloog met digitale data?

Integriteit en duurzaamheid van onderzoeksdata staan hoog op de agenda van beleidsmakers en financiers. Maar hoe gaan bijvoorbeeld sociologen in de praktijk met data om?
Marion Wittenberg

Bij de meeste afdelingen sociologie aan Nederlands universiteiten is, mede naar aanleiding van fraudezaken als de Stapel-affaire, beleid geïntroduceerd om de controleerbaarheid van onderzoek te vergroten. Door middel van *publication packages* - het opslaan van de data en syntax waarop een publicatie gebaseerd is - probeert men op een meer transparante manier met data om te gaan wanneer een onderzoek is afgerond.

Dit komt naar voren in een onderzoek van Beau Oldenburg (Rijksuniversiteit Groningen), in opdracht van DANS. Haar onderzoek onder 14 sociologen - zes programmaleiders en acht promovendi - van zes Nederlandse universiteiten laat verder zien dat er echter geen beleid is hoe je als onderzoeker tijdens je onderzoek met data moet omgaan. Hoewel alle universiteiten beschikken over een beveiligde digitale omgeving worden veel data uitgewisseld via Dropbox, USB-sticks of e-mail.

Onbekend maakt onbemind

Het uitwisselen van data wordt bij voorkeur gedaan met onderzoekers die men al kent. Vooral promovendi staan terughoudend tegenover het delen van hun data met anderen. Oldenburg: "Alhoewel ze aangeven zich bewust te zijn van deze onwetenschappelijke houding, vinden vier van de zes promovendi die zelf data verzamelen, het oneerlijk als anderen 'zomaar' gebruik zouden kunnen maken van hun data". De zes programmaleiders staan positiever tegenover het delen van data, meer in lijn met het beleid van financiers zoals NWO, KNAW en ZonMw. Zij geven aan hun data liever bij een landelijke repository zoals DANS te archiveren dan bij een lokale voorziening als een universiteitsbibliotheek. In de praktijk blijkt dat onderbrengen bij DANS echter niet altijd te gebeuren.

Data begrijpen en gebruiken

Om data te kunnen hergebruiken, is het van belang dat deze goed gedocumenteerd worden. Hulpmiddel hierbij is een datamanagementplan (DMP). In een DMP wordt de context van het onderzoek zo omschreven dat anderen, die niet bij het onderzoek betrokken zijn geweest, de data ook kunnen begrijpen en gebruiken. Om goed datamanagement te bevorderen kunnen onderzoekers de kosten hiervoor, sinds kort, financieren vanuit hun eventuele NWO-subsidie. Het is echter opmerkelijk dat vrijwel geen van de geïnterviewden bekend is met de term datamanagementplan. De onderzoekers blijken gaandeweg het onderzoek wel over deze onderwerpen na te denken, maar zetten deze niet op papier. Ook hier is dus nog veel winst te behalen.

Benieuwd naar het volledige onderzoek? De white paper 'Integriteit en duurzaamheid in het digitale tijdperk' is te downloaden via de DANS-website.
dans.knaw.nl

SINDS KORT BESCHIKBAAR

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

CentERdata

• Reizend Nederland



foto Paul Voorham, CC BY-NC-ND 2.0 NL

Deze studie meet het reisgedrag van Nederlanders. Wanneer gebruikt iemand de auto, wanneer kiest men voor het openbaar vervoer of pakt men de fiets? Gekeken is naar welke invloed het reisgedrag van een individu heeft op zijn of haar welbevinden en hoe het woon-werk reisgedrag zich verhoudt tot reizen in vrije tijd. Het onderzoek is verricht door Susan Handy (University of California), Bert van Wee (TU Delft), en Maarten Kroesen (TU Delft). De data zijn verzameld in juli 2013 en juli 2014 in het LISS panel. De dataset 'Travel behavior, well-being and transport-related attitudes' is beschikbaar via LISS Data Archive.

lissdata.nl/dataarchive

Ook sinds kort beschikbaar:

Studies LISS panel

- Hugenholtz, B.; Helberger, N.; Poort, J.; Balázs, B.; Handke, C.; Quintas, J.P.; Leenheer, J.; Elsen, M., 2014, Alternative Compensation Systems to compensate for and legalize infringing online practices - Part 2
- Van Kippersluis, H.; Koc, H., 2014, Food Choice - Part 1 - 2
- Korndörfer, M.; Krumpal, I.; Jann, B.; Schmukle, S., 2014, Measuring quantitative sensitive behaviors
- Regt, S. de; Lippe, T. van der; Jaspers, E., 2014, Freedom and Liberation Day in the Netherlands
- Naumann, E.; Ebbinghaus, B., 2013 - 2014, A Cross-National Survey Experiment on Support for Pension Reform - Wave 1 - 2 - 3
- Kalmijn, M.; Bol, T., 2013 - 2014, Heterogeneity in the Grandparent Effect
- Robertson, S.; Wildschut, T., 2012 - 2014, Nostalgia - Wave 1 - 2
- Pieters, R.; Leenheer, J., 2014, Tilburg Consumer Outlook Monitor

Studies Immigrant panel

- CentERdata, 2014, Social Integration and Leisure - Wave 2
- Slootjes, J., 2014, (shortened) Sense of Coherence
- Regt, S. de; Lippe, T. van der; Jaspers, E., 2014, Freedom and Liberation Day in the Netherlands
- Stupar, S.; Vijver, A.J.R. van de; Fontaine, J.R.J., 2014, The Interpersonal Conflicts influence Verbal and Indirect Aggression through Anger and Emotion Suppression



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

DANS

• Herinneringen aan Indië naar DANS

Yvonne van Genugten, directeur van het Indisch Herinneringscentrum: "Herinneringen aan Indië is een interviewproject waarbij mensen worden geïnterviewd, die voor, tijdens en na de Tweede Wereldoorlog in Nederlands-Indië verbleven. De ooggetuigen dragen met hun verhalen bij aan de herinnering van een essentiële periode uit de Nederlandse geschiedenis. De interviews zijn onlangs overgedragen aan DANS. Door deze samenwerking blijven de interviews duurzaam gearchiveerd en vrij beschikbaar voor geïnteresseerden en kunnen de bijzondere persoonlijke ervaringen met een breed publiek worden gedeeld. Daarnaast is de collectie interessant voor wetenschappelijk onderzoek als aanvulling op de schriftelijke bronnen uit die periode."

<http://dx.doi.org/10.17026/dans-x4y-zww4>
indischherinneringscentrum.nl



De overdracht wordt met handtekeningen bekrachtigd foto Lucas Pasteuning

Ook sinds kort beschikbaar:

- Van Campenhout, K. (ADC ArcheoProjecten) (2013): Harderwijk Waterfront-Zuid Maritieme Opgraving. DANS. <http://dx.doi.org/10.17026/dans-zbz-26ba>
- Centraal Bureau voor de Statistiek (CBS); Rijkswaterstaat (RWS) (2015): Onderzoek Verplaatsingen in Nederland 2014 - OViN 2014. DANS. <http://dx.doi.org/10.17026/dans-x95-5p7y>
- Centraal Bureau voor de Statistiek (CBS) (2011): Onderzoek Jeugd en Opgroeien - OJO 2011. DANS. <http://dx.doi.org/10.17026/dans-2ah-u5zb>
- Centraal Bureau voor de Statistiek (CBS) (2014): Consumentenconjunctuuronderzoek - CCO 2014. DANS. <http://dx.doi.org/10.17026/dans-x7x-ahkb>
- Dekker, A.P. (University of Groningen) (2015): Leven op krediet (Indebted life) - Anton Kröller (1862-1941). DANS. <http://dx.doi.org/10.17026/dans-23b-n6bf>
- During, Dr B.S. (Faculty of Archaeology, Leiden University); Glatz, Dr C (University of Glasgow); Wilkinson, Dr T; Serifoglu, Dr TE (Bitlis Eren University) (2015): Cide Archaeological Project. DANS. <http://dx.doi.org/10.17026/dans-xrc-uvww>
- Esveldt, Drs. I (Nederlands Interdisciplinair Demografisch Instituut, NiDi) (1995): Migratiemotieven, Migratienetwerken en Partnerkeuze van Turken en Marokkanen in Nederland. DANS. <http://dx.doi.org/10.17026/dans-zww-f88x>
- Hoogland, drs. K. (APS) (2013): Rekenen in Beeld. DANS. <http://dx.doi.org/10.17026/dans-za6-5q6c>
- Inspectie van het Onderwijs (2015): Toezichtkaart (voorheen Kwaliteitskaart) Voortgezet Onderwijs 2015. DANS.

<http://dx.doi.org/10.17026/dans-2c2-crdm>

- Koetsenruijter, J. (Radboudumc) (2013): EU-WISE: a patient observational study. DANS. <http://dx.doi.org/10.17026/dans-x5s-qq46>

- Snijder, MSc. R. (OAPEN Foundation) (2015): Revisiting an Open Access monograph experiment: measuring citations and tweets five years later. DANS. <http://dx.doi.org/10.17026/dans-x6m-67b2>

<http://dx.doi.org/10.17026/dans-x6m-67b2>

- Wit, drs. M.J.M. de (MUG Ingenieursbureau) (2015): Archeologische opgraving Terborgseweg Doetinchem, gemeente Doetinchem (GLD). DANS. <http://dx.doi.org/10.17026/dans-znw-vsrb>



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

Huygens ING

• Kennis over Tartarije digitaal

Op 24 september werd in Teylers Museum in Haarlem de digitale versie gepresenteerd van het boek Noord en Oost Tartarije. Het boek werd ruim driehonderd jaar geleden geschreven door de Amsterdamse burgemeester en amateurgeleerde Nicolaas Witsen. Nooit eerder bracht iemand zoveel kennis bijeen over 'Tartarije', het tegenwoordige Eurazië. Daarvoor kreeg Witsen vooral van Russische geleerden veel waardering. In Nederland bleef zijn werk echter lange tijd onderbelicht. De digitale uitgave bevat naast een inleiding in het Russisch en een toelichting in het Nederlands ook kaarten, een zo uitvoerig mogelijk register van zaken, persoonsnamen, geografische begrippen en etnografische namen.

resources.huygens.knaw.nl/retroboeken/witsen



Nicolaas Witsen, auteur van Noord en Oost Tartarije. Gravure naar een portret door Michiel van Musscher bron Europeana

Ook sinds kort beschikbaar:

- Plakaatboek Guyana 1670-1816. resources.huygens.knaw.nl/retroboeken/guyana



Deze publicaties zijn beschikbaar via historici.nl. Bezoek deze site of scan de QR code.

AGENDA

2 november • Den Haag

NCSRA-symposium

NCSRA staat voor Nationale Cyber Security Research Agenda. Dit symposium biedt de mogelijkheid tot kennisdeling en community building op het gebied van cybersecurity, onderzoek en innovatie. alertonline.nl

2 - 6 november • Chapel Hill, USA

iPRES-conferentie

De jaarlijkse iPRES-conferentie heeft als centraal thema: onderzoek op het gebied van digitale archivering. ipres2015.org

12 november • Utrecht

Netwerkdag Oorlogsbronnen

Het Netwerk Oorlogsbronnen organiseert zijn jaarlijkse Netwerkdag. Het thema is 'Samenwerken en verbinden'. oorlogsbronnen.nl

16 -17 november • Luxemburg

European Data Forum (EDF)

Het EDF is een ontmoetingsplaats voor industrie, onderzoek, beleidsmakers en maatschappelijke initiatieven om de uitdagingen van Big Data en de opkomende 'Data Economie' te bespreken. 2015.data-forum.eu

19 november • Amersfoort

Pump your Career

Een door NWO en het Landelijk Netwerk Vrouwelijke Hoogleraren georganiseerd congres voor vrouwen in de wetenschap. nwo.nl/actueel/evenementen/pump+your+career

24 november • Den Haag

Gebruikersbijeenkomst DANS CBS-bestanden

Het CBS en DANS organiseren een gebruikersbijeenkomst. Tijdens deze middag wordt onder andere de huidige beschikbaarstelling van beveiligde microdatabestanden samen met gebruikers geëvalueerd. dans.knaw.nl

KORT

Recensies BMGN ook op recensio.net

BMGN Low Countries Historical Review is het eerste Nederlandse geschiedwetenschappelijke tijdschrift dat zijn recensies niet alleen in Open Access beschikbaar stelt via de eigen website bmgn-ichr.nl, maar ook full-text online aanbiedt via recensio.net. Via deze meertalige Europese online database met recensies over historische literatuur hoopt het tijdschrift het lezerspubliek te vergroten en de uitwisseling van kennis gemakkelijker te maken. [Recensio.net](http://recensio.net) biedt daarvoor een uitgelezen mogelijkheid: niet alleen ontsluit het recensies van ruim veertig historische tijdschriften, het biedt gebruikers ook de mogelijkheid om met elkaar in debat te treden over monografieën en wetenschappelijke artikelen via het publiceren van presentations. (FG) recensio.net

Hoogleraar Digitale methoden en geschiedwetenschappen Charles van den Heuvel:

‘Soms gebruik je digitale middelen, soms ook niet’

Wat kan de bijdrage van Digital Humanities aan geesteswetenschappelijk onderzoek zijn? Aan het woord is Charles van den Heuvel, hoogleraar Digitale methoden en geschiedwetenschappen. *Ilja Nieuwland*

Een half jaar geleden sloot je jouw oratie af met de hoop UvA-studenten te motiveren de combinatie te verkennen van Digital Humanities, onderzoek en samenwerking met instellingen en bedrijven. Zie je daar de vruchten al van?

“De Universiteit van Amsterdam heeft samen met de Vrije Universiteit een Digital Humanities-minor opgezet. Een onderdeel hiervan is *Digital Humanities in Practice* waar ik samen met Lora Aroyo (VU) verantwoordelijk voor ben. En ja, hierin zie ik dit al terug. Ik

sprak ook de hoop uit, zelf te kunnen fungeren als verbindende factor tussen Huygens ING, erfgoedinstellingen, private partijen en onderzoekers. Ook dat lukt: ik ben stagebegeleider binnen deze minor en breng studenten onder bij projecten om ze kennis te laten

maken met de praktijk. Een voorbeeld is de samenwerking met Lab1100, waar studenten met de tool Nodegoat echt werken aan een product dat vervolgens bij onderzoek weer wordt ingezet. Daarnaast onderhouden we contacten in de wereld van het cultureel erfgoed voor potentiële stageplekken. Die link is niet nieuw voor me; ik heb zowel binnen instellingen voor cultureel erfgoed als de academische wereld gewerkt.”

Jullie samenwerking met private partijen lijkt interactiever dan gebruikelijk.

“Dat is ook zo. Ik schrijf bijvoorbeeld nu een artikel met de ontwikkelaars van Lab1100, waarin we hun tool zowel positief als negatief analyseren. De resultaten worden verwerkt in nieuwe software. Dus in die zin is er wel degelijk een interactie tussen de ontwikkeling en de onderzoekspraktijk: zij ondersteunen ons technisch en wij zien dingen waar zij niet aan hebben gedacht - en die ze gebruiken voor de verdere ontwikkeling van hun tool.”

Spelen studenten daar ook nog een rol bij?

“Ja, studenten kunnen gedetailleerde feedback geven omdat ze dagelijks met dergelijke tools werken. Omgekeerd delen ontwikkelaars hun praktijkervaring door les te geven aan mijn studenten.”

Soms lijken Digital Humanities, Digital History in het bijzonder, te lijden aan definitieproblemen en ook aan een minderwaardigheidscomplex. Vind je dat de praktische toepassing van Digital Humanities tot een beter theoretisch kader moet leiden, of werkt dat andersom?

“Ik vind dat er veel getheoretiseerd wordt en ik ben een beetje klaar met de vraag wat Digital Humanities is. Tegelijkertijd moet iets dat je ontwikkelt, of het nu een tool of een benadering is, ook herbruikbaar zijn in andere omgevingen. Het experiment Digital Humanities is pas geslaagd als je een methode of tool overhoudt die een toepassing vindt in de geesteswetenschappen. Er zijn ook mensen

die vinden dat Digital Humanities belangrijk genoeg zijn om er een aparte discipline van te maken, maar daar geloof ik helemaal niet in. Ik geloof in humaniora waarin je af en toe digitale middelen gebruikt - maar soms ook niet. We kunnen via Digital

Humanities wel bijdragen aan het veranderen van de bedrijfscultuur binnen de humaniora. Het is vooral belangrijk dat een cultuur van experimenteren ontstaat - dat gebeurt veel te weinig in de geesteswetenschappen.

We hoeven niet alle methoden uit de natuurwetenschappen over te nemen - in het verleden heeft dat tot een soort reductionisme geleid. De conclusie die daar door sommigen uit is getrokken, dat we onze eigen taal moeten ontwikkelen, vind ik niet altijd overtuigend. Dat werkt misschien nog wel voor kleine datasets, maar bij grote datasets kun je niet die grip krijgen, die we nu met patroonherkenning wel hebben.

Waar we kunnen leren van de natuurwetenschappen, is de traditie van het opzetten van experimenten die ook mis kunnen gaan - en dat je daar over rapporteert. Daar hebben we heel weinig ervaring mee in de geesteswetenschappen, binnen en buiten het digitale

ren met een ERIC (European Research Infrastructure Consortium) status samen: projectcoördinator European Social Survey (ESS), de Survey for Health, Ageing and Retirement in Europe (SHARE) en het Consortium of European Social Science Data Archives (CESSDA).

Het project wordt gefinancierd door het Europese Horizon2020 programma voor Onderzoek en Innovatie. (EB) europeansocialsurvey.org



Charles van den Heuvel

Charles van den Heuvel studeerde kunstgeschiedenis aan de Rijksuniversiteit Groningen en promoveerde in 1991 op een proefschrift over de introductie van Italiaanse steden- en vestingbouw in de Nederlanden tijdens de Tachtigjarige Oorlog en het gebruik van tekeningen in de overdracht van kennis.

INTERVIEW

‘We slagen als we onze methodes en tools hergebruiken’

Van den Heuvel: “Alle projecten moeten slagen, en in rapportages wordt vrolijk geschreven over wat goed is gegaan - en de rest mag dan later uitgezocht worden, maar niemand weet hoe. Laten we ook publiceren over minder geslaagde experimenten. Dan lever je echt een belangrijke bijdrage aan de Digital Humanities” foto Milo van de Pol

veld. Alle projecten moeten slagen, en in rapportages wordt vrolijk geschreven over wat goed is gegaan - en de rest mag dan later uitgezocht worden, maar niemand weet hoe. Met ons ePistolarium-project hebben we bewust wel volledig gerapporteerd, zodat mensen ook met conclusies uit minder geslaagde experimenten verder kunnen werken. Dat moeten we veel meer doen, bij voorkeur met onderzoekers en andere partijen. Van de feedback kan iedereen leren. En als je daar dan vervolgens over rapporteert en dat ook publiceert, lever je echt een belangrijke bijdrage aan de Digital Humanities.”

Hoe neem je zulke feedback zelf mee in het ontwerpproces van tools?

“Analyse en design zijn in feite niet van elkaar te scheiden. Ik probeer een situatie te creëren waarin je tegelijkertijd analyseert en ontwerpt. En als je dat niet allemaal zelf kunt, breng dan mensen bij elkaar die dat wel kunnen, die ook open staan voor de input. Dat is voor mij de ideale situatie - samenwerking in kleine groepjes van historici en ontwikkelaars die elkaars taal verstaan.”

oratiereeks.nl/upload/pdf/PDF-4371web-oratie_Van_den_Heuvel_-_DEF.pdf
ckcc.huygens.knaw.nl/epistolarium/

SERISS-project geeft boost aan onderzoek

Afgelopen zomer is het SERISS-project gestart. SERISS staat voor *Synergies for Europe's Research Infrastructures in the Social Sciences*, het project wil bestaand sociaalwetenschappelijk onderzoek in Europa versterken. Het project brengt drie Europese infrastructu-

Data kiesgedrag beschikbaar via Survey Data NL

De data van het Nationaal Kiezersonderzoek (NKO) zijn nu ook beschikbaar via Survey Data Nederland (SDN). Dit onderzoek naar het kiesgedrag van Nederlanders wordt steeds gehouden rond de verkiezingen voor de Tweede Kamer. Vraagteksten en variabelen

van de afgelopen zeven NKO's zijn nu online doorzoekbaar. Opname van onderzoek in SDN vergroot de vindbaarheid van de data, verbetert de disseminatie en draagt zorg voor de langetermijnarchivering. NKO is de vierde grote studie binnen de SDN-portal. SDN is een initiatief van DANS en Centerdata om alle belangrijke Nederlandse surveydata onder één dak te brengen en op vraagniveau doorzoekbaar te maken. (EB) surveydata.nl

Virtuele collectie inspireert vele wetenschappers

Meestal bekijken archeologen de buitenkant van artefacten. Dominique Ngan-Tillard en haar collega Wim Verwaal maakte micro-CT-scans om ook de binnenzijde virtueel te conserveren. Annemiek van der Kuil

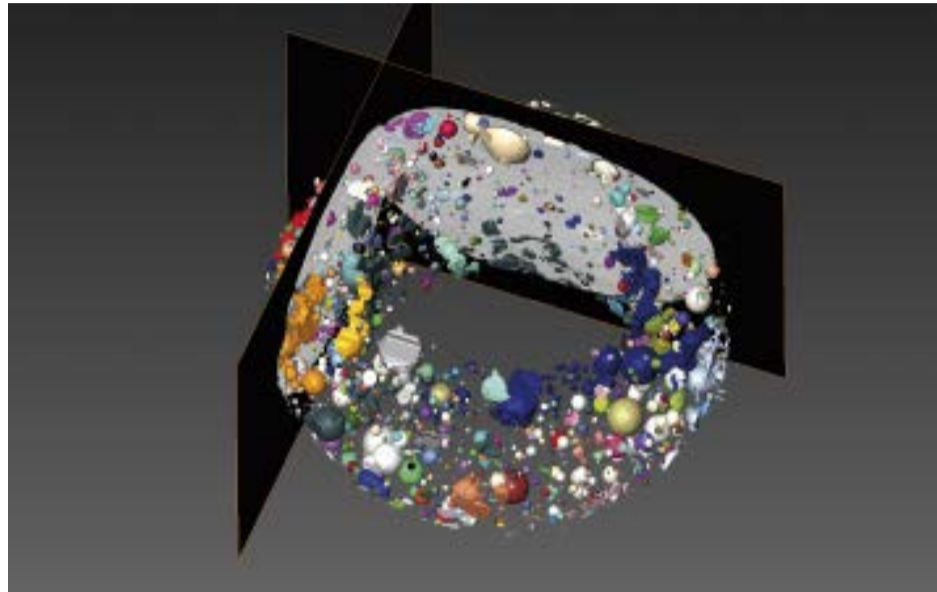
In de vroege ijzertijd was glas zeldzaam in Nederland. Dat maakt de twee gevonden kralen van bobbelig glas in een urnenveld in Zutphen bijzonder. Dominique Ngan-Tillard, onderzoeker bij de sectie Geo-Engineering van de Technische Universiteit Delft, vervolgde samen met haar collega Wim Verwaal het onderzoek van Hans Huisman, Bertil van Os en Ineke Joosten van de Rijksdienst voor het Cultureel Erfgoed en Joshua van der Laan van de Vrije Universiteit Amsterdam. Ze maakten microstructurele analyses aan de hand van micro-röntgentomografie. Ook onderzochten ze de inwendige structuur van de archeologische artefacten.

Wereldwijde toegang

Alle bevindingen zijn opgeslagen in het digitale archief van 3TU.Datacentrum en beschikbaar gesteld voor andere onderzoekers.

3D-visualisatie

De meeste Nederlandse archeologische gegevens worden bewaard in het E-depot voor de Nederlandse archeologie (EDNA) bij DANS. Deze micro-CT-scans en de analyse zijn, vanwege de omvang, beschikbaar in het data-archief van 3TU.Datacentrum. DANS en 3TU.Datacentrum werken nauw samen binnen Research Data Netherlands (RDNL), zodat de gegevens in beide databanken gemakkelijk gecombineerd konden worden. Kijk voor een video over de samenstelling van de kralen op de site van 3TU.Datacentrum.



Bubbels in kralen uit de IJzertijd

Uit de micro-CT-scans blijkt dat het schuimende aard van het glas wordt veroorzaakt door de aanwezigheid van een groot aantal gasbellen van verschillende afmetingen. Ze zijn waarschijnlijk ontstaan door het versmelten van verschillende glasscherven. (Huisman et al., 2012, Proceedings van de 39e Internationale Symposium voor Archeometrie, Leuven)

De dataset ZAP 70, opgeslagen bij DANS, omschrijft de archeologische context van de vondsten en geeft informatie over de chemische samenstelling. Door deze nieuwe manier van werken, kunnen archeologen uit de hele wereld de virtuele collectie online raadplegen. Artefacten kunnen worden vergeleken en verdere analyses kunnen worden uitgevoerd, zonder een bezoek aan een museum of depot. Dominique: "De micro-CT-modellen van de structuur van de binnen- en buitenkant van de ornamenten kunnen bijvoorbeeld worden gebruikt om semitransparante meerkleurige 3D-prints van deze ornamenten te maken. Dat kan dan leiden tot bijvoorbeeld meer inzicht in de kunst van het maken van glas en sieraden. Een virtuele collectie wordt zo een inspiratiebron voor geïnteresseerden in oude lokale sieradenstijlen en de culturele diversiteit in

Europa. Door de digitalisering en het online beschikbaar stellen van gegevens, begrijpen we samen verschillende aspecten van ons cultureel erfgoed beter."

Kennis door digitale collecties

Dominique concludeert: "Het opslaan van digitale collecties van archeologische resten in het 3TU.Datacentrum in combinatie met de al beschikbare data bij DANS en de meta-gegevens en interpretatie van de datasets zal het (her)gebruik bevorderen van wetenschappelijke technieken in de kunst en het modern archeologisch onderzoek. De digitale collecties zullen nieuwe, innovatievere uitdagingen voor onze kennis opleveren!"

DOI:

<http://dx.doi.org/10.4121/uuid:a4162b99-5110-4b89-bcc8-68985e7de32f>

KORT

Open access tot en DOI's voor alle datasets in EASY

Per 1 september worden nieuw gedeponeerde datasets in het online archiveringssysteem EASY standaard ingesteld op Open Access - CC0 Waiver. Geïnteresseerden kunnen datasets in deze toegangscategorie downloaden zonder registratie. Indien gewenst kunnen deponeerders de categorie aanpassen in EASY. Ook nieuw in EASY is dat aan alle datasets een unieke DOI als persistent identifier is toegekend. DOI staat voor Digital Object Identifier. DOI's kunnen worden gebruikt bij het citeren van de betreffende dataset. Bovendien bevat elke dataset een instructie (een 'Cite as') om datacitatie voor hergebruikers van de dataset mogelijk te maken. Hiermee ondersteunt DANS het citeren van datasets volgens de richtlijnen van DateCite en FORCE11. (HB) dans.knaw.nl

Datamanagement in vier eenvoudige stappen

Een goed datamanagementplan geeft tijdig inzicht in de voorzieningen en kennis die tijdens en na onderzoek nodig zijn. Handig voor onderzoekers, handig voor leidinggevenden, dataondersteuners en onderzoeksfinanciers.

Een goed plan zorgt er bovendien voor dat onderzoeksdata ook na afloop herbruikbaar zijn. Open als dat kan, beschermd als het moet.

De vier stappen naar een goed datamanagementplan zijn:

1. Spreek af wie wat mag met de data;
2. Documenteer het onderzoek en de data;
3. Regel opslag en toegang tijdens en na het onderzoek;
4. Houd het datamanagementplan actueel.

Een discipline-onafhankelijk sjabloon voor een datamanagementplan staat onder datamanagementbeleid op de website van DANS. (HB) dans.knaw.nl

Een onderzoek naar het Nederlandse big data-landschap

De vele gezichten van Big Data

De Utrecht Data School onderzocht Big Data en vat samen: het label wordt overal opgeplakt, maar achter Big Data schuilen vooral nog grote verschillen. Sander Prins en Yvonne van der Wal

In het afgelopen jaar onderzocht de Utrecht Data School het big data-landschap van Nederland. Het resultaat is een indrukwekkende lijst van big data-professionals en -organisaties actief op het terrein van datagerelateerde praktijken. Bovendien geeft het onderzoek inzicht in de beloftes en knelpunten van big data en laat het zien hoe het vage begrip 'big data' in verschillende domeinen kan worden opgesplitst.

Middels Named Entity Recognition is gekeken of er namen van personen, organisaties, steden of gemeen-

tes veel voorkomen in het Nederlandse big data-landschap. Dit zou duiden op een prominente positie in het werkveld. Vervolgens is via de Delphi-methode, waar een groep experts herhaaldelijk wordt gevraagd op een lijst met namen te reflecteren en hier toevoegingen aan te doen, een groep van 118 experts naar voren gekomen. Hiervan zijn 16 experts geïnterviewd (met het oog op privacy worden de namen van deze personen niet openbaar gemaakt).

Meer dan privacy

Uit de interviews kwam duidelijk naar voren dat het label big data (door de media) overal op wordt geplakt, maar dat achter het label verschillende diensten, kennis en vaardigheden te vinden zijn. Dit maakt het voor professionals moeilijk om voor complexe projecten de juiste

partner te vinden. Terwijl samenwerking juist in big data-projecten relevant is omdat verschillende expertises nodig zijn, van data management en softwareontwikkeling tot het oplossen van juridische vraagstukken.

Veel vragen

Wat ook duidelijk uit de interviews naar voren kwam, is dat de meningen van de professionals over de uitdagingen en beloftes van big data, verschillen met de berichtgeving in de massamedia. Beide zien privacy als het grootste probleem omtrent big data. Maar de professionals kaarten ook andere thema's aan, zoals kwesties over eigendom: wie is bijvoorbeeld de eigenaar van data? Degene die de data aanlevert? Degene die de data beheert? Of degene die de data gebruikt om diverse dien-

Utrecht Data School

Studenten van de Universiteit Utrecht onderzoeken 'big data' in opdracht van bedrijven, non-profit organisaties en overheden in het onderzoeksprogramma Utrecht Data School. Daarbij experimenteren ze met verschillende analyse- en visualisatietechnieken, aangevuld met essentiële kwalitatieve analyses om de data in de juiste context te plaatsen en betekenis te geven.

sten te faciliteren? Al deze personen hebben toegang tot de data, maar hoe voorkom je misbruik? Hoe kan competent toezicht gewaarborgd worden? Ook toegang tot de data wordt als grote uitdaging gezien: als een gemeente haar data -zoals per-

soonsgegevens- bij een commercieel data warehouse of een data broker opslaat, hoe kan er dan gegarandeerd worden dat deze data niet in de handen komen van, bijvoorbeeld, verzekeraars? Met het big data-panel heeft de Utrecht Data School experts verzameld om antwoord te kunnen geven op de vele vragen rondom data-analyse en -praktijken.

Later meer

Nog niet alle resultaten van het onderzoek zijn bekend. Zo wordt nog gewerkt aan de resultaten van een sentimentanalyse middels text mining op een selectie uit het Lexis-Nexis-krantenarchief en edities van E-data & Research. Zodra beschikbaar wordt het volledige rapport op de website van de Utrecht Data School geplaatst. dataschool.nl

Rijksdienst voor het Cultureel Erfgoed en Rijkswaterstaat delen kaarten Maasdal

Archeologische verwachtingen digitaal in kaart gebracht

Stel, vanwege waterveiligheid moet een hoogwatergeul worden gegraven. Hoe zorg je ervoor dat archeologisch erfgoed in een oud rivierenlandschap niet verloren gaat door deze ingrijpende verandering?

Heidi Berkhout

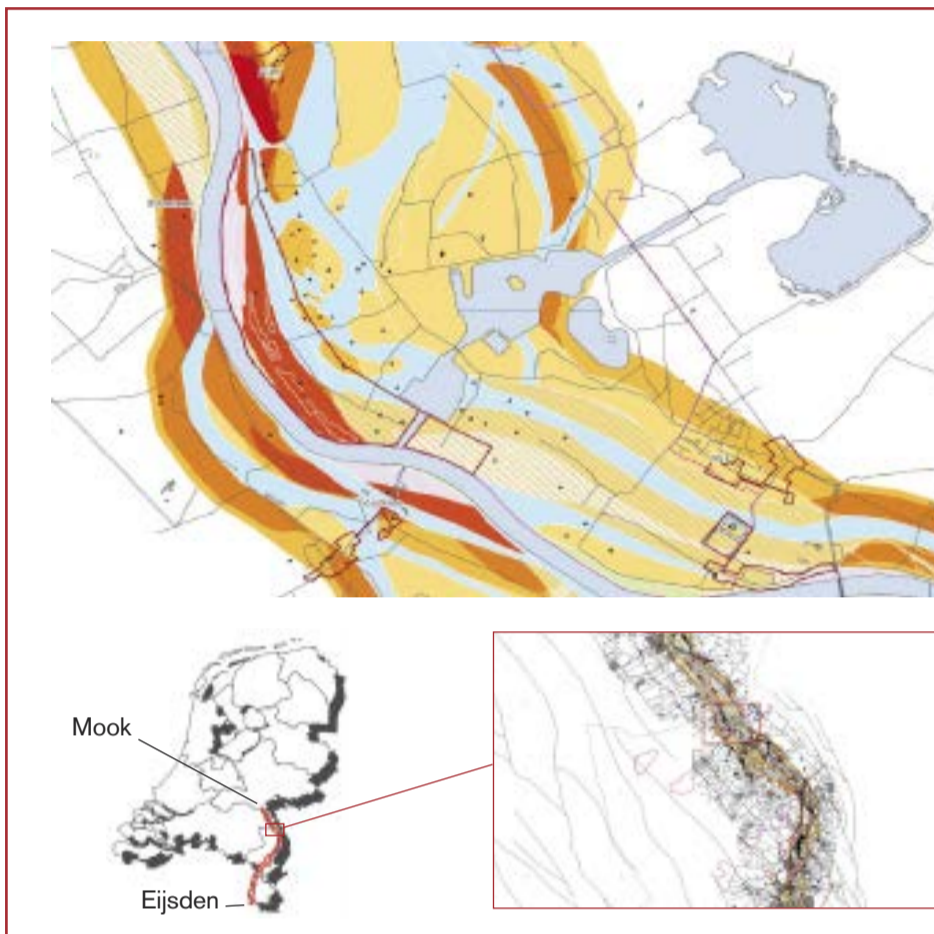
Dijken, deltawerken, havens en nevengeulen, vier voorbeelden uit een typisch Hollands landschap. Bij het voorbereiden van ruimtelijke plannen moet rekening worden gehouden met ons rijke culturele erfgoed. Want waardevolle (en vaak onbekende) archeologische resten kunnen anders onopgemerkt verloren gaan. De reden om verwachtingskaarten te maken.

Kans op resten

Een verwachtingskaart toont gebieden en zones waar kans is om archeologische resten aan te treffen. Een recent voorbeeld is de Archeologische Verwachtingskaart Maasdal tussen Mook en Eijsden (zie afbeelding). De kaarten, gemaakt in opdracht van de Rijksdienst voor het Cultureel Erfgoed (RCE) en Rijkswaterstaat – Maaswerken, zijn tot stand gekomen vanuit de ‘Visie Erfgoed en Ruimte; programma Eigenheid en Veiligheid, cultuurhistorie in wateropgaven’ van de RCE. Hierin werken RCE en Rijkswaterstaat samen.

Signalerende waarde

De RCE heeft in juli 2015 de verwachtingskaarten digitaal beschikbaar gesteld met het oog op toekomstige ontgrondingen en andere ruimtelijke ontwikkelingen in het Maasdal.



Archeologische Verwachtingskaart Maasdal tussen Mook en Eijsden

Uitsnede uit de alle-perioden-kaart van het Maasdal tussen Well en Aijen in Noord-Limburg.

De pastelkleuren geven aan waar in het Maasdal welke archeologische resten worden verwacht (basisverwachting). De aanvullende verwachting wordt weergegeven met de kleuren licht oranje tot donkerrood.

Deze alle-perioden-kaart is een cumulatieve kaart en toont de archeologische verwachting voor alle perioden. Voor een meer gedetailleerd beeld van de (onderliggende) verwachtingszones voor vier onderscheiden archeologische hoofdperiodes kunnen vier periodekaarten worden geraadpleegd.

Kijk online voor meer kaarten en de legenda.

Credits: R. Isarin, E. Rensink, R. Ellenkamp en E. Heunks, 2015, Rijksdienst voor het Cultureel Erfgoed, Amersfoort.

Dr. Eelco Rensink, projectleider namens de RCE: “De kaarten hebben een signalerende waarde. Wanneer grootschalige ingrepen in het Maasdal op stapel staan, denk aan: de aanleg van nevengeulen, weerdverlaging, dijkverplaatsingen en grind- of zandextractie, laten de kaarten zien in welke gebieden en zones de kans op het aantreffen van archeologische resten het grootst is. De kaarten zijn voor diverse betrokkenen interessant. Zo kunnen ze worden gebruikt door bijvoorbeeld beleidsbepalers bij provincies, ontgronders en (geo)archeologische professionals. Maar ook

beleidsmakers bij gemeenten kunnen de kaarten gebruiken als aanvulling op de eigen beleids- en verwachtingskaarten.”

Nieuwe kaartproducten

Rensink vervolgt: “De kaarten sluiten in het noorden bij Mook aan bij ‘Archeologische verwachtingskaart uiterwaarden rivierengebied’. Een andere kaart waarvan het Limburgse Maasdal deel uitmaakt, is de ‘Cultuurhistorische elementenkaart uiterwaarden rivierengebied’ die onlangs beschikbaar is gesteld via EASY. Deze en andere nieuwe

kaartproducten zullen in de komende jaren een belangrijke rol spelen bij de actualisatie van de Archeologische Verwachtingskaart Maasdal.

Online beschikbaar

De geomorfogenetische kaart, vier periodekaarten en alle-perioden-kaart zijn via EASY en archeologieinederland.nl beschikbaar als PDF-documenten. De volledige dataset en kaarten zijn te vinden bij DANS.

DOI:

<http://dx.doi.org/10.17026/dans-xbe-977w>

Sterke authenticatie, autorisatie en accreditatie

Veilig datagebruik binnen handbereik

Directe toegang tot data, zonder belemmeringen. Dat is de kern van het Open Access principe. Maar hoe weet je zeker, dat de persoon die jouw data wil gebruiken, is, wie hij/zij zegt te zijn?

René van Horik

Als eigenaar van data wil je toegang geven tot jouw (gevoelige) data. Aan onderzoekers die je persoonlijk kent, aan onderzoekers met een bepaalde expertise, of bijvoorbeeld aan personen die kunnen aantonen, dat ze betrouwbaar met het materiaal omgaan. Maar hoe weet je zeker, dat de persoon degene is, die hij/zij zegt te zijn? En hoe weet je zeker, dat deze persoon geen misbruik maakt van de data?

Velen kennen eduroam. Met eduroam krijgen studenten en onder-

zoekers wereldwijd toegang tot internet en informatiebronnen op basis van de combinatie ‘identiteit’ en ‘instelling’. Alhoewel dit een goed begin is, is dit niet veilig genoeg. We betreden het domein van sterke authenticatie, autorisatie en accreditatie.

Drie pijlers

Sterke authenticatie houdt in dat naast wachtwoorden zogenaamde tokens (zoals een in te voeren code uit een sms) of biometrische gegevens gebruikt worden om zekerder te zijn van de identiteit van de persoon die toegang vraagt tot een dienst of data.

Autorisatie heeft betrekking op het op maat toekennen van gebruiksrechten op het benaderen van data en systemen.

Accreditatie heeft te maken met het geven van toegang tot data aan ge-

bruikers die over specifieke vaardigheden en expertise beschikken. Het kan zijn dat gebruikers hiervoor een bepaalde training moeten volgen.

Aantal initiatieven

Veilig datagebruik komt binnen handbereik, mede door een aantal initiatieven waarbij aandacht is voor deze drie pijlers. Voorbeelden zijn de SURFconext Sterke Authenticatie dienst voor clouddiensten op instellingsniveau. Authentication and authorisation wordt onderdeel van onderzoeksinfrastructures als EUDAT en is ook opgenomen in het projectplan van CLARIAH. Verder heeft bijvoorbeeld het AARC Project (Authentication and Authorisation for Research and Collaboration) als doel om een geïntegreerde veilige toegang op Europees niveau te harmoniseren.



Elke dag geeft SURFconext tienduizenden studenten, docenten en onderzoekers toegang tot clouddiensten voor online samenwerking, onderzoek, video en het delen van bestanden. Vanaf augustus is de nieuwe functionaliteit SURFconext Sterke Authenticatie beschikbaar. Hiermee kunnen instellingen de toegang tot op SURFconext aangesloten clouddiensten beter beveiligen. Dit is vooral van belang voor clouddiensten met meer gevoelige data. Deze functionaliteit is tegen een aanvullend tarief af te nemen. Kijk voor meer informatie en een animatie over deze dienst op surfconext.nl/sterke-authenticatie. credits SURF

surfconext.nl/sterke-authenticatieaarc-project.eu

Alle Nederlandse universiteiten - op één na - ontwikkelen MOOC's. Wat betekenen deze Massive Open Online Courses voor het universitair onderwijs?

Erica Renckens

Nu ook Tilburg University onlangs is begonnen met de ontwikkeling van een MOOC Data Science voor Alfa en Gamma, begeven zo goed als alle Nederlandse universiteiten zich op het terrein van de grote, vrij toegankelijke online cursussen. Alleen de Radboud Universiteit investeert niet in MOOC's.

Impuls voor vernieuwing

“De Radboud Universiteit is positief over het ontwikkelen van digitale toepassingen die aansluiten op de onderwijsfilosofie, maar niet in de vorm van MOOC's,” aldus de woordvoerder. “De M van massive en het eenrichtingsverkeer van deze

Data-MOOC's

Tilburg University ontwikkelt samen met de Open Universiteit een MOOC Data Science voor Alfa en Gamma die naar verwachting in augustus 2016 afgerond zal zijn. In deze MOOC leren studenten en onderzoekers door middel van data analytics en datamining inzicht te krijgen in grote databestanden. De tot nu toe beschikbare MOOC's op dit gebied zijn met name gericht op de exacte disciplines.

Gratis onderwijs voor iedereen

Digitaal leren over data



onderwijsvorm past niet in onze filosofie. De Radboud Universiteit is een campusuniversiteit, waar studenten ook komen om gezamenlijk te studeren en elkaar te ontmoeten.” Marja Verstelle, programmamanager ICT en Onderwijs aan de Universiteit Leiden, denkt dat MOOC's juist een impuls geven aan vernieuwing van het campusonderwijs. “Toen de Universiteit Leiden in 1575 werd opgericht, las de docent tijdens het college voor uit een boek,” vertelt Verstelle, die de ontwikkeling van MOOC's aan haar universiteit coördineert. “Straks kunnen we ons ook niet meer voorstellen dat een docent basisstof uitlegt die je veel beter met animaties en interactieve opdrachten online kunt leren.” De tijd die overblijft,

kan de docent gebruiken om studenten te coachen en te inspireren. “De MOOC's van 2020 zullen nog interactiever en meer op maat zijn, met de beste docenten die niet alleen als wetenschapper aan de top staan, maar ook didactisch.”

Iedereen kan nakijken

Technisch is er al steeds meer mogelijk. “Docenten uit de alfa- en gamma-domeinen denken vaak dat MOOC's niet mogelijk zijn voor hun vakgebied, omdat de geschreven opdrachten niet automatisch na te kijken zijn, zoals bij de bètavakken,” geeft ze als voorbeeld. “Maar peer assessment blijkt heel effectief te zijn. De docent zorgt voor een nakijkmodel waarin de nakijkende peer kan aangeven in hoe-

MOOC's zijn gratis universitaire cursussen die online voor iedereen te volgen zijn via platforms als Coursera, edX en FutureLearn. Deelnemers kunnen bij goed resultaat na afloop een certificaat ontvangen. MOOC's worden ook ingezet voor blended learning: een door de docent samenstelde mix van online en offline activiteiten waar studenten studiepunten voor krijgen. Op master-niveau worden MOOC-deelnemers soms geselecteerd en gecombineerd met studenten. Verstelle: “Een unieke kans voor studenten om met professionals samen te werken.” foto Leidraad/Hielco Kuipers

verre de deelnemer aan de criteria voldoet. De technologie zorgt ervoor dat elke deelnemer een aantal peers nakijkt, en dat je zelf pas je cijfer krijgt als je dat gedaan hebt.”

Hogere percentages

Met een percentage van 85 tot 95% lijkt de uitval binnen MOOC's doorgaans hoog, maar dit komt grotendeels door mensen die zich alleen aanmelden als toehoorder of voor een specifiek onderdeel van de MOOC. Verstelle: “We zien dat een lager niveau - propedeutisch - en een juiste duur - vier weken - leiden tot hogere slagingspercentages. Verder besteden we veel aandacht aan het didactisch ontwerp, anderstaligen, community management en het managen van verwachtingen.”

GELEZEN

Big data, little data, no data - Scholarship in the networked world, Christine L. Borgman, 2015

René van Horik



De waarde van onderzoeksdata wordt bepaald door het gebruik van de data door onderzoekers.

Klinkt logisch, maar in de praktijk variëren gebruikers per wetenschapsgebied en veranderen deze ook nog in de loop der tijd. Christine Borgman heeft deze, en andere factoren voor een aantal wetenschapsgebieden bestudeerd en daarover een helder boek geschreven.

Data zonder context zijn waardeloos, zo stelt de auteur. Op basis van zes provocaties wordt de discussie over de rol van data in onderzoek geschetst. De provocaties hebben betrekking op aspecten zoals: het delen en hergebruiken van data, de essentiële eigenschappen van data, de relatie tussen publicaties en datasets, de kennisinfrastructuur die nodig is om data-intensief onderzoek mogelijk te maken en de lange-termijnvisie die er voor moet zorgen dat data bruikbaar en toegankelijk blijven. Big data of little data, daar gaat het niet om, aldus Borgman. Als er geen consensus is over welke data bewaard moeten worden, onder welke voorwaarden en door wie, zal er helemaal geen data overblijven.

<https://mitpress.mit.edu/big-data>

COLUMN

Zwagerman digitaal

Stel, je krijgt de opdracht om de biografie van Joost Zwagerman te schrijven. Hoe zou je de documentatie bij elkaar brengen?

Zwagerman heeft veel geschreven: fictie en non-fictie, brieven en e-mails, stukken in kranten en tijdschriften, enzovoorts. Er is ook veel over hem geschreven: recensies, portretten en – recent – een hele reeks necrologieën. Daarnaast is er nog allerlei visueel en audiovisueel materiaal: voor je het weet heb je een enorme bronncollectie bij elkaar. En daarmee ben je er nog niet, want er zijn een hoop mensen die Zwagerman hebben gekend en die van alles over hem te vertellen hebben – de laatste tijd zijn ze op veel plekken aan het woord gekomen.

Kortom: het bijeenbrengen van de bronnen voor zo'n biografie is

een hele klus. Vervolgens moet het 'echte' werk nog beginnen: het analyseren en structureren van alle informatie.

Hoe zal de toekomstige biograaf dit allemaal gaan aanpakken? De kans is vrij groot dat hij of zij dit zal doen zoals de meeste biografen die ik ken momenteel te werk gaan. Ze beluisteren, bekijken en lezen alles en maken daar aantekeningen van. Ik ken biografen die hun aantekeningen op papier maken en biografen die dit digitaal doen, bijvoorbeeld in een Wordbestand. Maar ik ken slechts heel weinig biografen die volop gebruikmaken van moderne digitale hulpmiddelen. Vaker is het tegenovergestelde het geval. Althans ik heb geregeld te maken met biografen die hun computer alleen als tekstverwerker gebruiken.

Is dat een gemiste kans? In mijn ogen wel. Als je een biografie van een schrijver gaat maken, zou je toch op z'n minst al zijn of haar publicaties digitaal willen hebben. Waarom? Je bent al snel jaren bezig met een gedegen biografie en de kans is groot dat je halverwege een nieuw thema ontdekt. Je herinnert je vagelijk dat je hoofdperson over dit thema heeft geschreven, maar wáár ook alweer? En heeft hij of zij er ook iets over ge-



foto Leo van Velzen

zegd in een brief, gesprek of interview?

“Wat zou ik graag alle boeken van mijn schrijver digitaal hebben”, heb ik diverse biografen horen verzuchten. “Dan zou ik ze steeds op andere steekwoorden kunnen doorzoeken. Maar ja, ze zijn niet digitaal beschikbaar.”

Als je vervolgens zegt dat je van een boek van 200 bladzijden binnen 5 minuten een doorzoekbare pdf plus een Wordbestand kunt maken, en dat je iemands oeuvre dus vrij snel kunt digitaliseren, kijken ze je vol ongeloof aan. Terwijl dat echt al ruim tien jaar kan. Gewoon met een snelle scanner en ocr-software.

Als je daarna vertelt dat je ook heel makkelijk handschriften, foto's en audiovisueel materiaal van steekwoorden kunt voorzien, dat

je met een druk op een knop bronnen op relevantie, chronologisch, omgekeerd chronologisch en thematisch kunt sorteren, of alles tegelijk kunt doorzoeken, dan krijg je antwoorden als: helaas, ik ben niet zo technisch. Ik ook niet. Maar met een kleine tijdsinvestering zijn deze technieken goed te begrijpen en toe te passen. Ze leveren niet alleen veel tijd maar ook beter werk op. En juist daarom vind ik het echt doodzonde dat dergelijke basistechnieken nog zo weinig bekend zijn bij onderzoekers die zich met veel inzet en passie op een onderwerp storten.

Ewoud Sanders

Taalhistoricus en journalist.

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