data & research

Jaargang 3 nummer 4 / maart 2009

Kwartaalblad over data en onderzoek in de alfa- en gammawetenschappen

Mogelijk gemaakt door: Data Archiving and Networked Services, Centraal Bureau voor de Statistiek, Huygens Instituut, Internationaal Instituut voor Sociale Geschiedenis, Koninklijke Bibliotheek en de Vereniging voor Geschiedenis en Informatica

Pagina 3 • Datakeurmerk • Nu de ontwikkeling van het Datakeurmerk een eind op weg is geeft DANS het beheer uit handen aan een internationale Board met leden uit Europese landen en de Verenigde Staten.

Pagina 4 • Lange tijdreeksen • Het CBS heeft een speciaal Expertisecentrum opgericht om lange tijdreeksen samen te stellen en vooral om de breuken daarin te repareren, ontstaan door bijvoorbeeld veranderingen in definities.

ELMER SPAARGAREN

Pagina 5 • John Nerbonne • De Groninger hoogleraar John Nerbonne brengt de verspreiding van dialecten in kaart met behulp van grote tekstcorpora.

Pagina 6 • FOCUS op EDSC • De bibliotheek van de Erasmus Universiteit kent een uniek service centre voor wetenschappelijke data voor economen, bedrijfskundigen en sociale wetenschappers.

EN VERDER

Digitale duurzaamheid wordt verkend

Sinds begin dit jaar is een projectteam van vijf mensen bezig de Nederlandse situatie te verkennen op het gebied van digitale archivering in de publieke sector. Het ministerie van Onderwijs, Cultuur en Wetenschap subsidieert het onderzoek, dat onder leiding staat van de Nationale Coalitie voor Digitale Duurzaamheid (NCDD).

De verkenning moet al in de eerste helft van dit jaar een beeld opleveren van de risico's van verlies van digitaal materiaal in de publieke sector. Tegelijk wordt door het Utrechtse strategisch adviesbureau Thaesis een inventarisatie uitgevoerd bij een aantal commerciële bedrijven: hoe gaan zij om met data die langdurig bewaard moeten worden en wat kan de publieke sector van het bedrijfsleven leren?

Op basis van de nationale verkenning gaat de coalitie een strategie ontwikkelen om in Nederland tot een nationale infrastructuur voor digitale duurzaamheid te komen. Rond het verschijnen van deze *e-data&research* zal de eerste fase van het onderzoek zijn afgerond. In die fase zijn de leden van het projectteam 'het veld' in geweest om een algemeen beeld te krijgen van de situatie in hun sector. Onderscheiden worden de sectoren overheid/archieven, cultuur/erfgoed en wetenschap. In dat eerste algemene beeld zal plaats zijn voor aspecten als de rol van digitale informatie, de bestaande wetten en regels, de

Het onderzoeksteam van de Nationale Verkenning: vanaf linksboven, met de klok mee: René van Horik, Annelies van Nispen, Ingrid Dillo, Inge Angevaare en Petra Helwig.

routes waarlangs digitale informatie wordt doorgegeven, de partijen die daarbij een rol spelen, de kosten en de partijen die daarvoor instaan. Ook aan best en worst practices zal volgens de NCDD aandacht worden besteed, en aan maatregelen die nodig zijn om de toegang tot digitale informatie te waarborgen. Ten slotte zal de rapportage na afloop van de eerste fase inzicht moeten geven in de maatregelen die per sector moeten worden genomen of juist sectoroverschrijdend. Op basis van de ervaringen van de eerste twee maanden wordt een plan gepresenteerd voor het vervolg over de periode van maart tot en met mei, wanneer de nationale verkenning moet zijn afgerond. In het projectteam voor de verkenning in de publieke sector is DANS-medewerker René van Horik verantwoordelijk voor de sector wetenschap (zie interview pagina 3). Annelies van Nispen (Digitaal Erfgoed Nederland) onderzoekt in het bijzonder de erfgoedsector en Petra Helwig (Nationaal Archief) de overheidsarchieven. Ingrid Dillo (Koninklijke Bibliotheek) en Inge Angevaare (NCDD) voeren het projectmanagement. (MdG)

Europese subsidie voor onderzoek sociale mobiliteit

De European Research Council (ERC) heeft eind vorig jaar een van zijn prestigieuze Advanced Investigator Grants toegekend aan Marco van Leeuwen (UU en IISG). Het is voor het eerst dat het ERC, een soort NWO op Europees niveau, deze beurzen uitdeelt.

'De Advanced Investigator Grant van de ERC is te vergelijken met een VICI subsidie van NWO,' aldus Van Leeuwen, die in het project met Ineke Maas (UU) samenwerkt. 'We krijgen voor vijf jaar financiering voor ons onderzoek naar sociale mobiliteit.' In totaal gaat het om een bedrag van twee miljoen euro. Het onderzoeksproject heet 'Towards open societies? Trends, variations and driving forces of intergenerational social mobility in Europe over the past three centuries'. Het richt zich op de mate waarin kinderen los kunnen komen van het sociale milieu van hun ouders door een ander beroep te kiezen. Welke instituties en andere factoren bevorderen die sociale mobiliteit en welke juist niet? Is het onderwijs, of industrialisatie en andere veranderingen in de beroepsstructuur? Gaat het om wetgeving en overige vormen van overheidsingrijpen of juist om veranderende opvattingen, bijvoorbeeld over de beroepskeuze van vrouwen? Of zijn oorlogen die grote keerpunten?

Dit onderzoek gebruikt een unieke database van ongeveer vier miljoen huwelijksaktes uit de periode 1680-1970, verbonden met gegevens uit hedendaagse surveys (1950-heden), met data uit personeelsadvertenties en andere bronnen. Naast de Historische Steekproef Nederland (HSN) worden diverse andere HSN-achtige databases uit verschillende Europese landen gecombineerd. Het project maakt direct gebruik van de internationale historische beroepsindeling HISCO, waarvan Van Leeuwen, Maas en Andrew Miles (University of Manchester) de belangrijkste ontwikkelaars zijn. (LS)

Data-college op toernee

Voor een collegezaal met zo'n 125 masterstudenten Communicatiewetenschappen gaf DANS-medewerkster Marion Wittenberg op 5 februari in de Amsterdamse Oudemanhuispoort een gastcollege over het vinden en gebruiken van data voor wetenschappelijk onderzoek. Ze was daar op uitnodiging van de Universiteit van Amsterdam, in het kader van een toernee langs verschillende universiteiten in het land. DANS wil met de gastcolleges studenten attenderen op de mogelijkheden van het gebruik van bestaande data voor studenten bij het maken van hun scripties, maar ook op het nut van opslaan en hergebruiken van data in het algemeen. Wittenberg liet naast de door DANS aangeboden archieven EASY en NESTAR ook andere mogelijkheden de revu passeren zoals die worden geboden door bijvoorbeeld het Centraal Bureau voor de Statistiek, het European Social Survey en de Europese archievenkoepel CESSDA.

DIEDERIK VAN DER LAAN

Studente Effie Beumer, zojuist begonnen aan haar master Commerciële Communicatie en Voorlichting, oordeelde na afloop dat ze veel nieuws had gehoord: 'Ik had nog nooit van DANS gehoord en wist ook niet dat je zoveel databestanden nog terug kon vinden. Ik dacht altijd dat je het met de gepubliceerde artikelen van de auteurs moest doen. Maar kennelijk heb je dus ook toegang tot de databases waarmee ze hebben gewerkt.' Net als andere aanwezige studenten verwachtte Beumer met de informatie van het college vooral iets te kunnen 'tegen de tijd dat ik aan mijn scriptie begin. De timing kon dus beter. Maar als het zo ver is weet ik waar ik de informatie weer kan opvragen' (MdG)

AGENDA

5 maart
Rotterdam, World Trade Centre
Kennis delen met web 2.0-tools in de organisatie
Workshop voor professionals die binnen de eigen organisatie beleidsmatig bezig zijn of willen zijn met web 2.0-toepassingen: HR-managers, beleidsfunctionarissen, kennismanagers en informatiespecialisten. http://essentials-media.matchpoint.nl/index.php?item=479

5 maart
Utrecht, SURFnet
Gevorderdenworkshop Inrichten van SURF-groepen voor onderwijs
Workshop voor bijvoorbeeld ICTO adviseurs en docenten uit het hoger onderwijs, met als doel een leeromgeving in te richten die verschillende onderwijsvormen kan ondersteunen. Nadruk op communicatie, organisatie, informatie, aanbieden van leermiddelen en samenwerking.
www.surfnet.nl/nl/bijeenkomsten/

16-17 maart
London, The British Library Conference Centre
Unlocking Audio 2: Connecting with Listeners
A key event exploring the use of sounds online. The conference will focus on ways in which researchers and other audiences expect to discover, browse, audition and analyse archival audio resources.
www.bl.uk/unlockingaudio

27 maart
Amsterdam, Trippenhuis KNAW
Biografisch Portaal Nederland als nieuw onderzoeksinstrument.
Symposium van DANS en ING over het Biografisch Portaal Nederland. Dat Portaal wil een zo volledig mogelijke en duurzame website opbouwen, ontwikkelen en onderhouden met biografische informatie over de bewoners van Nederland, van de vroegste tijden tot nu.
www.dans.knaw.nl/nl/dans_symposia/

1-3 april
Chapel Hill, NC USA
DigCCurr 2009: Digital Curation Practice, Promise and Prospects
Conference, focussing on current practice and research, surrounding digital curation with a look towards the future, and trends in preparing digital curation professionals.
www.ils.unc.edu/digccurr2009/

15-17 april
Amsterdam, Westergasfabriek,
The Next Web Conference
Three-day conference about the future of the web. A blend of decision makers from the European & American Internet scene, technology entrepreneurs, start-ups, innovators, along with venture capitalists, industry journalist, bloggers, and senior level executives are expected. Thought leaders and Internet entrepreneurs are invited.
www.2009.thenextweb.com

17 april
Den Haag, Koninklijke Bibliotheek
Curating Research 2009
Workshop, aiming to develop a basic understanding of the issues presented by long-term digital curation and preservation of resources, (to be) deposited in institutional and subject-based repositoriesState of the art in digital curation, best practices, including possibilities for outsourcing. For policy makers, research librarians, directors of research institutions, repository managers and middle management and publishers.
www.kb.nl/hrd/congressen/index-en.html

26-29 mei
Tampere, Finland, Tampere Hall
IASSIST/IFDO 2009 Conference – Mobile data and the life cycle
The annual conference of IASSIST and IFDO for all professionals working in and with information technology, data services and data archiving.
www.fsd.uta.fi/iassist2009/

15-19 juni
Rotterdam, Beurs-WTC Congress Center
GSDI11 – Spatial data infrastructure convergence, building SDI bridges to address global challenges.
Internationale conferentie over geo-informatie en geodata infrastructuur, georganiseerd door GSDI Association, Geonovum, TU Delft, Eurogi, JRC en RGI.
www.gsdi11.nl

Göttingen: TextGrid Summit

Op 21 en 22 januari vond in Göttingen de TextGRID summit plaats. Tussen de honderd en honderdvijftig mensen kwamen bijelkaar in de indrukwekkende Pauliner Kirche, nu een onderdeel van de gastorganisatie, de Niedersächsische Staats- und Universitätsbibliothek Göttingen. Deze instelling was ook projectleider van het nu afgeronde TextGRID project. Deelnemers waren afkomstig uit tekstwetenschappen en academische bibliotheken, uit informatica en softwareontwikkeling en van Duitse overheden en onderzoeksfinanciers.

Volgens de gelijknamige website beoogt TextGRID een netwerk (grid) te zijn voor het gezamenlijk bewerken, annoteren, analyseren en publiceren van specialistische teksten. Zo vormt het een hoeksteen van de e-Humanities, die zich steeds meer manifesteren. Achter die definitie gaan nogal wat elementen schuil, die tijdens de summit uitgebreid aan bod kwamen in presentaties en paneldiscussies en aan goedgevulde borreltafels.

Zo verwijst de term GRID naar een infrastructuur met een hoog technisch gehalte en een geweldige performance. In Duitsland heet dit het D-GRID. Daarop bouwen allerlei disciplines hun eigen structuren, zoals nu dus het TextGRID. Er staat al ruim twintig terabyte aan mooi materiaal op, zoals het Wörterbuchnetz van Trier. Anderen overwegen, zo bleek, hun materiaal erop te zetten, zoals Münster dat de 5500 handschriften van het Nieuwe Testament wil digitaliseren. Er zijn nog niet veel tools, maar er is al wel een online XML-editor. Op een editor om facsimiles en transcripties te verbinden hoeven we nog maar een paar weken te wachten. Op de hands-on workshop wordt met de aanwezigen een applicatie in elkaar gesleuteld die eigennamen uit teksten herkent (gebaseerd op een tool van het Huygens Instituut); deze worden in Wikipedia opgezocht en hun beschrijvingen worden naast de originele tekst gepresenteerd, dit alles *on the fly*.

Verder gaat het niet om grote stromen getallen, maar om vaak klassieke teksten in al hun verschijningsvormen. Om winst te halen uit samenwerking moet je die teksten in standaardvormen representeren. Text GRID leunt daarbij sterk op het Text Encoding Initiative (TEI).

Ten slotte is er de samenhang met de grote Europese projecten voor infrastructuren in de gebieden cultureel erfgoed, geesteswetenschappen en sociale wetenschappen: DARIAH, CLARIN en CESSDA. Die kanaliseren nieuwe digitale methoden naar een klassiek wetenschapsbedrijf. De uitdaging is om het volle potentieel van deze methoden te aan te boren door goede onderzoeksvragen. De vraag naar de duurzaamheid van de infrastructuur zelf werd in allerlei toonaarden gesteld, waarbij gewezen werd naar pioniers in het buitenland zoals DANS, en op de rol van (universiteits)bibliotheken. Maar op dit moment is het bedje voor TextGRID nog bepaald niet gespreid. (Dirk Roorda)

www.textgrid.de/en/startseite.html
www.tei-c.org/index.xml

PETER DOORN

Bezoekers van de TextGRID summit in de Pauliner Kirche

Gezocht: serieus spelletje

Om games te introduceren in de erfgoedsector en om de Vlaamse game-industrie in contact te brengen met die sector organiseerde FARO (Vlaams steunpunt voor cultureel erfgoed) in samenwerking met enkele partners op 5 december de conferentie 'Games & Erfgoed'.

De impact van computergames op de Vlaamse erfgoedsector is zeer gering. Games worden echter wel beschouwd als cultureel erfgoed en er wordt steeds vaker verwezen naar de ongebreidelde mogelijkheden die games kunnen bieden voor het ontsluiten van erfgoed. Bij erfgoed en games wordt al snel de link gelegd naar *serious games* of *meaningfull play*. Bij deze categorie van games ligt de nadruk meer op het *firsthand learning* dan op het spelelement. De Nederlandse onderzoeker Jeroen van Mastright liet zien dat games echter veel meer zijn dan alleen een communicatie-instrument. Volgens hem raken games namelijk aan de essentie en doelstellingen van erfgoedinstellingen.

De relatie tussen erfgoed en games gaat verder dan het 'gebruik' van games door erfgoedinstellingen of games die op hun beurt erfgoed gebruiken. Het computerspel is inmiddels zelf al erfgoed geworden en dus dringt zich de vraag op wat de sector moet aanvangen met games als cultureel erfgoed. Het bewaren van games is namelijk veel meer dan louter het bewaren van digitale media. De onderzoeksleider van het Preserving Virtual Worlds project, Jerome McDonough, presenteerde de wijze waarop dit project oplossingen tracht te vinden op de vragen die het bewaren van games met zich meebrengt.

Jeroen van der Vliet en Connie Veugen lieten zien dat ontwikkelaars van commerciële geschiedenisgames (waarbij de nadruk ligt op entertainment) veel gebruik maken van historische bronnen uit erfgoedinstellingen. Er is dus sprake van een directe link met de erfgoedsector. Deze commerciële historische games worden soms gebruikt als leermiddel in het onderwijs, maar het spanningsveld tussen de 'historische wereld' (docenten, onderzoekers, erfgoedinstellingen) en de commerciële historische games blijft bestaan.

De relatie tussen de erfgoedsector en de game-industrie stond in de sessie 'Een vat vol content' centraal. Een serie van praktijkvoorbeelden liet zien op welke manieren erfgoedinstellingen gebruik kunnen maken van games of gametechnologie. Het gaat dan bijvoorbeeld om op een unieke manier verhalen te vertellen over de Eerste Wereldoorlog of om het stimuleren van de beleving bij een museumbezoek. Op die wijze kan ook het publieksbereik van musea vergroot worden. FARO hoopt met deze conferentie games op de Vlaamse erfgoedkaart te zetten en blijft het thema volgen. (Bart de Nil)

http://www.faronet.be/faceted_search/results/taxonomy%3A322

Naar Buiten! met digitaal erfgoed

'Naar buiten!' was het thema van de vijfde Digitaal Erfgoed-conferentie die op 9 en 10 december 2008 in de Rotterdamse Doelen werd gehouden. Het thema had in dit lustrumjaar een uitroepteken en de organisatie had kosten noch moeite gespaard om dat aan te zetten met een flink aantal sprekers dat vooral motivatie in de aanbieding had. Motivatie om los te komen van de gehechtheid aan de fysieke objecten en maximaal gebruik te maken van de mogelijkheden die digitale informatie biedt om alle uithoeken van de wereld te bereiken en virtuele bezoekers in hun eigen domein te bedienen.

Meer dan vierhonderd toehoorders uit allerlei geledingen van het cultureel erfgoed werden aan het begin van de conferentie op het goede spoor gezet door twee keynotes uit de Verenigde Staten. Diane M. Zorich behandelde de dynamiek van samenwerkingsverbanden tussen bibliotheken, archieven en musea aan de hand van het rapport 'Beyond the Silos of the LAMs' dat zij vorig jaar schreef. Hoewel die samenwerking niet altijd gemakkelijk tot stand komt, ontstaan volgens Zorich juist daar mogelijkheden om de online ruimtes te bereiken waar de gebruikers zich bevinden zoals YouTube, Flickr en iTunes. Nancy Proctor van het *Smithsonian American Art Museum* erkende dat het omarmen van dergelijke media musea soms een ongemakkelijk gevoel geeft. Het museum verliest immers de controle over wat er met hun erfgoed gebeurt. Maar zij daagde de conferentiegangers uit om het museum te zien als een 'agora', een marktplaats, met een community, mogelijkheden voor mash-up, en vooral ook 'fun'.

Dat niet alleen Amerikanen kunnen innoveren en motiveren werd de volgende dag duidelijk toen George Freriks van iMMovator 'The Netherlands: Digital Gateway to Europe' presenteerde en Ranti Tjan van museumgoudA het Laboratorium Waterwolf dat op zoek gaat naar 'lokale betekenis'.

Tussen de keynotes door was er sprake van een overweldigend aanbod aan workshops en parallelsessies die hier niet allemaal beschreven kunnen worden. De rode draad was echter duidelijk: allerhande 21steeeuwse technieken maken het mogelijk om steeds meer bezoekers steeds verder weg te bereiken met het Nederlands erfgoed – mits de hoeders van dat erfgoed bereid zijn de digitale uitdaging aan te gaan. (IA)

www.den.nl/innovatie/de_conferentie2008/

Datakeurmerk krijgt internationale erkenning

Het datakeurmerk, bedoeld om het duurzaam opslaan en hergebruiken van onderzoekgegevens te stimuleren, krijgt een internationaal bestuur. Initiatiefnemer DANS (Data Archiving and Networked Services) treedt op 1 mei terug als beheerder van het keurmerk. Het instituut heeft zich als eerste zelf aan een 'assessment' onderworpen.

De Data Seal Board in januari in Den Haag bijeen voor een kick-off meeting: Natascha Schumann (NESTOR), Paul Trilsbeek (MPI Nijmegen), Olivier Rouchon (CINES), Henk Harmsen (DANS), Mary Vardigan (ICPSR), Hans Pheiffenberger (Alfred Wegener Institute), Lisa de Leeuw (DANS), Laurents Sesink (DANS) en Matthew Woollard (UKDA)

Ook andere Nederlandse data-instituten zullen binnenkort, bij wijze van 'pilot', zo'n keuringsproces doorlopen. Met verschillende instituten en instellingen zijn daarover besprekingen gaande. Het keurmerk is bedoeld om gebruikers en financiers van het onderzoek te garanderen dat de gegevens altijd vindbaar blijven en zorgvuldig worden beschermd tegen verval door veroudering van soft- en hardware. Onderzoekorganisaties als NWO vragen om zo'n garantie.

In april wordt in Luxemburg een Europese workshop gehouden met deelnemers uit Nederland, Frankrijk, Duitsland, het Verenigd Koninkrijk en de Verenigde Staten. Ook de Europese Commissie zal vertegenwoordigd zijn. Op 1 mei zal DANS het Data Seal of Approval officieel overdragen aan de DSA Board.

Intussen wordt de instelling voorbereid van de Board van het Data Seal of Approval (DSA), het internationale bestuur dat verantwoordelijk wordt voor de assessments. Een verkennende bijeenkomst voor de installatie van die Board werd eind januari in Den Haag gehouden met DANS als gastheer. Deelnemers in de Board zijn, naast DANS, gereputeerde internationale instituten zoals het UK Data Archive (UKDA), het Duitse datapreserveringsproject NESTOR, het Franse data-instituut CINES, het Nijmeegse Max Planck Instituut, het Alfred Wegener Instituut in Duitsland en het grootste data-archief ter wereld ICPSR in het Amerikaanse Michigan. De eerste release van het Data Seal of Approval vond plaats in januari 2007. Het commentaar en de suggesties die sindsdien binnenkwamen, zijn gebruikt om het keurmerk verder te verbeteren. De komende instelling van de internationale Board betekent voor het Datakeurmerk een sprong in de ontwikkeling tot internationale standaard op het gebied van duurzame data.

'Hoe maken we het delen van onderzoekdata gewoon?'

In januari ging de Nationale Verkenning Digitale Duurzaamheid van start. Met de resultaten van de verkenning in de hand zal de Nationale Coalitie Digitale Duurzaamheid beleidsaanbevelingen doen aan onder meer het ministerie van Onderwijs, Cultuur en Wetenschap. Voor de sector wetenschap is René van Horik aangetrokken als onderzoeker. *e-data&research* vroeg hem naar zijn plannen.

De wetenschap is groot en divers. Hoe gaat u dat onderzoek aanpakken?

Van Horik: 'Ik richt me op drie groepen: onderzoekers, ondersteunende organisaties, zoals data-archieven, en financiers en beleidsmakers van wetenschappelijk onderzoek. Door middel van desk-research en door te spreken met vertegenwoordigers van elke groep hoop ik een goed beeld te krijgen van de stand van zaken en vooruit te kijken naar de toekomst. Daarbij probeer ik een antwoord te krijgen op een aantal vragen. In de eerste plaats: wat voor soort digitale objecten speelt een rol in het wetenschapsbedrijf? Verder: welke digitale objecten dienen ook op lange termijn toegankelijk en bruikbaar te zijn, en: hoe moeten die bewaard en beschikbaar gesteld worden en hoe kun je dat goed organiseren? En ten slotte: wie is waarvoor verantwoordelijk, en wie gaat dit betalen?'

Wat hoopt u aan het eind van het onderzoek in kaart te hebben?

Van Horik: 'Het beste argument om te investeren in digitale duurzaamheid is aan te tonen dat toekomstig hergebruik van onderzoeksdata de wetenschap vooruit helpt. Dit inzicht bestaat al bij een aantal wetenschappelijke disciplines, maar ik hoop dat de NCDD-verkenning er voor zorgt dat het een nog breder draagvlak krijgt, zowel bij de wetenschappers als bij de financiers. Het delen en hergebruiken van onderzoeksdata zou de gewoonste zaak van de wereld moeten zijn. Ik wil onderzoeken wat ervoor nodig is om dat mogelijk te maken: bij de onderzoekers, bij de instellingen waarvoor ze werken, bij de archieven, en bij de financiers.'

Veel onderzoek wordt uitgevoerd in internationale netwerken. Hoe verhoudt dat zich tot een Nationale Verkenning?

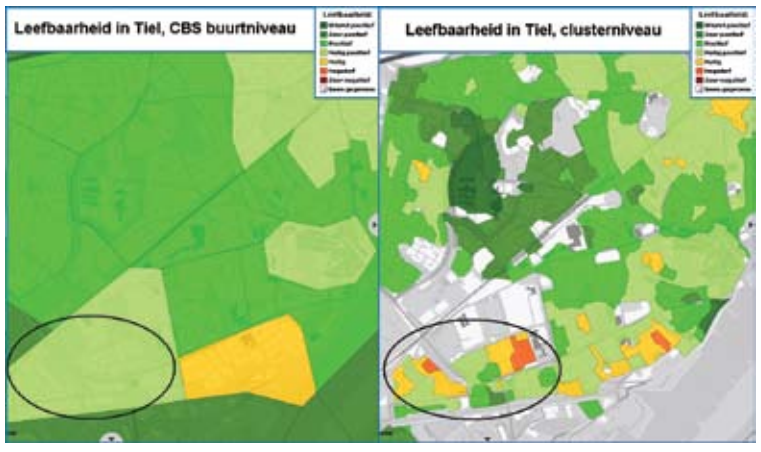
Van Horik: 'Het spreekt vanzelf dat internationale ontwikkelingen en initiatieven een grote rol spelen bij de situatie in Nederland. EU-projecten op het gebied van digitale duurzaamheid, zoals Planets en Caspar, zijn belangrijk voor ons. Hetzelfde geldt voor initiatieven op het gebied van *research infrastructures*, zoals de CLARIN, DARIAH en CESSDA. Dit onderzoek kijkt naar het Nederlandse aandeel in deze internationale initiatieven.'

Het onderzoek gaat uit van de NCDD, waarin digitale archieven als DANS en de KB deelnemen; kijkt het ook naar de behoeftes en eisen van onderzoekers?

Van Horik: 'De behoeftes en eisen van onderzoekers zijn maatgevend, want zij produceren de wetenschappelijke informatie én zij zullen degenen zijn die de informatie hergebruiken. Maar voor de opslag van de informatie hebben zij gespecialiseerde dienstverleners nodig als bijvoorbeeld DANS, de KB, en Narcis. Hun aanbod moet goed aansluiten op de behoeftes van de wetenschappers. Die behoeftes verschillen natuurlijk per wetenschappelijke discipline, maar ik probeer toch een zo volledig beeld te schetsen. (IA)

http://www.ncdd.nl/activiteiten-natverkenning.php

Leefbaarheidssituatie in kaart



Eind vorig jaar heeft de minister voor Wonen, Wijken en Integratie de Leefbaarometer geïntroduceerd. Deze barometer geeft de leefbaarheidssituatie in alle woongebieden in Nederland op verschillende ruimtelijke schaalniveaus weer: vanaf het gemeenteniveau, via wijken en buurten tot aan clusters van enkele postcodegebieden. Daardoor kunnen eventuele kleinere probleemgebieden die op hogere schaalniveaus onzichtbaar zouden blijven, nu ontdekt en gelokaliseerd worden.

Naast de leefbaarheid op een bepaald moment, brengt de Leefbaarometer ook de ontwikkelingen sinds 1998 in kaart. Daardoor kan het effect van gebiedsgericht beleid worden gevolgd, maar ook vroegtijdig signaleren van nieuwe problemen wordt mogelijk. Bovendien kan de leefbaarheidsscore 'uitgeklapt' worden naar onderliggende dimensies: gaat het bijvoorbeeld vooral om veiligheidsproblemen of is het voorzieningenniveau aan het afnemen? De Leefbaarometer maakt tijdig ingrijpen mogelijk, waardoor negatieve ontwikkelingen omgebogen kunnen worden en verder afglijden van gebieden voorkomen. (Sinisa Boksic)

www.vrom.nl/leefbaarometer

Biografisch Woordenboek

Eind vorig jaar hebben het Instituut voor Nederlandse Geschiedenis (ING) en de Digitale Bibliotheek voor de Nederlandse Letteren de gedigitaliseerde versie gepresenteerd van het Nieuw Nederlandsch Biografisch Woordenboek (NNBW). Het oorspronkelijke NNBW verscheen in de jaren 1911-1939 in tien delen met meer dan 22 duizend beknopte biografieën van belangrijke en opvallende Nederlanders. Er wordt nog steeds veel uit geput door wetenschappers, genealogen en journalisten. Volgens een gezamenlijk persbericht betekent de afronding van het project voor het ING een belangrijke stap op weg naar het Biografisch Portaal, in de voorbereiding waarvan het ook samenwerkt met de DBNL. Laatstgenoemde meldt dat het gereedkomen van het gedigitaliseerde woordenboek ervoor zorgt dat nu op haar website van bijna twee keer zoveel personen biografische informatie is te vinden.

IISG start Content Mashup Platform

De subsidieregeling 'Digitaliseren met beleid' van het ministerie van OCW heeft een projectvoorstel gehonoreerd voor een *Content Mashup Platform: een web 2.0 platform voor het erfgoed*. Het Internationaal Instituut voor Sociale Geschiedenis (IISG), dat het voorstel indiende, wil met het platform meer en makkelijker gebruik van collecties bevorderen. Het Platform moet eind 2010 een toolkit en best practices opleveren voor (her)gebruik van erfgoed collecties in andere webomgevingen, voor het vastleggen en uitwisselen van kennis over collecties en voor het bouwen van interactieve websites met collectiepresentaties. De technische oplossingen, gebaseerd op web 2.0 technologie en open standaarden uit de erfgoed sector, zullen generiek van opzet zijn en worden in open source ontwikkeld. Doel van het project is dat zowel experts als geïnteresseerd publiek de mogelijkheid krijgen zelf aan het werk te gaan met erfgoedcollecties. Het platform moet standaard een aantal mogelijkheden voor presentatie geven zoals bijvoorbeeld Google maps of timeline toepassingen. Binnen het project worden pilots uitgevoerd met IISG-collecties die op de nieuwe websites gepresenteerd en van extra informatie voorzien worden. Met het Content Mashup Platform moet bijvoorbeeld een biograaf die onderzoek bij het IISG heeft gedaan naar Pieter Jelles Troelstra snel en gemakkelijk de gebruikte bronnen samen met zijn eigen materiaal publiceren op een website. Ook moet het mogelijk worden dat onderzoekers een portalwebsite maken op basis van gegevens van het IISG en andere instellingen. (Eric de Ruijter)

Memoria in beeld: showcase van nieuwe mogelijkheden

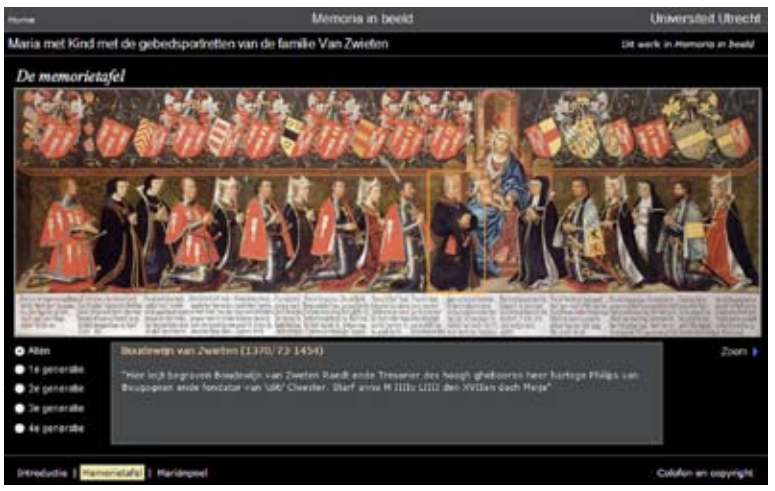
LEEN BREURE

Begin januari is de website 'Memoria in beeld' officieel opengesteld voor publiek. Op de website kan worden gezocht in afbeeldingen en beschrijvingen van ruim vijfhonderd beeldhouwwerken en schilderijen waarvan een groot deel als memorievoorstelling een functie had in de middeleeuwse dodengedachtenis. Door het sterk visuele karakter is de site anders dan veel andere in de kunsthistorische sfeer.

De website en de onderliggende database zijn in samenwerking met DANS tot stand gekomen als onderdeel van het onderzoeksproject *De functies van kunst, ritueel en tekst in de memoria in de Middeleeuwen*, waarvan dr. Truus van Bueren (Universiteit Utrecht) de projectleider is. De database bevat een schat aan fotomateriaal, dat een gedetailleerd beeld geeft van de memorievoorstellingen.

Door de grote aandacht voor het visuele onderscheidt *Memoria in beeld* zich van de meerderheid van de kunsthistorische websites, die vaak te herleiden zijn tot de klassieke genres van geïllustreerde kunsthistorische boeken en tijdschriften, met een nadruk op tekst.

Slechts een enkele website, zoals die van het *Metropolitan Museum of Art*, gaat verder en heeft een sectie *Explore & Learn*. Beeld wordt daar gecombineerd met geluid en film, de bezoeker krijgt een actieve rol, kan ieder moment kiezen welke informatie hij of zij wil zien en wordt uitgenodigd om zelf het beeldmateriaal te onderzoeken. Dit soort webapplicaties staat bekend als *Rich Internet Applications* (RIAs), vanwege de rijkdom aan ervaringsmogelijkheden voor de gebruiker.



Fragment uit de Rich Internet Application over de memorietafel van de familie Van Zwieten.

RIAs worden meestal gecreëerd met een educatief doel, voor een groot publiek. Maar ook in onderzoek en daaraan gekoppeld onderwijs kunnen ze een belangrijke rol vervullen. Op het onlangs gehouden symposium *ICT in de mediëvistiek: het memoria-onderzoek in Nederland* is een eerste versie getoond van een RIA voor memoria-onderzoek, die over enkele maanden zal worden toegevoegd aan *Memoria in beeld*.

Eén schilderij met zijn historische context staat centraal in deze RIA: de memorietafel *'Maria met Kind met de gebedsportretten van de familie Van Zwieten'* en het klooster Mariënpoel bij Leiden, waar het kunstwerk oorspronkelijk hing. De verschillende verhaallijnen over de familie Van Zwieten, het klooster en de politieke gebeurtenissen van die tijd zijn daarin onderling vervlochten, waardoor de gebruiker gemakkelijk van het ene onderwerp naar het andere kan overstappen zonder voorkennis te missen, terug te hoeven bladeren om de draad weer op te pakken of zoekvragen voor de database te hoeven formuleren.

Ook hoe een RIA kan bijdragen aan een beter gebruik van een database, is te demonstreren met behulp van *Memoria in beeld*. Want hoe goed de zoekfaciliteiten van een database ook mogen zijn, de gebruiker moet eerst de nodige achtergrondkennis hebben om de juiste vragen te stellen en de resultaten goed te interpreteren. Dat blijkt alleen al uit de uitgebreide verantwoording en aanwijzingen waarvan *Memoria in beeld* vergezeld gaat.

Deze RIA is een showcase: hij toont het kunstwerk als rijke primaire bron in relatie tot andere bronnen, zowel geschreven bronnen als kunstwerken. Het (letterlijk) dichter bij elkaar brengen van beeld en tekst helpt een kunstwerk sneller en gemakkelijker te 'lezen' en demonstreert hoe de combinatie van verschillende bronnen heeft geleid tot de uiteindelijke interpretatie.

Eeuwenlang ontbraken de technische middelen om zo direct en concreet te communiceren. Daardoor zijn wij verknocht geraakt aan de tekstuele vorm en moet het nieuwe genre nog even wennen. Gedragen door een internetcultuur in 2D (en steeds meer 3D) is het visuele in opmars. Ook in het geesteswetenschappelijk onderzoek kunnen het abstracte lezen en het in gedachte reconstrueren meer en meer plaats maken voor een directer beleven, dat zowel bij het grote publiek als bij studenten interesse en betrokkenheid vergroot.

www.let.uu.nl/memorie/

Opwaardering lange tijdreeksen bij het CBS

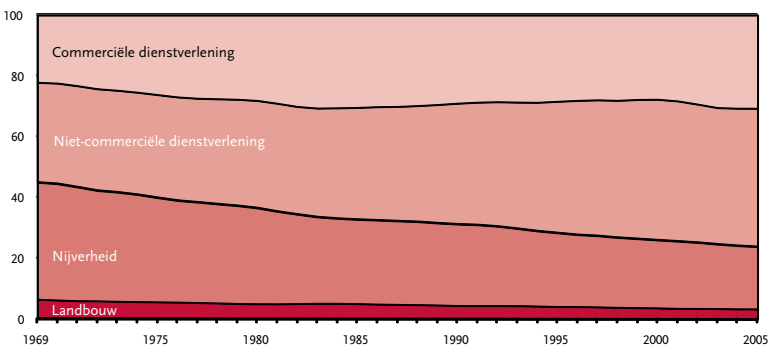
RUURD SCHOONHOVEN

Het Centraal Bureau voor de Statistiek (CBS) beschikt over een goudmijn aan historische gegevens. Het is echter moeilijk om die gegevens zo te ontsluiten dat ze in de tijd vergelijkbaar zijn. Gelukkig krijgt dit onderwerp de laatste jaren meer aandacht van de statistici van het CBS.

Historische gegevens zijn niet altijd makkelijk toegankelijk voor de buitenwereld. Lange tijdreeksen bevatten bovendien soms lacunes omdat bepaalde gegevens niet doorlopend zijn verzameld. Ook komen er breuken in voor door veranderingen in definities of in de opzet van een steekproef. Die gaan ten koste van de vergelijkbaarheid in de tijd.

Om dit probleem structureel aan te pakken is enkele jaren geleden bij het CBS een Expertisecentrum Lange Tijdreeksen opgericht. Dat ondersteunt statistiekmakers bij het samenstellen van de reeksen en vooral bij het repareren van breuken daarin. Ook krijgen volledigheid en samenhang van historische tijdreeksen, zoals die worden gepubliceerd op de elektronische database StatLine, meer aandacht.

Een belangrijke tijdreeks die inmiddels is voltooid is die van de Nationale Rekeningen: de boekhouding van Nederland. Met ingang van 2001 zijn daarin nieuwe concepten, definities en werkwijzen gehanteerd. In een groot project zijn de oude cijfers terug tot en met 1969 consistent gemaakt met deze nieuwe situatie.



Procentuele verdeling arbeidsvolume over bedrijfstakken (CBS, 2008)

Figuur 1 laat aan de hand van de zo ontstane reeksen zien hoe de verdeling van de arbeidsinzet over de verschillende bedrijfstakken in deze periode is veranderd. In 1969 zorgde de nijverheid (industrie, bouw etc) nog voor de meeste werkgelegenheid, maar dit aandeel is in 2005 gehalveerd terwijl de commerciële dienstverleners de grootste bron van werkgelegenheid zijn geworden. Binnen de niet-commerciële dienstverlening is vooral het aandeel van de zorg sterk toegenomen.

Een andere belangrijke tijdreeks die met hulp van het expertisecentrum consistent is gemaakt is die van de gezonde levensverwachting: de 'klassieke' levensverwachting minus het aantal jaren dat niet in (goede) gezondheid wordt doorgebracht. Voor dat laatste worden definities gehanteerd als de zelf ervaren gezondheid, het lijden aan chronische ziekten of het hebben van lichamelijke beperkingen. Om tijdreeksen van deze variabele te kunnen maken moesten methodebreuken worden gerepareerd in enquêtedata op het gebied van gezondheid. De veranderingen die in de loop der tijd in de opzet van de enquête waren doorgevoerd, gaven op zichzelf geen indicatie voor de grootte van de breuk die ze veroorzaakten. Daarom zijn wiskundige methoden gebruikt om de die grootte te schatten. Vervolgens kon daarop weer een correctie worden gebaseerd om de gehele reeks consistent te maken met de actuele werkwijze. Figuur 2 geeft een illustratie van de methode aan de hand van een set denkbeeldige data. Figuur 3 geeft de tijdreeks Gezonde Levensverwachting weer zoals die onlangs op Statline is gepubliceerd.

In de loop van 2009 zullen tijdreeksen worden gepubliceerd op het gebied van bedrijvenconjunctuur vanaf de jaren vijftig, en van beroepsbevolking en arbeidsparticipatie. Voor de komende jaren staan nieuwe projecten op het programma, waarin onder meer vooruitgelopen wordt op te verwachten nieuwe reeksbreuken, die nu eenmaal onvermijdelijk met het vak van statistiek maken verbonden zijn.

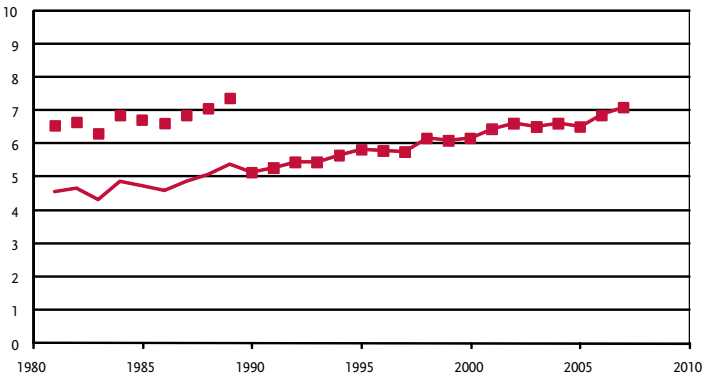


Fig. 2 Reparatie van reeksbreuken in statistische gegevens. De variabele is met ingang van 1990 onderzocht met een veranderde vraagstelling. Met een wiskundig model wordt de breuk in de data (▪) gerepareerd tot een consistente reeks (–).

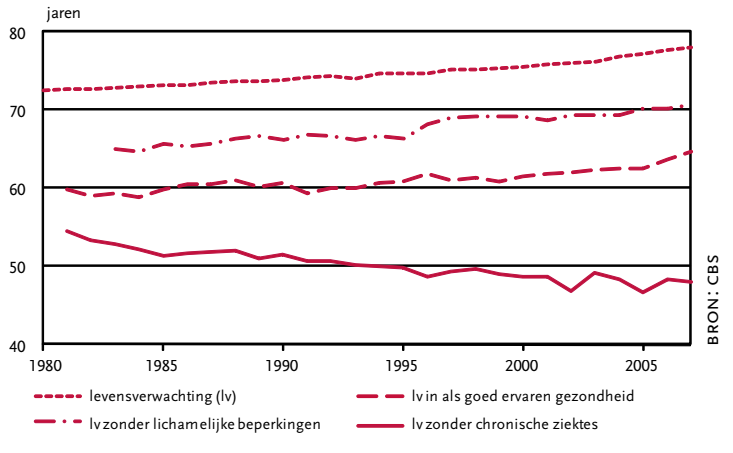


Fig. 3 Gezonde levensverwachting van mannen bij geboorte, 1981-2007.

Alfa-informaticus John Nerbonne:

‘Ik vind dat iedereen zijn data beschikbaar moet stellen’

MARTIJN DE GROOT

Op het kruispunt van exacte wetenschappen en humaniora. Zo zou je het beste de plek kunnen beschrijven waar John Nerbonne als wetenschapper is neergestreken. Je zou ook kunnen zeggen: Groningen, want voordat de geboren Amerikaan daar hoogleraar alfa-informatica werd had hij al werkplekken in verschillende staten en landen achter zich gelaten.

‘Een echte alfa,’ noemt Nerbonne zichzelf. Maar, zegt hij er meteen bij, ‘je moet in dit vak wel een programma zelf kunnen ontwerpen als je in het bestaande aanbod niet vindt wat je nodig hebt.’ En een diepgewortelde belangstelling hebben voor wiskunde, moet daar toch aan worden toegevoegd in het geval van Nerbonne die als veertienjarige jongen een biografie van de Britse getalstheoreticus G.H. Hardy las en daar nu nog een deel van zijn fascinatie op terugvoert. Maar een feit is dat zijn vak computationele taalkunde in Groningen als een van de weinige plekken in Nederland onder de Letterenfaculteit is ondergebracht. ‘Bij de meeste universiteiten hier valt het onder informatica, net als in de Verenigde Staten. In Duitsland is het weer meestal onderdeel van taalkunde. Zelf ben ik blij dat ik hier mag werken, maar een informaticus wil niet graag aan de Letterenfaculteit promoveren.’

‘Die vrijheid is uniek. Je vindt iets interessants en dan ga je daar iets mee doen’

Duitsland en de Verenigde Staten. Het zijn de twee landen waar Nerbonne verbleef voordat hij in 1993 in Groningen terecht kwam. ‘De liefde bracht me naar Nederland’ vat hij zijn zwerftocht samen. Hij werd geboren in de Amerikaanse staat Massachusetts en kreeg als 22-jarige de kans om een aantal jaren in Duitsland te studeren: *Germanische Philologie* aan de universiteit van Freiburg. ‘Daar heb ik mijn huidige vrouw leren kennen. We zijn toen samen naar Amerika gegaan, waar ik eerst in Ohio ben gepromoveerd en later in California bij Hewlett Packard werkte. Maar na een jaar of tien wilde zij graag terug, en ik had nooit met tegenzin in Europa gewoond dus dat hebben we toen gedaan.’ Nerbonne nam een baan aan bij het Duitse bedrijf DFKI, dat zich met kunstmatige intelligentie bezighoudt, en dat bleek achteraf het voorportaal van het hoogleraarschap dat hem in Groningen wachtte. Nederland, ‘het hoofdland van de logica in de wereld met op dat gebied een topwetenschapper als Johan van Benthem’, dat leek hem ook een prima bestemming. De wetenschap was toch wel zijn eerste keus, weet hij nu. ‘De vrijheid die je aan een universiteit hebt, dat is uniek. Je vindt iets interessants en dan ga ja daar iets mee doen!’

ELMER SPAARGAREN

Die laatste zinsnede moet bij Nerbonne heel letterlijk worden opgevat, zo blijkt. ‘Vaak wordt de vraag ‘fundamenteel of toegepast’ als een polariteit gezien maar ik zie dat niet als een tegenstelling. Computational Linguistics wordt wel een ingenieurswetenschap genoemd. Dat zou ik zelf niet doen maar een feit is dat er veel aan toepassingen wordt gedaan en dat is ook leuk. Het brengt bovendien geld op.’ Technieken om hardop te lezen voor blinden en dyslectici en andere voorleestoepassingen zijn daar voorbeelden van. Of het werk aan computerondersteund taalonderwijs, bijvoorbeeld met een elektronisch woordenboek achter de tekst zoals in samenwerking met Van Dale is gerealiseerd (zie www.let.rug.nl/glosser). Of, ander voorbeeld, software van de hand van een collega van Nerbonne die producenten van medicijnen helpt om namen uit te kiezen die niet te veel lijken op reeds gangbare namen van andere medicijnen. ‘Zanax en Sanex bijvoorbeeld. Onderschat dat niet, dergelijke namen worden in de medische praktijk vaak door elkaar gehaald en daar gebeuren ernstige ongelukken mee.’

‘Ik denk dat wij de eerste goede visualisatie hebben gemaakt van de verspreiding van dialecten’

Is computationele taalkunde nu eigenlijk een discipline of niet? ‘Het is een soort verbond, een federatie van verschillende disciplines,’ antwoordt de linguïst, die met gemak voorbeelden uit de mouw schudt van samenwerkingsvormen met andere wetenschappen. ‘Ik heb heel vruchtbaar samengewerkt met genetici, om de verbanden te onderzoeken tussen de herkomst van bevolkingsgroepen en uitspraakverschillen. Er zijn ook collega’s bezig in de musicologie om in overeenkomsten in de verspreiding van volksmuziek en taalverschillen te zoeken. Er zijn projecten met fysici, met archeologen. En ik kan me voorstellen dat je architectuurpatronen over de Atlantische oceaan heen gaat volgen om die weer in verband te brengen met vestigingspatronen en taalvariatie.’

‘Ik zie de vraag ‘fundamenteel of toegepast’ niet als een tegenstelling’

De laatste jaren is de belangstelling van de Groninger Amerikaan steeds meer de kant van de dialectkunde op gegaan, het vak dat de ontwikkeling, geografische spreiding en variatie van dialecten bestudeert. Hij liet zijn wetenschappelijk oog al vallen op Nederlands, Duits, Noors, Sardijns, Bulgaars en Bantu. Maar ook de tweede rijkstaal, het Fries, en een officiële minderheidstaal van de Europese Unie, het Nedersaksisch, mogen zich in zijn belangstelling verheugen. Daarbij gaat het Nerbonne vooral om uitspraakverschillen en de wiskundige formules om die *en masse* te meten. Want dat is ook volgens Nerbonne de ‘allerbelangrijkste ontwikkeling’ van de laatste jaren: ‘De gebruikelijke manier was altijd: je vindt een moedertaalspreker en daar stel je vragen aan. Die vertelt je dan hoe woorden worden uitgesproken en welke woorden wel en niet gangbaar zijn. Maar je was daarmee aangewezen op de tamelijk willekeurige weergave van degenen met wie je in contact kwam. Nu we over steeds meer en grotere corpora kunnen beschikken – en dat neemt echt een steeds grotere vlucht – kunnen we zien òf een woord of uitspraak wordt gebruikt en hoe vaak. En je kan natuurlijk veel beter naar patronen gaan zoeken in de verspreiding van taalvariaties.’

Nerbonne laat met zichtbaar plezier een tweetal kleurrijke kaarten zien van het Nederlandse taalgebied, die een paar jaar geleden het resultaat waren van een samenwerkingsproject met het Meertens Instituut. ‘Hierbij hebben we *multidimensional scaling* toegepast, een techniek die het mogelijk maakt om de verschillen in uitspraak en vocabulaire weer te geven als continuüm, in geleidelijke kleurovergangen. Dat was echt een grote stap. Ik denk dat Peter Kleiweg en ik in dit project, waaraan vanuit het Meertens ook Wilbert Heringa deelnam, de eerste goede visualisatie hebben gemaakt van de verspreiding van dialecten. Op dezelfde manier hebben we een weergave gemaakt (haalt een andere kaart tevoorschijn) waarin de afstand van een bepaalde streektaal tot het Algemeen Beschaafd Nederlands is weergegeven’.

Zulke vormen van onderzoek zouden ondenkbaar zijn zonder de beschikbaarheid van enorme databestanden, weet Nerbonne. ‘Mensen die in dit vakgebied specifieke studies willen doen hebben behoefte aan heel veel data. Daarom is de functie van een instituut als DANS ook van onschatbare waarde. Het zelf verzamelen van data is duur en die kosten moet je besparen voor betere doelen. Maar daarvoor is wel nodig dat onderzoekers hun data ook beschikbaar stellen, en dat gebeurt nog lang niet altijd van harte. Er zijn er nog steeds die daar eigenlijk niet aan willen: ‘Ja, we zijn er nog mee bezig, het is nog niet definitief…’, zulke argumenten. Ik vind dat iedereen zijn data beschikbaar moet stellen via DANS of een instelling zoals DANS. We horen nog niet zo veel van de activiteiten van DANS op het gebied van taal en taalkunde. De populatiegenetici hebben het nu bijna zo goed voor elkaar dat ze op websites die hele grote, belangrijke bestanden kunnen zoeken en vinden. Met projecten als CLARIN en DARIAH gaan we hopelijk in de taalkunde ook die kant op.’

John Nerbonne is sinds 1993 hoogleraar Alfa-Informatica aan de Rijksuniversiteit van Groningen. Daarvoor studeerde hij onder meer Germanische Philologie aan de Universiteit van Freiburg en werkte hij (aan zijn proefschrift) aan de universiteit van Ohio (VS), bij Hewlett Packard in California en bij het Duitse kunstmatige-intelligentiebedrijf DFKI. Sinds 1999 is hij ook directeur van het Groninger Centrum voor Taal en Cognitie. Hij leidde projecten over computer-assisted taalonderwijs, grammatica ontwikkelen met behulp van machinaal leren en het terugvinden van handgeschreven documenten. Hij bestudeerde uitspraakverschillen binnen de Nederlandse, Duitse, Noorse, Engelse en Bulgaarse taalgebieden en het Bantu. Hij diende als president van de Association for Computational Linguistics (2000 leden) in 2002 en werd in 2005 lid van de Koninklijke Nederlandse Akademie van Wetenschappen.

Focus

Erasmus Data Service Center

LUUK SCHREVEN

Het groeiende belang van kwantitatief onderzoek en de toegenomen vraag naar gegevens van databanken leidden bij de universiteitsbibliotheek van Rotterdam tot de oprichting van een speciaal service centre voor data, het EDSC. 'Wij bieden iedere bezoeker toegang tot financiële en sociaal-wetenschappelijke databanken, we ondersteunen studenten en medewerkers en we geven workshops. Dat is uniek in Nederland,' zegt Paul Plaatsman, hoofd van het centrum.

e-data & research vroeg aan hem en Gusta Drenthe, lid van het EDSC adviesorgaan, naar de functie en de taken van het centrum. Het EDSC is in 2007 toegevoegd aan de dienstverlening door de Universiteitsbibliotheek, lichten ze toe, maar de ontwikkelingen in die richting zijn al in de jaren negentig in gang gezet. De Erasmus Universiteit heeft van oudsher een sterke economische faculteit die zich op statistiek en kwantitatief onderzoek richt. Ook is de Universiteitsbibliotheek in Rotterdam als een centrale bibliotheek opgezet, met een beperkte en afnemende rol van faculteitsbibliotheken. 'Binnen de bibliotheek bestond een studiezaal statistieken met oorspronkelijk alleen papieren publicaties waaruit het EDSC heeft kunnen groeien', aldus Plaatsman. 'Toen in de jaren negentig steeds meer gegevens digitaal werden aangeboden ontstond ook steeds meer vraag naar ondersteuning.' Datastream, een database van Thomson-Reuters met bedrijfsinformatie en financiële data, was een van de eerste commerciële producten die de UB aanschafte. 'Datastream bevat heel veel informatie, maar was zeker niet gebruikersvriendelijk opgezet. In het begin brachten we vijf gulden kosten in rekening voor de ondersteuning bij het gebruik. Later deden we dat gratis voor iedereen met een relevante vraag.'

WIEBE KIESTRA

Gusta Drenthe en Paul Plaatsman van het EDSC

Door het groeiende belang van kwantitatief onderzoek en de groeiende vraag naar en beschikbaarheid van elektronische gegevens kwam er ook steeds meer vraag naar ondersteuning. Er werden licenties gekocht op uitgebreide datasets, maar het ontbrak binnen de faculteiten aan structurele dienstverlening. Drenthe: 'Het besef begon door te dringen dat je met enkel de aanschaf van dure datasets geld over de balk smijt. Onderzoekers en studenten hebben behoefte aan voldoende ondersteuning om met zo'n dataset te werken.' De UB nam in 2006 het initiatief om het EDSC op te richten, eerst in samenwerking met de faculteiten bedrijfskunde en economie, maar sinds de zomer van 2007 neemt ook de faculteit sociale wetenschappen eraan deel. Inmiddels heeft het centrum een Datateam van vijf part-time specialisten, waaronder twee aio's.

Volgens Plaatsman kunnen de faculteiten nog wel een stap zetten wat betreft de inbedding in het onderwijs: 'Soms begeleid ik een student die al dagen bezig is met het verzamelen van data via allerlei websites en jaarverslagen. Als ik dan in een keer alle benodigde gegevens voor verschillende bedrijven uit onze databases trek, dan vragen ze zich hardop af waarom ze niet eerder op het bestaan van onze club geattendeerd zijn.' Maar de bekendheid bij studenten neemt toe; het EDSC geeft regelmatig workshops in de masterfase van hun opleiding.

Een wens voor de toekomst is verdere uitbreiding van het werkgebied. 'Ook op het gebied van de gezondheidszorg en de ziektekosten zijn veel datasets te ontsluiten,' aldus Drenthe, 'bijvoorbeeld van de Organisatie voor Economische samenwerking en Ontwikkeling van de Verenigde Naties'. Daarnaast wil het centrum verder werken aan het verzamelen van tijdreeksen en zijn er plannen om een Remote-Access werkomgeving in te richten voor onderzoek op CBS-microdata. Drenthe: 'Er is een grote vraag naar die microdatabestanden. Als het Centraal Bureau voor de Statistiek ook verder geanonimiseerde en beveiligde microdata beschikbaar zou stellen voor studenten in het onderwijs zou dat voor veel faculteiten interessant zijn.'

www.eur.nl/edsc

Databanken, bij het gebruik waarvan het Erasmus Data Service Centre helpt:

Financiële databanken	Sociaalwetenschappelijke databanken
Audit Analytics	DANS
Bankscope	Eurostat
Company.info	FAO
Compustat	Global Development Finance
CRSP	ICPSR
Datastream	IMF: BOP, DOTS and IFS
DealScan	Lexis Nexis Statistical
Execucomp	OECD Health Data
I/B/E/S	Premium China Database / CEIC data
IMF: IFS	SourceOECD
Market Insight	Statline
Mutual Fund Link	Steinmetz-archief (DANS)
Option Metrics	Unctad handbook of statistics
REACH	United Nations
SDC	World Development Indicators
Thomson ONE Banker	World Database of Happiness
Thomson Research	
WRDS	
Worldscope	
Zephyr	

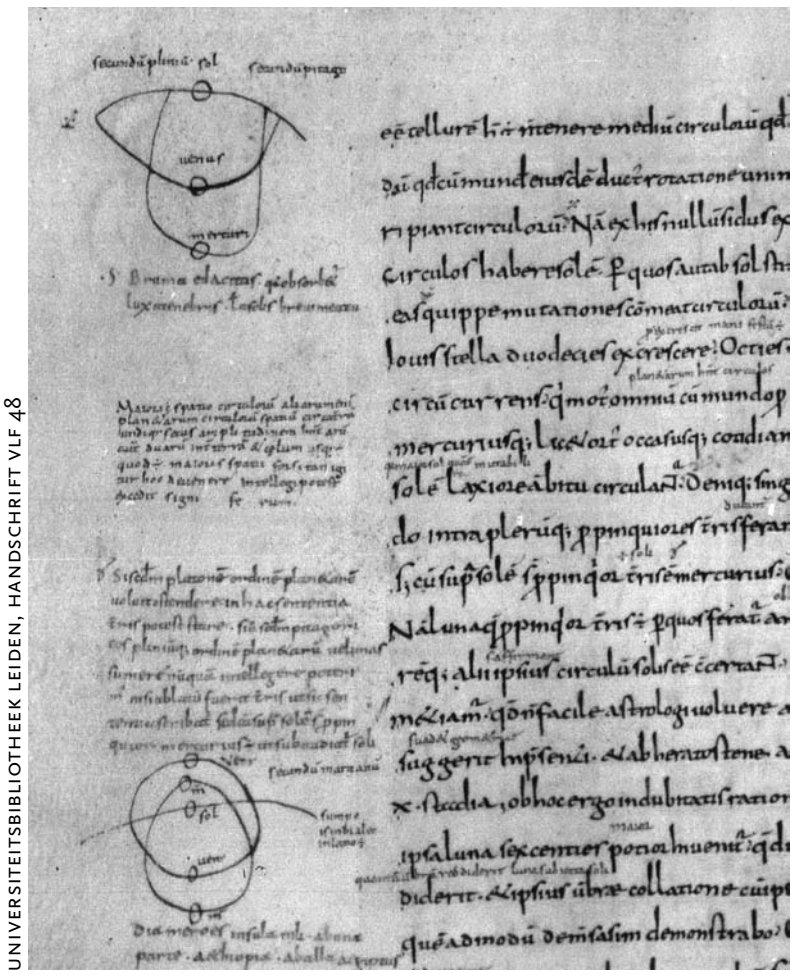
Middeleeuwse kijk op antieke kennis: een online editie

Als Harry Potter in J.K. Rowlings Harry Potter en de Halfbloed Prins een oud receptenboek leent voor het bereiden van toverdranken ontdekt hij tot zijn ergernis dat het volgekliederd staat met aantekeningen van de vorige eigenaar. Al snel komt hij er echter achter dat de aantekeningen uiterst waardevol zijn, en gaat hij eerder op de handgeschreven kriebels af dan op de gedrukte aanwijzingen.

Ook in de negende eeuw gebruikten geleerden hun afschriften van oude teksten om er nieuwe tekst aan toe te voegen. Ze kozen zelfs speciaal voor een lay-out met grote marges en wijde interlinie om hun aantekeningen (glossen) bij de tekst kwijt te kunnen, en net als de hoofdtekst werd ook de commentaartekst overgeschreven van exemplaar naar kopie.

Een van de teksten die in de negende eeuw met uitzonderlijk enthousiasme bestudeerd werd door de intellectuele elite is het werk van de vijfde-eeuwse Noordafrikaanse auteur Martianus Capella, die een handboek schreef over de zeven Vrije Kunsten (de *artes liberales*): *De nuptiis Philologiae et Mercurii (Over het huwelijk van Philologia en Mercurius)*. De negende-eeuwse bestudering van de antieke wetenschappelijke traditie is in grote mate bepaald door dit werk. De commentaartraditie op dit werk geeft inzicht in het allereerste contact tussen middeleeuwse geleerden en de antieke erfenis op het terrein van bijvoorbeeld logica, wiskunde en sterrenkunde.

In november 2008 is op het web een editie gelanceerd van het negende-eeuws materiaal dat Karolingische geleerden hebben geproduceerd rondom deze vijfde-eeuwse tekst. Voor het project heeft senior onderzoeker dr. M.J. Teeuwen van het Huygens Instituut een online werkomgeving gecreëerd met behulp van eLaborate, een tool waarmee onderzoekers uit de geesteswetenschappen op eenvoudige wijze een collaboratory kunnen opzetten. Digitale foto's van de belangrijkste manuscripten zijn online geplaatst, geflankeerd door panelen met een semi-diplomatische editie van hoofdtekst en commentaartekst. De zeldzame specialisaties die nodig zijn voor het editeren van een tekst over, bijvoorbeeld, wis- of sterrenkunde in de negende eeuw, konden op deze wijze eenvoudig aan het project verbonden worden, ook al moesten zij gezocht worden in allerlei hoeken van de wereld. Vanaf hun eigen werkplek in Ierland, Groot-Brittanië, Frankrijk, de Verenigde Staten of Nederland konden specialisten in hun eigen tempo de foto's van vier handschriften raadplegen, transcripties toevoegen en/of corrigeren, en voortdurend op de hoogte blijven van elkaars vorderingen en vragen. Voor het project is steeds nauw samengewerkt met de afdeling e-Research van het Huygens Instituut, die een geheel nieuwe online omgeving heeft ingericht voor de editie. (Mariken Teeuwen)



UNIVERSITEITSBIBLIOTHEEK LEIDEN, HANDSCHRIFT VLF 48

Detail uit het boek over sterrenkunde. De diagrammen illustreren de bewegingen der planeten

http://martianus.huygensinstituut.nl.

Preferred formats

DANS publiceerde in 2008 het Datakeurmerk (zie ook pagina 1 'Datakeurmerk gaat internationaal'). Een van de richtlijnen van het Datakeurmerk dat DANS vorig jaar publiceerde, geeft aan dat een dataproducent zijn gegevens moet aanleveren in een door het archief voorgeschreven formaat. In de afgelopen decennia zijn er echter heel wat bestandsformaten gekomen en weer (bijna) verdwenen. Een bekend voorbeeld is Word Perfect, ooit razend populair. Voor DANS zelf was het publiceren van de richtlijnen aanleiding om nog eens goed te kijken naar haar eigen lijst van 'preferred formats'.

Een werkgroep binnen DANS heeft deze lijst geactualiseerd. De werkgroep heeft dertien bestandssoorten onderkend die voor het beheren en beschikbaar stellen van onderzoeksgegevens in de alfa- en gammawetenschappen van belang zijn. Enkele daarvan zijn opgemaakte tekst, spreadsheets, images, statistische bestanden en databases. Voor elke bestandsoort heeft de werkgroep vastgelegd in welk formaat of welke formaten op dit moment het meeste vertrouwen bestaat als het gaat om de bruikbaarheid op de lange termijn. Aanvullend zijn enkele andere gangbare bestandsformaten genoemd die op betrouwbare wijze om te zetten zijn naar een duurzaam formaat.

Een bestandsformaat dat de laatste jaren wat betreft duurzaamheid zeer veel vertrouwen heeft gekregen is PDF/A, dat echter als nadeel heeft dat de erin verpakte data maar zeer beperkt herbruikbaar zijn. Ook voor het Open Document Format (ODF) zijn de verwachtingen hoog. ODF kent een goede balans tussen duurzaamheid en herbruikbaarheid. Daarom hebben PDF/A en ODF eveneens een plaats gekregen in de lijst. In de komende tijd zal de lijst van preferred formats worden gepubliceerd.

Daarbij heeft DANS enerzijds op het oog dataproducenten bewust maken van de noodzaak om de bruikbaarheid op lange termijn in het oog te houden en daar ook een extra inspanning voor te verrichten. Anderzijds wil het instituut de producenten niet ontmoedigen om data die ze al beschikbaar hebben bij haar te deponeren. Andere actiepunten zijn eventuele aanwijzingen voor het aanmaken van preferred formats en het aanbieden van gereedschappen hiervoor. Wie geïnteresseerd is in de lijst van preferred formats, of daarover graag wil meedenken, kan contact opnemen met henk.koning@dans.knaw.nl. (Henk Koning)

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
Werkloosheidsuitkeringen (WW)	1e helft 2008
Bijstandsuitkeringensatistiek (BUS en BUS-TRANS)	1e helft 2008
Bijstandsfraude en bijstandsdebiteuren statistiek (BFS-BDS)	1e helft 2008
Registratie Arbeidsongeschiktheid (AO)	1e helft 2008
Statistiek reïntegratie gemeenten (SRG)	1e helft 2008
AOW-uitkeringen (AOW)	1e helft 2008
Pensioenaanspraken (PA)	2005
Inkomens panel onderzoek (IPO)	2006
Consumenten prijzen index (CPI)	2007
Enquête Beroepsbevolking (EBB)	2007
Landbouwtelling (LBT)	2007
Milieukosten van bedrijven (MKB)	2006
Productiestatistieken van diverse bedrijfstakken	2000-2006
Beschikbaar via DANS EASY	
Archeologie	
Wonen en begraven nabij Elst (Gld.)	2006
Archeologisch onderzoek te Opperdoes Kluiten-Zuid	2007
Romeinen aan de Ring	2006
Neolithische bewoningsresten te Leidschendam	2006
Prehistorische bewoning op het World Forum gebied – Den Haag	2007
Erven uit de vroege ijzertijd en de Late Middeleeuwen	2007
Heesche landweren	2007
(Allen Archeologisch Onderzoeksbureau Leiden, Archol b.v.)	
Geschiedenis	
Census Nederlands Toneel (CENETON, dr A.J.E. Harmsen)	1500-1803
Sociale wetenschappen	
Arbeidsaanbodpanel (OSA – Universiteit Tilburg)	2004
High-school pupils about themselves and their contacts (Radboud Universiteit, Hidde Bekhuis)	2007
ICT gebruik – POLS ICT Consumenten Conjunctuur Onderzoek – CCO POLS (Beide CBS – beveiligde microbestanden)	2008

Dataverse maakt delen eenvoudig

Eind vorig jaar gaf projectleidster dr. Merce Crosas van de Harvard University bij DANS in Den Haag een presentatie over het Dataverse Network Project. Dataverse biedt instellingen, onderzoeksgroepen en onderzoekers de mogelijkheid om onderzoeksdata op een eenvoudige manier te archiveren en te delen met anderen. Het gebruik van de software is gratis en de gebruiker kan de onderzoeksdata op de Harvard-server opslaan, een eigen versie van Dataverse installeren of volstaan met een beschrijving van een databestand met een link naar het bestand.

LUCAS PASTEUNING

Dataverse projectleidster Merce Crosas op bezoek bij DANS

De aanwezigen bij de presentatie waren getuige van een inspirerende bijeenkomst met een levendige discussie over alle ins and outs van het archiveren en delen van onderzoeksdata. Uit het groeiend aantal (gerenommeerde) instellingen dat gebruik maakt van Dataverse blijkt dat Dataverse snel aan populariteit en bekendheid wint.

Een van de verklaringen van dit succes is dat het team achter Dataverse goed heeft gekeken naar de eisen en wensen van onderzoekers. Onderzoekers willen namelijk controle houden over wie de data gaat gebruiken en voorwaarden kunnen stellen aan het gebruik van de data. Merce Crosas liet zien dat dit in Dataverse eenvoudig in te stellen is en dat downloadstatistieken van de databestanden standaard beschikbaar zijn. Een belangrijk onderdeel van Dataverse is dat een onderzoeker geciteerd kan worden met een databestand. Verder kan elk databestand van een echtheidsstempel worden voorzien.

Een van de instellingen die gebruik maakt van Dataverse is de Universiteit van Tilburg, die het gebruikt voor de opslag en het delen van bestanden binnen het Network of European Economists Online (NEEO, www.nereus4economics.info/neeo.html). NEEO is een Europees project waarin zestien universiteiten uit acht landen samenwerken om publicaties en de onderzoeksdata die aan de basis liggen van de publicatie, open access beschikbaar en toegankelijk te maken. Doorslaggevend voor de keuze voor Dataverse was de eenvoud in het gebruik en de mogelijkheid om de databestanden via een internationale standaard voor sociaalwetenschappelijk onderzoek (Data Document Initiative, DDI) te kunnen beschrijven. (Rob Grim)

http://thedata.org/

Workshop over Trusted Digital Archives

Hoe kunnen de kwaliteit en betrouwbaarheid van digitale onderzoeksarchieven gegarandeerd kan worden en wanneer is een instelling een Trusted Digital Archive? Dat waren op 30 januari belangrijke vragen in een workshop die bij DANS in Den Haag werd gehouden. Er is de laatste tijd een aantal richtlijnen en *best practices* ontwikkeld die heel verschillend zijn en variëren van breed geformuleerd tot zeer gedetailleerd. Ze hebben echter met elkaar gemeen dat ze criteria bevatten waaraan een *trusted digital repository* zou moeten voldoen.

De workshop was specifiek gericht op de toepasbaarheid van al deze richtlijnen in de sociale wetenschappen en in de humaniora. Na korte inleidingen over de richtlijnen werden deze kritisch onder de loep genomen door onderzoekers uit de genoemde wetenschapsgebieden en informatiespecialisten werkend bij (data)archieven of andere repositories.

De aandacht ging daarbij vooral uit naar het door DANS ontwikkelde 'Data Seal of Approval' (DSA), of

Datakeurmerk. De richtlijnen van het Datakeurmerk werden door hun ruime formulering en flexibiliteit als breed toepasbaar beoordeeld. Als belangrijke pluspunten werden bovendien het creëren van bewustwording van het digitale duurzaamheidsprobleem genoemd en het scheppen van een algemeen referentiekader. Daardoor is het Data Seal bruikbaar voor uiteenlopende onderzoeksgroepen en organisaties. Wel is in de praktijk nadere operationalisering nodig. Methoden voor digitale data-archivering zoals TRAC, DRAMBORA of NESTOR kunnen daarbij van pas komen, ook al zijn deze technischer, specifieker en gedetailleerder van aard.
(Heiko Tjalsma)

www.datasealofapproval.org/node/6

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 Internationaal Instituut voor Sociale Geschiedenis, 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: 7500. 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: Peter Boot, Ivo Gorissen, Martijn de Groot (hoofd/eindredacteur), Inge Angevaare, Jetske van der Schaaf, Luuk Schreven

Aan dit nummer werkten mee: Sinisa Boksic, Leen Breure, E.H. Dooijes, Rob Grim, Thijs Hermsen, Henk Koning, Jan Kooistra, Bart de Nil, Dirk Roorda, Eric de Ruijter, Ruurd Schoonhoven, Mariken Teeuwen, Heiko Tjalsma.

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Vormgeving en opmaak: Ellen Bouma

Productie: Uitgeverij Aksant, Amsterdam

Druk: PlantijnCasparie, Almere

ISSN: 1872-0374

INGEZONDEN

Emulatie vraagt behoud originele hardware

In *e-data&research* van december 2008 las ik het artikel 'Pleidooi voor een softwarearchief' van Jeffrey van der Hoeven en Frank Houtman. In dit artikel wordt bepleit systeem- en toepassingsprogrammatuur uit het verleden in een bruikbare toestand te bewaren, omdat alleen dan de mogelijkheid bestaat oude databestanden te 'beleven' op de manier waarop dat destijds bedoeld was. Het is – zo stellen de auteurs – dan niet nodig ook naar de oude hardware terug te grijpen, omdat deze immers geëmuleerd kan worden op moderne computers.

Met het uitgangspunt van de auteurs ben ik het volledig eens. Maar zij zien een niet te onderschatten moeilijkheid over het hoofd. Van der Hoeven en Houtman noemen enkele obstakels, zoals het feit dat wat je aan software moet bewaren de neiging heeft zich als een olievlek uit te breiden, en dat je een gebruikerslicentie nodig hebt. Maar het meest voor de hand liggende probleem vermelden zij niet: hoe lees je de originele drager als het een 8-inch floppy disk is, of een pak ponskaarten?

Het zal toch duidelijk zijn dat de hiervoor benodigde hardware – die een fysiek proces, anders dan een verzameling logische operaties moet implementeren – niet geëmuleerd kan worden. Er zit dus niets anders op, dan voor het inlezen de originele leesapparaten te gebruiken. Die zullen in de meeste gevallen geenszins *plug-and-play compatible* zijn met een moderne computer, al dan niet voorzien van emulatiesoftware.

De meest efficiënte oplossing is in veel gevallen om niet alleen de originele leesapparatuur maar ook de daarbij behorende (oude) computers te gebruiken. Dit veronderstelt natuurlijk wel de beschikbaarheid van beide. Overigens speelt hetzelfde probleem ook – zij het in mindere mate – bij het tot leven brengen van de data. Specifieke uitvoerapparaten zoals vectordisplays en plotters zijn tegenwoordig met een kaarsje te zoeken.

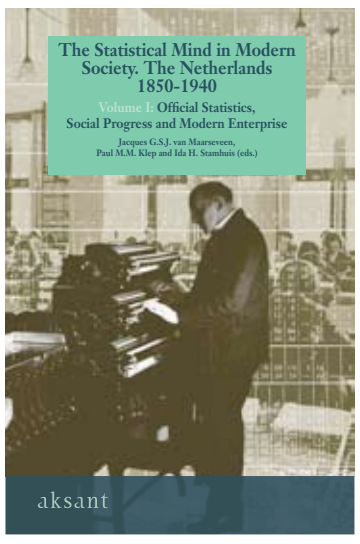
FOTO E.H. DOOIJES

De Ducumation M300 (ca 1975) leest, aangesloten aan een Digital PDP11 minicomputer, 300 pons- of schrapkaarten per minuut.

Om aan zulke moeilijkheden tegemoet te komen is het wenselijk een verzameling antieke apparaten in bruikbare toestand te houden. Dit kan (om technische redenen) hoogstens enkele tientallen jaren worden volgehouden, maar aangenomen mag worden dat in zo'n tijdsbestek de interessante databestanden wel geïdentificeerd en ingelezen zijn. Dit nog afgezien van het feit dat de meeste dragers een zeer beperkte levensduur hebben. Het Computermuseum van de Universiteit van Amsterdam beschikt over zo'n 'data-atelier' dat ondanks bescheiden middelen zijn nut in heel wat gevallen al heeft bewezen. Ik zou er voor willen pleiten dat ook deze kant van de data-conserveringsproblematiek ruimere aandacht en steun verkrijgt.

Dr. E.H. Dooijes
Conservator UvA Computermuseum

Gelezen



***The Statistical Mind in Modern Society; The Netherlands 1850-1940.* Vol.I: J.G.S.J. van Maarseveen, P.M.M. Klep & I.H. Stamhuis – eds., Official Statistics, Social Progress and Modern Enterprise; Amsterdam, Aksant, 2008. ISBN 987-90-5260-321-6**
Vol. II: I.H. Stamhuis, P.M.M. Klep & J.G.S.J. van Maarseveen – eds., *Statistics and Scientific Work*; Amsterdam, Aksant, 2008. ISBN 978-90-5260-322-3
Statistiek als nieuwe en succesvolle combinatie van denkstijl en sociale praktijk, dat is het thema van deze twee bundels. Omstreeks 1850 nog een uitzondering, was deze nieuwe habitus in 1940 een geaccepteerde 'objectieve' weergave van de 'werkelijkheid'. Op basis hiervan werden in toenemende mate beslissingen genomen. Dat gold voor de beleidsvoorbereiding, het parlementair debat, de staatscontrole op de burgers en het in kaart brengen van sociale problemen, maar ook voor productieorganisatie en investeringsbeslissingen in het bedrijfsleven. Ook in de wetenschappen drongen kwantificatie en statistiek door. Binnen medicijnen gaven in 1940 statistische analyses van het succes van behandelmethodes en medicijnen de doorslag, iets wat medici van rond 1850 zich nooit hadden kunnen voorstellen. Deze publicaties verklaren het maatschappijbrede succes van deze nieuwe habitus.

Irma Moi-Reci: *Unemployed and scarred for life; Longitudinal analysis of how employment and policy changes affect re-employment careers and wages in the Netherlands, 1980-2000.*
Dissertatie Vrije Universiteit Amsterdam, onder andere gebaseerd op de data van het 'Arbeidsaanbodpanel 1985-2000' van de Organisatie voor Strategisch Arbeidsmarktonderzoek OSA (DANS databestand – Persistent Identifier: urn:nbn:nl:ui:13-4js-jl3)

ABC-DE: Woordenboek voor het digitaal erfgoed; Digitaal Erfgoed Nederland, december 2008
Publicatie met terminologie voor erfgoedinstellingen die te maken hebben met digitalisering. Online versie: www.den.nl.

Centraal Bureau voor de Statistiek: *Kaarten regionale indelingen 2009*
Als hulpmiddel voor gebruikers van regionale gegevens en indelingen heeft het CBS een viertal kaarten uitgegeven: Provincies 2009, COROP-gebieden 2009, Stadsgewesten en grootstedelijke agglomeraties 2009, en Economisch-geografische gebieden 2009. Op elke kaart staat een regionale indeling met de gemeentelijke indeling als ondergrond. Per regionale indeling is tevens aangegeven welke gemeenten daartoe behoren. De kaarten zijn op papier verkrijgbaar in klein formaat (1:800 000) en in groot formaat (1:400 000), ze kosten respectievelijk € 8,– en € 16,10 excl. verzendkosten. De kaart van stadsgewesten en grootstedelijke agglomeraties is alleen op klein formaat op papier verkrijgbaar. Deze en eerdere kaarten met regionale indelingen zijn ook in pdf-formaat elektronisch beschikbaar. www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/publicaties/archief/2009/2009-regionale-kaarten-pub.htm

DRIVER Guidelines 2.0, guidelines for content providers – exposing textual resources with OAI-PMH. Digital Repository Infrastructure Vision for European Research, November 2008
Guidelines for repository managers and administrators on how to expose digital scientific resources using OAI-PMH and Dublin Core metadata, creating interopearbility by homogenising the repository output.

Sustaining the digital investment: issues and challenges of economically sustainable digital preservation; Interim Report of the Blue Ribbon Task Force on Sustainable Digital Preservation and Access; December 2008.
Een brede Task Force met een sterke vertegenwoordiging uit de wetenschap heeft zich gebogen over de economische aspecten van langetermijnbewaring van digitale gegevens. Wat maakt digitale informatie zo anders dan informatie op papier? Wie neemt verantwoordelijkheid voor langetermijnopslag en wie financiert die? Dit eerste interim report legt de theoretische basis voor meer concrete aanbevelingen die de Task Force eind 2009 zal doen. Het bevat een goede samenvatting van alle rapporten die tot nu toe zijn verschenen over de kosten van digitale duurzaamheid. Het identificeert bovendien de belangrijkste factoren die duurzame oplossingen momenteel in de weg staan, waarvan de belangrijkste wel is dat projectmatige financiering slecht is voor digitale data omdat die ononderbroken zorg vereisen van wieg tot graf.
http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf

Column
Jan Kooistra

Kind reminder

Het verzoek een column te schrijven voor e-data & research bereikte mij pas toen ik het had opgevist uit mijn spambox. Dat het daar terecht kwam was niet zomaar. Ik heb namelijk mijn mailbox dichtgespijkerd. Alles wordt tegengehouden tenzij ik het zelf doorlaat. Baas in eigen box. Als je dat lang genoeg volhoudt vormt het een goede bescherming tegen het virus dat is meegelift met het succes van email: het idee dat wat ons per post wordt bezorgd belangrijk is. Overblijfsel uit de tijd dat we elkaar brieven schreven. De metacommunicatieve elementen die de geschreven boodschap zo kenmerkten (handschrift, vlekken, doorhalingen e.d.) zijn vervangen door manipulatieve functies als cc. en bcc. Er heeft Aufhebung (Hegel) van het schrijven plaatsgevonden. De boodschap is het feit dat het bericht is gestuurd.

In het begin kostte het mij veel tijd. Nu gooi ik de spambox gemiddeld eens per week met grote halen leeg. Ik kijk dan met een geoefend oog naar de kenmerken (code) van de mail, als aan de sorteerband in de fabriek waar ik als jongen vakantiewerk deed. Ook toen ging dat trouwens niet vanzelf. Volgestopt met het onderscheid tussen 'goed' en 'kwaad' vond ik het een hele verantwoordelijkheid om iets van de band te verwijderen. Ik herinner mij uit die tijd een televisie uitzending. Ergens in de Achterhoek speelde een compleet dorp in een toneelstuk waarin het Laatste Oordeel werd verbeeld. En natuurlijk kwam er iemand in de verkeerde rij. Drama, al kwam het goed. Last minute. De hemel is tenslotte waterdicht. Bij mij is dat anders. Er raakt nog wel eens een zieltje onterecht doorgedraaid. Zo miste ik bijvoorbeeld recent de uitnodiging een keynote te houden op een toch wel prestigieuze summer school. Gewoon weggegooid op basis van een onbekend adres en de welluidende vrouwennaam van de afzender. Hoe verkeerd kan men bezig zijn!

Het concept van virtuele kenniscentra zoals we dit aan de Universiteit Utrecht uitwerken (http://partner.library.uu.nl/Pages/default.aspx), draait om de combinatie van openzetten en dichttimmeren. Enerzijds moet men de informatiestromen leren verwerken, anderzijds mag men zichzelf beschermen. Het werkt via een trapsgewijs opgebouwd systeem van portals. De universiteitsbibliotheek is de centrale portal. Vrije toegang tot de wetenschappelijke bronnen vormt de basis van het systeem. Men verkrijgt de toegang als student en behoudt deze na het afstuderen in de vorm van het lidmaatschap van het virtuele kenniscentrum van het eigen vakgebied. Participatie in feitelijke projecten vormt de toegang tot een volgende trede in het systeem. De deelnemer treft aldaar een portal die is gestructureerd als een online digitale werkomgeving met een veelheid aan functionaliteiten. De inrichting van deze collaboraty-repository omgeving vindt bottom-up plaats. Men bouwt met elkaar op locatie het vaartuig om de informatiezee te trotseren. En zo wordt het dichtspijkeren de kunst van te blijven drijven om de benodigde informatie te verzamelen.

En wat betreft de keynote: de mevrouw met de welluidende naam stuurde mij gewoon een kind reminder.

Jan Kooistra is is senior onderzoeker aan de universiteit van Utrecht en gastlector aan de faculteit Architectuur van de Technische Universiteit Delft.