

Pagina 3 • Booming archeologie •
Het archeologisch depot EDNA groeide in twee jaar tijd spectaculair als gevolg van het terugwerkende-kraft-project EDNA II. Naast het aantal deponeringen stegen ook de downloads.

Pagina 4 • Verrijkte publicaties •
Niet de techniek maar de organisatie blijkt de belangrijkste bottleneck bij verrijkte publicaties.

HENK VAN RUITENBEEK



Pagina 5 • David Hoover • 'Niet iedereen zit te wachten op onderzoek dat mooie theorieën ontkracht,' zegt de Amerikaanse tekstanalyticus David Hoover.

Pagina 6 • NSCR • De rubriek Focus stelt scherp op het Nederlands Studicentrum Criminaliteit en Rechts-handhaving, dat 'het allerbeste' onderzoek wil doen.

Pagina 8 • Meningonderzoek •
Hoe kan het dat er zulke grote verschillen zijn tussen de voorspelde verkiezingsuitslagen van allerlei opiniebureaus? Hans Schmeets geeft het antwoord.

EN VERDER

Agenda	2
Nieuws	3
Achtergrond	4,6
Sinds kort beschikbaar	7
Column	8
Gelezen	8



Biografisch portaal maakt 'martelaars en misdadigers' vindbaar

Op 17 februari is de website van het Biografisch Portaal van Nederland officieel gelanceerd met een muisklik op www.biografischportaal.nl door Alexander Rinnooy Kan, bestuurder van het financierende Prins Bernhard Cultuurfonds.

Het portaal zal alle biografische informatie bij elkaar brengen en doorzoekbaar maken die in oude en nieuwe biografische woordenboeken van Nederland is verschenen. Het initiatief werd genomen door tien culturele en wetenschappelijke instellingen, die daartoe een gelijknamige stichting in het leven hebben geroepen. Ook die stichting, met als voorzitter prof. J.C.H. Blom, ging op 17 februari officieel van start.

De website is bedoeld voor iedereen die informatie zoekt over een persoon of groepen van personen uit de Nederlandse geschiedenis. Dat kunnen wetenschappelijke onderzoekers zijn maar ook scholieren, journalisten, genealogen of mensen die gewoon willen bladeren. Het Biografisch Portaal wordt een index waarin

alle bewoners van Nederland te vinden zijn aan wie ooit in een van de bestaande naslagwerken een lemma (een korte levensbeschrijving) is gewijd: van staatslieden tot verzetshelden van prinses tot stakingsleiders, van sporthelden tot toneelspelers en van martelaars tot misdadigers. Alle personen in het Biografisch Portaal krijgen een *persistent identifier*, een duurzame internetverwijzing. Dat maakt verwijzen betrouwbaar en verspreide informatie over één en dezelfde persoon kan zo eenvoudig worden gekoppeld.

Het portaal digitaliseert ook naslagwerken die nog niet digitaal raadpleegbaar zijn. Dit worden zelfstandige collecties op internet. Alle collecties met biografische informatie blijven waar ze zijn. Het portaal

ontsluit ze met een index en met kerngegevens: pagina's met korte, feitelijke gegevens per persoon. Momenteel zijn de gegevens van zeven collecties opgenomen, met in totaal zo'n veertigduizend personen in meer dan vijftigduizend levensbeschrijvingen.

Omdat er veel verschillende bronnen worden gebruikt moet er ook rekening worden gehouden met verschillen in betrouwbaarheid en volledigheid. Ook minder betrouwbare informatie, bijvoorbeeld uit negentiende-eeuwse naslagwerken, blijft beschikbaar. Op de pagina met kerngegevens wordt de informatie naar betrouwbaarheid gerangschikt.

Met het oog op het Biografisch Portaal is een standaard afgesproken voor de uitwisseling van bio-

grafische informatie, BioDes. De daarbij gebruikte software is gedocumenteerd en met voorbeelden beschikbaar: www.biografischportaal.nl/biodes. Doorzoekbaarheid en selectiemogelijkheden vereisen data aan alle opgenomen personen metadata worden toegekend zoals leefjaren, sekse, geboorte- en sterfplaats, beroep. Dit enorme karwei vraagt nog veel handwerk. Ook in dit opzicht verwachten de makers van het Biografisch Portaal de komende jaren nog aanzienlijke groei.

Behalve van het Prins Bernhard Cultuurfonds kwam een startsubsidie voor het Portaal van NWO en financiële steun van SenterNovem. (Rik Hoekstra)

DANS zoekt uitbreiding naar bètawetenschappen

Data-instituut DANS wil zijn werkerterrein verbreden. Minder versnippering, beter gebruik van deskundigheid en schaalvoordelen zijn belangrijke overwegingen. Moederorganisaties NWO en KNAW reageren positief.

'Uitbreiden naar meer (of alle) wetenschapsgebieden is zinvol' zegte een visienota die DANS (Data Archiving and Networked Services) eind vorig jaar heeft aangeboden aan de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) en de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW). Beide organisaties richtten in 2004 DANS op als data-instituut voor de alfa- en gammawetenschappen. Een verbreed DANS zou volgens de nota 'belangrijk in het voordeel zijn om tot een van de sterkste datavoorzieningen in Europa en in de wereld uit te groeien'. DANS-directeur Peter Doorn wijst erop dat het idee niet alleen uit de koker van zijn instituut komt. 'Het rapport *Toekomst van ons*

Digitaal Geheugen van de Nationale Coalitie voor Digitale Duurzaamheid wees vorig jaar op het belang van samenwerking voor een gezamenlijke basisinfrastructuur. De gedachte kwam al aan de orde in gesprekken met wetenschappelijke organisaties, de Stichting Nationale Computerfaciliteiten (NCF) en het datacentrum van de drie Technische Universiteiten.' De beperking tot de alfa- en gammawetenschappen is nu soms een belemmering in de samenwerking met organisaties zoals SURF die wél alle disciplines in hun portefeuille hebben.

De sprong naar discipline-onafhankelijk werken is groot, bevestigd Doorn. 'De humaniora en de sociale wetenschappen bestrijken nu onge-

veer een kwart van het hele wetenschapsgebied. Er zou dus drie keer zo veel bij komen. Maar veel van onze huidige activiteiten zijn al discipline-onafhankelijk.' Het Datakeurmerk en de *persistent identifier*, ontwikkeld om bestanden met een tijdbestendig adres op het internet bereikbaar te maken, zijn voorbeelden. Toch zal uitbreiding van de beschikbare expertise nodig zijn. Dat hoeft echter niet in een keer en evenmin op alle gebieden: 'We doen nu ook niet niet alles op alle gebieden binnen alfa en gamma. Bovendien is er in de bètawetenschappen al veel activiteit op het gebied van data-archivering en uitwisseling. Doorn: 'We gaan niet dubbelen. We gaan koppelen of aanvullen net zoals we nu al doen'.

Ron Dekker, directeur Instututen, Financiën en Infrastructuur van NWO, bevestigt dat zijn organisatie heil ziet in de ideeën van DANS: 'Wij hebben DANS gevraagd wat er nodig is voor Open Access op het gebied van data, zowel wat betreft de duurzame opslag als wat betreft toegankelijkheid. Dit is het antwoord daarop.' De DANS-notitie werd half februari besproken in het NWO-bestuur, dat er een basis voor verder gesprek in zag. NWO-voorzitter Jos Engelen na afloop: 'We willen de infrastructuur en voorzieningen voor duurzame opslag en toegang tot data faciliteren. Daarover gaan we met DANS overleggen'.

Ook directeur Onderzoek Theo Mulder van de KNAW staat positief tegenover de verbredingsgedachte: 'De discussie over *e-research* en hergebruik van data is nu actueel. Dan lijkt het alleen maar logisch om de positie van DANS daarbij te betrekken'. Dat het instituut niet eerder

met een wetenschapsbrede opdracht werd belast verklaart Mulder uit de geschiedenis: 'Het waren destijds instituten uit de alfa- en gammasfeer die in een nieuw verband ondergebracht moesten worden. Maar als je dan ziet dat de bètawetenschappen voor een groot deel dezelfde problemen kennen, moet je naar synergie streven'.

Voor de verbredingsambities is natuurlijk geld nodig. Maar minder dan de gebiedsuitbreiding zou doen denken, benadrukt Doorn. 'We denken aan een verdubbeling of verdrievoudiging van het budget'. De KNAW geeft op dit gebied even niet thuis. Mulder: 'Wij zijn nu bezig met ècht pijnlijke bezuinigingsmaatregelen'. DANS-directeur Doorn laat zich daardoor niet afschrikken. 'Bij infrastructurele werken moet je vaak in trajecten van jaren denken'. (MdG)

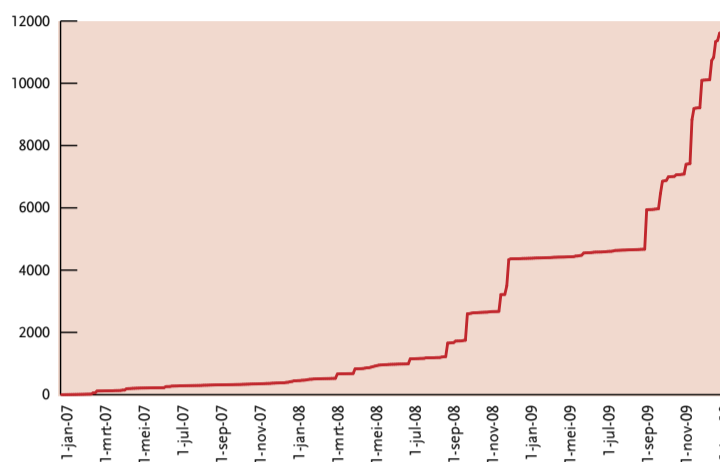
Archeologisch archiefproject leidt tot sterke stijging van bezoek

Een enorme groei van het aantal gearchiveerde opgravingen en een sterke stijging van het gebruik. Dat zijn de opvallende resultaten van het EDNA II-project (Elektronisch Depot voor de Nederlandse Archeologie) dat dit voorjaar wordt afgesloten.

Na het eerste EDNA-project, dat was bedoeld om het archeologische e-archief van de grond te krijgen, was deze tweede fase vooral bedoeld om het archief te vullen met data en verslagen van opgravingen die in de voorafgaande vijftien jaar zijn uitgevoerd. Veilig stellen en beschikbaar houden van die informatie was het voor de hand liggende eerste doel. Daarnaast zou het archiveren met terugwerkende kracht ook moeten helpen om de loop erin te krijgen bij het nieuwe elektronische archief. De resultaten liegen er niet om. EDNA zorgde voor een spectaculaire groei van het doe-het-zelf e-archief EASY van DANS

‘Een van de ideeën achter het project was dat er een zekere basishoeveelheid informatie in zo’n archief te vinden moet zijn om te zorgen dat er meer bezoek komt. Is het een keer zo ver, dan zullen bezoekers terugkomen en ook zelf datasets gaan deponeren,’ zegt projectleider Milco Wansleben. ‘Nu we bezig zijn met het afronden van het project lijkt het erop dat het precies zo heeft gewerkt’.

Het aantal gedeponeerde datasets en rapporten steeg van een paar honderd aan het begin van het project in de zomer van 2007 tot ruim elf duizend nu. Tegelijk zagen we het aantal raadplegingen stijgen tot 30 duizend per maand aan het eind van vorig jaar. Ook het aantal downloads nam sterk toe. ‘In 2009 werden 180 duizend files uit archeologische datasets gedownload. Daarbij gaat het vooral om foto’s en rapporten,’ merkt Wansleben op. Dat laatste is opmerkelijk omdat EASY eigenlijk vooral voor het opslaan en beschikbaar maken van data is bedoeld. ‘Het plan was om de data hier onder te brengen en de rapporten bij de Rijksdienst voor het Cultureel Erfgoed. Uit oogpunt van tempo en efficiency hebben we gaandeweg besloten om beide bronnen in DANS EASY op te slaan en dat heeft goed gewerkt’, aldus Wansleben. Nu de achterstand is ingehaald willen de beheerders van EDNA hun aandacht richten op het archiveren van huidige en toekomstige opgravingen. (MdG)



Aantallen datasets in EDNA 2007-2009.

<http://easy.dans.knaw.nl/>humanities>archaeology>

Alle parlementaire stukken 20ste eeuw digitaal

Sinds medio januari zijn alle stukken van de Eerste en Tweede Kamer uit de vorige eeuw online toegankelijk en full-text doorzoekbaar. Daarmee ligt het project op schema sinds de lancering van de website www.statengeneraaldigitaal.nl in de zomer van 2007, toen alleen nog de jaargangen 1990-1995 waren in te zien (*e-data* 2-2, september 2007). Aan de eerder gedigitaliseerde collectie zijn onlangs de *Handelingen*, *Bijlagen* en *Aanhangsels* van de voorgaande jaren tot 1917 toegevoegd. Het

materiaal uit de periode 1814-1900 zal in de loop van dit jaar gepubliceerd worden.

De website is een samenwerking tussen de Tweede Kamer en de Koninklijke Bibliotheek. Het doel is om de complete set *Handelingen* uit de periode 1814-1995 door microfilm te behouden voor het nageslacht en door digitalisering voor iedereen gratis toegankelijk te maken op het internet. (IA)

www.statengeneraaldigitaal.nl

Website en boek over GIS

Op 11 februari werd in Den Haag een website geopend die op een historische kaart verschillen in gemeentelijke gegevens weergeeft. Ook werd er een boek gedoopt: *Tijd en Ruimte; Nieuwe toepassingen van Geografische Informatiesystemen (GIS) in de alfawetenschappen*.



DAGMAR STIEBRAL

De website is een gezamenlijk product van DANS (Data Archiving and Networked Services) en de Radboud Universiteit. Gebruikers kunnen gemeentelijke data van elk jaar tussen 1812 en 2000 uploaden en zichtbaar maken. Het boek *Tijd en Ruimte*, uitgegeven door Matrijs in Utrecht, komt voort uit het AlfaGeo-project om het gebruik van GIS in de geesteswetenschappen te stimuleren. Het toont de nieuwste ontwikkelingen op het gebied van GIS in de alfawetenschappen. Op de foto bereiden de redacteurs Onno Boonstra (links) en Anton Schuurman (rechts) en DANS-medewerker René van Horik zich voor op de boekpresentatie.

<http://nlgis.dans.knaw.nl>

Walewein ende Keye

Eind 2009 verscheen een digitale editie van de dertiende-eeuwse Middelnederlandse Arturroman *Walewein ende Keye*, bezorgd door Marjolein Hogenbirk en Wim Gerritsen in samenwerking met de afdeling ICT & Teksten van het Huygens Instituut. Met de editie wordt de roman toegankelijk voor een groter publiek. De tekst is voorzien van foto's uit het handschrift en van uitgebreide annotaties en woordverklaringen die naar believen opgeroepen kunnen worden of uitgeschakeld. Zo is de editie geschikt voor zeer divers gebruik. In een inleiding wordt de roman toegelicht. Dankzij tweemaandelijks updates kunnen nieuwe inzichten verwerkt worden. (Marjolein Hogenbirk) www.waleweinenendekeye.huygens.knaw.nl/

Middelnederlandse encyclopedie

Op 5 maart lanceert het Huygens Instituut officieel het digitaal beschikbaar gemaakte werk van Bartholomeus Engelsmans, *Van den proprieteyten der dinghen*. De uitgave van deze Middelnederlandse vertaling (1485) van de dertiende-eeuwse encyclopedie *De proprietatibus rerum* is gemaakt door de Werkgroep Middelnederlandse Artesliteratuur (WEMAL). De digitale teksteditie is al beschikbaar bij het Huygens Instituut op <http://bartholomeusengelsman.huygens.knaw.nl>. De tekst komt nu ook beschikbaar via de Digitale Bibliotheek voor de Nederlandse Letteren (www.dbnl.org) en het Instituut voor Nederlandse Lexicologie (www.inl.nl) (Karina van Dalen-Oskam)

Immigranten in archief

Het landenvergelijkend databestand IFIS (International File of Immigration Surveys) wordt momenteel klaargemaakt voor hergebruik door derden. Het werd in de jaren 2000-2005 samengesteld door prof. Frank van Tubergen, die er aan de universiteit van Utrecht een veelgeprezen dissertatieonderzoek mee deed naar de integratie van immigranten en hun kinderen in westers landen. Aan de vele verzoeken om de data te mogen gebruiken kan tot op heden niet worden voldaan, onder meer omdat sommige metadata ontbreken en onzekerheid bestaat over de toestemming voor zulk gebruik. Met subsidie van DANS wordt dat nu verholpen zodat IFIS binnenkort via het DANS-archief EASY kan worden geraadpleegd.

Audiovisuele samenwerking

AVANET is de naam van een nieuw samenwerkingsverband van audiovisuele archieven in Nederland. Het is een vervolg op AVA, de werkgroep die functioneerde binnen de opgeheven erfgoedkoepel DIVA. Archiefinstellingen met veel audiovisueel materiaal werken erin samen, zoals Beeld en Geluid, verschillende stedelijke archieven en DANS. Behalve het ontsluiten van cultureel erfgoed speelt wetenschappelijk gebruik bij verschillende deelnemers een rol. De leden gaan kennis uitwisselen en gezamenlijke projecten en workshops organiseren. AVANET is medio februari officieel gelanceerd.

Een hitparade van repositories

Rank	Repository	Country	Open Access	Full Text	Non-Open Access
1	Arxiv.org e-Print Archive	USA	4	2	3
2	Citeseer	USA	3	1	413
3	Scientific and Technical Information Network	USA	7	7	1
4	Social Science Research Network	USA	6	3	31
5	Research Papers in Economics	USA	2	6	142
6	Smithsonian/NASA Astrophysics Data System	USA	3	4	520
7	HAL-INSPIRE	FR	10	9	5
8	Institut National de Recherche en Informatique et en Automatique Archives Ouvertes	FR	10	11	35
9	ORNL Document Service	USA	11	10	40
10	MIT DSpace	USA	14	34	7
11	Centre Polytechnique Fédérale de Lausanne Infoscience	CH	8	14	48
12	Agricon Search Research in Agricultural and Applied Economics	USA	21	45	4
13	University of Michigan Deep Blue	USA	27	28	13
14	Hyper Archive en ligne Sciences de l'Homme et de la Société	FR	40	12	122
15	Igitur Archive Universiteit Utrecht	NL	26	60	3
16	Harvard Personal Robot Archive	USA	47	30	14
17	Universitat de València Departament d'Informàtica i Ciències de la Comunicació	ES	25	13	37
18	University of Southampton Department of Electronics and Computer Science	GB	40	20	32

Vijf Nederlandse archieven telt de laatste Top Honderd van het Ranking Web of World Repositories. Het Cybermetrics Lab van de Spaanse organisatie voor wetenschappelijk onderzoek CSIC, stelt de hitlijst op 'om vrije toegang tot elektronische publicaties en ander academisch materiaal te bevorderen.' Het Igitur Archief van de Universiteit Utrecht haalt met een vijftiende plaats de hoogste nationale ranking, op korte afstand gevolgd door het Groninger dissertatie-archief met een notering als negentiende.

De eerste plaats wordt bezet door arXiv.org van de Amerikaanse Cor-

nell University. In de Top Vijftien zijn verder acht andere in de VS beheerde repositories te vinden, naast drie Franse, twee Zwitserse en één van de Verenigde Naties.

De positie op de ranglijst wordt onder meer bepaald door de score op omvang, webzichtbaarheid en het aantal beschikbare pdf's. Excel- of andere databestanden worden nog niet mee gerekend. 'We concentreren ons nu nog op artikelen,' licht coördinator Isidro Aguillo van het Lab toe. 'Maar in de toekomst worden data-archieven zeker belangrijker en dan gaan we die ook opnemen'.

<http://repositories.webometrics.info>

Tekstanalyticus David Hoover:

Literaire computeranalyse als modern handwerk

HERMAN BRINKMAN

Dit jaar is David L. Hoover, professor of English aan de New York University, te gast in Nederland als visiting KNAW-professor. Hoover is een toonaangevend wetenschapper in de analyse van literaire verschijnselen met behulp van statistische methoden. Begin januari verbleef hij op het Huygens Instituut voor een eerste bezoek, een goede gelegenheid voor een gesprek over zijn werk en de ontwikkelingen in zijn vak.

Voor dat onderzoeksgebied zijn verschillende termen in omloop. Het brede onderzoeksveld wordt wel *digital humanities* genoemd, of *humanities computing*. Een deelgebied daarvan, waar de meeste van Hoovers publicaties zijn te situeren, wordt aangeduid als *computational stylistics*. Daarbinnen heeft *authorship attribution* zijn speciale belangstelling. Auteursonderscheiding (of auteursherkenning) kent een lange geschiedenis. Hoewel de wortels ervan reiken tot in de oudheid, en humanisme en positivisme sterk hebben bijgedragen aan de ontwikkeling van toeschrijvingsmethoden, vond er vooral in de twintigste eeuw een grote doorbraak plaats met de toepassing van statistiek op taalkundige verschijnselen. Later nam die een grote vlucht met de inzet van computerprogramma's en de vorming en beschikbaarstelling van grote tekstcorpora.

Vergelijken

Gevraagd naar de belangrijkste verschillen tussen de meer traditionele methoden en de methoden die hij in de praktijk brengt zegt David Hoover: 'Bij de tradionele methoden zien we vaak minutieuze analyses van één afzonderlijke tekst. Het is echter erg gevaarlijk om je op slechts één auteur te richten, omdat je bij gebrek aan vergelijking geen goede inschatting kunt maken over wat specifiek is voor een bepaalde auteur en wat niet. Het grote verschil met de technieken uit de *digital humanities* is dat er nu enorme tekstcorpora bestaan die dienst kunnen doen als vergelijkingsmateriaal. Vergelijken bestond in het verleden slechts als theoretische mogelijkheid; praktisch gezien was het ondoenlijk. Een ander belangrijk verschil is de houding ten opzichte van de bewijsvoering. Bij de traditionele methoden, die overigens nog altijd worden beoefend, is die voor een groot deel retorisch, op overtuiging gericht, terwijl de bewijsvoering in de *digital humanities* veel meer vraagt van de kwantitatieve aspecten van de bewijslast.'

Argwaan

Ondanks het feit dat er in de laatste jaren opmerkelijke resultaten zijn geboekt, kent nog maar een beperkt deel van de onderzoekers in de geesteswetenschappen de



HENK VAN RUITENBEEK

nieuwe methoden een plaats toe in het onderzoek. Volgens Hoover zijn daar verschillende redenen voor. Allereerst heeft dat te maken met de wijze waarop in de *humanities* wordt gekeken naar meer empirisch gefundeerde manieren van kennisverwerving. Er is volgens hem nog altijd argwaan tegenover een positivistische, op bewijzen gegrondveste tekstanalyse binnen de menswetenschappen.

Veel wetenschappers vrezen aangezien te worden voor handswerksman in plaats van theoreticus. Het aanzien van de theorie blijft het allerhoogste'

Tevens vermoedt hij de invloed van het postmodernisme, *new historicism* en context-gerelateerde benaderingen, waarbij het door de onderzoeker gehanteerde concept dikwijls belangrijker wordt geacht dan het onderzoeksobject. 'Er is over het algemeen nog steeds weinig interesse voor onderzoek dat louter tekstueel is gericht.'

Postkoloniale vooroordelen

De universeel toepasbare analysemodellen liggen bovendien klaar om te worden toegepast. 'Met een hamer in de hand voelt alles om je heen aan als een spijker. Ik heb meegemaakt dat een onderzoekster de stelling wilde verdedigen dat Virginia Woolfs *To the Lighthouse* uit 1927 gebaseerd was op postkoloniale vooroordelen. De enige feitelijke bewijsgrond die ze had was

dat over een van de personages ik meen vijf keer werd gezegd dat ze met oriëntaalse spleetogjes ergens naar keek. En dat terwijl die roman alleen gaat over een groep blanke mensen die zich ophouden aan de Schotse kust. Het gaat me niet om dit specifieke voorbeeld, maar om het feit dat teksten niet zo 'open' zijn als de postmodernisten graag zouden willen. Tekstinterpretatie zal zijn fundament toch moeten vinden in de tekst zelf; de tekst zelf schept de grenzen van de interpretatie ervan.'

Een concreet voorbeeld van de manier waarop *computational stylistics* een toegevoegde en corrigerende waarde heeft zie je bij analyse van het werk van Henry James. 'We weten dat hij op zekere leeftijd vanwege reumatische problemen zijn romans ging dicteren in plaats van ze zelf te schrijven. Daarbij is er al vaak geconstateerd dat er een aanzienlijk verschil bestaat tussen zijn vroege en late stijl.

Nu hebben sommigen beweerd dat die stijlontwikkeling is toe te schrijven aan de overgang van schrijven naar dicteren. We kunnen het niet bewijzen aan de hand van manuscripten, want die zijn er niet. Maar als we nu een digitale stijlanalyse maken van zijn romans, blijkt er nergens een scherpe overgang. Wel kunnen we een duidelijke ontwikkeling laten zien, maar die is overal gradueel. Dat betekent dat James stijlontwikkeling geheel losstaat van de manier waarop hij zijn romans schreef.'

Vruchtbaar samengaan

Hoover is niet van mening dat de corpus-gerelateerde onderzoeksmethoden de traditionele, niet-statistische analysemodellen overbodig maken. Er is alle ruimte voor een vruchtbaar samengaan, met name wanneer het gaat om de vervolgvraag. 'Zoals Hugh Craig (professor aan de University of Newcastle, Australia - HB) zegt: "Als je eindelijk zekerheid hebt over het auteurschap, wat ben je dan eigenlijk te weten gekomen, wat is het precies dat je hebt geleerd?"'

Politieke lading

De invloed van de *digital humanities* blijft, vreemd genoeg, achter bij de successen die er duidelijk worden geboekt. Anders dan je misschien zou denken is het in de Verenigde Staten, maar ook elders in de wereld, niet eenvoudig om voor projecten in de *digital humanities* onderzoeksfinanciering te krijgen. Hoe valt dat te verklaren? 'Niet iedereen zit te wachten op onderzoek dat mooie

theorieën ontkracht. Sommige theorieën hebben soms ook een politieke lading. Zo doe ik op het ogenblik onderzoek naar een anonieme tekst waarvan wel wordt beweerd dat hij door Edgar Allen Poe is geschreven. Als dat zo zou zijn, zou hij te boek komen te staan als een 'race theorist'. Zoals het er nu naar uitziet is Poe geenszins de auteur van die tekst, maar sommigen hebben moeite om dat te accepteren.'

Teveel theoretiseren

Dikwijls zijn de verwachtingen over de uitkomsten van onderzoeken ook te hoog gespannen. Men verlangt, zeker als de computer wordt ingezet, een ondubbelzinnig oordeel maar een oordeel kan soms niet anders dan onder het nodige voorbehoud worden gegeven.

'Vergelijken bestond in het verleden slechts als theoretische mogelijkheid; praktisch gezien was het ondoenlijk'

Hoovers indruk van het huidige debat onder zijn collega's, met name onder degenen die werken met computeranalyses, is dat men over het algemeen teveel blijft steken in het theoretiseren over wat er zou moeten gebeuren in plaats van het implementeren van wat er al mogelijk is. 'Veel wetenschappers vrezen het nu eenmaal aangezien te worden voor een handswerksman in plaats van een theoreticus. Hetaanzien van de theorie blijft het allerhoogste.'

Historische letterkunde

Waar de *digital humanities* echt voet aan de grond hebben gekregen, in het bijzonder doordat er duidelijke resultaten zijn geboekt, is het gebied van de historische letterkunde en met name de achttiende en negentiende eeuw. Ook in mediëvistische kringen groeit de belangstelling. 'Misschien komt dat doordat bij onderzoekers in deze specialismen er altijd al wel een besef is geweest dat men hoe dan ook op teksten gericht moet blijven. Daar ligt de kern van ons vakgebied.'

David Hoover zal in mei en juni van dit jaar terugkeren naar Nederland voor een tweede en derde termijn van gastaanstelling.

David L. Hoover is professor in de Engelse Taal aan de Universiteit van New York, waar hij sinds 1981 les geeft. Hij haalde zijn bachelors in Engels en Filosofie in 1971 aan het Manchester College en zijn masters in 1974 aan de Universiteit van Indiana, waar hij in 1980 promoveerde op een proefschrift over het oud-Engelse dichtmetrum. Zijn onderzoek richt zich behalve op dat metrum op digital humanities, computational stylistics, corpus stylistics, auteursherkenning. Recente boeken: *Stylistics: Prospect and Retrospect* (2007) en *Language and Style in The Inheritors* (1999).

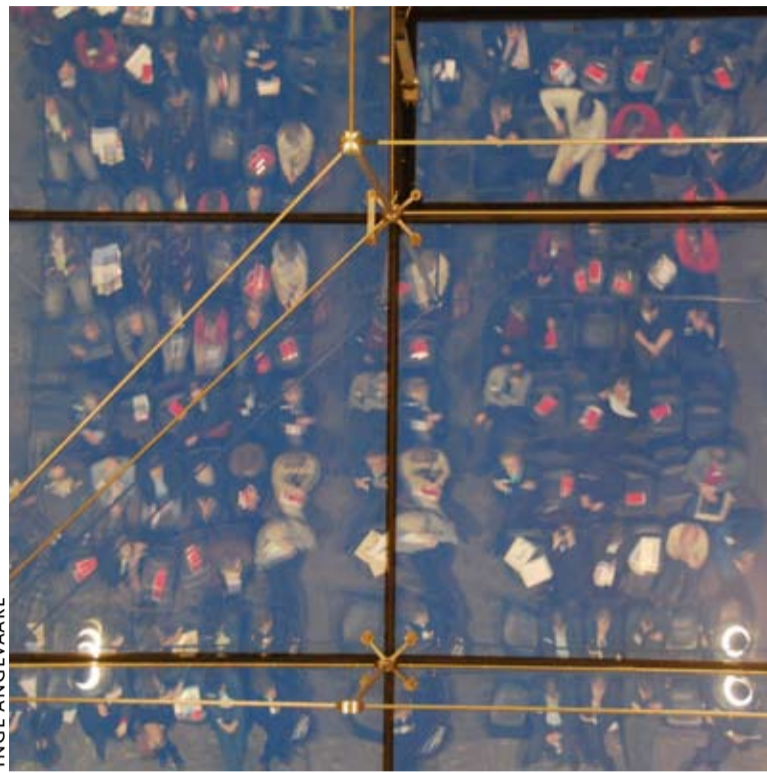


Een verjaardagsfeestje met vergezichten

'Het verhaal van een ooievaar die de spoorwegen gebruikt om naar huis te komen; van het parlements lid dat het vaakst het woord "ik" gebruikt; van Mies Bouwman en *Open het dorp*; en van de vele persoonlijke geheimen die liggen opgeslagen bij het CBS.' Zo schilderde datablogster en e-redacteur Inge Angevaare op haar weblog de vergezichten die het jubileumsymposium van DANS op 2 december te bieden had.

Ze trof daarmee ook de bedoeling van de organisatoren, wie het duidelijk niet om de vijfenvestig jaar ging die verstreken sinds het eerste databestand werd gearchiveerd door de Steinmetz Stichting, maar om de innovatieve mogelijkheden voor de toekomst. Duurzaamheid van data komt de wetenschap ten goede, was de stelling waarmee DANS-directeur Peter Doorn het symposium in de Haagse Glazen Zaal opende. De andere presentaties onderstreepten die uitspraak.

Journalist en taalhistoricus Ewoud Sanders bood de toehoorders een blik op zijn digitale bibliotheek met ruim twintigduizend banden, 4,5 miljoen pagina's en twee miljard woorden. Zelf kent hij



INGE ANGEVAARE

Het Glazen plafond boven de gelijknamige zaal.

geen belemmeringen door de *Auteurswet*, maar bij wetenschappelijke toepassing wordt dat anders, zo werd vanuit de zaal opgemerkt. Willem Bouten (Universiteit van Amsterdam) liet zien hoe in zijn Virtual Lab for eScience (VLe) velddata van allerlei groepen biologen en vrijwil-

ligers uit Europa, aangevuld met gegevens van militaire radarsystemen of satellieten, als één geheel kunnen worden geanalyseerd.

Henk den Heijer (Universiteit Leiden) stelde aan de hand van nieuwe data ons beeld bij van Nederland als belangrijke slaventransporteur in de zeventiende eeuw. Esther Jansma (Rijksdienst voor Cultureel Erfgoed) liet zien dat hout niet alleen informatie oplevert over herkomst en datering van voorwerpen maar ook over de economie, technologie, landschappelijke veranderingen en het klimaat in het verleden.

Johan van Rooijen (CBS) wierp licht op het Sociaal Statistisch Bestand (SBB) waar het CBS tegenwoordig zo veel mogelijk haar gegevens voor onderzoek uit probeert te putten, waarna Maarten Marx (Universiteit van Amsterdam) liet zien welke vragen beantwoord kunnen worden met een eenvoudige tekstanalyse van de Handelingen der Staten Generaal. Frank van Harmelen (Vrije Universiteit) beschreef het web als 'wetenschapsversneller' en schetste hoe de volgende fase, web 3.0 (het semantische web), er een 'netwerk van de data' van maken; door computers voor computers en daardoor nuttig voor mensen. (Marion Wittenberg)

Software voor automatisch zoeken in afbeeldingen en video

Onderzoekers van de Universiteit van Amsterdam (UvA) hebben de afgelopen maanden drie prijzen in de wacht gesleept met software om afbeeldingen en video direct te doorzoeken zonder dat er een beroep op tekst hoeft te worden gedaan. Hierdoor wordt het makkelijker om grote databases als YouTube, Flickr en Facebook te doorzoeken op eigenschappen die niet in de tekst worden genoemd.

Het team van de UvA nam deel aan drie grote competities waarin automatische zoekmachines werden getest en vergeleken. Bij de wedstrijden wordt gekeken welke zoekme-

thode het beste voorwerpen en concepten kan herkennen in beelden en filmpjes. Alle drie de keren werd de eerste prijs in de wacht gesleept. De Amsterdamse onderzoekers wonnen *ImageCLEF* in Griekenland (in oktober), *PASCAL VOC* in Japan (gedeelde eerste plaats, in oktober) en *TrecVID* in de VS (in september). Het beeld- en videozoekstelsel maakt gebruik van geavanceerde beeldherkenning, snelle indexering en intelligente classificatietechnieken. Om de grote hoeveelheden data aan te kunnen, wordt er gebruik gemaakt van supercomputers. Informatie: th.gevers@uva.nl. (IA)

Aandacht voor datawensen van jeugdonderzoekers

Sinds in 2007 een minister voor Jeugd en Gezin in de regering werd opgenomen, is er meer aandacht ontstaan voor jeugdonderzoek. Er blijkt weinig zicht te zijn op de kwaliteit van het beschikbare onderzoek en onderzoeksdata worden niet of nauwelijks gedeeld. Voor DANS was dat reden om de wensen van jeugdonderzoekers te verkennen en daarover op 25 november een discussiemiddag te organiseren.

Uit de verkenning kwam naar voor dat er een versmalling in het jeugdonderzoek gaande is. Veel aandacht gaat naar de effectiviteit van interventies om de kennis van professionals in de jeugdsector te ontwikkelen. Daarnaast wordt er nog steeds fundamenteel onderzoek gedaan, gericht op de onderliggende

mechanismen van opgroeien en opvoeden. Juist voor dit soort onderzoek denkt DANS een rol te kunnen spelen.

De discussiemiddag ging vooral over de voor- en nadelen van het delen van data. Onderzoekers blijken bang dat anderen over hun data gaan publiceren. Opdrachtonderzoekers aarzelen uit concurrentieoverwegingen. Toch waren de deelnemers aan de discussie het er over eens dat data gedeeld zouden moeten worden als aan een aantal voorwaarden zou worden voldaan. DANS wil met de resultaten verder gaan en meer gaan doen aan het verwerven van jeugddata. Mogelijk kan hieruit een jeugdportaal ontstaan. De 'Verkenning Jeugddata' is te bestellen via info@dans.knaw.nl. (Marion Wittenberg)



DANS-directeur Peter Doorn in geanimeerd gesprek met jeugd- en tieneronderzoekster prof. Wilma Vollebergh

Sinds kort beschikbaar

Het overzicht toont een aantal databestanden die recent voor onderzoekers beschikbaar zijn gekomen bij CBS en DANS. Een volledig overzicht van de CBS-bestanden is te vinden op www.cbs.nl/microdata. De bij DANS beschikbare databestanden komen

van diverse andere onderzoeksinstellingen. Deze kunnen kosteloos worden gedownload vanuit DANS EASY: <http://easy.dans.knaw.nl>. Via DANS kunnen ook de beveiligde microdata van het CBS kosteloos geleverd worden: www.dans.knaw.nl/nl/data/cbs/overzicht/

Centraal Bureau voor de Statistiek	Periode
Productiestatistiek	
Bouwnijverheid	1993-1999
Ambulancediensten	2006-2007
Trombosediensten	2006-2007
Welzijnswerk-Kinderopvang	2006-2007
Groothandel	1993-2007
Detailhandel	1993-2007

Via DANS EASY	
<i>Archeologie</i>	
Project Vleuten-Geldermalsen/Randstadspoor	2005
Negen projecten: BAAC bv, ADC, AAC, ACVU-HBS (RCE)	
Graven in de Grote- of Sint Jacobskerk (Gemeente Den Haag)	2009
Linne, Ossenbergweg-Mergelweg (BAAC bv)	2009
Aanleg riolering Wijk bij Duurstede (Grontmij Nederland B.V.)	2006
Built Environments, Constructed Societies (Vis, B.N., Sidestone Press)	2009
<i>Geschiedenis</i>	
Papua's in diaspora. Erfgoed van de Oorlog; interviews afgenomen in het kader van het project Getuigen Verhalen; (Stichting Papua Cultureel Erfgoed)	1940-1945
Gewoon doen en fier rechtop staan. Vier verzetsvrouwen. Erfgoed van de Oorlog (Stichting Stichting Cultuur & Co)	1940-1945
Samen leven en samen werken in het Bakkumse Sperrgebied. Erfgoed van de Oorlog; (Caleidoscoop Film)	1940-1945
Indonesisch en Indisch verzet in Nederland tijdens WOII. Erfgoed van de Oorlog (Artesteem)	1940-1945
<i>Sociale wetenschappen</i>	
Dutch Parliamentary Election Studies – Cumulative file (Kees Aarts, Bojan Todorosijevitz, University of Twente)	1971-2006
Computer Use (Harm Slijper – Erasmus MC – University Medical Center Rotterdam)	2002-2008
Tijdsbestedingsonderzoek – TBO	2006
Vrouwen in deeltijdarbeid (Sociaal en Cultureel Planbureau)	2008
Nulmeting van de publieke opinie over nanotechnologie (Commissie Maatschappelijke Dialoog Nanotechnologie)	2009
Consumenten Conjunctuur Onderzoek CCO (CBS beveiligd microbestand)	2009
Sociale Integratie Minderheden – SIM (CBS beveiligd microbestand)	2006
<i>Geodata</i>	
Bestand Bodemgebruik (CBS en Kadaster)	2006

COLOFON

e-data@research is het kwartaalblad in Nederland over data en onderzoek in de alfa- en gammawetenschappen. Het verschijnt onder auspiciën van DANS, het Huygensinstituut, het Centraal Bureau voor de Statistiek, de Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica. Toezending kosteloos aan relaties van de stakeholders en op verzoek aan studenten in de alfa- en gammarichtingen.
Oplage: 9100.
e-data@research is online te raadplegen op www.edata.nl

Uitgever: Stichting Uitgeverij *e-data@research*, Postbus 93067, 2509 AB Den Haag

Redactieadres: Postbus 93067, 2509 AB Den Haag; t (070)3494450 f (070)3494451 e edata@dans.knaw.nl

Redactie: Inge Angevaare, Ronald van der Bie, Peter Boot, Martijn de Groot (hoofd/eindredacteur), Thijs Hermsen, Jetske van der Schaaf

Aan dit nummer werkten mee: Herman Brinkman, Karina van Dalen-Oskam, Paul Doorenbosch, Friedel Grant, Rob Grim, Rik

Hoekstra, Marjolein Hogenbirk, Hans Schmeets, Barbara Sierman, Harm Slijper, Jeroen van der Vliet, Marion Wittenberg

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Vormgeving en opmaak: Ellen Bouma

Productie: Uitgeverij Aksant, Amsterdam

Druk: Ten Brink, Meppel

ISSN: 1872-0374

INGEZONDEN

Naar een betrouwbaar permanent meningenonderzoek

Op 3 maart worden de gemeenteraden gekozen. De ervaring leert dat zo'n zestig procent van het electoraat naar de stembus zal gaan. De winst- en verliescijfers van de politieke partijen zijn lastiger te voorspellen. De diverse opiniebureaus verschillen immers keer op keer in hun voorspellingen. Wel staat vast op welke manier we moeten stemmen. Na het verbod op het stemmen per stemmachine in 2008, gebeurt dat weer met het rode potlood. De Organisatie voor Veiligheid en Samenwerking in Europa, de OVSE, had in haar rapport over de Tweede-Kamerverkiezingen een aantal verbeterpunten opgenomen. Zo adviseerde de organisatie om het stemmen met de stemcomputer te heroverwegen, mede doordat de privacy van de kiezers niet is gegarandeerd. Daarnaast verbaasde de OVSE zich over het stemmen per volmacht. Dit druipt in tegen internationale maatstaven waar democratische verkiezingen aan moeten voldoen, waaronder die dat elke kiezer één stem mag uitbrengen. In Nederland is dat echter niet zo. Liefst 12 procent van het electoraat geeft een volmachtstem af.

Deze laatste informatie is afkomstig van het Nationaal Kiezersonderzoek (NKO) dat rondom de Tweede Kamerverkiezingen van 22 november 2006 onder zo'n 2.500 stemgerechtigden is uitgevoerd. Het NKO geeft de verdeling over de partijen, en daarmee de zetelverdeling, nagenoeg perfect weer. Op zich is dat nu ook weer niet zo vreemd. Je trekt een steekproef uit het electoraat zodanig dat elke persoon dezelfde kans heeft om daarin terecht te komen. Idealiter heb je dan alleen fouten die veroorzaakt worden door toevalligheden in de getrokken steekproef. Dat zijn de betrouwbaarheidsmarges, die echter nogal klein zijn bij 2.500 personen. Maar hoe kan het dan dat er grote discrepanties zijn met de voorspellingen van de

Zetelverdeling volgens opiniebureaus, Nationaal Kiezersonderzoek en stembus, 22-11-2006.

	Politieke Barometer	Peil.NL	De Stemming	Nationaal Kiezersonderzoek	Stembusuitslag
CDA	41	42	41	41	41
PvdA	37	38	31	33	33
VVD	23	22	21	22	22
SP	23	23	32	26	25
PVV	4	5	6	8	9
GroenLinks	7	8	5	7	7
D66	3	2	1	3	3
CU	6	6	8	6	6
SGP	2	2	3	2	2
Pvd Dieren	2	1	2	2	2
Eén NL	1	1	0	0	0
LPF	1	0	0	0	0

Bron: Schmeets, J.J.G. (2009.) 'Vertrouwen in onderzoek naar verkiezingen. Enkele opmerkingen over kiezersonderzoek en het beoordelen van verkiezingen.' Inaugurale rede, uitgesproken op 6 maart 2009 (Maastricht University: Maastricht), p. 10.

opiniebureaus in de peilingen zoals 'De Stemming', 'Peil.NL', en 'De Politieke Barometer' kort na sluiting van de stemlokalen? En bovendien: hoe is het te verklaren dat dergelijke bureaus steevast ook onderling grote verschillen laten zien? De oorzaak ligt zeer waarschijnlijk in de steekproef, of beter geformuleerd: het ontbreken van een steekproef die een afspiegeling is van de doelgroep. De opiniebureaus maken immers veelal gebruik van een groep personen die zich vrijwillig hebben gemeld om aan de onderzoeken mee te doen. Een selectie uit die groep wordt via het internet benaderd. Sommigen doen heel vaak mee, maar velen niet. Aan noodzakelijke voorwaarden om de kwaliteit van een onderzoek te beoordelen – zoals het (kunnen) vaststellen van responscijfers of de over- en ondervetegenwoordiging van bevolkingsgroepen in de steekproef – wordt niet voldaan. De kosten voor dergelijke onderzoeken die niet voldoen aan kwaliteitscriteria zijn weliswaar laag, maar daar betaal je dan

ook een prijs voor. Dat blijkt telkens weer als de bevindingen geïkt kunnen worden met officiële cijfers zoals de stembusuitslag. Daarentegen is een NKO duur: per vraaggesprek zo'n 150 euro. Dat komt ook doordat in bijna alle gevallen de gesprekken bij de personen thuis plaatsvinden en veel geïnvesteerd wordt om de respons – in 2006 was dat 72 procent – te maximaliseren.

Bijna dagelijks wordt niet alleen de bevolking via de media geconfronteerd met meningen en peilingen, maar de onderzoeksresultaten worden ook steeds vaker aan het kabinet aangeboden. Dat is vreemd. Het CBS heeft immers in het maken van 'onbetwiste, samenhangende en actuele statistieken die relevant zijn voor de praktijk, beleid en wetenschap', een duidelijke taakstelling. En daar hoort ook het peilen van de meningen van de bevolking bij. Sinds 1 januari heeft het CBS op dit gebied een nieuwe speerpunt onder de naam 'Belevingen'. Na een pilot in 2010 moet daaruit een permanent onderzoek voortkomen over de 'Meningen van de Nederlandse bevolking'. Uiteraard met een goed steekproefontwerp, hoge responscijfers, en een goede afspiegeling van de bevolking. Een dergelijk ontwerp impliceert tevens snelle cijfers om de meningen weer te geven over actuele kwesties die in de samenleving spelen. En dat wordt nog een hele uitdaging.

Hans Schmeets
Centraal Bureau voor de Statistiek,
Universiteit Maastricht



TWAN WIJERMANS

Column
Martijn de Groot

Voorhoedegevecht

Soms heb je wel eens meningen of doelen die zo onaantastbaar, verheven of gewoon goed zijn dat niemand zich er tegen durft te verzetten. Verkeersveiligheid is een voorbeeld. Minder doden in het verkeer, daar moet je vóór zijn. Of dierenleed. Daar ben je tegen.

Open Access lijkt wel wat op zo'n doel. Kan het mooier? Dat iedereen, rijk of arm, oost of west, zonder belemmeringen kan profiteren van kennis uit wetenschappelijk onderzoek? Dat willen we echt allemaal, en dat blijkt ook wel uit de sinds 2003 almaar groeiende lijst met namen onder de Verklaring van Berlijn. Vrijwel geen Nederlandse universiteit of academische club ontbreekt daarop en zo hoort het ook. 'Sinds de uitvinding van de openbare bibliotheek in de negentiende eeuw zijn we eraan gewend dat iedereen recht heeft op informatie; niet alleen de mensen die het kunnen betalen,' motiveerde Sijbalt Noorda, baas van de universiteitenvereniging VSNU, in een SURF-filmpje op YouTube. En de Utrechtse decaan Wiljan van den Akker in datzelfde filmpje: 'Ieder argument tegen Open Access is een (...) instrument in een volstrekt achterhoedegevecht'.

Zelfs de uitgevers, waarvan de woekerwinsten nogal eens werden aangevoerd als bewijs dat er toch echt een vijand te verslaan was, monstersen aan bij het leger van de goede zaak. Zie Rafael Sidi van Elsevier vorig najaar in de Open Access Week (www.opendataspeakerscorner.nl) en Eelco Ferwerda van Amsterdam University Press, die vorige maand de Open Access Award in ontvangst mocht nemen.

Die strijd voor Open Access lijkt dus een strijd zonder vijand, zolang het tenminste gaat over toegang tot publicaties en niet tot de data waarop ze zijn gebaseerd. Nu is het aan één kant wel een veilig gevoel om aan zo'n strijd deel te nemen, vooral in het zicht van de victorie. Bij open publicaties ligt de winst immers voor het oprapen want daar gaat het om het overwinnen van papieren en financiële ongemakken. Maar aan de andere kant vraag ik me af of het niet beter zou zijn om de aandacht te verleggen naar een front waar nog veel te bevechten en veel te winnen valt.

Ik weet wel zo'n front. Denk maar eens aan de data. Aan die enorme hoeveelheden data die al wél zijn geoogst en tot publicaties hebben geleid, maar nog niet door andere onderzoekers kunnen worden gebruikt. Daar is niet de vraag aan de orde van meer of minder betalen, maar van wel of niet ermee aan het werk kunnen. Als die data vrij beschikbaar komen is de winst groter, de vooruitgang sneller en de maatschappelijke impact dieper dan wanneer datzelfde zou worden bereikt voor onderzoekpublicaties. Oude antwoorden kunnen worden aangescherpt, nieuwe vragen kunnen worden gesteld, nieuwe antwoorden gevonden en blootgesteld aan de toets van de wetenschappelijke kritiek. Want dat is nog een oud principe waar we dan eindelijk werk van kunnen maken: de mogelijkheid van kritiek door herhalingsonderzoek die tot nu toe vooral theoretisch bleef.

Hier lonkt een voorhoedegevecht. Steek de trompetten!

Martijn de Groot is hoofdredacteur van e-data@research

Gelezen

Saskia Woutersen-Windhauer et al.: *Enhanced Publications: Linking Publications and Research Data in Digital Repositories*; Amsterdam, Amsterdam University Press, 2009; ISBN 9789089641885

According to this publication, publishers and repositories have the building blocks and the tools, but in general do not use them to create 'Enhanced Publications'. Publishers and repositories should offer the service and tools to add research data, extra materials and post-publication data to the publications. Researchers should be responsible for the content.

Zie ook pagina 4. Meer informatie in twee andere publicaties van de Amsterdam University Press uit 2009: 'The European Repository Landscape 2008' en 'Emerging Standards for Enhanced Publications and Repository Technology: Survey on Technology'.

O. Boonstra, A. Schuurman (red.): *Tijd en ruimte. Nieuwe toepassingen van GIS in de alfawetenschappen*. Utrecht, Matrics, 2010. ISBN 978-90-5345-388-9

Geografische Informatiesystemen (GIS) eisen een steeds grotere rol op in onze sa-



menleving. De digitale representatie van ruimte is bijvoorbeeld al door 'TomTom' en 'GoogleMaps' dicht bij de doorsnee gebruiker gebracht. Het gebruik van 'geoinformatie' maakt een opvallende bloei door, maar er is nog maar bij een beperkte groep onderzoekers kennis over aanwezig.

Deze uitgave in samenwerking met DANS (Data Archiving and Networked Services) laat zien dat GIS op meerdere onderzoeksvelden toepasbaar is. Zo wordt GIS nu ook binnen onderzoeksgebieden als geschiedenis, archeologie, kunstgeschiedenis en de taal- en letterkunde toegepast, wat interessante nieuwe vraagstukken en resultaten kan opleveren.

Henk Scholten, Rob Velde, Niels Mamen, (eds.): *Geospatial Technology and the Role of Location in Science*; Springer, Geojournal Library, Vol. 96, 2009; ISBN: 978-90-481-2619-4

Globalisation has not led to the 'death of geography'. Intensified relations between communities in different parts of the world have only highlighted the need for understanding and managing phenomena on a variety of geographic scales. From global warming to credit crunch, and from epidemics to terrorism, causes and solutions are sought on local, regional, national as well as inter-continental levels. With the advent of Geospatial Technology, scholars, policymakers and entrepreneurs have

valuable tools in hand to proceed.

This book offers the first systematic account of the science behind this mental and technological revolution. Tracing the adoption and dissemination of Geospatial Technology in a range of disciplines, it examines the impact this technology has had, and is likely to have, on the explanation of spatial behaviour, phenomena and processes. At the same time, stressing innovative usage, it explores scientific contributions to technology advancement.

Tom Kuipers, Jeffrey van der Hoeven: *Survey – Insight into digital preservation of research output in Europe; Permanent Access to the Records of Science, december 2009*

This report describes the results of the surveys conducted by PARSE.Insight (Permanent Access to the Records of Science).

Major surveys were held within three stakeholder domains: research, publishing and data management. In total, almost 2,000 people responded and thus provided interesting insights in the current state of affairs in digital preservation of research data, the outlook of data preservation, data sharing, roles and responsibilities of stakeholders in research and funding of research. The PARSE.Insight project is concerned with the preservation of digital information in research, from primary data through analysis to the final publications resulting from this research. Ultimately, the project will develop a roadmap for an e-Science infrastructure, intended to guide the European Commission's strategy about research infra-structures. www.parse-insight.eu/