

NWO haalt de teugels weer strakker aan

# ‘Die data zijn niet van jou’

**De regels van Nederlands grootste onderzoeksfinancier voor open data en open access worden steeds strenger.**

*Warna Oosterbaan*

NWO wordt mede-eigenaar van de data die in een gesubsidieerd onderzoek worden geproduceerd. De instelling kan die data dan weer aan andere onderzoekers ter beschikking stellen. Dit staat in de nieuwe Regeling Subsidieaanvragen van NWO, die op 1 mei van kracht werd en die voor alle wetenschappelijke disciplines geldt. De onderzoeksfinancier wil op die manier het hergebruik van data bevorderen.

Volgens Ron Dekker, directeur Instituten bij NWO leidt hergebruik tot een beter rendement op wetenschappelijke resultaten, en voorkom je zo dat voor het produceren van onderzoeksgegevens “twee keer betaald moet worden”. NWO verstrekt per jaar ruim 550 miljoen euro aan onderzoekssubsidies.

Bij het beoordelen van onderzoeksvorstellen zal er voortaan ook strenger op gelet worden of

de data die de onderzoekers willen verzamelen, misschien al zijn te vinden in de dataset van een ander onderzoek. “Als dat zo is, dan is het einde oefening”, zei Dekker op de ‘Open Onderzoeks-Data Dag’ die op 18 mei door Surffoundation was georganiseerd.

Dekker maakte op die dag bekend dat NWO naast het invoeren van deze regels een campagne gaat starten waarin onderzoekers erop wordt gewezen dat hun onderzoek met publieke middelen wordt gefinancierd. “En we gaan ze ervan bewust maken: die data zijn niet van jou. Ze zijn van de instelling en van NWO. En daarom moeten die data goed en toegankelijk worden gedocumenteerd en opgeslagen”.

## Steun in de rug

Peter Doorn, directeur van data-archief DANS zegt in een reactie de nieuwe bepaling “een steun in de rug van DANS” te vinden. “Wij kunnen NWO helpen door data-overeenkomsten met de subsidieontvangers te sluiten. Dat doen we al voor de investeringssubsidies bij de gebieden Geesteswetenschappen en Maatschappij- en Gedragwetenschappen, maar we breiden dat met liefde uit naar de andere vakgebieden. Daarbij

zullen we het credo hanteren: ‘open als het kan, beschermd als het moet’.”

In de nieuwe voorwaarden zijn ook de regels die betrekking hebben op de vrije toegang tot wetenschappelijke publicaties (*open access*) verder aangescherpt. Auteursrechten mogen niet worden overgedragen als open access daardoor onmogelijk wordt. Een andere regel stelt dat NWO ter bevordering van de kennisoverdracht “nadere voorwaarden kan stellen aan de auteursrechten”.

In een toelichting laat de voorlichter van NWO weten dat hier gedacht kan worden aan een boek voor een breder publiek over een bepaald onderwerp. “NWO dient dan te beschikken over de nodige licenties om de door haar gesubsidieerde publicaties te mogen gebruiken voor het uit te geven boek.” Impliceert ‘gebruiken’ dan letterlijke overname? NWO: “Ja, overname van het artikel in zijn oorspronkelijke staat.” En als de uitgever van het oorspronkelijke artikel dat niet toestaat? NWO: “Als er uit een programma een artikel voortkomt dat ter publicatie wordt aangeboden, dan moet op dat moment al duidelijk zijn dat er op het gebruik van de inhoud een claim ligt. Uiteraard heeft de uitgever dan de mogelijkheid om af te zien van publicatie.”

**Zoekmachine vindt relevante WO2-verwijzingen in Handelingen der Staten Generaal**

## Dat doet denken aan de oorlog

**Verwijzen naar de Tweede Wereldoorlog is een beladen kwestie in het politieke debat. Het lijkt wel of politici dat steeds vaker doen. Maar is dat ook zo? Erica Renckens**

Het NIOD Instituut voor Oorlogs-, Holocaust- en Genocidestudies ontwikkelt samen met de Universiteit van Amsterdam een intelligente zoekmachine die automatisch relevante verwijzingen naar de Tweede Wereldoorlog vindt in de Handelingen der Staten Generaal. Zo kunnen onderzoekers precies zien wanneer een politicus in welke context en met welk doel de oorlog aanhaalde.

“Tot nu toe was dat niet goed mogelijk”, zegt Hinke Piersma, onderzoekster bij het NIOD. Een normale zoekmachine gaat alleen uit van het ingetypte trefwoord en geeft daarvoor ook veel resultaten die niet relevant zijn. Bijvoorbeeld dat het brood sinds de Tweede Wereldoorlog veel duurder is geworden. “Deze zoekmachine kan dat wel, omdat hij alle woorden uit het document meeneemt in zijn analyse. Zo kunnen we inzicht verschaffen in de wijze waarop de Tweede Wereldoorlog doorwerkt in de Nederlandse politieke arena.” Het zoekprogramma wordt ontwikkeld in het kader van CLARIN, de Europese digitale onderzoeksomgeving voor geesteswetenschappers.

“Onze zoekmachine kijkt naar alle



**Mei 1945. Een lid van de Binnenlandse Strijdkrachten gebruikt een foto van Hitler als schietschijf** foto Beeldbank WO2, Verzetsmuseum

woorden die eerder door training met de Tweede Wereldoorlog geassocieerd zijn”, aldus Piersma.

## Wel of niet relevant

In deze training hebben historici voor een klein gedeelte van de Handelingen der Staten Generaal aangegeven of de documenten wel of

niet verwijzingen naar de oorlog bevatten. Op basis hiervan is een onderwerpspecifiek model getraind dat automatisch van de overige documenten kan bepalen of zij wel of niet relevant zijn voor het onderwerp Tweede Wereldoorlog. “Bij deze analyse gebruiken we alle woorden die tijdens de training in de

“Als het hebben van een dubbele nationaliteit de weg zou openen naar verraad en disloyaliteit, dan stel ik daar tegenover dat iemand als Mussert, voor zover ik weet, niet de Duitse nationaliteit had.”

*Door de zoekmachine gevonden citaat van Mohammed Rabbae (GroenLinks) tijdens een debat op 22 februari 1995.*

als relevant geannoteerde documenten voorkwamen én de woorden uit de niet-relevante documenten”, legt Piersma uit. “Een woord dat in beide categorieën ongeveer even vaak voorkomt beschouwen we als ‘neutraal’. Komt een woord onevenredig vaak voor in één van beide collecties dan zien we dat als een sterke aanwijzing om het document als ‘relevant’ of juist als ‘niet-relevant’ te categoriseren.”

Een veelgehoord geluid is dat het hedendaagse politieke debat veruwt, met meer verwijzingen naar de Tweede Wereldoorlog. “Deze vergelijkingen zijn er echter altijd geweest”, zegt Piersma. “Maar door de hele naoorlogse periode in het onderzoek te betrekken is het wellicht mogelijk om trends en patronen hierin te laten zien.”



**E-DATA & RESEARCH**

Jaargang 6 | nummer 2

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**  
Strektalen met Gabmap in beeld gebracht

**4**  
De ideale digitale bibliotheek in tien stellingen

**4**  
500 veteranen leggen ziel en zaligheid bloot

**5**  
20.000 krantenartikelen over kunst in de database

**6**  
Tapes en ponskaarten in het computermuseum

**7**  
Hollandse graven hebben hun oorkonden op het net

**7**  
Jon Orwant van Google: ‘We willen iets terugdoen’

**8**  
Statistiek en sushi op Summer School

## 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 Hermesen, Erica Renckens, Jetske van der Schaaf

**Redactiesecretariaat:** Lucas Pasteruning, Jetske van der Schaaf

**Aan dit nummer werkten mee:** Jan Burgers, Ingrid Dillo, Marieke van Erp, Maria Fleischmann, Barend van der Meulen, Hilje Oosterbaan Martinius, Leo van Velzen, Steamwork Graphics

**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 gamma-richtingen.



## Bibliotheken op zoek naar hun toekomst

Inge Angevaere

Sinds de komst van internet is de positie van de universiteitsbibliotheek als spil van het academische leven niet langer vanzelfsprekend. De toegevoegde waarde van de universiteitsbibliotheek voor onderwijs en onderzoek stond centraal tijdens de jaarlijkse conferentie van LIBER, de vereniging van Europese wetenschappelijke bibliotheken, van 29 juni tot 2 juli in Barcelona.

In haar openingstoespraak zette Eurocommissaris Neelie Kroes stevig in op de digitale agenda. Open access, open data en digitale cultuur zijn de speerpunten van haar beleid. In een breed scala aan presentaties uit de bibliotheken werd duidelijk dat die digitale agenda door de bibliotheken is omarmd: van gebruikersdiensten via de iPhone via digitaliseringsprojecten tot data-ondersteuning aan de wetenschap. Maar zijn de bibliotheken snel genoeg? Een aantal sprekers meende van niet. Marco de Niet van Digitaal Erfgoed Nederland (DEN) verwees naar de enorme hoeveelheid *born digital* materiaal (websites, wiki's, forums, etc) die nog niemand verzamelt. Herbert van de Sompel presenteerde technische ontwikkelingen die momenteel grotendeels buiten de bibliotheken om gaan, zoals nanopublicaties (publicatie van afzonderlijke bevindingen) en *Linked Data*. *Linked Data* is een systeem waarbij inhoudelijke informatie uit vele disciplines aan elkaar gekoppeld kan worden door een gestandaardiseerde manier van aanbieden. Omdat deze data door machines kunnen worden geanalyseerd (in plaats van mensen), is het onderzoekspotentieel enorm. Maar, zo benadrukte Alma Swan van Enabling Open Scholarship, de meeste universitaire repositories slaan vooral informatie op in PDF-formaat – dat weliswaar duurzaam is, maar



Rick Luce (links) en Herbert van de Sompel

meestal niet geschikt om door machines geanalyseerd te worden. Kurt de Belder, bibliothecaris van de UB Leiden, vroeg zich af of bibliotheken wel de meest geschikte organisaties zijn om de grote technische ontwikkelingen in de wetenschap op te pakken. Rick Luce van Emory University was duidelijk in zijn antwoord: bibliotheken die hier niets aan doen zullen in 2020 geen reden van bestaan meer hebben.

[www.liber.eu](http://www.liber.eu); <http://biblioteca.upc.edu/LIBER2011/>

### Probleem van de vindbaarheid is oplosbaar

Inge Angevaere

Het internet is per definitie veranderlijk en flexibel. Dat heeft veel voordelen, maar ook nadelen, zoals de foutmelding “404 page not found” als een brok informatie is verplaatst naar een andere website. Op 14 en 15 juni kwam een internationale groep experts in Den Haag bijeen om te bespreken hoe de zogenaamde persistent identifiers – (onveranderlijke) namen in plaats van (veranderlijke) adressen – kunnen bijdragen aan betere vindbaarheid van informatie, met name op de lange termijn.

De bijeenkomst was georganiseerd door het internationale Knowledge Exchange, SURFfoundation en Data Archiving and Networked Services (DANS). Al snel bleek dat het pro-

bleem van vindbaarheid technisch oplosbaar is, maar dat het instellen nog vaak ontbreekt aan beleid, aan duidelijke wil, om zich aan de spelregels te houden en de dienst die de nieuwe verbindingen legt (de zogenaamde resolver) te onderhouden. Bovendien is de afgelopen jaren een aantal verschillende standaarden ontwikkeld (URN, DOI, PURL) die niet op elkaar aansluiten, en die zich slecht verhouden tot het systeem van *Linked Data* dat het mogelijk moet maken dat computers informatie uit heel verschillende bronnen bij elkaar brengen. Aan het eind van de bijeenkomst stelde Andrew Treloar van de Australian National Data Service een ‘Den Haag manifest’ voor, waarin



Workshop met dr. Tulp op de achtergrond

alle partijen worden opgeroepen de door hen gehanteerde systemen transparant te maken en de gemeenschappelijke noemer te zoeken die uitwisseling mogelijk moet maken. Maar nog belangrijker acht hij het slaan van een brug naar *Linked Data*.

[www.knowledge-exchange.info/Default.aspx?ID=440](http://www.knowledge-exchange.info/Default.aspx?ID=440)

### WebScience bestudeert web en samenleving

Marieke van Erp

Web Science is een jong vakgebied dat het Web en de invloed van het Web op de samenleving wil begrijpen en daarmee het Web tot een nog

nuttiger medium wil maken. Van 15 tot 17 juni 2011 kwamen onderzoekers uit de hele wereld samen in Koblenz voor de ACM WebScience'11 conferentie.

De conferentie begon en eindigde met uitstekende lezingen. De aftrap was voor Jamie Teevan van Microsoft Research. Zij houdt zich onder meer bezig met de vraag hoe en hoe snel websites veranderen (nieuwswebsites veranderen hun content bijvoorbeeld dagelijks, bedrijfswebsites hebben een veel tragere cyclus) en hoe je mensen kunt helpen om snel nieuwe content te vinden op een website. De afsluitende lezing was van Barry Wellman. Hij bekijkt het Web vanuit het perspectief van de socioloog en zijn onderzoek duidt erop dat het Web sociale netwerken in de *real world* (familie, gemeenschap en werk) beïnvloedt. Tussen de twee keynotes kwam een grote verscheidenheid aan onderzoek voorbij. Een interessant thema bleek het waarborgen van kwaliteit in user-generated content (bijdragen van web gebruikers op nieuwswebsites, Wikipedia etc.). Wikipedia gebruikt bv. een democratisch model, maar het effect kan zijn dat de mening van de meerderheid overheerst. Ook de vraag hoe meerdere perspectieven toch een plaats kunnen krijgen kwam aan de orde. In de postersessie passeerden de relatie van het Web met ethiek, gezondheid en economie de revue.

[www.websci11.org/home/](http://www.websci11.org/home/)  
[http://videolectures.net/acmwebsci2011\\_koblenz/](http://videolectures.net/acmwebsci2011_koblenz/)

### Tijdsdruk en tijdarmoede bij IATUR

Anna Grigolon e.a.

Alle aspecten van tijdsbestedingsonderzoek kwamen aan de orde op de conferentie van de International

Association for Time Use Research (IATUR) die 1 tot 3 augustus 2011 plaats vond aan de Universiteit van Oxford. Een centraal onderwerp voor de 189 deelnemers uit 39 landen is dagelijkse tijdsdruk en tijdarmoede. Met vraagstukken als: hoe ouders tijd investeren in het onderwijs aan hun kinderen; hoe werkende vrouwen hun dagelijks levens organiseren; etc.

Een terugkerend punt was de relatie tussen tijdsbesteding en welzijn, met verschillende toepassingen, zoals balans tussen werk en leven, opvoeding en kinderopvang, gezondheid en tijdsbesteding, en vrije tijd. Dit is duidelijk een thema dat snel in belangstelling toeneemt. Er waren ook sessies over verschillen in tijdsbesteding naar geslacht, GPS en tracking, nationale onderzoeksrapporten, ontwikkelingen in enquêtemethoden en nog veel meer. De deelnemers aan de conferentie hebben zeer gevarieerde achtergronden (sociologie, economie, media-agentschappen, nationale bureaus voor statistiek, stedenbouw en milieu, en psychologie). Nuttige informatie werd gegeven over de online en gratis databases zoals Time-scapes en de Multinational Time Use Survey (MTU), die meer dan 60 gegevensbestanden uit 25 landen dekt.

De organisatoren legden een grote gastvrijheid aan de dag en boden een prachtige plek en sfeer voor communicatie en idee-uitwisseling. Interessant en nastrevenswaardig voor andere congressen is dat de belangrijkste resultaten van de presentaties na de sessies onmiddellijk beschikbaar kwamen op Twitter, waardoor de Tijdsbestedings Conferentie inderdaad een bijdrage leverde aan de bestrijding van tijdarmoede.

(Anna Grigolon, Feixiong Liao, Anastasia Moiseeva, Dajuan Yang, Urban Planning Group, TU/e)  
<http://iatur2011.timeuse.org/>

## AGENDA

### 10 - 16 okt • University of Würzburg Philology in the Digital Age: TEI members' meeting

The program features keynote lectures by Edward Vanhoutte and Andrea Rapp, parallel sessions of papers and round table discussions, the annual TEI business meeting, a poster session/tools demonstration, and special interest group (SIG) meetings.

[www.zde.uni-wuerzburg.de/tei\\_mm\\_2011/](http://www.zde.uni-wuerzburg.de/tei_mm_2011/)

### 18 - 21 okt • University of Prato International Symposium on Virtual Reality, Archaeology and Cultural Heritage

Symposium concerning the application of information technologies to Cultural Heritage, in particular data acquisition and processing, metadata handling, presentation and practitioner's experience.

[www.vast-conference.eu/](http://www.vast-conference.eu/)

### 19 - 20 okt • Sofitel Hamburg Digital Preservation Summit

Experts exchange their practical knowledge and experiences on digital preservation.

[www.digitalpreservationsummit.de/invitation.html](http://www.digitalpreservationsummit.de/invitation.html)

### 25 - 28 okt • University of Cologne From Cordoba to Cologne, transformation and translation, transmission and edition of Averroes's works

International Conference organized by Digital Averroes Research Environment (DARE), first international DARE conference.

<http://dare.uni-koeln.de/?q=node/44>

### 26 - 27 okt • Koninklijke Bibliotheek Den Haag

The KEEP approach to Digital Preservation KEEP (Keeping Emulation Environments Portable) is developing emulation services to enable accurate rendering of both static and dynamic digital objects: text, sound, and image files; multimedia documents, websites, databases, videogames etc. The overall aim of the project is to facilitate universal access to our cultural heritage by making flexible tools for accessing and storing a wide range of digital objects.

<http://keep-den Haag.eventbrite.com/>

### 9 - 10 nov • Howard Hughes Medical Institute - Washington DC

### Berlin 9 Open Access Conference - The impact of Open Access on Research and scholarship

The Berlin Open Access Conference Series convenes leaders in the science, humanities, research, funding, and policy communities around The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities. Berlin 9 is the first of the annual meetings to take place in North America.

[www.berlin9.org/](http://www.berlin9.org/)

### 17 - 18 nov • University of Copenhagen, Center for Language Technology Supporting Digital Humanities, answering the unaskable

Digital technologies have the potential to transform the types of research questions that we ask in the Humanities, allowing us both to address traditional questions in new and exciting ways but also to answer questions that we were not even aware we could ask. How can digital humanities help us not just to find the answers to our research questions more easily, but also to formulate research questions we would never have been able to ask without access to large quantities of digital data and sophisticated tools for their analysis?

<http://cst.ku.dk/sdh2011/>

### 22 - 23 nov • University of Tromsø Campus The 6th Munin conference 2011 – Enhancing publications

Annual conference on scientific publishing, with an Open Access approach.

[www.ub.uit.no/](http://www.ub.uit.no/)

### 5 - 6 dec • Gotheborg

#### 3rd Annual European DDI Users Group Meeting (EDDI11): DDI - The Basis of Managing the Data Life Cycle

Presentations, talks, papers, posters on the Data Documentation Initiative: case studies, mature implementations, early implementations, interplay of DDI with other standards or technologies, projects in early phases in which DDI is under consideration, and critiques of DDI.

[www.iza.org/conference\\_files/EDDI2011/](http://www.iza.org/conference_files/EDDI2011/)

### 6 - 9 dec • WTC - Rotterdam

#### Digital Strategies for Heritage 2011 – DISH 2011

DISH is a bi-annual conference that unites digital strategies with international professionals in heritage.

[www.dish2011.nl/](http://www.dish2011.nl/)

Gabmap brengt taalvariaties in kaart

# Het beeld en de streektaal

**Groningse onderzoekers zijn erin geslaagd taalvariaties te meten en ze overzichtelijk te presenteren.**

*Erica Renckens*

“Alles wat deze nieuwe webapplicatie kan, konden wij technici eigenlijk al. Maar we willen het juist voor anderen óók mogelijk maken. Pas als niet-technisch onderlegde wetenschappers het ook gebruiken is dit project geslaagd.”

Aan het woord is John Nerbonne, taaltechnoloog aan de Universiteit Groningen. Tijdens een informatiemiddag van CLARIN geeft hij een presentatie over Gabmap, online software die kwantitatieve analyse van taalvariatie mogelijk maakt. De gebruiker kan met dit programma zelf op een laagdrempelige manier de verschillen tussen taalvarianten in kaart brengen.

## Bagage

Onderzoek naar taalvariatie, en met name dialecten, is voornamelijk kwalitatief. De onderzoekers beschrijven de verschillen tussen de taalvarianten, maar kunnen niet nauwkeurig meten in hoeverre deze op verschillende gebieden van elkaar afwijken of juist op elkaar lijken. De software om zulke objectieve vergelijkingen mogelijk te maken bestaat al een tijdje, maar de ge-



*Illustratie Hilje Oosterbaan Martinius*

bruiker moet flink wat technische bagage hebben om ermee uit de voeten te kunnen.

“De gebruiker van Gabmap kan zijn data aanbieden in een eenvoudig document waarin de verschillende variabelen met een tab van elkaar gescheiden zijn.” aldus Nerbonne. “Vaak is dit de uitspraak van een bepaald woord, zoals ‘hoes’ en ‘huus’ voor ‘huis’. Maar het kan ook gaan om verschillende woorden voor hetzelfde begrip, zoals ‘jij’, ‘gij’ en ‘du’.”

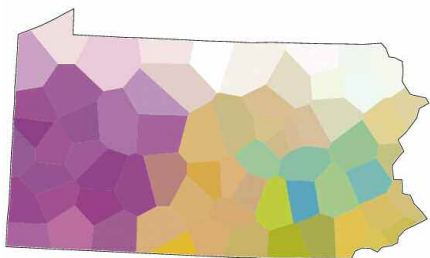
## Kinderliedjes

Gabmap genereert vervolgens automatisch gestructureerde overzichtjes van deze gegevens. De gebruiker kan aangeven van welke variabelen hij de verschillen geanalyseerd wil zien, waarna de software de resultaten van deze analyse ook weer overzichtelijk in kaart brengt, onder andere met behulp van kaartjes uit Google Earth. Op deze kaartjes zie je dan precies welke variant in welk gebied voorkomen. Bovendien is het mogelijk om meerdere variabelen in één kaartje op

te nemen, zodat eventuele verbanden makkelijker zichtbaar worden.

“In de toekomst hopen we ook een vergelijking met andere kaarten mogelijk te maken”, vertelt Nerbonne. “Dan zouden we bijvoorbeeld ook culturele kenmerken in de analyse kunnen meenemen, zoals religie of welke kinderliedjes men in een bepaald gebied zingt.” Maar voorlopig is Nerbonne al erg tevreden met wat de software nu allemaal kan. “We wilden het grootschalig meten van dialectverschillen mogelijk maken, onder andere om de wetenschappelijke discussie over deze totaalaanpak te stimuleren”, verklaart hij. “Het team van Gabmap heeft er hard aan gewerkt om ook de faciliteiten voor de traditionele dialectologie beschikbaar te maken. Dat vergemakkelijkt de vergelijking tussen de kwalitatieve en kwantitatieve methode.”

*Gabmap is te gebruiken via [www.gabmap.nl](http://www.gabmap.nl). Op deze site zijn ook demonstraties en handleidingen te vinden.*



*In deze Gabmap van Pennsylvania zijn drie variabelen weergegeven. Het blauw/groen duidt op de aanwezigheid van ‘Pennsylvania Dutch’, het wit/grijs op het gebruik van ‘midlands’ dan wel een noordelijke variant van het Amerikaans, en in het paarse gebied spreekt men ‘car’ uit met een /r/, in tegenstelling tot het oranje gebied.*

# Slagroomtaart voor Friese galgen

**Dit jaar reikt SURF foundation elke maand een prijs uit voor de ‘verrijkte publicatie van de maand’.** *Peter Boot*

Verrijkte publicaties zijn digitale wetenschappelijke publicaties waaraan niet-traditionele elementen zijn toegevoegd, zoals toegang tot de onderzoeksdata, interactieve illustraties, commentaar of andere bestanden.

In augustus ging de maandelijkse prijs (een slagroomtaart) naar de publicatie *Geen Paniek! Riscocommunicatie door de ogen van de burger*, geschreven door Jan M. Gutting, Marlies Baan, Milou Kievik en Karin Stone van de Universiteit Twente en Deltares. In juli ging de taart naar ‘Gallows in medieval Fri-

sia’, een publicatie die de geografische verspreiding van galgen in middeleeuws Friesland behandelt. Aan de publicatie zijn tabellen, grafieken en een interactieve kaart toegevoegd.

## Xpos’re

De galgenpublicatie werd gemaakt met het tool Xpos’re (XML-based Publications On Scientific Research), ontwikkeld door Leen Breure (Universiteit Utrecht). Xpos’re is gebouwd om een demonstratie te kunnen geven van het werk met Geografische Informatie Systemen. Maar de ambities reiken verder, verklaart Breure. “Xpos’re is nu meer een prototype, maar moet een echt product worden, dat vooral voor de auteur gebruiksvriendelijk

is.” Breure wil in het najaar een *requirements study* starten, en ziet de huidige versie van Xpos’re vooral ook als een hulpmiddel om ideeën te concretiseren. Het streven is dat auteurs met enige kennis van ICT zelf met behulp van Xpos’re hun verrijkte publicatie kunnen maken. Maar, zegt Breure, een andere mogelijkheid is dat de verrijkte publicatie gemaakt wordt door bijvoorbeeld bibliotheekmedewerkers of de redactie van een tijdschrift. De bibliotheek van de VU is voornemens een pilot te doen, en Xpos’re zou bijvoorbeeld ook inpasbaar moeten zijn in de workflow die ontwikkeld wordt door het SURF-project ‘Enhanced Journals Made Easy’. Een belangrijke wens is om meer interactieve componenten toe te voe-

gen. Breure: “Op de verlanglijst staan tabellen met filtermogelijkheid en interactieve grafieken. Bij voorkeur maken we daarbij gebruik van bestaande software.”

Waarom maakt Xpos’re gebruik van Flash-software? Breure verwijst naar de verschillen tussen browsers (zoals Internet Explorer en Firefox), waarvan hij bij Flash geen last heeft. “Flash is geschikt voor een product waarbij ‘mooi’ en ‘pixel-nauwkeurig’ essentieel is.” Overigens kan Xpos’re ook uitvoer aanmaken voor andere platforms: iPads, smartphones en e-readers.

*[www.cs.uu.nl/research/projects/i-cult/xposre/demo/gallows/](http://www.cs.uu.nl/research/projects/i-cult/xposre/demo/gallows/)  
[www.cs.uu.nl/research/projects/i-cult/xposre/](http://www.cs.uu.nl/research/projects/i-cult/xposre/)*

## KORT

### ERIC-status voor ouderenonderzoek

Dit voorjaar heeft de Europese onderzoeksfaciliteit SHARE als eerste de status van een ‘European Research Infrastructure Consortium’ (ERIC) gekregen. De SHARE ERIC is op 25 mei 2011 officieel op het ministerie van OCW geïnstalleerd. SHARE, de ‘Survey of Health, Ageing and Retirement in Europe’ doet in twintig landen langlopend onderzoek naar de bevolking van 50 jaar en ouder. De ERIC is een nieuwe Europese rechtsvorm en is bedoeld om pan-Europese onderzoeksfaciliteiten beter te laten functioneren. Een belangrijk pluspunt is dat de ERIC is vrijgesteld van BTW over diensten en dat scheelt veel geld bij het uitbesteden van het veldwerk.

*(Eric Balster)*

[www.share-project.org](http://www.share-project.org)

### Breed verlangen naar CLARIAH

Een zeer breed consortium binnen de geesteswetenschappen heeft eind augustus een voorstel ingediend in het kader van de call Nationale Roadmap voor Grootschalige Onderzoeksfaciliteiten van NWO. Het voorstel met de naam CLARIAH (Common Lab Research Infrastructure for the Arts and Humanities) bouwt voort op de Europese projecten CLARIN en DARIAH. Het landelijke consortium bestaat uit alle faculteiten geesteswetenschappen, alle humaniora instituten van de KNAW, het Max Planck Instituut voor Psycholinguïstiek, het Instituut voor Nederlandse Lexicologie, het Nederlandse eScience Center, het instituut voor Beeld en Geluid, de Koninklijke Bibliotheek en het Nationaal Archief. In het CLARIAH project moet een faciliteit voor eHumanities research komen, dat onderzoekers toegang biedt tot een grote hoeveelheid digitale bronnen. *(Ingrid Dillo)*

### Betere data met de smartphone

Het Sociaal en Cultureel Planbureau en CentERdata gaan een experiment met smartphones uitvoeren onder leden van het LISS-panel. LISS staat voor Longitudinal Internet Studies for the Social sciences. Het LISS-panel bestaat uit mensen uit alle lagen van de bevolking die via internet vragenlijsten invullen. In het experiment wordt onderzocht of bestaande vragenlijsten met een smartphone kunnen worden afgenomen. Dit is interessant voor bijvoorbeeld het bekende Tijdbestedingsonderzoek (TBO). Als respondenten bereid zijn op hun smartphone een dagboekje bij te houden, kan dat de kwaliteit van de data ten goede komen. Verder wordt bekeken of nieuwe soorten gegevens kunnen worden verzameld: locatiegegevens bijvoorbeeld, of scans van kassabonnen. *(Warna Oosterbaan)*



De ideale digitale bibliotheek in tien stellingen

# Volledig, actueel en uniek

Hij is er nog steeds niet, de ideale digitale bibliotheek. Peter Boot, onderzoeker aan het Huygens Instituut en redacteur van E-data, geeft nog eens zijn verlanglijstje, in tien stellingen.

**1. De ideale digitale bibliotheek bevat een exemplaar van alles wat in Nederland of in het Nederlands is gedrukt of geschreven.**

Dus niet alleen boeken, maar ook tijdschriften en kranten, gelegenheidsdrukwerk, wetenschappelijke literatuur, 'grijze' literatuur (zoals rapporten) en handschriftelijk materiaal van de Middeleeuwen tot heden. De beperking tot Nederlands materiaal is geen principiële maar een praktische keuze: dat materiaal is onze verantwoordelijkheid. Ook omdat onderzoek steeds vaker gebruik zal maken van automatische tekstanalyse, is er alle reden de collectie van bruikbare documenten zo groot mogelijk te maken.

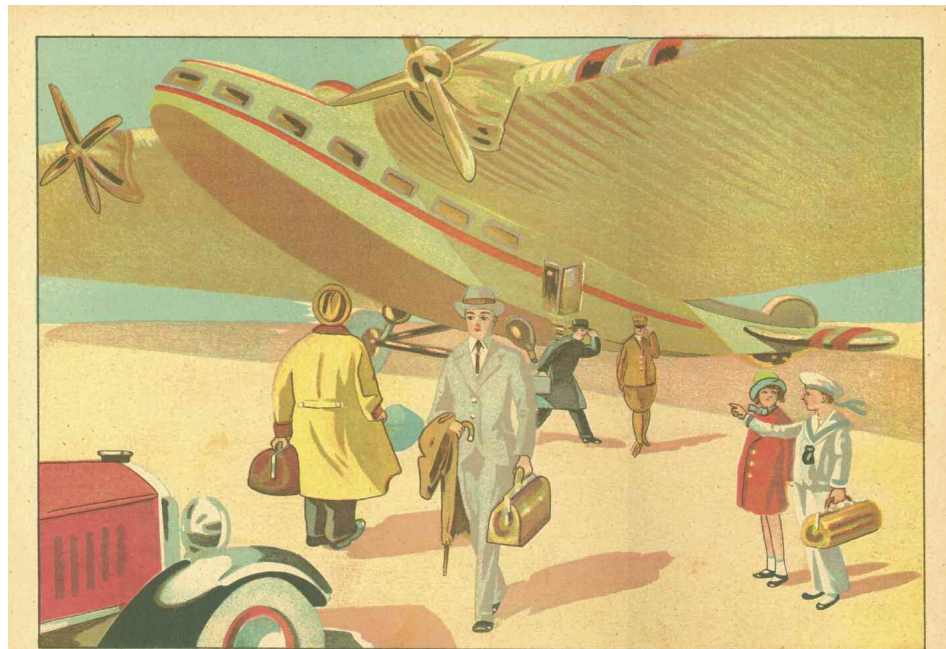
**2. De bibliotheek geeft een volledige en foutloze tekst.**

Waar een foutloze tekst onmogelijk is (zoals bij kranten, vanwege de aard en omvang van het materiaal, of bij middeleeuwse handschriften, omdat een transcriptie daar vakwerk is), moet die beperking duidelijk worden aangegeven.

**3. De bibliotheek bevat facsimileafbeeldingen van elke pagina, ook van recente werken.**

Vorm bepaalt mede hoe de inhoud wordt gelezen en moet dus in de bibliotheek worden gereproduceerd (in kleur, op ware grote, in hoge resolutie).

**4. De bibliotheek biedt teksten aan in een veelheid aan (aanpasbare) uitvoerformaten, waaronder tenminste een prettig leesbaar formaat (bijvoorbeeld HTML), een gestructureerd formaat**



In een vliegtuig zweven wij omhoog. Boven al de wolken veiligjes en droog. Zijn wij weer beneden slaken wij een zucht. Veiliger is het op den grond, dan daar in de lucht

Prentenboek 'Hoera! Wij gaan op reis', vermoedelijk uit het begin van de jaren dertig van de vorige eeuw, nu als digitaal 'bladerboek' op de website van de KB in te zien

**(XML), een afdrukbaar formaat (bijvoorbeeld PDF), en platte tekst.**

Lezen van het scherm is maar één van de gebruiksmogelijkheden. Onderzoekers willen ook een concordantie kunnen maken (lijst van woorden in context) of andere analyses kunnen uitvoeren. De bibliotheek drukt op verzoek een niet meer leverbaar boek, levert zijn inhoud ook aan e-book lezers en tablets en maakt het mogelijk dat via een Application Programming Interface (API) andere sites nieuwe diensten kunnen aanbieden. op het materiaal van de bibliotheek.

**5. De bibliotheek biedt uitgebreide zoekmogelijkheden in titels en in volledige teksten.**

Het is tenminste mogelijk om te zoeken op auteur, titel, volledige tekst en periode. Het is mogelijk om zoekopdrachten te combineren met AND, OR en NOT. Het is mogelijk te zoeken op frases. De bibliotheek biedt adequate metadata, waardoor het tenminste mogelijk is van alle werken van een auteur of alle delen van een werk. op te vragen. Het is mogelijk te bladeren van hoofdstuk naar hoofdstuk (en niet alleen van bladzij naar bladzij).

**6. De bibliotheek bevat niet overwegend ouder materiaal.**

En bevat dus ook materiaal dat volgens de huidige wetgeving valt onder het auteursrecht. Auteurs die een aanmerkelijk financieel belang hebben bij de verkoop van gedrukte exemplaren van hun boek zullen wellicht aarzelen, maar alle andere auteurs winnen bij de digitale bereikbaarheid van hun werken. De bibliotheek zet zich in voor inperking van het auteursrecht. Er is geen maatschappelijk belang gediend bij de huidige termijn van het auteursrecht (zeventig jaar na de dood van de auteur). Twintig jaar na publicatie zouden alle rechten moeten vervallen en zou een auteur niet meer de mogelijkheid moeten hebben opname van zijn werk in een digitale bibliotheek tegen te houden

**7. Er is maar één digitale bibliotheek.**

Bij elke begrenzing (op periode, thema, genre, of wat dan ook) is er onderzoek denkbaar dat geheel of gedeeltelijk onmogelijk wordt. Er moet dus één infrastructuur zijn waarbinnen alle materiaal valt te raadplegen. Digitale bibliotheken gericht op een enkel thema (een auteur, een maatschappelijk verschijnsel, een

periode) kunnen op basis van de API van de overkoepelende bibliotheek (stelling 4) worden ontwikkeld.

**8. De bibliotheek is een centrum voor onderzoek naar (semi-)automatische elektronische verrijking van teksten.**

De bibliotheek experimenteert met technieken als naamherkenning, analyse van citaten en intertekstualiteit, herkenning van thema's, en zoeken van qua betekenis verwante teksten. Verwijzingen binnen een boek (bijvoorbeeld in een index) en naar andere boeken worden door de bibliotheek al dan niet automatisch aanklikbaar gemaakt.

**9. De bibliotheek is ook een platform voor werken en samenwerken.**

In de bibliotheek kunnen gebruikers annotaties maken, kunnen docenten studiepaden definiëren voor hun studenten en kunnen groepen samenwerken aan de studie van (deel)collecties.

**10. De ideale digitale bibliotheek is een publieke nutsvoorziening.**

De bibliotheek biedt de neerslag van ons verleden en moet toegankelijk zijn voor iedereen met een interesse in onze cultuur, binnen en buiten de landsgrenzen.

Deze vrije toegankelijkheid hoort niet alleen een recht om te lezen te zijn, maar ook een recht om door te geven en om in technisch opzicht te bewerken. Derden kunnen dan op basis van het materiaal in de digitale bibliotheek bijvoorbeeld een hoogwaardige digitale editie maken, of experimenteren met een intuïtieve interface voor de collecties.

**Tenslotte**

Het is goed dat Nederlandse bibliotheken en andere instellingen vaart maken met het digitaliseren van ons geschreven en gedrukte erfgoed. Het is jammer dat niet wat meer verbeeldingskracht wordt gestoken in de toegankelijkheid van die teksten en in de hulpmiddelen waarmee ze worden aangeboden.



## Inzoomen op vijfhonderd veteranenlevens

**Duizend veteranen zijn de afgelopen jaren geïnterviewd over hun ervaringen in oorlogsgebied. Vijfhonderd van die verhalen staan nu online.**

Warna Oosterbaan

Het is een bijzondere ervaring. Achter de computer ondergedompeld te worden in de levensgeschiedenissen van honderden Nederlanders. Ze hebben in eigen land of in het buitenland gevochten, of ze waren betrokken bij een vredesmissie. Sommigen vertellen over de meidagen in 1940, anderen over hun uitzending naar Libanon of Afghanistan. En alles daartussen: Indië,

Korea, Nieuw-Guinea, Cambodja en Bosnië. De ondervraagden vertellen hun hele leven, en zoomen in op die maanden of jaren dat ze die vaak zo beslissende ervaringen opdeden. Ze vertellen over de barre omstandigheden in het veld, over hun terugkeer in de burgermaatschappij, over herinneringen die hen nog kwellen.

Soms vertellen ze hun relaas voor het eerst. Sommige gesprekken duren vele uren en strekken zich uit over meerdere sessies. Enkele respondenten zijn de negentig al gepasseerd, anderen zijn ruim een halve eeuw jonger. Al met al een fraai voorbeeld van *oral history*.

Sinds kort is het voor iedereen mogelijk die verhalen te beluisteren, want van de interviews die het Veteraneninstituut met duizend veteranen hield, staan er inmiddels vijfhonderd op de site van dat instituut. Na eenmalige registratie kan de bezoeker in een venster een zoekterm intikken – vriendschap, post, bom, trauma of welk ander woord dan ook – en even later verschijnt er een lijst van de samenvattingen van interviewfragmenten waarin dat woord voorkomt. Doorklikken naar het bijbehorende audiogedeelte maakt het mogelijk de oorspronkelijke geluidsband te beluisteren. Biografische gegevens als naam,

geboortjaar, opleiding, datum van uitzending, rang etc. worden van elke geïnterviewde in tekst weergegeven.

**Magisch getal**

Over het project verscheen vorig jaar een verrijkte publicatie: *Wat veteranen vertellen*, met veel informatie over de opzet van het project en de gehanteerde werkwijze. Daar valt bijvoorbeeld te lezen dat er in Nederland 120.000 veteranen zijn en dat voor duizend gesprekken werd gekozen om zo een zekere representativiteit te bereiken, maar ook wegens het 'magische getal'. Van videoregistratie werd afgezien.

Niet alleen bleek dat vier keer zo duur te zijn, ook werd de aanwezigheid van een camera, een cameraman en een geluidstechnicus als een bezwaar gezien. De kleine digitale geluidsrecorder met ingebouwde microfoon waarvoor gekozen werd en die tijdens het gesprek op tafel werd gelegd, was veel minder ingrijpend.

De interviews worden gearchiveerd door DANS en de metadata van alle interviews zijn in EASY te bekijken.

<http://interview.veteraneninstituut.nl>  
[www.watveteranenvertellen.nl/](http://www.watveteranenvertellen.nl/)  
 WatVeteranenVertellen.pdf  
<http://easy.dans.knaw.nl>



Susanne Janssen en Marc Verboord:

# ‘Jammer dat de kranten van de laatste halve eeuw niet zijn gedigitaliseerd’

**Krantenartikelen over cultuur in vier landen maken een sterke trend zichtbaar: populaire cultuur is salonfähig geworden.**

Warna Oosterbaan

Er hebben 14 codeurs aan gewerkt en het heeft meer dan 10.000 menur gekost. Nu zitten er ruim twintigduizend artikelen in het databestand van Susanne Janssen, hoogleraar Media en Cultuur aan de Erasmus Universiteit. Twintigduizend artikelen over kunst en cultuur uit acht internationale kranten in de afgelopen vijftig jaar.

Het is een wetenschappelijke goudmijn gebleken. Janssen baseerde haar oratie erop, drie aio's schreven er hun proefschrift over. Twee dissertaties zijn in de maak en voor vier themanummers en ruim vijftig artikelen in voor- aanstaande wetenschappelijke tijdschriften legde het project de fundamenten.

Janssen en haar team zagen hoe de vroegere scherpe statusverschillen tussen hoge en lage cultuur vervaagden en dat het domein van de ‘legitieme’ cultuur steeds meer genres ging omvatten. Over film, fotografie, popmuziek, thrillers en strips schrijven gevestigde kranten inmiddels volop, en dat is een teken van een ingrijpende verschuiving in het culturele domein: het langzamerhand salonfähig worden van de populaire cultuur. Je kon het vermoeden, maar het is nu onweerlegbaar vastgesteld. “Dat is het belangrijkste, maar we hebben nog veel meer ontdekt, en we kunnen nog wel een aantal jaren doorgaan. Ook met vragen waar we helemaal niet aan dachten toen we het onderzoeksplan maakten.”

We spreken Janssen, een energieke vrouw met blonde krullen, in een lunchroom op de uitgestrekte campus van de Erasmus Universiteit. Naast haar is Marc Verboord aangeschoven, een onderzoeker die vanaf het begin bij het project betrokken was. Hij was een van de twee postdocs die Janssen kon aanstellen toen ze in 2003 in het kader van de Vici-projecten een flinke NWO-onderzoekssubsidie in de wacht sleepte. Er kwamen ook drie promovendi en halverwege het project nog eens twee.

## Hoge en lage cultuuruitingen

Janssen: “Het gaat ons om de veranderingen in de classificatie van cultuur, de sociale waardering van cultuuruitingen, zou je kunnen zeggen. We proberen daarin ontwikkelingen in kaart te brengen en verschillen tussen landen op het spoor te komen. Dus hoe verandert de waardering voor opera, voor popmuziek, voor beeldende kunst? In hoeverre verschilt Frankrijk van Nederland, of Europa van de VS in de waardering voor hoge of lage cultuur? Wat zijn de effecten van globalisering op de aandacht voor lokale cultuuruitingen en in welke richting ontwikkelt de culturele journalistiek zich?”

Janssen en haar onderzoekers kozen ervoor artikelen over cultuur in vooraanstaande kranten in vier landen te gaan turven. In Nederland de *Volkskrant* en de *NRC*. In Frankrijk *Le*



Susanne Janssen en Marc Verboord foto Leo van Velzen

### Culturele Classificatie

**Susanne Janssen** (1963) is sinds 2007 hoogleraar Sociologie van Media en Cultuur aan de Erasmus Universiteit. Zij studeerde taal- en literatuurwetenschap in Tilburg en promoveerde in 1994 op een studie naar literatuurkritiek. Voor haar project Cultural Classification Systems in Transition kreeg ze in 2003 een VICI-subsidie van NWO.

**Marc Verboord** (1971) studeerde cultuurwetenschappen in Tilburg en promoveerde in 2003 aan de Universiteit van Utrecht op de invloed van ouders en leraren op het leesgedrag van opgroeiende kinderen. Sinds 2004 werkt hij aan de Erasmus Universiteit - vanaf 2011 als universitair hoofddocent - en participeert hij in het Cultural Classification Systems in Transition-project.

*Monde* en *Le Figaro*, in Duitsland de *Frankfurter Allgemeine Zeitung* en de *Süddeutsche Zeitung*, in de VS *New York Times* en de *Los Angeles Times*. Ze bekeken de periode 1955 - 2005, waarbij ze om de tien jaar één jaargang geheel inventariseerden. Van elk artikel werd vastgesteld of het misschien over cultuur ging, welke cultuuruiting erin werd behandeld, hoe groot het stuk was, op welk gedeelte van de pagina het stond, of er een foto bij stond. “En wat voor een stuk het was”, zegt Verboord.

“Een interview, een recensie of een nieuwsbericht. Dat waren allemaal belangrijke gegevens voor ons.”

### Voorrang historische kranten

Voor het verzamelen van de data moesten de onderzoekers daarom de complete kranten zien en de originele krantenpagina's. Dat was niet altijd eenvoudig. Verboord: “Op de Koninklijke Bibliotheek kregen we toestemming om maandenlang in het magazijn te bivakkeren om de leggers te bekijken van de Nederlandse, Duitse en Franse kranten. Bij de *New York Times* en de *Los Angeles Times* hadden we het geluk dat alle afleveringen die gedurende de twintigste eeuw zijn verschenen volledig zijn gedigitaliseerd en beschikbaar zijn gemaakt voor onderzoekers in de vorm van een online database waarop je een abonnement kunt nemen.”

“Hadden we dat ook maar in Nederland”, zegt Janssen. “De Koninklijke Bibliotheek is wel bezig met de digitalisering van kranten, maar heeft voorrang gegeven aan historische kranten uit de periode voor 1945. Het zou een enorme stimulans voor Nederlands onderzoek zijn als we gedigitaliseerde kranten van de laatste zestig jaar zouden hebben. Niet alleen voor media- en cultuurwetenschappers, maar

### De KB en de kranten

Waarom zet de KB de scans van Nederlandse kranten na 1950 niet op het net? Jasper Faase, manager van het krantenproject op de KB: “Het is een rechtenkwestie en daarover moeten we met de krantenuitgevers onderhandelen. NRC Media heeft geen belangstelling en met de Persgroep zijn we aan het praten, maar er is nog geen doorbraak. Intussen zijn we wel aan het scannen, ook kranten na 1945. Binnenkort komen de eerste plukjes op het net.” Van welke kranten? “Het Nieuwsblad van het Noorden, en mogelijk ook wat jaargangen van *De Telegraaf*.”

## Lekenoordelen worden belangrijker, de invloed van kranten geringer

ook voor bijvoorbeeld bestuurskundigen, economen, historici en taalkundigen.”

Met het honderd pagina's dikke codeboek ging een legertje codeurs aan de slag. Maar bij de krantendata bleef het niet. “We hebben ook gegevens verzameld over de mediawereld in die landen, over cultuurparticipatie en cultuurbeleid. Je moet die data ook kunnen interpreteren. Neem Frankrijk. Daar kun je zien dat er in verhouding veel aandacht is voor de eigen Franse cultuur. Maar dan moet je weten dat dat samenhangt met het cultuurbeleid daar: dat is nogal protectionistisch, er zijn allerlei regels die gericht zijn op de bevordering van het Franse cultuuroed.”

Krijgt Janssen wel eens onderzoekers aan de telefoon met de vraag of ze het databestand mogen gebruiken?

“We krijgen wel vragen over ons onderzoek en die mensen helpen we waar we kunnen. Maar een verzoek om het hele bestand op te sturen, nee, dat niet. We gaan ons bestand natuurlijk wel deponeren als we er klaar mee zijn. Maar nu zou ik het niet zo gauw opsturen. Er lopen nog onderzoeken van onszelf, en we willen wel de eersten zijn die erover publiceren.”

2005 was het laatste peiljaar. Zou het niet aantrekkelijk zijn om er in 2015 weer een update aan te geven? Janssen: “Ik denk het wel. Maar zonder meer het lijntje doortrekken is niet zo interessant. Je zou bijvoorbeeld meer aandacht aan internet moeten geven. De sociale media zijn pas onlangs opgekomen, en die zijn voor de sociale waardering van cultuur erg belangrijk.” Verboord: “We hebben nu een student die aan het afstuderen is op de vraag of er een

relatie is tussen twitterberichten over een film en het succes van die film. Die relatie blijkt er inderdaad te zijn: positieve tweets hangen samen met een hoge kaartverkoop. Dus lekenoordelen lijken steeds belangrijker te worden, en de invloed van kranten wordt misschien wel geringer.”

Dus misschien toch niet zo'n gek idee, om met 2005 af te sluiten?

Janssen en Verboord grinniken. “Ja, misschien is dat wel zo.”

## INTERVIEW



**Het computermuseum in het Amsterdamse Science Park maakt tapes, floppies en ponskaarten weer leesbaar. *Warna Oosterbaan***

“Ja, waarom doet iemand dit?” Edo Dooijes vindt het een moeilijke vraag. “Ik ben een oorlogskind”, zegt hij dan. “Ik vind het zonde om dure dingen in de gracht te gooien. Prachtige apparaten, die miljoenen hebben gekost. Zo is het begonnen. Maar in de loop der tijd is belangstelling voor de geschiedenis van de informatica de belangrijkste reden geworden.”

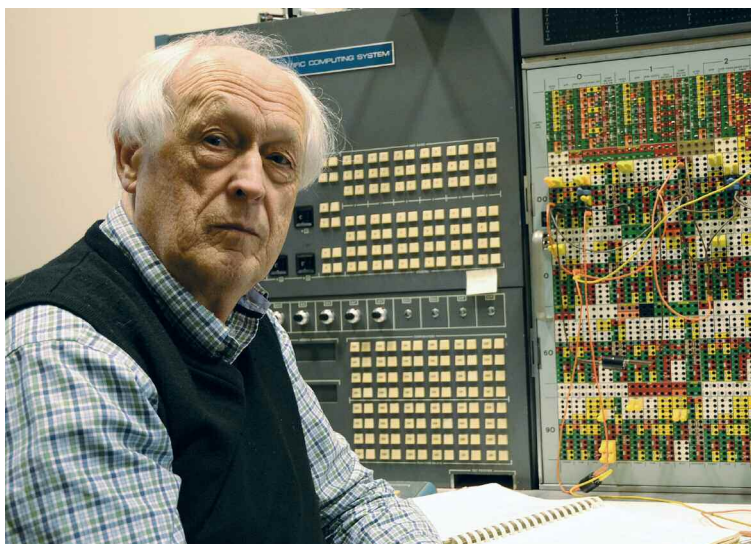
Dooijes loopt naar een ouderwets uitzijnde typemachine. Hij typt er iets op en met veel geratel komt er aan de zijkant een smalle strook papier uit waarin de machine gaatjes heeft geponst. Een offline-terminal uit de jaren vijftig van de X1 van Electrologica, ooit een roemruchte Nederlandse computerfabrikant.

Dooijes is al tien jaar gepensioneerd, maar nog steeds de drijvende kracht achter het Computermuseum van de Universiteit van Amsterdam.

We bevinden ons in een raamloze ruimte in een van de gebouwen in

het Science Park in Amsterdam. In kasten aan de muur staan ponskaartenlezers, beeldschermen en verouderde desktopcomputers. Op de grond staan tapedrives en computers op ijskastformaat: minicomputers uit de jaren zeventig. “Dit is maar een klein deel van de collectie”, zegt Dooijes. “Het depot bevindt zich in een ruimte van de UB en daar staat het grootste deel.”

Het begon toen in de jaren tachtig de pc opkwam. Onderzoekers wilden allemaal wel zo'n ding op hun bureau, want dan kon je zelf je berekeningen maken en hoefde je niet



Edo Dooijes voor een vroege, analoge computer

## ‘Prachtige machines moet je bewaren’

meer met je tape of je ponskaarten naar de balie van het rekencentrum. Na een dag of wat kon je dan de geprinte output ophalen. Dooijes: “En dan had je vaak alleen maar een foutmelding, kon je opnieuw be-

ginnen”.

De pc's bleken niet te stuiten, de balies van de rekencentra werden opgeheven en peperdure apparatuur dreigde op de mestvaalt van de geschiedenis te belanden.

Dooijes, van oorsprong opgeleid en gepromoveerd in de natuurkunde, maar daarna overgezwaaid naar de informatica, kon het niet aanzien en samen met zijn toenmalige directeur René Jongerius stichtte hij in 1991 het Computermuseum, een opvanghuis voor afgedankte rekenapparatuur.

En voor alles wat erbij hoort. Tapedrives bijvoorbeeld, de apparaten waarop de nu in onbruik geraakte

computertapes werden afgespeeld. Zo af en toe komt er iemand aanzetten met zo'n brede spoel. Acht-honderd meter tape, met 120 megabyte aan gegevens. Maar welke? “Als de band nog in goede staat is, kunnen we hem nog afspelen en de gegevens op een cd zetten, zegt Dooijes. De tape wordt eerst door een poetsmachine gehaald, om te voorkomen dat de leeskop van de tapedrive dichtslibt met plastic residu. “Als de filterpapierjes bruin werden, dan wist je: o, daar zitten de data. Hoefde je verder geen moeite meer te doen.”

Maar vaak lukt het ook wel. Archeoloog Esther Jansma was als een kind zo blij toen Dooijes haar boomringgegevens die op antieke Apple IIe-floppies stonden, kon overzetten naar een gangbaar formaat. Een huisarts die zijn complete administratie op een zeer incompatible Philipscomputer had gezet, kreeg na veel gepuzzel zijn brieven en noti-

ties weer terug. Een fabrikant van aluminiumprofielen zag zijn productie stranden toen de computering van de freesmachine crashte. Dooijes maakte van de ponsband die nog ergens lag een floppy en de fabriek kon weer verder.

En daar staat een grote doos met een verzameling ponskaarten op verwerking te wachten. Het zijn gegevens uit oude enquêtes naar de arbeidsomstandigheden van textielarbeiders. Het computermuseum heeft nog een goed werkende ponskaartlezer. De rubber wietjes voor het transport van de kaarten heeft Dooijes net met behulp van een bedrijf dat raamrubbers maakt gerestaureerd. Nu moet er nog wat gesleuteld worden aan het apparaat dat de gegevens naar pc-formaat overzet.

“Ja, en daarom moet er eigenlijk altijd zo iets als een computermuseum zijn”, zegt Dooijes. “Je moet machines hebben om die oude formaten te kunnen lezen”. Hij moet het tot zijn verdriet zonder veel budget doen. “Deze ruimte is natuurlijk te klein. En wat gaat er gebeuren als ik ermee ophoud?”

Maar vlak voor E-data ter perse gaat belt Dooijes op met goed nieuws: “We gaan verhuizen naar een mooiere en grotere ruimte, hier tegenover. Half oktober kunnen we open!”

Informatie over de collectie en openingstijden: [www.science.uva.nl/museum](http://www.science.uva.nl/museum).



De offline-terminal X1 van Electrologica, met links de ponsstrook

## Vaker raak door registers

*Eric Balster*

Nederland beschikt over zeer goede registerdata (administratieve gegevens die zich veelal in overheids-systemen bevinden). Voor wetenschappelijk onderzoek heeft het Centraal Bureau voor de Statistiek (CBS) een zorgvuldig beveiligde infrastructuur beschikbaar om deze data te gebruiken en ze te koppelen met surveydata. Voor de onderzoeker betekent het veel nauwkeuriger informatie en minder ‘missings’ (weet niet/wil niet zeggen). Voor de respondent betekent het minder vervelende en tijdrovende vragen (bijvoorbeeld over pensioenrechten). Hogere datakwaliteit is het gevolg. Klaas de Vos, senior onderzoeker bij CentERdata heeft in een pilot-study surveydata van de DNB Household Survey Bank (DHS) gekoppeld aan het Sociaal Statistisch Bestand (SSB) van het CBS waarin informatie uit verschillende registers is verwerkt. De pilot toonde aan dat een koppeling haalbaar is en betere data oplevert. In dit geval is specifieke informatie over loon en inkomen vergeleken.

Toch is koppelen met registerdata niet altijd mogelijk omdat registerdata meestal met vertraging beschikbaar komen. Tevens kost koppelen geld en zit er een ethische en juridische (privacy) kant aan het koppelen met registerdata. Omdat het ondoenlijk is om toestemming te vragen aan de respondenten bij iedere koppeling zal dit op survey-niveau moeten worden verankerd. En dat is moeilijker dan het lijkt.

<http://www.centerdata.nl/dhs>

### FOCUS

**Het overzicht toont een aantal databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS, CentERdata, Huygens ING (Huygens Instituut en Instituut voor Nederlandse Geschiedenis) en DANS.**

#### Centraal Bureau voor de Statistiek

Een volledig overzicht van de CBS-bestanden staat op [www.cbs.nl/microdata](http://www.cbs.nl/microdata)

- Statistiek financiën niet-financiële ondernemingen, 2008-2009
- Statistiek financiën grote niet-financiële ondernemingen, 2008-2009
- Enquête beroepsbevolking, 2010
- Inkomens Panel Onderzoek, 2009
- Regionaal inkomensonderzoek, 2005, 2008
- Algemeen bedrijfsregister, 2006-2010
- Pensioenaanspraken, 2005-2008
- Arbeidsongeschiktheidsuitkeringen, 2010
- Bijstandsuitkeringenstatistiek, registraties, 2010
- Woonruimteregeerterij, 2009, 2010

#### CentERdata LISS Data Archive

De LISS-bestanden zijn kosteloos beschikbaar via [www.lissdata.nl/dataarchive](http://www.lissdata.nl/dataarchive)

#### Studies LISS panel

- Do voters learn where parties stand for? A study on the effect of the 2010 election campaigns on issue knowledge (Aelst, P. van), mei 2010

## SINDS KORT BESCHIKBAAR

- LISS Core study - Politics and values, wave 4 (CentERdata), dec 2010 /jan 2011
  - National elections for the house of representatives 2010 (Aelst, P. van), jun 2010
  - Local elections 2010, (Aelst, P. van), mrt 2010
  - Enrichment experiment (CentERdata), mrt 2011
  - Initial questionnaire 2010 (CentERdata), maandelijks
  - Self-regulatory orientation: addressing a basic aspect of the self and its relation to social indicators and life-outcomes (Keller, J.), febr 2011
  - Time use and consumption, wave 2 (Vermeulen, F., Soest, A. van) sept 2010
- Studies allochtonen panel**
- Health, wave 1 (CentERdata), nov 2010

#### Huygens ING

De bestanden staan op [www.historici.nl](http://www.historici.nl)

- Bronnen tot de geschiedenis van den handel met Engeland, Schotland en Ierland. 4 delen, 1150-1585
- Bronnen tot de geschiedenis der Leidsche universiteit. 7 delen, 1574-1811
- Bronnen betreffende de Midden-Molukken. 4 delen, 1900-1942
- Romeinse bronnen voor de Nederlandse

geschiedenis, 20 delen

- De dagregisters van het kasteel Zeelandia, Taiwan, 4 delen, 1629-1662
- Pauselijke aflaten (Codex documentorum sacratissimarum indulgentiarum Neerlandicarum), 1 deel, 1300-1600

#### DANS EASY

De DANS-bestanden komen van diverse onderzoekinstellingen en zijn kosteloos beschikbaar via <http://easy.dans.knaw.nl>

#### Archeologie

- Reeks: Rapportages Archeologische Monumentenzorg (RAM), Rijksdienst voor het Oudheidkundig Bodemonderzoek, 114 delen) o.a.: RAM125: De Romeinse villa Meerssen-Onderste Herkenberg (T. de Groot), 2005
- Scheepswrak Muiden, inventariserend veldonderzoek (onderwaterfase) door middel van een duikinspectie. Periplus Archeomare (S. v.d. Brenk en W.B. Waldus), 2008
- Best centrum, archeologisch en cultuurhistorisch bureauonderzoek SRE Milieudienst (J. Bosman en R. Berkvens), 2011
- Inventariserend veldonderzoek door middel van een proefsleuf De Kikkerpit te Domburg, gemeente Veere. Walcherse Archeologische Dienst (B.H.F.M. Meijlink), 2007
- Landgoed Ekenstein bij Appingedam, ge-

meente Appingedam. Libau (A. Mennens-van Zeist), 2009

- Abrikoosstraat, Utrecht, gemeente Utrecht. IDDS Archeologie B.V. (M. Berkhout), 2011
- Archeozoologisch onderzoek Achlum 2009 (masterscriptie A.G.J. Hullegie), 2010

#### Sociale Wetenschappen

- Culturele veranderingen in Nederland (Sociaal en Cultureel Planbureau), 2008
- Permanent Onderzoek van de Leefsituatie-POLS Module POLS Basis (beveiligd microbestand van het CBS), 2009
- Permanent Onderzoek van de Leefsituatie-POLS Module POLS Gezondheid (beveiligd microbestand van het CBS), 2009
- Arbeidsaanbodpanel - met COROP variabelen (M. de Voogd-Hamelink, Sociaal en Cultureel Planbureau), 1985-2008
- Arbeidsvraagpanel - Aanvulling I en II (M. de Voogd-Hamelink, Sociaal en Cultureel Planbureau), 1989-2008
- Doorsnee Nederlanders in de Oorlog (W.C. Ultee, Radboud Universiteit Nijmegen), 2006-2010
- Labour participation of single parents (Dr. F.W. van Wel, Universiteit Utrecht), 1999
- Working and caring by mothers with young children (Dr. F.W. van Wel, Universiteit Utrecht), 1999
- Working and caring by single parents in the Netherlands (Dr. F.W. van Wel, Universiteit Utrecht), 1999

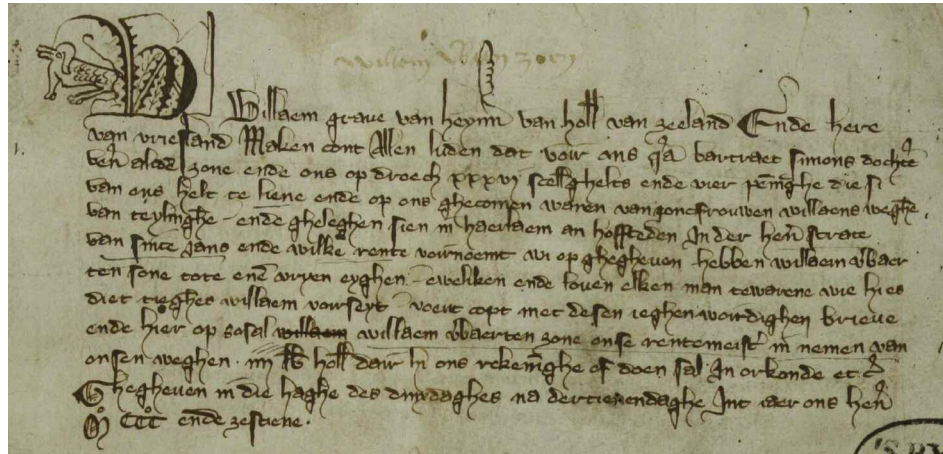


Ze werden in de de veertiende eeuw samengesteld, en ze zijn nu bijna allemaal op het wereldwijde web te bekijken en te lezen: de registers van de Hollandse grafelijkheid. *Jan Burger*

Bij het Huygens Instituut voor Nederlandse geschiedenis (Huygens ING) wordt sinds na-jaar 2006 gewerkt aan een digitale editie van die registers, die door de klerken van de graven van Holland, Zeeland en West-Friesland werden bijgehouden. Dit project nadert zijn voltooiing: inmiddels zijn 2953 documenten gepubliceerd. In totaal worden zo'n 3600 documenten uit de periode 1316-1345 ontsloten. Het zijn voornamelijk oorkonden, meestal door de graaf uitgevaardigde stukken met betrekking tot zijn diverse territoria, terwijl er ook registers zijn met teksten die de relaties met het buitenland betreffen.

Oorkonden vormen voor de mediëvist een belangrijke bron. Ze bieden een schat aan feitelijke informatie over allerlei aspecten van maatschappij en bestuur. Daarom worden de oorkonden al sedert de 18e eeuw grote aantallen uitgegeven. Vanaf de 19e eeuw gebeurt dat op een wetenschappelijk verantwoorde wijze, met de Duitse Monumenta Germaniae Historica als lichtend voorbeeld. In Nederland zijn vanaf het begin van de 20e eeuw oorkondenedities verschenen van bijvoorbeeld Holland en Zeeland, van Utrecht, van Gelre en van Noord-Brabant, elk met daarin alle over-

# De Hollandse graven hebben hun oorkonden nu op het net



Oorkonde uit 1317, waarin graaf Willem III aan zijn rentmeester een renteleen schenkt van 36 schelling en 4 penning

geleverde oorkonden over het betreffende ter-ritoir. Deze edities eindigen rond 1300, omdat de oorkonden zo talrijk worden dat volledigheid in een wetenschappelijk verantwoorde uitgave onmogelijk wordt.

## Geretrodigitaliseerd

Digitalisering opende nieuwe mogelijkheden voor oorkondenedities, en daarbij worden internationaal verschillende strategieën gevolgd. Om te beginnen worden vele van de (honderden!) bestaande papieren edities geretrodigitaliseerd; dergelijke plannen bestaan ook voor de oorkondenboeken die eertijds zijn uitgege-

ven door het Huygens ING. Een koppeling van dergelijke edities over de landsgrenzen is dan een logische stap, omdat tot in de 13e eeuw oorkonden een grotendeels uniforme Europese cultuur weerspiegelen, en zij mede-endeels zijn geschreven in dezelfde taal, het Latijn.

Een tweede optie is om door te gaan met het uitgeven van materiaal op de oude, wetenschappelijk verantwoorde en integrale editie-wijze, maar nu digitaal: het Oorkondenboek van Noord-Brabant is daarvan een voorbeeld ([www.donb.nl](http://www.donb.nl)).

Blijft het probleem van de grote massa's oor-

konden van na 1300. Een mogelijkheid is het om afbeeldingen van zoveel mogelijk oorkonden online te zetten, zonder de pretentie van wetenschappelijk editeren. Voorbeelden hiervan zijn het Gronings-Drentse Cartago ([www.cartago.nl](http://www.cartago.nl)), en vooral het Monasterium-project ([www.monasterium.net](http://www.monasterium.net)), dat diverse midden-Europese regio's bestrijkt (inmiddels meer dan 220.000 oorkonden).

Voor de digitale editie van de Hollandse grafelijke registers is weer een andere aanpak gekozen, namelijk een wetenschappelijke uit-gave, maar dan niet meer gericht op regionale compleetheit. Anders dan bij de papieren voorganger, het Oorkondenboek van Holland en Zeeland tot 1299, worden hier niet alle oorkonden uitgegeven, maar is dit een integrale editie van een afzonderlijk, maar om-vangrijk bronnencomplex. Tegenover de drie-kwart eeuw van het Oorkondenboek, zal het registerproject uiteindelijk zes jaar vergen. En daarbij komen de voordelen van een digi-tale uitgave: er zijn uitgebreide zoekmogelijkheden in en door de teksten, en de registers zijn compleet afgebeeld. De onderzoeker krijgt op deze wijze een optimale toegang tot het materiaal.

[www.historici.nl/Onderzoek/Projecten/RegistersVanDeHollandseGrafelijkheid1299-1345](http://www.historici.nl/Onderzoek/Projecten/RegistersVanDeHollandseGrafelijkheid1299-1345)

## Google wil graag iets terugdoen

Jon Orwant is engineering manager bij Google en werkt vooral aan Google Books. Hij was in Nederland op uitnodiging van NWO. E-data sprak met hem. *Peter Boot*

Eén ergernis voor de niet-Amerikaanse gebruikers van Google Books is dat we sommige boeken niet te zien krijgen, terwijl het auteursrecht toch echt verstreken is. Orwant: "We kennen niet altijd de precieze publicatiedatum van een boek – we verzamelen boekgegevens uit heel veel bronnen, en het is een heel werk om die te matchen. Het volgende probleem is dat we geen betrouwbaar register van overlijdensdata van auteurs hebben. We moeten voorzichtig zijn, want mensen procederen graag tegen Google – zeker omdat in de VS bij schendingen van het auteursrecht schadevergoedingen tot \$150.000 worden toegekend."

Bij veel digitaliseringsprojecten



Jon Orwant

laat de kwaliteit van de OCR (Optical Character Recognition) te wensen over. Werken jullie daar aan?

"Ja, we hebben een OCR-team, dat trouwens de software als open source ter beschikking stelt. Bij de Gotische letters en klein afgedrukte tekst hebben we de laatste tijd veel vooruitgang geboekt."

Werk Google ook aan manieren om tekststructuur in een boek te analyseren?

"Ja. Ik denk dat we onze boeken al kunnen taggen in TEI Lite (TEI: Text Encoding Initiative, standaard voor codering van teksten). Veel verder in de analyse zouden we nu niet durven gaan. We zouden liever een basistekst leveren, en dan de wetenschapper de mogelijkheid geven om de tekst te annoteren. Maar de infrastructuur daarvoor vergt nogal wat. Hoe weten we dat iemand geen onzin typt? Moeten we anoniem commentaar toestaan? Wat moet de eenheid van annotatie zijn: een boek, een bladzijde, een hoofdstuk, een woord? Mag iemand zijn eigen annotaties nog wijzigen?"

Wordt dat een betaalde dienst?

"Nee, het gaat Google goed, en willen iets teruggeven aan de wereld. We verdienen nu wel iets aan de boeken, door advertenties en het verkopen van digitale versies, maar dat haalt het niet bij wat het scannen

gekost heeft. In dezelfde spirit zijn de corpora van onze n-gram viewer (een website waar het gebruik van woorden in de tijd kan worden gevolgd – PB) open source beschikbaar. Iedere onderzoeker kan de bestanden downloaden en herpubliceren met een eigen interface."

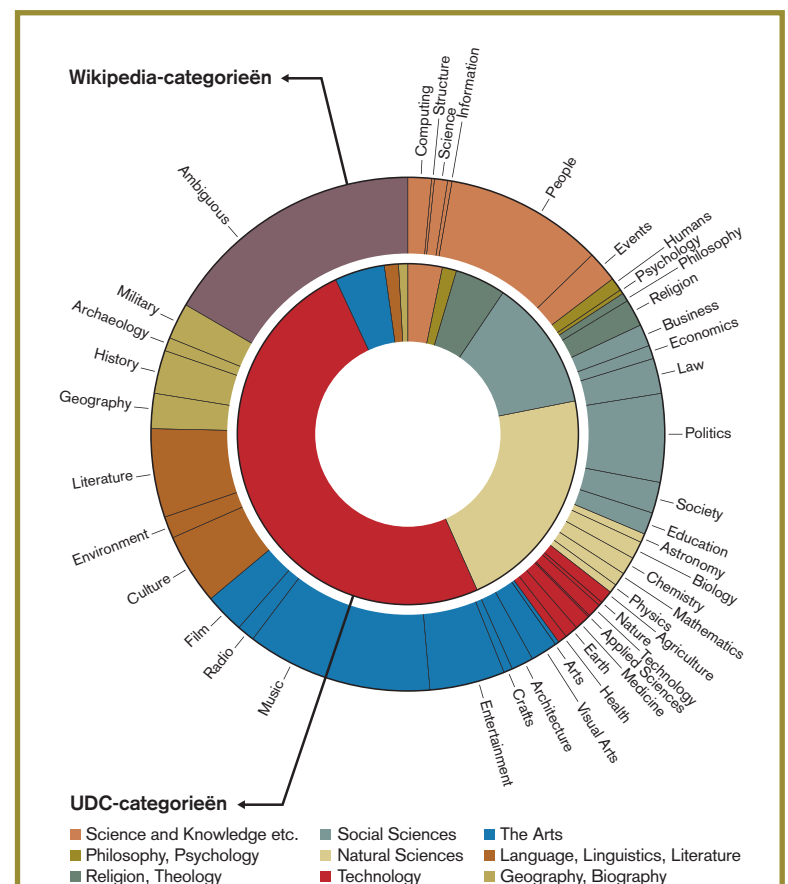
Wanneer komt de Nederlandse versie?

"Die komt, maar wanneer weet ik nog niet. We willen eerst zeker weten dat de collectie voldoende representatief is. Van de 3,1 miljoen Nederlandstalige boeken hebben we er nu 168.000 gescand, maar overwegend ouder materiaal. Dat is nog niet genoeg. Houd er trouwens rekening mee dat we voor deze bestanden regelmatig met nieuwe versies zullen komen, omdat we nieuwe boeken hebben gescand en onze software weer beter is geworden. Analyses op de bestanden zullen dus altijd moeten aangeven met welke versie ze zijn uitgevoerd. Daarmee wordt onderzoek in de digitale geesteswetenschappen repliceerbaar. En omdat de zichtbaarheid van onderzoek met digitale bronnen enorm toeneemt, willen we er zeker van zijn dat het onderzoek is dat de toets der kritiek kan doorstaan."

<http://books.google.com/>

<http://ngrams.googlelabs.com/>

<http://googlebooks.byu.edu/>



## UDC en Wikipedia: ontworpen versus gegroeid

De twee schijven hierboven stellen twee indelingen van informatie voor: de categorie-indeling van de Universal Decimal Classification (UDC, binnenste ring) en die van Wikipedia (buitenste ring). In de twee ringen wordt het contrast zichtbaar tussen het betrekkelijk formele UDC (ontwikkeld aan het begin van de twintigste eeuw en veel gebruikt in bibliotheken) en Wikipedia, dat meer een gegroeid, sociaal systeem is. In UDC is veel ruimte voor technologie (toegepaste wetenschappen), terwijl in Wikipedia kunst, entertainment en sport dominant zijn. Voor een belangrijk deel komt dat doordat Wikipedia alle informatie behandelt, niet alleen de wetenschap. De tekening is een bewerking van een beeldgrafiek van het Knowledge Space Lab project, dat wordt gesubsidieerd door het Strategiefonds van de KNAW. Deelnemers zijn de Virtual Knowledge Studio, the e-humanities group, DANS, de Erasmus Studio Rotterdam en BigGrid NL. De beeldgrafiek is onderdeel van een prachtig gekleurde kaart die onlangs is geselecteerd als een van de tien beste in de jaarlijkse wedstrijd van Places & Spaces.

De originele kaart, met nog veel meer graphics, is te zien op [http://scimaps.org/maps/map/design\\_vs\\_emergence\\_127/](http://scimaps.org/maps/map/design_vs_emergence_127/)



## Statistiek en sushi

De Rotterdamse aio Maria Fleischmann bezocht de Summer School in Ann Arbor.

Deze zomer ben ik twee keer vier weken in Ann Arbor, Michigan in de VS geweest. Ann Arbor is een klein stadje dichtbij Detroit en zou niet echt bekend zijn als er niet twee dingen waren: de University of Michigan met al zijn sportverenigingen en het ICPSR summer program voor kwantitatieve methoden.

Voor mij, aio bij de afdeling Sociologie aan de Erasmus Universiteit Rotterdam, waren de cursussen in statistische methoden de reden voor dit verblijf. In mijn dissertatie houd ik me bezig met arbeidsmarkttransities. Het paper dat ik recentelijk heb geschreven gaat over de beslissing om met pensioen te gaan in 21 Europese landen. Ik kijk daarbij naar individuele kenmerken, maar ook naar landenkenmerken.

ICPSR staat voor Inter-university Consortium for Political and Social Research, maar de letters zouden voor mij ook kunnen staan voor:

### International

Sinds de jaren zestig wordt dit *summer program* aangeboden en het staat hoog aangeschreven. Mensen uit alle continenten komen naar Ann Arbor om een methodische opleiding te volgen en als het over statistiek gaat begrijp je elkaar snel. Ver-

schillen in alledaagse dingen leiden dan juist tot spannende discussies, en je leert meteen nog meer.

### Classes

De programmaleiders ontraden het om meer dan twee workshops tegelijk te volgen. Maar het is erg moeilijk om je tot twee cursussen te beperken als je eigenlijk iedere cursus wel spannend vindt. En op school of op de universiteit heb je nog nooit zulke goede docenten gehad! Ik koos voor vier workshops: Regression II en Maximum Likelihood Estimation (MLE) in de eerste vier weken, en Regression III en Advanced Methods of MLE in het tweede gedeelte. In Regression II ging het om lineaire modellen, in Regression III onder meer om Model Specification en Outlying Observations. Het is erg nuttig om tot een betere modelspecificatie te komen of om te leren wat je moet doen met onverwachte observaties. Voor mij zal dat een rol spelen in bijvoorbeeld landenkenmerken die extreme waarden hebben. MLE was erg nuttig om survival analyse te doen. Met betrekking tot de pensioensleeftijd zou je kunnen vragen welke individuele kenmerken tot een jongere pensioensleeftijd leiden.



### Picknick

Na een week hard werken moet je even wat anders doen. Iedere twee weken organiseert het ICPSR een picknick. De bedoeling is om iedereen bij elkaar te krijgen en een middag te kletsen, te netwerken of samen te sporten. Volleybal, baseball, basketbal of voetbal, het is er alle-

maal en docenten en studenten genieten van een middagje niets statistiek-achtigs doen.



### Sushi

Eten in de VS was een onderwerp waarover je het met iedereen kon hebben. Met de supermarkten ver buiten de stad en wij allemaal zonder auto. Gelukkig is het restaurant-aanbod in Ann Arbor erg groot en er waren veel alternatieven naast burgers en sandwiches: Sushi, Koreaans, Veggie. The sky is the limit en het is goed betaalbaar!



### Reis

Tussen de cursussen en opdrachten door blijven dan maar enkele dagen over om nog wat van het land te zien. Een auto huren is goedkoop, maar ook de trein en de bus brengen je naar Chicago of Detroit. Wil je geen steden meer om je heen, dan kun je naar het Noorden van Michigan, naar de Sleeping Bear Dunes en Lake Michigan. Alleen, je merkt het al, aan je dissertatie kom je tijdens deze weken dus echt niet meer toe.

Het was een ontzettend leuke, leerzame en belevingsvolle zomer. Ik ben erg blij dat ik met de financiële steun van DANS en mijn begeleiders naar Ann Arbor kon gaan.

Maria Fleischmann

## COLUMN

### Drie netelige datakwesties

Open Acces van onderzoeksdata biedt nieuwe mogelijkheden voor wetenschappelijk onderzoek. Het is echter nog maar de vraag of in het huidige academisch klimaat deze vorm van e-science wel tot zijn recht kan komen.

Er is een standaardmodel van wetenschapsbeoefening aan de Nederlandse universiteiten. Wetenschappelijk onderzoek wordt verricht door excellente onderzoekers, die samenwerken in onderzoeksgroepen en onderzoeksprogramma's en beoordeeld worden op wetenschappelijke publicaties. Onderzoeksprojecten duren vier jaar, de periode waarbinnen een beginnend onderzoeker een proefschrift kan schrijven. Wetenschappelijke carrières worden gebouwd op individuele subsidies. Alle prestatiecijfers van de Nederlandse wetenschap laten zien dat dit model zeer succesvol is.

In haar boek *Zuiver om de wetenschap* verhaalt Patricia Faasse over het ontstaan van de levenswetenschappelijke instituten van de KNAW, waaronder het Herseninstituut, het Centraal Bureau voor Schimmelcultures (het ene CBS) en het Hubrechtlaboratorium. De oprichting van deze instituten aan het begin van de 20e eeuw laat een andere stijl zien van wetenschappelijk onderzoek. Met veel geduld bouwen de eerste directeurs van de instituten collecties op van onderzoeksmateriaal. Soms zonder dat duidelijk is wat de directe waarde van het verzamelde materiaal is voor het eigen onderzoek, en vaak om andere onderzoekers op zeker moment van dienst te kunnen zijn.

Het ontstaan van de instituten weerspiegelt een taxonomische stijl van onderzoek die in de loop van de 20e eeuw steeds meer is weggedrukt. Veelzeggend is bijvoorbeeld hoe in de jaren tachtig bij bezuinigingen de taxonomie vrijwel geheel verdween uit de biologische faculteiten. En misschien wel het meest taxonomische instituut dat we in Nederland op dit moment kennen, het Centraal Bureau voor de Statistiek (dat andere CBS), komt in overzichten van de Nederlandse wetenschappelijke faciliteiten nooit voor.

Maar de taxonomie is terug: in de vorm van e-science en Open Acces van databases.

Is de organisatie van het academisch onderzoek wel toegesneden op deze nieuwe stijl van onderzoek? Ik geef u drie netelige kwesties om de gedachten hierover te prikkelen.

- Promotie-eisen: Zijn we bereid onderzoekers te laten promoveren op een mooie database? Of leren we hen de data eerst alleen zelf te gebruiken voor publicaties en deze pas openbaar te maken als de eigen carrière op de rails staat?
- Databasebeheer: Zijn we binnen universiteiten in staat om de continuïteit van databases te organiseren: de opbouw, de opslag, het onderhoud en de vernieuwing van data? Of moeten we daarvoor nieuwe para-universitaire instituten oprichten?
- En als laatste kwestie: Zijn we in staat om het eigendom en de toegang te regelen? Van wie zijn de data: van de onderzoeker, de onderzoeksinstelling, de financier of misschien wel van de data zelf (lees: patiënten, geëquipteerden, organisaties)? En wie neemt dan de telefoon op als een onderzoeker de data graag wil gebruiken voor nieuw onderzoek?

Ik vrees dat bij veel organisaties deze vragen nog nauwelijks beantwoord zijn. Ze worden overgelaten aan de onderzoekers. In het wetenschapsbeleid heet dat dan 'bottom up' en dan zijn we er voor. In de praktijk betekent het dat onderzoekers moeten aanmodderen, en dan is het jammer dat we Open Access van onderzoeksdata nog niet goed geregeld hebben.

Barend van der Meulen  
hoofd afdeling Science System  
Assessment van het Rathenau  
Instituut.



## GELEZEN

David Giaretta [et al.]: *Advanced Digital Preservation*. Springer, 2011, ISBN 978 3 642 16808 6, €99,95

In dit boek legt Giaretta uit waarom gangbare technieken voor digitale duurzaamheid (migratie, emulatie) niet werken voor wetenschappelijke onderzoeksdata, die vaak bestaan uit complexe digitale objecten als databases en softwareprogramma's. In plaats daarvan ontwikkelt Giaretta met zijn team een techniek die draait om het netwerk van representation information - ieder digitaal object moet alle informatie in zich bergen die ooit nodig kan zijn om het object te hergebruiken. Naast deze les digitale duurzaamheid voor gevorderden bevat het boek veel basisinformatie over digitale duurzaamheid en kan het ook goed op een minder technisch niveau worden gelezen, bijvoorbeeld door managers van repositories en databibliothecarissen.

Marloes Scholtens: *Born-digital Bijzondere collecties*. Rapport van de werkgroep Born-digital van de UKB commissie Bijzondere Collecties. Leiden, 2011

Digitaal materiaal dat van oorsprong digitaal is, en dus geen analoge tegenhanger heeft, loopt nog steeds een groot risico om verloren te gaan. Vaak is niet duidelijk wie er verantwoordelijk is voor collectiëren en voor duurzaam bewaren. In Nederland heeft de Koninklijke Bibliotheek de verantwoordelijkheid genomen voor elektronische boeken en tijdschriften, en het Nationaal Archief ontfermt zich over de bedrijfskundige archieven van de universiteiten, maar veel ander materiaal valt nog tussen wal en schip. De persoonlijke archieven van hoogleraren bijvoorbeeld, of wetenschappelijke archieven van vakgroepen. Het zeer gedegen rapport inventariseert de soorten materiaal die bedreigd worden en de internationale praktijk op dit gebied, en geeft aanbevelingen. Het rapport is beschikbaar op de NCDD-website [www.ncdd.nl/](http://www.ncdd.nl/).

Madeleine de Cock Buning, Barbara van Dinther, Christina H. Jeppesen-de Boer, Allard Ringnalda: *The legal status of research data in the Knowledge Exchange partner countries*. Centre for Intellectual Property Law (CIER) and Molengraaff Institute for Private Law, Utrecht University, 2011

Knowledge Exchange is a cooperative effort between four European organisations that support the use and development of Information and Communications Technologies infrastructure for higher education and research. Based on the four national strategies the joint vision of the initiative is to make a layer of scientific and scholarly content openly available and re-usable on the internet. The central aim of this report is to identify whether Intellectual Property (IP) rights can form any obstacle to sharing publicly funded research data under open access terms. To the extent that such obstacles do indeed exist, the report makes some suggestions as to how these could be resolved. [www.knowledge-exchange.info/](http://www.knowledge-exchange.info/)