

ePistolarium maakt zeventiende-eeuwse correspondentie doorzoekbaar

Online grasduinen in brieven van geleerden

Binnenkort vindt de lancering plaats van het ePistolarium, een online onderzoeksomgeving rond ruim twintigduizend brieven van zeventiende-eeuwse geleerden uit negen verschillende verzamelingen. *Erica Renckens*

“Het begint eigenlijk pas net leuk te worden!” Charles van den Heuvel is vanuit het Huygens ING betrokken bij het onlangs afgeronde Geleerdenbrievenproject, waar de afgelopen vijf jaar ook het Descartes Centre, de Koninklijke Bibliotheek, DANS en de universiteiten van Amsterdam en Utrecht aan werkten. Gedurende die tijd is het karakter van het project iets bijgesteld. “Oorspronkelijk was het de bedoeling om een *collaboratory* te bouwen”, aldus Van den Heuvel. In een virtuele omgeving zouden wetenschaps-historici kunnen samenwerken en het materiaal verrijken. “Maar het bleek geen goed idee om het systeem direct open te stellen; de initiële interface en resultaten van het automatisch herkennen van *topics* in meertalige brieven waren nog niet goed genoeg.”

350 spellingsvarianten

De mogelijkheden van het ePistolarium zijn daarom uitgebreid met verschillende taaltechnologische methoden. “De meest geavanceerde is waarschijnlijk de taalafhankelijke *topic modeling*”, vertelt Walter Ravenek, die als ontwikkelaar vele uren in het project heeft gestoken. Op basis van het gezamenlijk voorkomen van woorden bepaalt deze methode de onderwerpen die in de brief behandeld worden. “Daarmee



De correspondentie van Constantijn Huygens, ruim 7.000 brieven, is één van de negen verzamelingen die momenteel zijn opgenomen in het ePistolarium credits The National Gallery, London

Antwoorden op veel zorgvragen

Gegevens over volksgezondheid en zorg zijn eenvoudig online te vinden. Maar door het enorme aanbod is het lastig om de betrouwbaarheid van deze gegevens in te schatten. Zorggegevens.nl geeft antwoorden. *Rutger Nugteren*

“Zorggegevens.nl is een online catalogus met informatie over gegevensbronnen die betrekking hebben op volksgezondheid en zorg.” Aan het woord is Laurens Zwakhals, projectleider Zorggegevens bij het RIVM. “Het gaat om beschrijvende informatie, de metagegevens, van gegevensbronnen zoals zorgregistraties, enquêtes, monitors en langlopende (cohort) onderzoeken. De catalogus biedt zoveel mogelijk in-

formatie zodat het vinden van de gegevens zo eenvoudig mogelijk wordt. De gegevens zelf kunnen onderzoekers vervolgens bij de bronhouders opvragen.”

Handige dossiers

Zwakhals vervolgt: “Onderzoekers vinden op de site informatie over bijvoorbeeld organisaties op het terrein van zorggegevens. Met korte teksten en links naar websites wordt een beeld geschetst wie welke gegevens verzameld en waarom. Handig als je wilt weten voor welke informatie je bij wie kan aankloppen. De site beschikt ook over handige dossiers zoals ‘Ouderen en gezondheid’ of ‘Infectieziektesurveillance’.

Gevraagd naar het belangrijkste onderdeel van de website antwoordt

Zwakhals: “Dat is de catalogus met beschrijvende informatie, de meta-database. Hierin staat algemene informatie zoals de naam van de gegevensbron. Je vindt er ook een beschrijving van de data, de gebruiksdoelen en de beschikbaarheid van de gegevens. Zo kun je bijvoorbeeld opzoeken of de bron al of niet online beschikbaar is en zo niet, met wie je dan contact kunt opnemen. Daarnaast is informatie over de inhoud van de bron beschikbaar zoals de geografische dekking, de steekproefgrootte en het responspercentage.”

Routekaart

“Ten slotte is er ook operationele informatie beschikbaar. Daarbij moet je denken aan het eerste en meest recente meetjaar, de op-

drachtgever en de belangrijkste financier van de dataverzameling.” Sinds begin dit jaar werkt Zorggegevens.nl samen met anderen (CentERdata, DANS, BBMRI en NIVEL) aan een datamanagement routekaart voor ZonMw. Deze routekaart is bedoeld om onderzoekers te helpen met een subsidieaanvraag bij ZonMw. “Voor de routekaart onderzoeken we bijvoorbeeld hoe bestaande databronnen beter vindbaar gemaakt kunnen worden en hoe de (duurzame) opslag van data gefaciliteerd kan worden. Ook beoordelen we het financieringsmodel voor (duurzame) opslag van data, bijvoorbeeld hoeveel ruimte (tijd en geld) projecten moeten reserveren voor adequaat datamanagement.”

zorggegevens.nl



E-DATA & RESEARCH

Jaargang 7 | nummer 3

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

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

INHOUD

2 **Gehoord en bijgewoond: events besproken**

3 **80.000 KB-boeken online in Google Books**

3 **Volkstellingen aan Linked Data onderworpen**

4 **Deze datasets zijn sinds kort beschikbaar**

4 **Agenda: interessante bijeenkomsten op een rij**

5



Interview directeur Center for Digital Humanities

6 **Utrecht Data School visualiseert data**

7 **3D-illustraties in Dynamic Drawings-project**

8 **Schatgraven in de statistieken van het CBS**

8 **Ewoud Sanders zoekt een digitale betaalkrant**



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

VERVOLG OP PAGINA 2

Actie bij deelnemers Beyond the PDF2

Lukas Koster

Nieuwe modellen van wetenschappelijke communicatie waren het onderwerp van informele en gepassioneerde discussies tussen onderzoekers, uitgevers, ontwikkelaars en bibliotheekmedewerkers tijdens Beyond the PDF2. Deze bijeenkomst vond plaats in Amsterdam op 19 en 20 maart en werd georganiseerd door FORCE11, “a virtual community working to transform scholarly communications toward improved knowledge creation and sharing”. Drie hoofdthema’s konden worden onderscheiden: *content creation*, *data sharing* en *content dissemination*. *Content creation* betreft het publiceren over resultaten van wetenschappelijk onderzoek. Centraal hierbij stonden Open Access, geïntegreerde en dynamische publicaties, en alternatieven voor het peer review systeem. Bij *data sharing* werd ingegaan op het publiceren en citeren van onderzoeksdatasets. *Content dissemination* is het verspreiden, vindbaar maken en consumeren van wetenschappelijke output, waarbij Open Access de verbindende schakel is met nieuwe vormen van publiceren. Opvallend was dat bètawetenschappers tijdens de bijeenkomst de interactie met alfa- en gammawetenschappers bijzonder vruchtbaar noemden. Even opvallend was dat er weinig bibliotheekvertegenwoordigers aanwezig waren, en dat bij de wetenschappers het cliché van bibliotheken als een boekenopslagplaats leek te overheersen. Karakteristiek was de actiebereidheid van de deelnemers. Het best kwam die tot uiting in de sessies *Making it happen* en *Visions for the future* waar de volgende initiatieven uit voortkwamen:



De essentie van het event in tekening foto Maurice vander Feesten

Scholarly Revolution (content creation), *The Amsterdam Manifesto on Data Citation Principles* (data sharing) en het *Open Alternative to Google Scholar* (content dissemination).

force11.org/beyondthepdf2

Web ART zoekt naar bruikbaar webarchief

Peter Boot

Hét terrein waar een ambitieuze jonge onderzoeker zich op zou moeten specialiseren, is volgens Jaap Kamps het onderzoek naar archivering van het web. Kamps is projectleider van het CATCH-project Web ART (Web Archive Retrieval Tools), en hij deed zijn uitspraak bij de bijeenkomst over webarchivering op 19 april in Den Haag. Centraal thema van de bijeenkomst was de bruikbaarheid van webarchieven voor onderzoek, en de conclusie was dat daarbij nogal wat obstakels in de weg staan. De sprekers waren Helen Hockx-Yu, hoofd webarchi-

vering bij de British Library, Bill LeFurgy, programmamanager digitale initiatieven bij de Library of Congress, en Bernhard Rieder, onderzoeker aan de UvA. De obstakels die zij signaleerden waren onder andere de te selectieve archivering door nationale bibliotheken (als je moet selecteren, doet de onderzoeker dat liever zelf), documentgebaseerde toegang tot het archief (per pagina, dus ongeschikt voor data-mining), en zeer restrictieve toegang (alleen vanuit het gebouw van de archiefinstelling). Bovendien doet archivering op basis van nationale domeinen (.nl, .uk) weinig recht aan de aard van het web. En misschien de meest prangende vraag: wat is de zin van webarchivering als steeds grotere delen van het publieke web de afgesloten speeltuintjes worden van commerciële giganten als Facebook? Inderdaad uitdagingen te over voor (jonge) onderzoekers.

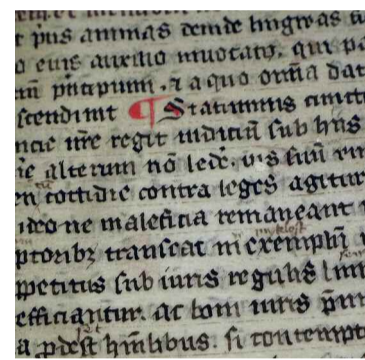
webarchiving.nl

Makkelijke tools voor moeilijke teksten

Mariken Teeuwen

Gemakkelijk te gebruiken computerprogramma’s die representatie, editie en analyse van de moeilijke bronnen van de Middeleeuwen ondersteunen, waren het centrale thema van een workshop gehouden op 18 en 19 april. Een gezelschap van mediëvisten, historici, literatuurhistorici, teksteditoren en software-ontwikkelaars kwam daarvoor bijeen in Den Haag, bij een bijeenkomst georganiseerd door Huygens ING in het kader van de COST Action Medieval Europe - Medieval

Cultures and Technological Resources. Een aantal sprekers presenteerden tools of digitale edities die al te vinden zijn op het web, anderen presenteerden wat er in het kader van concrete onderzoeksprojecten op dit moment ontworpen wordt. Projecten op het terrein van handschriftrepresentatie en digitale editie waren onder andere het Transcribe Bentham-project (waarin crowdsourcing een belangrijke rol speelt), de Elwood Viewer en eLaborate. CATMA en Shared Canvas waren voorbeelden van projec-



Tekst in middeleeuws handschrift foto Martin Muránsky

ten waarbij de verbinding van data en metadata centraal staat. In andere projecten (zoals Sharing Ancient Wisdoms en GEPHI-netwerkanalyses) ging het om de visualisatie van relaties die zonder tools voor de onderzoeker onzichtbaar blijven in de grote hoeveelheid data. Een voorbeeld van verrassend nieuw onderzoek was het HisDoc-project, dat automatische transcriptie van middeleeuwse tekst mogelijk wil maken. Een laatste tak van lezingen concentreerde zich op de vraag hoe de digitale methodes het onderzoek naar middeleeuwse (of

meer algemeen: geesteswetenschappelijke) bronnen beïnvloed heeft, en waarheen deze vernieuwing ons zal leiden. Bijzonder waardevol bleken de demonstratiesessies, waarin de deelnemers aan de workshop echt aan de slag konden met de gepresenteerde software. easytools.huygens.knaw.nl

GOR 2013:

Ratio wordt overschat

Arnaud Wijnant, Salima Douhou

“We think much less than we think we think”. Dat was de centrale boodschap van keynote-spreker John Kearon bij de GOR (General Online Research) conferentie afgelopen maart. Kearon maakte grote indruk met zijn manier van presenteren en zijn boodschap: mensen (en dus respondenten) reageren en nemen beslissingen op basis van intuïtie en niet op basis van ratio. De GOR is een vooral Europese conferentie over online, digitaal en sociale-media onderzoek. De locatie was dit jaar de ‘Duale Hochschule’ Baden-Württemberg te Mannheim. De parallelsessies hadden als thema’s Methodology & Surveys, Applied Research en Social Media. *Hot topic* tijdens de conferentie was de omgang met smartphones in online onderzoeken. Het onderwerp kwam terug in verschillende sessies. Naast de vraag of dit een geschenk of een vloek is voor de kwaliteit van data, was het ook niet duidelijk wat er allemaal gerekend kan worden tot ‘smartphonegebruik’ omdat de verschillen tussen smartphones, tablets en laptops/PC’s steeds kleiner worden. De GOR vond dit jaar plaats in samenwerking met Webdatanet, een multidisciplinair netwerk op het terrein van webgebaseerde dataverzameling. gor.de

VERVOLG VAN PAGINA 1

Brieven

kunnen we de gebruiker wijzen op vergelijkbare brieven, kunnen we zoeksuggesties geven en kan men zoeken naar vergelijkbare zinnen in de verzameling.”

Daarnaast kent het ePistolarium automatisch trefwoorden toe aan de brieven, herkent het de taal waarin de brief is geschreven, herkent het namen in de tekst en normaliseert het de spelling. Van den Heuvel: “Alleen al voor Christiaan Huygens vonden we meer dan driehonderdvijftig spellingsvarianten in de verzameling.”

Uitbreiding

Het is de bedoeling dat de database in de toekomst wordt uitgebreid met nieuwe verzamelingen. “Dat maakt niet alleen breder onderzoek mogelijk, maar verbetert ook de resultaten van de analyses”, aldus Ravenek. De onderzoekers zijn druk bezig met vervol-

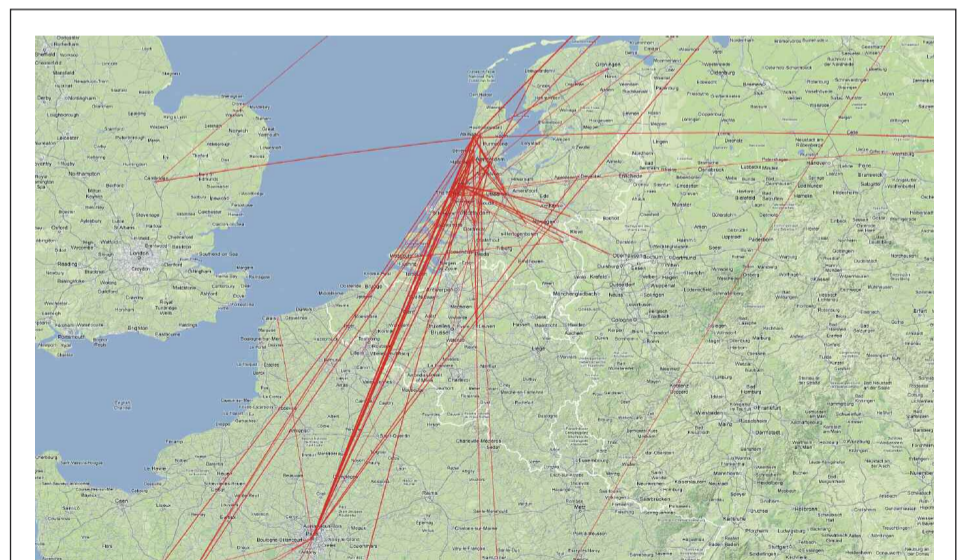
aanvragen met internationale partners, waarin ze ook de plannen voor het *collaboratory* meenemen. Van den Heuvel: “Het zou jammer zijn om nu te stoppen, terwijl het net spannend begint te worden.”

Op 13 juni vindt de presentatie plaats van het ePistolarium. Geïnteresseerden zijn welkom vanaf 15:00 uur in de Gertrudiskapel in Utrecht.

huygens.knaw.nl
descartescentre.com

Waarom brieven?

Tot halverwege de zeventiende eeuw bestonden er nog geen wetenschappelijke tijdschriften. De brief was het belangrijkste communicatiemiddel van geleerden. Via brieven deelde men de laatste ontdekkingen, nieuwste debatten, boekentips en roddels. Iedere geleerde van naam voerde een correspondentie van al snel duizenden brieven. Ze vormen daarmee een waardevolle bron voor wetenschapshistorici.

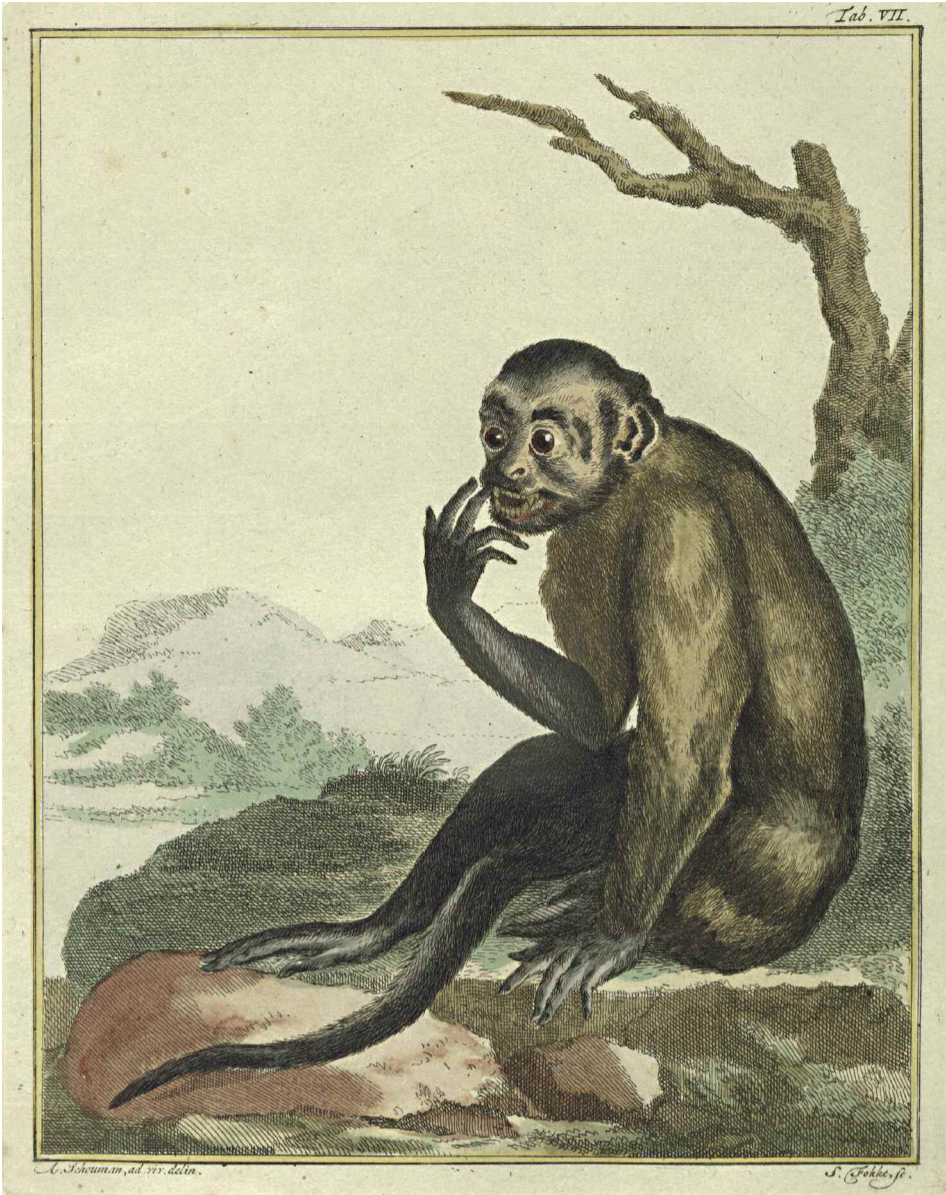


Verspreiding van kennis in beeld gebracht

In het ePistolarium is het mogelijk om full text te zoeken in de brieven, eventueel specifiek op jaar, locatie, verzender, ontvanger of genoemde personen. De zoekresultaten worden weergegeven in een overzicht, maar kunnen ook gevisualiseerd worden op landkaarten, in tijdbalken en in sociogrammen. Hiermee kan de verspreiding van kennis eenvoudig inzichtelijk gemaakt worden. Hier is de correspondentie van René Descartes in beeld gebracht. *Credit Bas Doppen en Google Maps*

KB en Google plaatsen tachtigduizend boeken uit bibliotheek online

‘Digitaal boek is godsgeschenk’



Natuurkundige afbeelding, één van de extra gedigitaliseerde beelden credits KB, Den Haag

Zoekgigant Google heeft met een aantal bibliotheken contracten gesloten om (een deel van) hun collectie online doorzoekbaar te maken. Zo ook met de KB. Peter Boot

Op 3 april presenteerde de Koninklijke Bibliotheek de eerste resultaten van haar samenwerking met Google: via boeken.kb.nl zijn de eerste tachtigduizend boeken beschikbaar. Het is de helft van de collectie van 1700 tot 1870 die de KB door Google laat digitaliseren. Tijdens deze presentatie maakte KB-directeur Bas Savenije duidelijk dat zonder de samenwerking met Google dit tempo nooit zou worden gehaald.

De boeken waren al beschikbaar op de site van Google Books. De presentatie van de boeken op de KB-site is een eerste stap naar een geïntegreerd aanbod van digitaal materiaal, waarbij verschillende boekencollecties, samen met de gedigitaliseerde kranten en tijdschriften, zullen worden aangeboden. Eerder dit jaar presenteerde de KB al een digitale collectie van tachtig gedigitaliseerde tijdschriften uit de periode 1850-1940.

Nieuwe vormen van onderzoek
Feestredenaar Inger Leemans (VU) benadrukte hoezeer het historisch onderzoek verandert door digitale beschikbaarheid van het onderzoeksmateriaal, en schetste hoe elektronische

beschikbaarheid geheel nieuwe vormen van onderzoek mogelijk maakt. Overigens is daarvoor vaak wel een betrouwbare elektronische tekst nodig, terwijl digitalisering à la Google niet verder gaat dan ongecorrigeerde OCR (*optical character recognition*). Wat dat betreft blijft er nog het nodige te wensen over.

Onderzoeker Willem Kuiper (UvA) is ook blij met Google Books en andere digitale bibliotheken. “Voor het REMLT (Repertorium van Eigennamen in Middelnederlandse Literaire Teksten) zoeken wij wekelijks het internet af en zien tot onze verbazing dat opeens een boek digitaal beschikbaar is. Je kunt zo’n boek online lezen, maar veel beter is de .pdf die je mag downloaden, waarop je vervolgens OCR loslaat. Vaak is het resultaat verbluffend. Het is een godsgeschenk om zo je eigen digitale handbibliotheek te kunnen opbouwen”, aldus Kuiper.

Boodschap voor de minister

Een schrijnend probleem, dat door de publicatie van deze collectie nog eens wordt benadrukt, is de tekortschietende digitale beschikbaarheid van modern materiaal, waar in veel gevallen nog auteursrecht op rust. In zijn toespraak beklemtoonde Bas Savenije nog eens hoe belangrijk het is de auteurswet zo aan te passen, dat digitale bibliotheken ook materiaal waarop nog rechten rust kunnen aanbieden. De staatssecretaris had zich bij de bijeenkomst laten excuseren, maar laten we hopen dat die boodschap bij het ministerie is aangekomen.

boeken.kb.nl

books.google.com

CEDAR-project onderwerpt Volkstellingen aan nieuwe techniek

Linked Data zorgen voor meer dan brood alleen

Het CEDAR-project probeert met Linked Data de historische volkstellingen beter bruikbaar voor onderzoek te maken. René van Horik

Hoeveel bakkers telde Nederland in 1795? Een interessante vraag voor sociaal-historisch onderzoek. Het antwoord ligt in de data van de Nederlandse Volkstellingen. Deze tellingen, van bevolkingsgrootte, leeftijden, geslacht en beroepsgroepen, zijn van 1795 tot 1971 gehou-

den. Inmiddels zijn de tabellen van deze zeventien tellingen grotendeels gedigitaliseerd met het project ‘Digitalisering Nederlandse Volkstellingen 1795-1971’.

De tellingen zijn een fantastische bron voor onderzoek. Maar de oorspronkelijke tabelstructuur kent ook beperkingen. Door de jaren heen veranderen bijvoorbeeld plaatsnamen, fuseren gemeenten en worden beroepen anders ingedeeld. Handmatig zijn datasets uit tabellen nog wel te analyseren, maar dit kost

veel tijd. Bieden Linked Data-technieken een uitkomst?

Gewenste koppelingen

Linked Data beschrijven de relatie tussen objecten. Zo kun je met Linked Data-technieken beschrijven dat het beroep ‘bakker’ in de ene telling een relatie heeft met het beroep ‘broodbakker’ uit een andere telling. Als alle variabelen van de Volkstellingen als Linked Data kunnen worden uitgedrukt (in een bestandsformaat waar een computer mee overweg kan), dan zijn in theorie de gewenste koppelingen te maken.

Het CEDAR-project onderzoekt de mogelijkheden om dit beschrijven van variabelen automatisch te harmoniseren en zodoende de tabellen van de tellingen beschikbaar te stellen als Linked Data. Belangrijk is dat de oorspronkelijke gegevens en tabelstructuur niet verloren mogen gaan. Bij de harmonisatie wordt gebruik gemaakt van externe classificaties, bijvoorbeeld voor histori-



Bakker met bakkershoorn door Adriaan de Lelie (1755 - 1820) credits Nederlands Bakkerijmuseum

sche beroepen. De eerste resultaten van deze omzetting van de Volkstellingen naar Linked Data-structuren staan op de CEDAR-website.

Fantastische methode

Het CEDAR-project laat zien dat de vele uitzonderingen, definitieveranderingen en vaagheden in de oorspronkelijke data verder onderzoek noodzakelijk maakt. Maar dat met

Linked Data voor onderzoekers een fantastische methode voor onderzoek beschikbaar komt, is wel bewezen. Handmatige voorbewerking van data wordt tot een minimum beperkt. Gegevens zijn veel gemakkelijker te verbinden met andere Linked Data-bronnen. En bovendien kunnen gegevens sneller worden gevisualiseerd.

cedar-project.nl

CEDAR (Census data open linked)

CEDAR staat voor ‘Census data open linked’, en is een samenwerkingsverband tussen de Vrije Universiteit en de KNAW-instituten IISG (Internationaal Instituut voor Sociale Geschiedenis) en DANS. CEDAR werkt aan het semantisch ontsluiten van de data van de Nederlandse Volkstellingen uit de laatste twee eeuwen. Deze data worden met behulp van RDF (*Resource Description Framework*) interoperabel gemaakt met andere hubs van historische socio-economische en demografische gegevens. Het hierdoor opgebouwde semantisch web-based datamodel maakt het mogelijk om vragen te beantwoorden op het gebied van regionale identiteit, technologische vooruitgang en geografische migratie.

SINDS KORT BESCHIKBAAR

Dit overzicht toont databestanden die recent beschikbaar zijn gekomen bij CentERdata en DANS.

CentERdata

• Kiezen tussen twee beschrijvingen

In maart 2012 hebben Maarten Berg en Paul de Beer de studie 'Solidarity' afgenomen in het LISS panel. Respondenten werden gevraagd om telkens te kiezen tussen twee beschrijvingen van mensen. Op deze manier werd informele (private) en formele (publieke) solidariteit onderzocht door middel van een vignetten-onderzoek. Het doel was om te kijken welke motieven, voorwaarden en omstandigheden de verschillende typen solidariteit aanwakkeren of juist ondermijnen. De data zijn beschikbaar via LISS Data Archive.



Onderzoek naar informele en formele solidariteit foto Yuralaïts Albert

Ook sinds kort beschikbaar:

Studies LISS panel

- Trautmann, S., Kuilen, G. van de., Zeckhauser, R.J., Betrayal, Beliefs, and the Demographics of Social Capital, maart 2011

- Aarts, K., Elections House of Representatives 2010, maart 2010
 - Houwen, K. van der., CBS, Role of the Government, mei 2012
 - Lugtig, P.J., Glasner, T.J., The use of dependent interviewing and event history calendars in recording monthly data in an annual panel study, januari 2012
 - Charness, G., Offerman, T., Villeval, M.C., Testing Mechanisms for Identifying True Risk Preferences, februari 2012
 - Veer, E. van de., Mindfulness, augustus 2012
- Studies Immigrant panel**
- Houwen, K. van der., CBS, Role of the Government, april 2012
 - CentERdata, Ecological Footprint, juli 2010
 - Celenk, O., Vijver, F. van de., Relationship Diary, mei 2012
 - Chasiotis, A., Bender, M., Vijver, F. van de., Relationship Context Scale, juli 2012
 - Chasiotis, A., Bender, M., Vijver, F. van de., Mini-K Short Form, juli 2012



Deze en andere bestanden zijn beschikbaar via www.liss-data.nl/dataarchive. Scan de QR-code en bezoek de site.

DANS

• WoonOnderzoek Nederland 2012

In het WoON komt informatie over huishoudenssituatie, huidige en gewenste woonsituatie, woonlasten en inkomens samen. Het is een onderzoek naar woonkwaliteit en woonbehoefte vanuit het Ministerie van Binnenlandse Zaken en Koninkrijksrelaties en het Centraal Bureau voor de Statistiek en dient ter ondersteuning van het overheidsbeleid op het gebied van wonen. Voor het onderzoek werden bijna 70.000



Ieder zijn eigen woonbehoeften, Den Haag in aanbouw foto Jan Kranendonk

huishoudens geëncquêteerd, waarna deze gegevens zijn verrijkt met gegevens van onder andere de Gemeentelijke Basis Administratie en de Belastingdienst. De data kunnen worden gebruikt voor onderzoek en statistiek ten behoeve van het agenderen en vormgeven van beleid, het volgen van de beleidsuitvoering en het verantwoorden van beleid, op de thema's bouwen en wonen. De dataset is kosteloos te downloaden in EASY.

urn:nbn:nl:ui:13-60fd-6i

Ook sinds kort beschikbaar:

Archeologie

- Archol en ADC. Mesolithisch gebruik van een duin ten zuiden van Swifterbant (8300-5000 v.Chr.). Een archeologische opgraving in het tracé van de N23/N307, Provincie Flevoland
- ROB. De Horden near Wijk bij Duurstede: plant remains from a native settlement at the Roman frontier: a numerical approach

- ADC. Het domein van de boer en de ambachtsman. Een opgraving op het terrein van de voormalige fruitveiling te Wijk bij Duurstede: een deel van Dorestad en de villa Wijk archeologisch onderzocht
- RCE. Onderzoek naar de context van een laat-Romeinse muntschat in Sint Anthonis (Noord-Brabant)
- Gemeente Dordrecht. In de voetsporen van de gebroeders De Witt. Archeologisch onderzoek aan de Nieuwstraat 60-62

Sociale Wetenschappen

- Centraal Bureau voor de Statistiek. Gezondheidsenquête 2010 en 2011 (beveiligd microbestand)
- Sociaal en Cultureel Planbureau. Arbeidsaanbodpanel 2010
- Ministerie van Sociale Zaken (Bureau ASTRI): Onderzoek Evaluatie WIA
- Kohnstamm Instituut Universiteit van Amsterdam. COOLspeciaal cohortonderzoek in speciaal (basis)onderwijs
- SEO Economisch Onderzoek. Technici, mobiel en toch honkvast

Historische wetenschappen

- Universiteit van Amsterdam; DANS. Verrijkt koninkrijk, het koninkrijk der Nederlanden in de Tweede Wereldoorlog in XML en gemanifesteerd

Taalwetenschappen

- Stichting Vrijwilligersnetwerk Nederlandse Taal. Hidde Dirks Kat: Dagboek eener reize ter walvisch- en robbenvangst (1818)

Geodata

- Kadaster. Kadastrale Kaart update februari 2013
- Kadaster. Top10NL update update februari 2013



Deze en andere bestanden zijn beschikbaar via easy.dans.knaw.nl. Scan de QR-code en bezoek de site.

AGENDA

26 - 29 juni • München

LIBER 2013

De associatie van Europese onderzoeksbibliotheken LIBER organiseert haar jaarlijkse conferentie met als thema: Research Information Infrastructures and the Future Role of Libraries. De presentaties gaan over open data, digitalisering en digitale duurzaamheid. liber2013.de

1 - 2 juli • Londen

Internationale OA-conferentie

OAPEN (Open Access Publishing in European Networks) organiseert een conferentie voor Europese onderzoeksfinciers, uitgever, bibliothecarissen en onderzoekers. Thema's zijn onder meer: Open Access businessmodellen, de kwaliteit en toekomst van peer review en ontwikkelingen op het gebied van creatieve commons licenties. project.oapen.org/

8 - 12 juli • Charlottetown (Canada)

Open Repositories 2013

Deze conferentie richt zich op de ontwikkeling, implementatie en management van digitale bewaarplaatsen en is bedoeld voor onderzoekers, bibliothecarissen en uitgever. Tijdens de conferentie staat de informatielevenscyclus centraal. or2013.net/

16 - 19 juli • Nebraska (USA)

Digital Humanities 2013

Het thema van deze jaarlijkse conferentie van de ADHO (Alliance of Digital Humanities Organizations) is Freedom to explore. Tijdens de conferentie zal er veel aandacht zijn voor interdisciplinair onderzoek. dh2013.unl.edu/

4 - 9 augustus • Sofia (Bulgarije)

ACL 2013

De jaarlijkse conferentie van de Association for Computational Linguistics richt zich op onderzoek op het gebied van computationele verwerking van natuurlijk taal. De conferentie bevat een aantal workshops, waaronder de workshop 'Language Technology for Cultural Heritage, Social Sciences, and Humanities'. acl2013.org/site/

17 - 23 augustus • Singapore

IFLA World Library and Information Congress

Het IFLA WLIC is het belangrijkste congres binnen de internationale bibliotheek- en informatiecentra-sector. Het thema is Future Libraries: Infinite Possibilities. conference.ifla.org/ifla79

2 - 5 september • Lissabon

iPRES 2013

De iPRES (International Conference on Preservation of Digital Objects) vindt gelijktijdig plaats met de Dublin Core conferentie (zie hierna). iPRES richt zich op digital preserva-

tion en biedt een podium voor uitkomsten van wetenschappelijk onderzoek, tutorials en panels.

ipres2013.ist.utl.pt/

2 - 6 september • Lissabon

DC-2013

De International Conference on Dublin Core and Metadata Applications (DC-2013) heeft als thema Linking to the Future. dcevents.dublincore.org/IntConf/dc-2013/

18 - 21 september • Veliko Tarnovo (Bulgarije)

DiPP2013

Deze conferentie richt zich op het presenteren van innovatieve uitkomsten, onderzoeksprojecten en toepassingen op het gebied van digitaliseren, documenteren en archiveren van culturele en wetenschappelijke bronnen. dipp2013.math.bas.bg/

22 - 26 september • Malta

TPDL 2013

Het thema van de conferentie TPDL (Theory and Practice of Digital Libraries) is Sharing meaningful information. De conferentie richt zich op onderzoekers, ontwikkelaars en content providers. tpdl2013.info



U.S. immigration service credits Europeana.eu

Op zoek naar een nieuw leven

Heidi Berkhout

Wat dreef dertig miljoen emigranten tussen 1820 en 1920 van Europa naar Amerika? Bijna een eeuw na de grootste piek in de emigratie lanceren de Digital Public Library of America (DPLA) en Europeana de online tentoonstelling Leaving Europe: A new life in America. De tentoonstelling staat op Europeana.eu, een portal met (delen van) de collectie van musea, bibliotheken, archieven en galeries uit Europa. Leaving Europe: A new life in America is het startpunt van een duurzame samenwerking tussen de Digital Public Library of America and Europeana. De bibliotheken delen de missie om cultureel erfgoed voor iedereen toegankelijk te maken.

www.europeana.eu

Rens Bod, directeur Center for Digital Humanities:

‘Er is een wereld te winnen in de geesteswetenschappen’

Het Center for Digital Humanities coördineert onderzoeksprojecten waarin geesteswetenschappers samenwerken met private partijen. Een gesprek met directeur Rens Bod.

Erica Renckens

Sinds de start in 2012 heeft het CDH (Center for Digital Humanities), dat mogelijk wordt gemaakt door de Universiteit van Amsterdam, in samenwerking met de KNAW en de Vrije Universiteit, dertien projecten gehonoreerd die resulteren in concrete producten of demo's. Stuk voor stuk voorbeelden van samenwerkingen tussen publieke en private partijen. "Zo heeft onderzoekster Aline Honingh samen met het bedrijf ElephantCandy een app ontwikkeld die de muzikale overeenkomst berekent tussen muziekstukken", aldus Bod. "Deze app kan onder meer vloeiend van een muziekstuk naar een andere overgaan door naar gelijksoortig ritme, metrum, melodie, harmonie of instrumentatie te zoeken. Dit is interessant voor dj's, maar ook voor ballroomdansers, en niet te vergeten voor joggers die hun hartslag het volgende muziekstuk willen laten bepalen. Dit project bouwt voort op onderzoek waaraan al tien jaar werd gewerkt en dat nu wordt gevaloriseerd. Voordat de app daadwerkelijk kon worden gebouwd, moesten de ontwikkelaars eerst allerlei muziektheoretische problemen oplossen. Maar nu de app er is, kan hij ook weer worden gebruikt in het onderzoek van Aline Honingh. Het geeft weer hoe nauw en interessant dit soort samenwerkingen zijn."

Interessant erfgoed

Andere projecten hebben betrekking op onder andere de kennisstructuur van Amsterdam in de Gouden Eeuw, het schatten van je leeftijd op basis van je tweets en het detecteren van emoties in films. "Deze projecten laten zien dat vrijwel alle geesteswetenschappen geschikt zijn voor valorisatie en publiek-private samenwerking", onderstreept Bod. "Lange tijd dachten we dat vooral de computerlinguïstiek commerciële toepassingen had, maar dit blijkt dus ook te gelden voor muziekwetenschap, geschiedenis, kunstgeschiedenis en filmwetenschap. Hier is een wereld te winnen!"

Hierbij hoeft het volgens Bod niet altijd om commercieel bruikbare kennis te gaan. "Het gaat om samenwerkingen tussen bedrijven en universiteiten waar beide iets van kunnen leren en iets aan hebben. Het is sowieso zelden het geval dat een gevaloriseerd stuk kennis ook tot een vermarktbaar product leidt. Maar de projecten van het CDH laten



"Het gaat om erfgoed waar bijna iedereen in geïnteresseerd is" foto Bart van Vliet

zien dat zeer veel kennis valoriseerbaar is, simpelweg omdat het alfa-materiaal zo interessant is. Het gaat immers om erfgoed, waar bijna iedereen, van toeristen tot liefhebbers, in geïnteresseerd is."

Overdonderende rijkdom

Zelf heeft Bod in ieder geval een brede interesse in de geesteswetenschappen. "Na mijn studie Natuurkunde in Utrecht kreeg ik een beurs voor een studie in Rome, aan de Università La Sapienza. Daar kon je alle geesteswetenschappen door elkaar studeren, dat leek me wel iets. Zodoende heb ik vakken als taalkunde, muziekwetenschap, geschiedwetenschap, Latijn en kunstgeschiedenis gevolgd. Het

was prachtig, en het gaf me een gevoel van een overdonderende rijkdom en verscheidenheid in al die geestesproducten en -wetenschappen. Maar goed, ik zat dan ook in het rijkste land ter wereld op het gebied van erfgoed."

Toch denkt Bod dat de Nederlandse faculteiten Geesteswetenschappen daar iets van kunnen opsteken. "De verkoking en fragmentatie van de alfastudies kan wel degelijk worden opgelost. Feitelijk zou iedereen een brede basis moeten krijgen in 'alle' geesteswetenschappen, van taalkunde tot muziekwetenschap en van filologie tot geschiedwetenschap, waarna specialisatie volgt. Men moet zich dan wel vooral concentreren op de onderliggende methodologieën, zoals formele grammatica, stemmatologie, narratologie, bronkritiek en dergelijke. Dat zijn uiterst belangrijke inzichten die voor alle geesteswetenschappers nuttig zijn."

Feitelijk zou iedereen een brede basis moeten krijgen in 'alle' geesteswetenschappen, van taalkunde tot muziekwetenschap en van filologie tot geschiedwetenschap, waarna specialisatie volgt. Men moet zich dan wel vooral concentreren op de onderliggende methodologieën, zoals formele grammatica, stemmatologie, narratologie, bronkritiek en dergelijke. Dat zijn uiterst belangrijke inzichten die voor alle geesteswetenschappers nuttig zijn."

Fuzzy en fragmentarisch

Digitalisering van de geesteswetenschappen is volgens Bod geen expliciete voorwaarde voor succesvolle samenwerking tussen publieke en private partijen. "Ver voor de introductie van de computer waren er al succesvolle publiek-private samenwerkingen. Muziekwetenschappers hebben al sinds de jaren vijftig contracten met platenmaatschappijen voor het schrijven van muziekhistorisch verantwoorde teksten.

Rens Bod

Rens Bod is naast directeur van het Center for Digital Humanities ook hoogleraar Computationale en digitale geesteswetenschappen aan de Universiteit van Amsterdam. In 2010 verscheen zijn boek 'De vergeten wetenschappen: Een Geschiedenis van de Humaniora'.

Kunsthistorici werken al meer dan een eeuw samen met kunstveilinghuizen om toeschrijvingen van schilderijen te verzorgen. En archeologen trekken samen op met bouwbedrijven voor archeologisch bodemonderzoek. Maar het gebruik van digitale technologie heeft de samenwerking tussen geesteswetenschappen en de private sector natuurlijk wel een boost gegeven. Enerzijds is dit te danken aan de explosie van digitaal materiaal dat beschikbaar is gekomen: veel historische, taalkundige, muzikale, kunsthistorische en literaire werken zijn nu digitaal toegankelijk. Anderzijds is het onmiskenbaar dat veel ICT-bedrijven recentelijk meer en meer gefascineerd zijn geraakt door het complexe materiaal waar geesteswetenschappers mee werken. Dat materiaal, zoals schilderijen, manuscripten en historische bronnen zijn fuzzy, complex en vaak ook nog fragmentarisch. Het is een enorme uitdaging hier mee aan de slag te gaan."

Behoeft aan matchmaking

Het CDH is in ieder geval niet de enige dat zich bezighoudt met publiek-private samenwerkingen. "Ook het door NWO uitgevoerde Topsectorenprogramma maakt er binnen de creatieve industrie gebruik van. We moeten nog afwachten of dit een succes wordt, maar de eerste stappen zijn gezet, en het opvoeren van een private partner is in elk geval vrij normaal geworden in de geesteswetenschappen. Denk ook aan het CATCH-programma dat een succes genoemd kan worden op dit terrein. Het zou wel goed zijn om iets als een forum te ontwikkelen om potentiële aanvragers met private partners in contact te brengen. Het valt me op dat daar behoefte aan is. Een eenvoudig matchmaking event kan wonderen doen."

Voorlopig lijkt het zwaartepunt van deze ontwikkeling in Amsterdam te liggen. "Er is een Amsterdam e-Humanities Alliance opgericht door UvA, VU en KNAW met een dozijn private partners. Het is de bedoeling dat deze Alliance in 2015 vloeiend overgaat in het Center for Humanities and Technology (CHAT). Ondertussen heeft het CDH een tweede *Call for Projects* de wereld ingestuurd. De Amsterdamse universiteiten en academia trekken dus met Amsterdamse bedrijven en de Amsterdam Economic Board gezamenlijk op. Hoewel de KNAW natuurlijk een nationaal instituut is, is het haar plan om de geesteswetenschappelijke instituten naar Amsterdam te halen, en kunnen we dus terecht spreken van een 'pan-Amsterdams zwaartepunt'."

cdh.uva.nl

INTERVIEW

OAPEN toont peer reviewed publicaties

Heidi Berkhout

Via OAPEN, een online bibliotheek en publicatieplatform, zijn peer reviewed Europese Open Access-boeken op het terrein van de geestes- en sociale wetenschappen beschikbaar. OAPEN staat voor Open Access Publishing in European Networks, een samenwerkingsverband tussen verschillende Europese universiteiten en uitgevers. OAPEN heeft verschillende Open Access publicatieprojecten voor wetenschappelijke boeken opgezet, zoals OAPEN-NL (in samenwerking met NWO) en OAPEN-UK (in samenwerking met JISC Collections). Naast het ontsluiten van de boeken organiseert OAPEN onder andere conferenties om kennis te delen en te vergroten. De eerstvolgende conferentie vindt plaats in samenwerking met JISC Collections op 1 en 2 juli 2013 in de British Library in Londen, en is bedoeld voor Europese onderzoeksfinanciers, uitgevers, bibliothecarissen en onderzoekers (zie ook de Agenda op pagina 4). oapen.org

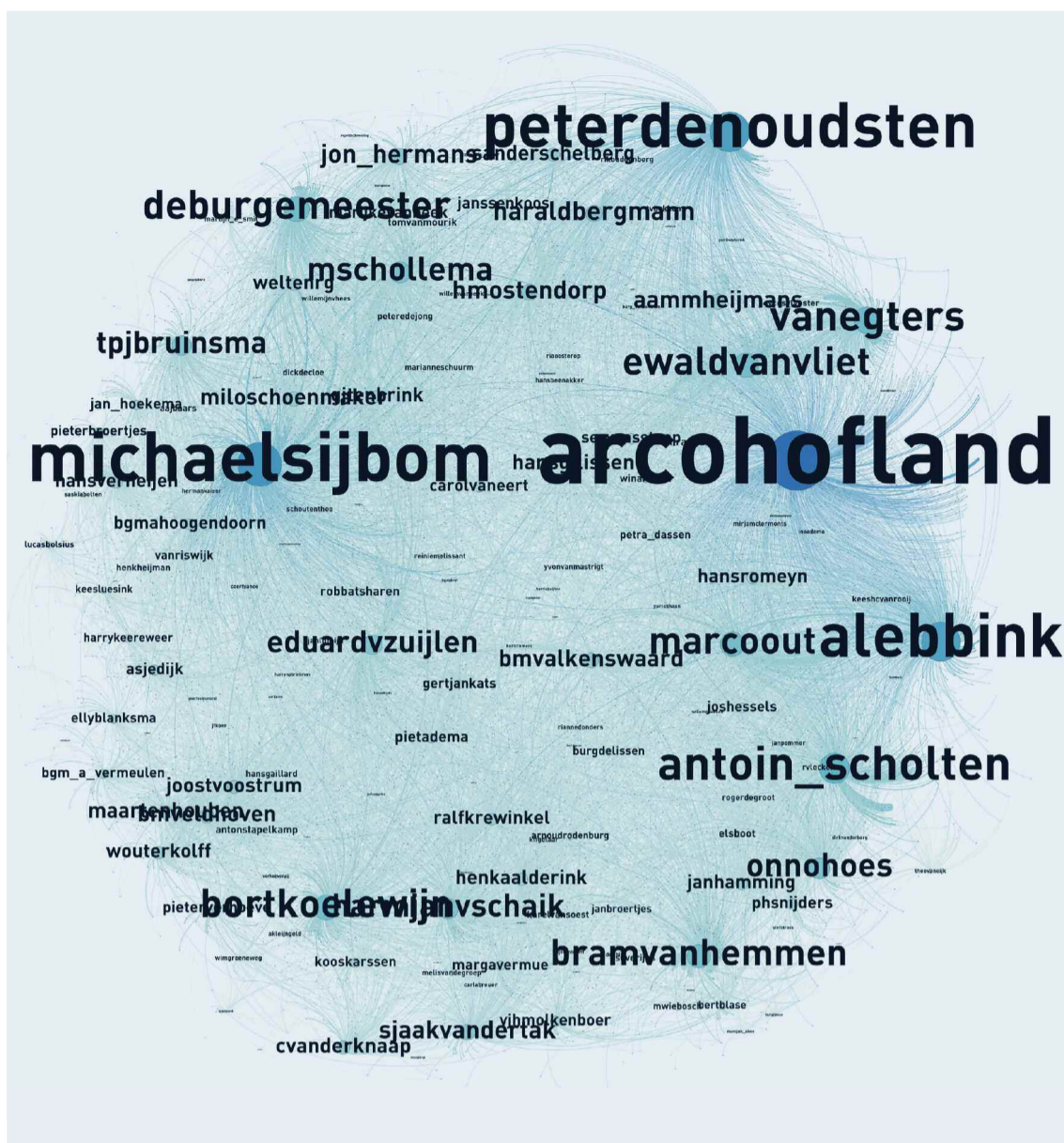


Data over beweeg-, zit- en slaapgedrag geven een unieke combinatie van biomarkers foto Ollyy

Data over inspanning en gezondheid

Joris Mulder

Hoe is lichamelijke inspanning gerelateerd aan gezondheid? Het LISS panel, American Life Panel en de ELSA studie onderzoeken dit door de lichamelijke activiteiten en het slaapgedrag van respondenten te verzamelen met een accelerometer. Dit apparaat, als een horloge om de pols gedragen, meet de versnellingsbeweging en lichaamstemperatuur van de drager. Daarnaast worden data verzameld over beweeg-, zit- en slaapgedrag via vragenlijsten, zodat de resultaten van de accelerometers kunnen worden vergeleken met subjectieve surveydata. Met deze en andere data ontstaat een unieke combinatie van meetbare biomarkers en gezondheidsgerelateerde surveydata waar onderzoekers gebruik van kunnen maken. De eerste data zijn naar verwachting beschikbaar in juli 2013. lissdata.nl



Van dataset tot visualisatie: welke burgemeesters worden door andere twitteraars genoemd in hun tweets? CDA-burgemeester Arco Hofland (Rijssen-Holt) is een actief twitteraar; hij krijgt ook veel 'mentions' credits N. Bergmann, A. Neven en A. Verheul

Visualisaties zijn meer dan mooie plaatjes

Het vinden van betekenisvolle patronen in grote hoeveelheden data is niet altijd eenvoudig. Op de Utrecht Data School leren studenten data analyseren en visualiseren. Erica Renckens

De hoeveelheid data beschikbaar voor onderzoek groeit met de dag. Om studenten met deze data te leren werken, is de Universiteit Utrecht gestart met het onderzoeksproject Utrecht Data School. Initiatiefnemers van het project zijn assistent-professor Mirko Tobias Schäfer, student Thomas Boeschoten en alumnus Lisette van Blokland. De eerste editie van het onderzoeksproject, voornamelijk gevolgd door studenten van de masteropleiding Nieuwe media en digitale cultuur, is onlangs afgerond.

Een frisse blik

“We zijn erg tevreden over deze eerste editie”, aldus een enthousiaste Schäfer. “Er liggen nu vijf onderzoeksrapporten over vijf verschillende projecten met verschillende

vraagstellingen en datasamples.” Voor de onderzoeksvragen benaderden de initiatiefnemers verschillende bedrijven en organisaties, waaronder Unicef, de PO-raad en het Ministerie van Infrastructuur en Milieu. In ruil voor een onkostenvergoeding voor de intensieve begeleiding door docenten gingen de studenten voor hen aan de slag. Schäfer: “Onze studenten leveren een frisse blik op hun data, stellen nieuwe vragen bij bestaande datasets en kijken kritisch naar de resultaten van de analyses. Ook de opdrachtgevers zijn tevreden.” Zo namen studenten de tweets en retweets van burgemeesters en gemeenten onder de loep. Hoewel de

media vaak veel waarde hechten aan zulke Twitter-accounts, blijkt uit de analyse dat ze vooral worden gebruikt voor persoonlijke profilering en niet voor het aangaan van de dialoog met burgers over inhoudelijke vraagstukken. De analyse en visualisatie van de data met programma's als Gephi leveren mooie plaatjes op. “Visualisaties maken patronen van interac-

tie of relatie goed zichtbaar”, vertelt Schäfer. “Maar vaak zijn deze netwerken in eerste instantie te complex om ze eenvoudig af te beelden. Daar is finetuning voor nodig: de studenten nemen beslissingen die tot de uiteindelijke visualisatie leiden.”

Mediawijsheid

Eén van de leerdoelen is dan ook om te leren hoeveel beslissingen en bewerkingen er nodig zijn om van een dataset een begrijpelijke visualisatie te maken. “Op die manier ontwikkelen de studenten ook een kritische houding en mediawijsheid in het werken met de tools en het lezen van visualisaties.” Begin september zal een nieuwe lichting studenten beginnen met de tweede editie van de Utrecht Data School. Hiervoor zijn de benodigde partners al gevonden. “Ons plan is om de Utrecht Data School uit te bouwen tot een interdisciplinair platform met cursussen en individuele onderzoeksprojecten omtrent data”, aldus Schäfer. “Daarnaast zal de Utrecht Data School ook als platform dienen voor internationale onderzoekers die in samenwerking met studenten en docenten projecten uitvoeren.” dataschool.nl

FOCUS

Te complex om eenvoudig af te beelden

KORT

Research Data Netherlands van start

Begin mei is het startschot gegeven voor het samenwerkingsverband Research Data Netherlands. Initiatiefnemers 3TU.Datacentrum en DANS bundelen met deze coalitie hun krachten op het gebied van duurzame data-archivering. De coalitie staat ook open voor andere partijen. Peter Doorn, directeur van DANS, benadrukt: “Research Data Netherlands brengt ons een stapje verder richting onze gedeelde missie: het bevorderen van duurzame toegankelijkheid en verantwoord hergebruik van wetenschappelijke onderzoeksgegevens. Daarvan profiteert uiteindelijk de hele wetenschap.” (CvZ) dans.knaw.nl

Laatste projecten

CLARIN-NL van start

Afgelopen voorjaar zijn elf projecten uit de vierde en laatste projectronde van CLARIN-NL van start gegaan. Twee van deze projectvoorstellen zijn geschreven op uitnodiging van CLARIN-NL om voldoende spreiding over de verschillende geesteswetenschappelijke disciplines te garanderen. In de projecten zullen bestaande tools en data geschikt gemaakt worden voor opname in de Europese onderzoeksinfrastructuur voor geesteswetenschappers die CLARIN ontwikkelt. CLARIN-NL, de Nederlandse tak van CLARIN, loopt nog tot eind 2014. De resterende tijd zal gebruikt worden aan de afronding van de infrastructuur. (ER) clarin.nl

Resultaten onderzoek gebarentaal online

Afgelopen voorjaar is de website gebareninzicht.nl gelanceerd, ontwikkeld door taalwetenschappers van de Radboud Universiteit. Op deze site worden de resultaten van onderzoek naar gebarentaal gepresenteerd in gebarentaal, zodat ook de doorgemeenschap kennis kan nemen van deze gegevens in haar moedertaal. Gebarentaalonderzoekers zullen de site doorlopend aanvullen met nieuwe onderzoeksresultaten. Gebareninzicht is het eerste project dat tot stand is gekomen met een MeerWaarde-subsidie van NWO. Deze subsidie heeft als doel onderzoeksresultaten toepasbaar te maken buiten de wetenschap en zo de betrokkenheid van derden bij onderzoek te vergroten. (ER) gebareninzicht.nl



Het gebaar 'binnen' credits gebareninzicht.nl

Onderzoekers en uitgeverij experimenteren met visualisaties

Bewegend beeld in oude drukken

Internet maakt interactieve, geanimeerde en driedimensionale illustraties mogelijk.

Maar hoe maak je ze deel van jouw publicatie?

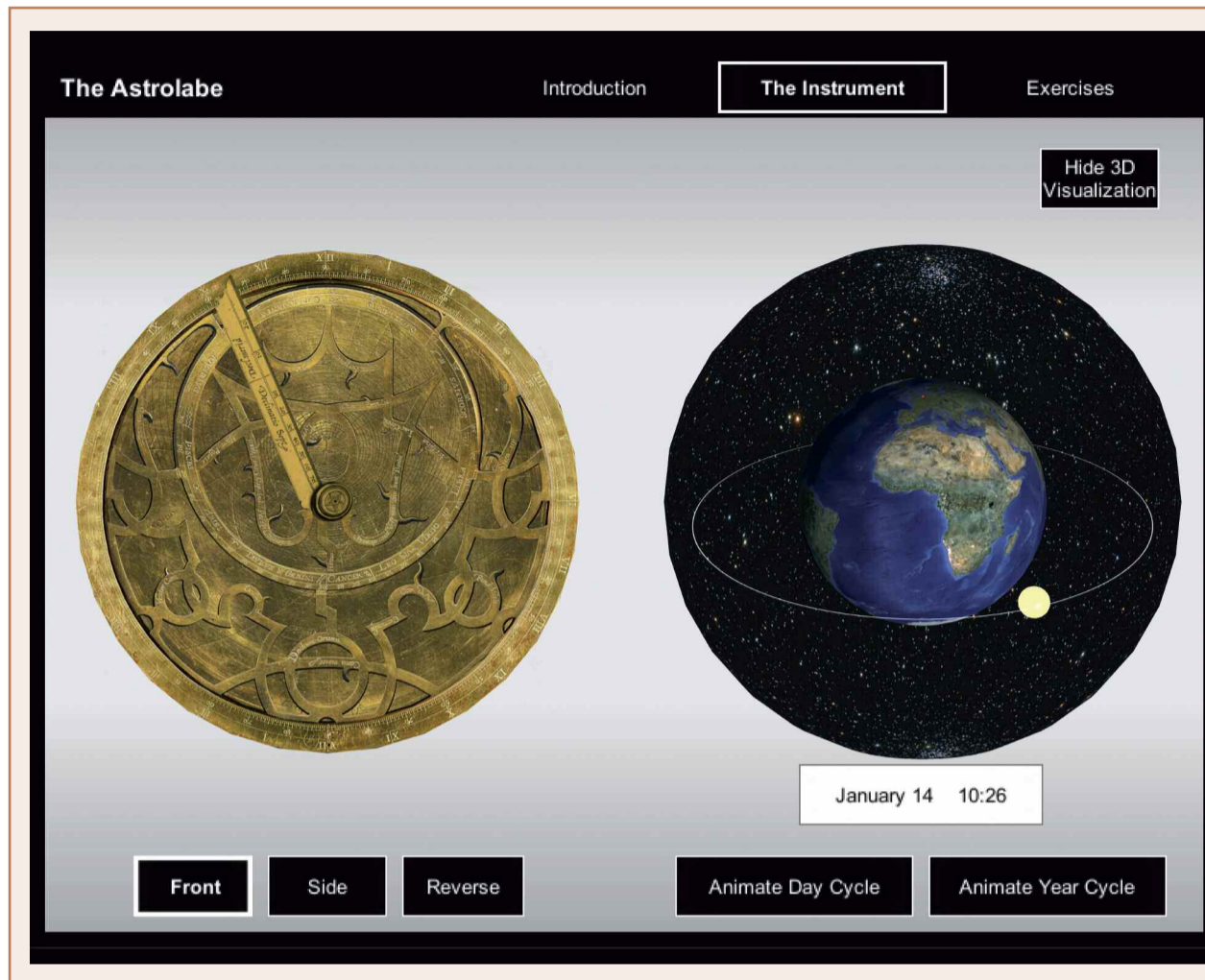
En zijn deze 'verrijkingen' over tien jaar nog te bekijken?

Maarten Hoogerwerf

Binnen de wetenschapsgeschiedenis loopt Tiemen Cocquyt regelmatig tegen dit probleem aan. Als curator bij Museum Boerhaave en als tijdelijk onderzoeker bij het Huygens ING tracht hij de historische context van wetenschappelijke instrumenten en hun principes op te helderen. "Bij het overbrengen van een verhaal wil je de technische achtergrond zo helder mogelijk aan de lezer overbrengen, of dit nu een gespecialiseerde onderzoeker of iemand uit een breder publiek is. Maar vaak loopt dat spaak op lastige beschrijvingen. Digitale middelen bieden daarin nu een uitkomst. Neem de astrolabe: de manier waarop dit vroegmoderne tijdmeetinstrument een model is van de hemelbewegingen, wordt meteen duidelijk wanneer instrument en hemelbeweging interactief naast elkaar worden gepresenteerd."

Extra dimensies

Ook uitgeverij Brill herkent de uitdaging van het overbrengen. Eén beeld zegt immers meer dan duizend woorden. "Brill helpt onderzoekers in de humaniora om online extra dimensies toe te voegen aan een boek, tijdschriftartikel of naslagwerk," zegt Michiel Thijssen van uitgeverij Brill. "Naast statische



In één oogopslag

Eén van de door het projectteam uitgewerkte zes casussen. Het betreft een astrolabe, een tijdmeetinstrument waarvan de werking uitvoerig aan bod kwam in de zestiende- en zeventiende-eeuwse sterrenkundige literatuur. Door het interactief te presenteren, wordt haar relatie met de planeetbewegingen zonder veel woorden duidelijk. *credits* Charles van den Heuvel, Huib Zuidervaart, Tiemen Cocquyt (Huygens ING), Michiel Thijssen (Brill), Dylan Nagel (Wild Card), Leen Breure, Heleen van de Schraaf (DANS)

tekst, datasets en illustraties, kunnen auteurs nu ook interactieve, geanimeerde en driedimensionale elementen aanbieden. Het Dynamic Drawings-project leert ons hoe we die kunnen publiceren."

Het gaat om het verhaal

Binnen dit project werkt Tiemen nauw samen met Dylan Nagel, game-ontwikkelaar bij Wild Card,

aan het ontwikkelen van de visualisaties. Dylan: "Techniek is belangrijk, maar vooral ondersteunend. Het gaat met name om het verhaal: wat wil je vertellen en op welke manier? Je moet een balans vinden tussen wetenschappelijke juistheid en lezersgemak. Binnen het project hebben we geëxperimenteerd met de mate van interactie: van simpele animatie tot volledig werkend

instrument."

De focus van het projectteam ligt op zes casussen welke in korte iteraties worden uitgewerkt. Een voorbeeld van deze casussen is de astrolabe. Bij het uitwerken van dergelijke casussen wordt snel duidelijk dat er een spanning bestaat tussen de oude teksten en een visualisatie. Is de visualisatie ondergeschikt aan één tekst, of wordt deze uit meerdere bronnen samengesteld waardoor de visualisatie zich ontwikkelt tot een zelfstandig product?

Duurzaamheid

DANS, als vierde en laatste partij betrokken bij het project, juicht het toevoegen van visualisaties en onderzoeksdata aan een publicatie toe,

maar waakt wel over de duurzaamheid: wanneer een driedimensionale visualisatie een essentieel onderdeel vormt van een publicatie, is deze dan over tien jaar nog terug te vinden en te bekijken? Samen onderzoeken de partijen hoe Brill geavanceerde visualisaties bij publicaties kan aanbieden en hoe DANS op de achtergrond kan zorgdragen voor duurzame koppelingen en voor duurzame archivering.

Het einde van het project is bijna in zicht. De eerste resultaten zijn reeds zichtbaar. Voor nu en de toekomst ligt de nadruk vooral op het maken van goede afspraken zodat de belangen van onderzoeker, uitgever én archief goed worden behartigd.

demo-brill.dans.knaw.nl

Dynamic Drawings

Het Dynamic Drawings-project wordt gefinancierd en uitgevoerd door Brill, Wild Card en KNAW (Huygens ING en DANS). Het is één van de projecten van het Centre for Digital Humanities van UvA, VU en KNAW waarmee publiekprivate samenwerking op het gebied van Digital Humanities wordt gestimuleerd. Het project loopt tot eind juli 2013.

cdh.uva.nl

Gericht zoeken naar die ene variabele

Hoe vind je die ene variabele in het enorme aanbod online data? 'Bladeren' door vele datasets werkt voor veel onderzoekers ontmoedigend. Lost een nieuwe manier van datadiseminatie dit op? *Marika de Bruijne*

Alleen een dataset op een website zetten, maakt de details van een studie nog niet vindbaar. Daarvoor dienen zowel de beschrijvingen van individuele variabelen als het achterliggende onderzoeksinstrument op een gestructureerde manier gedocumenteerd te worden. Daarna kunnen ze doorzoekbaar worden gemaakt, bijvoorbeeld met vrije zoektermen of op onderwerp.

Behalve voor het vinden, is het ook voor het begrijpen van de data van belang hoe diep

men kan 'meekijken' in de datasets zonder de data daadwerkelijk te downloaden. Analyses op de data zijn alleen goed uit te voeren als de variabelen in relatie tot de achterliggende vraag en in hun volledige context getoond worden.

Een eigen datamand

CentERdata, DANS en het SCP (Sociaal en Cultureel Planbureau) werken aan een nieuwe manier van datadiseminatie. Samen stellen ze de data van het onderzoek Culturele Veranderingen van het SCP via een online vragenbank beschikbaar zodat deze beter ontsloten worden. De data van deze (twee)jaarlijkse studie zijn momenteel al beschikbaar via het online archiveringssysteem EASY, maar de nieuwe vragenbank opent veel meer moge-

lijkheden. Waar het voor onderzoekers op dit moment veel tijd kost om te zoeken naar individuele vragen en variabelen, kan dit straks in een paar muisklikken.

Questasy, de applicatie hierachter, is ontwikkeld door CentERdata. "Typerend voor Questasy is dat je er ook longitudinale data in kunt onderbrengen omdat het gebaseerd is op de DDI data-lifecycle standaard", vertelt Marion Wittenberg, de betrokken projectleider bij DANS. Deze standaard legt de nadruk op het in kaart brengen van de hele levenscyclus van data: van het onderzoeksplan tot de uiteindelijke publicatie.

"Verder heeft de applicatie een basketfunctie, oftewel een digitale mand, waarmee onderzoekers hun eigen dataset kunnen samenstellen." In deze mand kan de onderzoeker de

voor hem relevante variabelen uit de vragenbank bijeenrapen en is het niet meer nodig gehele datasets te downloaden om deze vervolgens zelf te koppelen.

Voor de toekomst denkt Wittenberg aan een grote centrale vragenbank met meerdere bekende studies. "Ik zou het heel mooi vinden als meer Nederlandse surveydata op deze manier zouden worden gedissemineerd", zegt ze. "Als de vraagteksten vertaald worden, zou het materiaal ook internationaal gebruikt kunnen worden." Financiering voor een dergelijke uitbreiding is er echter voorlopig niet. Het blijft veel tijd en geld kosten om het zoekwerk voor onderzoekers makkelijker te maken. Maar met het gezamenlijke project is er wellicht weer een stap gezet.

centerdata.nl/nl/ict-toepassingen/questasy

Schatgraven in de statistieken

De website met historische CBS-publicaties is een half jaar online. Tijd voor de eerste gebruikservaringen en een blik op de toekomst.
Margriet van Stiphout

De historische collectie van de bibliotheek van het CBS (Centraal Bureau voor de Statistiek) bestaat uit zo'n 5 miljoen handgeschreven en gedrukte pagina's met statistische tabellen en bijbehorende toelichtingen. Al deze documenten worden stapsgewijs gedigitaliseerd en beschikbaar gesteld via cbs.nl/historisch. De collectie wordt hiermee toegankelijk voor een groter publiek en is tegelijkertijd beter doorzoekbaar dankzij onder andere full-text zoekmogelijkheden. Tevens kunnen tabellen gemakkelijker overgenomen worden.

Positieve ervaring

Dr. J(eroen) J. Euwe, werkzaam bij de Universit  Libre de Bruxelles, maakt voor zijn onderzoeken regelmatig gebruik van de CBS-collectie. Zo komen internationale handelsstatistieken, die teruggaan tot 1846, goed van pas bij zijn onderzoek naar de Nederlandse kunstmarkt in crisistijden in de eerste helft van de twintigste eeuw. Hij schetst grote voordelen van het digitaal brengen. Zo verkort de digitalisering de voorbereidingstijd van zijn onderzoek, doordat hij niet meer stad en land hoeft af te reizen



Het projectteam Digitalisering Historische CBS-publicaties v.l.n.r.: Henk Laloli, Margriet van Stiphout (projectcoördinator) en Alfred Moest foto Sjoerd van der Hucht

om archieven te raadplegen. Ook ervaart hij een drastische vermindering van papiergebruik en enorme ruimtewinst in zijn werkkamer, nu hij geen papieren kopieën hoeft te bewaren. De mogelijkheid tot delen van informatie met collega's via links en/of afbeeldingen ervaart hij ook als positief.

Digitaal hergebruik

Ook benadrukt Jeroen het digitaal

hergebruik van statistische tabellen, ook al is de techniek hiervoor nog niet optimaal. Verder roemt Jeroen het meedenken en persoonlijke advies van de medewerkers van de bibliotheek, die de collectie kennen en hem waardevolle informatie geven voor zijn onderzoek. Dat zal hij wel gaan missen als de volledige collectie over enkele jaren digitaal is.

Er lopen verschillende onderzoek-

ken naar de mogelijkheden om de data van bijvoorbeeld de volkstellingen als zogenoemde open en linked databestanden aan te bieden. Dat zou betekenen dat geïnteresseerden niet meer alleen de tabellen kunnen raadplegen, maar ook de afzonderlijke data kunnen gebruiken. Dat biedt nog meer mogelijkheden voor onderzoekers en andere gebruikers van CBS-data.
cbs.nl/historisch

COLUMN

Mijn eigen digitale kwaliteitskrant

In de loop van de tijd ben ik kranten steeds selectiever gaan lezen. Bestede ik vroeger zeker een uur per dag aan mijn krant, tegenwoordig vlieg ik er vaak binnen twintig minuten doorheen. Het belangrijkste nieuws heeft me in de loop van de dag al bereikt. Gehoord op de radio of gelezen op internet of op mijn smartphone. Laatst moest ik een stapel vakantiekranen bijlezen – de krant onderweg downloaden was niet gelukt. Met zoveel kranten voor je neus lees je nog selectiever, maar hoe dan ook vond ik het confronterend om te merken hoeveel bladzijden ik ongelezen omsloeg. Hele bijlagen legde ik ongelezen opzij. Al dat tijdrovende en zinloze geblader maakte fysiek duidelijk wat ik al langer vaststel: ik vind papieren kranten in een rap tempo uit de tijd raken. Nou horen we al langer dat de papieren krant op sterven na dood is, en inderdaad: de cijfers

geven een zeer somber beeld. In de VS verdween sinds 2007 ruim dertig procent van de banen in de krantenindustrie. En maar liefst 286 lokale kranten sloten daar hun deuren. Ook bij Nederlandse kranten gingen de afgelopen jaren duizenden banen verloren. Wat mij zo verbaast, is dat kranten nauwelijks lijken in te springen op een trend die nu al jaren zichtbaar is: personalisatie. Op internet kun je zo'n beetje alles op maat laten maken, maar veel digitale kranten zijn niet meer dan een replica van de papieren krant. Je kunt er niet doorklikken, niks highlighten en je kunt 'm niet aan je eigen wensen aanpassen. Wat ik zou willen is dit: een digitale betaalkrant voor op mijn pc, smartphone en tablet, met stukken uit de beste Nederlandse en buitenlandse kranten over onderwerpen en thema's die ik zelf kan bepalen door ze aan of uit te vinken. Er mag een algemeen gedeelte bij zitten, een keuze van de redactie om mij te verrassen,

maar verder bepaal ik graag zelf welke columnisten ik lees, welke bijvoegsels ik kan missen en in welke onderwerpen ik mij echt wil verdiepen. De voordelen lijken mij evident: geen hoge drukkosten meer, geen distributieproblemen, maar een flexibele, snel te actualiseren krant, die ik via mijn persoonlijk profiel doorlopend op maat kan snijden. Niet te dik, meer analyses en achtergronden dan nieuws, want voor veel nieuws ben ik al tevreden met de gratis nieuwsdiensten op internet. Ik zou willen kunnen highlighten, waar mogelijk wil ik worden verwezen naar digitale bronnen of dossiers



over hetzelfde onderwerp, ik zou graag makkelijk kunnen zoeken in het archief (geef mij alle stukken van de afgelopen drie maanden over het land waar ik nu ben of naartoe ga) en ik zou graag artikelen willen kunnen opslaan. Ik heb zelf niet zo'n behoefte om te reageren, maar ook dat is een trend, dus die mogelijkheid moet er zeker bij.

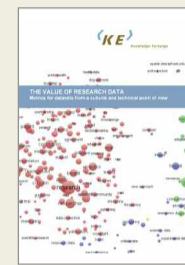
En ja, ik zou voor dit alles met liefde betalen. Waarschijnlijk net zo veel als voor mijn papieren krant, hoewel de beste op maat te snijden digitale krant niet van één krantenuitgever komt, maar per land of taalgebied van samenwerkende kwaliteitskranten. Het is te hopen dat de huidige kwaliteitskranten lang genoeg blijven bestaan om dit te realiseren.

Ewoud Sanders
Taalhistoricus en journalist.
Sanders is medewerker van NRC Handelsblad en Onze Taal.

foto Leo van Velzen

GELEZEN

Report 'The Value of Research Data. Metrics for datasets from a cultural and technical point of view' Rodrigo Costas, Ingeborg Meijer, Zohreh Zahedi and Paul Wouters of the Center for Science and Technology Studies (CWTS).
Leiden University
René van Horik



Gegevensuitwisseling is belangrijk voor wetenschappelijke vooruitgang en kennisbevordering. Er zijn echter nog steeds beper-

kingen en belemmeringen waardoor deze uitwisseling niet massaal plaatsvindt. Dit rapport geeft een aanzet om hier verbetering in aan te brengen. Centraal in het rapport staat een beloningssysteem op basis van datametrics, het stimuleren van het delen van onderzoeksdata. Ook behandelt het rapport de kosten van gegevenspublicatie. Het rapport adviseert een standaard te ontwikkelen voor het opslaan, publiceren, identificeren en citeren van datasets. Het rapport verkent de mogelijkheden om metrics toe te passen op datasets om een effectief beloningssysteem op te zetten. Dit beloningssysteem moet rekening houden met de belangen van degenen die bij het proces betrokken zijn. De publicatie bevat verder een overzicht van de huidige literatuur op het gebied van gegevensuitwisseling en data metrics. Ook de resultaten van gesprekken met de belangrijkste betrokkenen zijn opgenomen. Volgens zijn bestaande repositories en tools op het gebied van gegevensuitwisseling in kaart gebracht. Deze publicatie is interessant voor belanghebbenden in de wetenschap, voor onderzoeksfinanciers en wetenschappelijke instellingen. Het biedt inzicht in het instrumentarium dat beschikbaar is voor het stimuleren en belonen van gegevensuitwisseling en -publicatie binnen de eigen academische gemeenschap.

knowledge-exchange.info/datametrics

KORT

Nieuwe sponsor voor E-data&Research

De bestuurs- en redactieleden van dit blad verwelkomen het RIVM (Rijksinstituut voor Volksgezondheid en Milieu) als nieuwe sponsor van E-data&Research. Dit Nederlandse overheidsinstituut verzorgt informatie, monitoring en wetenschappelijke onderbouwing van het volksgezondheidsbeleid van de Nederlandse overheid. Wilt u ook meer weten over het sponsoren van dit blad? Kijk voor meer informatie op de website www.edata.nl of neem contact met ons op via edata@dans.knaw.nl.
(HB)
rivm.nl