



## E-DATA & RESEARCH

Jaargang 6 | 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: DANS, CentERdata, CBS, CLARIN-NL, Huygens ING, Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica

## INHOUD

- 3 Virtueel samenwerken steeds gemakkelijker
- 3 Loe de Jong binnenkort nog beter op het net
- 4 Een goede manier om data te bewaren is verboden
- 4 Oude korfbalclubs nu in handig register
- 5 Bas Savenije van de KB wil alles digitaal
- 6 De humanities slaan steeds meer het e-pad in
- 7 De hits van het Centraal Bureau voor de Statistiek
- 8 Ewoud Sanders wil iets doen voor de KB

### COLOFON

**Uitgever:** Stichting Uitgeverij E-data & Research Den Haag  
**Redactieadres:** Postbus 93067, 2509 AB Den Haag, 070-3446484 edata@dans.knaw.nl  
**Redactie:** Inge Angevaere, Eric Balster, Heidi Berkhout, Ronald van der Bie, Peter Boot, Warna Oosterbaan (hoofd-/eindredacteur), Thijs Hermsen, René van Horik, Erica Renckens  
**Redactiesecretariaat:** Lucas Pasteuning  
**Aan dit nummer werkten mee:** Ingrid Dillo, Ilya van Marle, Ewoud Sanders, Steamwork Graphics, Maarten Streefkerk, Milja van Tielhof, Leo van Velzen, Rosa Vitale  
**Opmaak:** Colette Sloots, Haarlem  
**Productie:** Amsterdam University Press  
**Druk:** Ten Brink, Meppel  
**Oplage:** 9100  
**ISSN:** 1872-0374  
 E-data & Research is online te raadplegen op [www.edata.nl](http://www.edata.nl). Toezending is kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen.

Inscannen van oude handschriften maakt ze voor iedereen toegankelijk. Maar de computer kan er dan nog niet in zoeken. Honderden vrijwilligers die de documenten woord voor woord overtypen lossen dit probleem op. Twee praktijkvoorbeelden.

# Zonder handwerk lukt het nog niet

**Handgeschreven Militie-registers worden thuis overgetikt.** Erica Renckens

De laatste jaren hebben archieven, bibliotheken en onderzoeksinstituten veel tijd en geld gestoken in het digitaliseren van hun collecties. Hoewel de resultaten van Optical Character Recognition (OCR) lang niet perfect zijn, is veel van het gedrukte materiaal al redelijk goed te doorzoeken met behulp van deze techniek. Voor oude drukken en handschriften is de situatie een stuk minder rooskleurig. Je kunt ze wel inscannen en daarna op een scherm bekijken, maar de computer kan er dan nog niet in zoeken. Verschillende nieuwe technieken en *crowd-sourcing*-projecten moeten hier verandering in brengen.

Afgelopen november lanceerde het



Henny van Schie van het Nationaal Archief foto Leo van Velzen

Stadsarchief Amsterdam de website VeleHanden.nl. Via deze website kan iedereen helpen om de Militie-registers online doorzoekbaar te maken. In deze handgeschreven registers staan de namen en gegevens

van alle dienstplichtige mannen van 1815 tot en met 1941. De gegevens van elke scan worden door twee vrijwilligers overgetypt, waarna een controleur zich ervan verzekert dat alles correct is overgenomen. Als

dank krijgen de vrijwilligers toegang tot informatie uit de Militie-registers over hun eigen familie. Eind december, amper twee maanden na lancering, had VeleHanden.nl al bijna 800 leden die gezamenlijk bijna 150.000 scans hadden ingevoerd. Het Stadsarchief is overweldigd door dit resultaat en is er van overtuigd dat deze vorm van *crowd sourcing* de toekomst is voor de culturele sector. Het Amsterdamse archief is niet de enige die de crowd inzet om data te ontsluiten; aan het Meertens Instituut lopen geregeld vrijwilligersprojecten, momenteel rond de Gekaapte Brieven.

### Vrijwel foutloos

Het grote voordeel van de inzet van een grote hoeveelheid vrijwilligers is dat de gegevens zonder veel kosten volledig beschikbaar worden, en toch vrijwel foutloos worden getranscribeerd. Nadeel van het woord voor woord overtypen is de grote inspanning die geleverd moet worden: als 'Jansen' tienduizend keer voorkomt, zal hij ook tienduizend keer ingetypt moeten worden. "Die arbeidsintensieve aanpak is voor zulke projecten onvermijdelijk, omdat je streeft naar volledigheid", aldus Henny van Schie, archivaris bij het Nationaal Archief. Zelf werkt hij binnen het CATCH-Plus-project Scratch4All aan een computerprogramma dat digitale handschriften met een relatief ge-

VERVOLG OP PAGINA 3

# De buit bestaat uit bits en bytes

**Gekaapte brieven en verhoren toegankelijk gemaakt voor onderzoek.** Ingrid Dillo

The National Archives (TNA) in Kew (Londen) bewaart het archief van de High Court of Admiralty. Dit archief bevat naar schatting veertigduizend in het Nederlands geschreven brieven en documenten. Dat is een deel van de buit die Engelse kapers in de zeventiende en achttiende eeuw veroverden op Nederlandse schepen. Die buit bevat brieven van en aan zeelieden en kooplieden en hun familieleden, scheepsjournalen, aanbevelingsbrieven, ladingboekjes en kwitanties. Eeuwenlang lagen ze in de donkere kelders van het High Court of Admiralty, zonder dat iemand er naar omkeek. Sommige brieven zijn tot op de dag van vandaag niet geopend.

### Papieren erfgoed

Een klein deel van dit materiaal is in het kader van het nationale programma voor het behoud van het papieren erfgoed, Metamorfoze, gedigitaliseerd. Dit heeft een kleine negenduizend scans opgeleverd.

Taalkundige Nicoline van der Sijs heeft samen met het Meertens Instituut en met subsidie van het Prins Bernard Cultuurfonds het initiatief genomen deze scans te ontsluiten en te transcriberen. "Het werk wordt uitgevoerd door een grote groep vrijwilligers. Een kleine honderdvijftig enthousiaste medewerkers is inmiddels aan de slag."

Volgens Van der Sijs is deze vorm van *crowd sourcing* een groot succes en zijn de resultaten van uitstekende kwaliteit, mede doordat vrijwilligers elkaars werk controleren. "Om de scans te ontsluiten voor wetenschappelijk onderzoek, worden ze in het project voorzien van inhoudelijke metadata. Daarnaast transcriberen de vrijwilligers zoveel mogelijk scans. Zo ontstaat er een interessant corpus voor onderzoek en wordt het bijvoorbeeld mogelijk het achttiende-eeuwse taalgebruik van de gewone man te vergelijken met de meer formele taal die werd gebezigd in bijvoorbeeld literaire teksten."

Waren de verschillen echt zo groot of hebben we dat wellicht altijd overschat? De waarheid kan binnenkort boven tafel komen", zegt Van der Sijs.

<http://tinyurl.com/cmwrvtv>

### Het Prize Papers Project

De Engelse kapers beperkten hun buit niet tot Hollandse schepen. In twee eeuwen zijn tienduizenden Franse, Spaanse, Portugese, Deense, Zweedse, Duitse, Italiaanse en Amerikaanse schepen in Engelse havens opgebracht. Om te bepalen of een schip rechtmatig was veroverd werd er een proces gevoerd en werd de bemanning ondervraagd. Functionarissen van het Britse High Court of Admiralty zorgden voor de administratieve begeleiding. De verhoren van de gevangen genomen bemanningsleden vonden plaats aan de hand van een standaard vragenlijst van

achtien en later zelfs vierendertig vragen over herkomst, route, bestemming, tonnage, lading en de eigenaren van het schip en over herkomst, nationaliteit, leeftijd en migratiegeschiedenis van de bemanning. De antwoorden werden door tolken vertaald en in het Engels genoteerd.

Dit materiaal levert een schat aan informatie op die zeer waardevol is voor de maritieme geschiedenis, maar bijvoorbeeld ook voor migratiestudies en de vroege Amerikaanse geschiedenis. De oudste Nederlandse academische uitgeverij, Koninklijke Brill NV, heeft het initiatief genomen deze internationale bron te digitaliseren en te voorzien van inhoudelijke metadata. Brill-directeur en maritiem historicus Perry Moree vindt dat de bron uitstekend past in Brill's online research collecties. In totaal zullen er ongeveer 300.000 scans worden gemaakt. Het Prize Papers Project zal in fasen worden uitgevoerd en start met de achttiende eeuw. Het is de bedoeling dat het eerste deel van het materiaal in de tweede helft van 2012 als betaalde online database beschikbaar is voor wetenschappelijk onderzoek.



De inhoud van één van de dozen uit het archief van de High Court of Admiralty. De documenten bevatten de verhoren van door Engelse kapers gevangen genomen zeelieden.