

# Gentse Boekentoren gescand door Google

De bibliotheek van de Universiteit van Gent was onlangs de eerste Nederlandstalige instelling die zich bij het Google Library programma aansloot. Bibliothecaris Sylvia van Peteghem is niet bezorgd over de kwaliteit van scans en metadata.

Zoekgigant Google is al lang niet meer alleen een zoekmachine voor het web. Naast populaire toepassingen als Google Earth en Google Maps werkt het bedrijf ook hard aan het toegankelijk maken van recente wetenschappelijke literatuur in Google Scholar. In het Google Print programma wordt samengewerkt met uitgeverij om hun uitgaven doorzoekbaar te maken. Voor geesteswetenschappers is daarnaast het Google Library programma van groot belang: daarin wordt het boekenbezit van grote bibliotheken gescand en met Optical Character Recognition (OCR) doorzoekbaar gemaakt. Oudere wetenschappelijke literatuur, primaire teksten, maar ook bijvoorbeeld veilingcatalogi, komen zo voor onderzoek ter beschikking.

Met de bibliotheek van de Gentse universiteit is één van de grootste bibliotheken van België nu toegetreden tot het Google Library project. Andere Europese partners zijn de universiteiten van Oxford, Lausanne en Madrid en de nationale bibliotheken van Catalonië (Barcelona) en Beieren (München). De universiteit van Keio (Japan) was onlangs de eerste partner

uit een niet-Europees taalgebied.

Alleen boeken waarop geen auteursrecht meer rust zullen worden gescand. Omdat de universiteit van Gent tot 1930 Franstalig was, is het Frans de belangrijkste taal (47% van de collectie van voor 1900), met Nederlands als goede tweede (22%). Voor de bibliotheek is digi-



JOHAN JACOBZ

De Boekentoren van de Gentse bibliotheek

tale beschikbaarheid van de collectie onder andere belangrijk omdat de hoofdvestiging, de door Henry van der Velde ontworpen Boekentoren, vanaf 2010 wordt gerenoveerd. Volgens hoofdbibliothecaris Sylvia van Peteghem zullen de digitale teksten rechtstreeks vanuit de catalogus toegankelijk worden gemaakt. Aanvankelijk worden alleen boeken uit de Boekentoren gescand; in een later stadium wellicht ook werken uit decentrale bibliotheken.

Het Google Library programma is niet onomstreden. De kwaliteit van de scans, de OCR en de metadata worden niet altijd positief beoordeeld. Van Peteghem heeft op dat punt geen zorgen, verzekert zij *e-data@research*: de Amerikaanse partnerinstellingen waarmee zij heeft gesproken zijn vol lof over Google en noemen zeer lage foutpercentages. De bibliotheek

houdt het recht de belangrijke stukken opnieuw te scannen wanneer een meer dan standaard kwaliteit gewenst is. De metadata zijn afhankelijk van wat de bibliotheek zelf aanbiedt. In eerste instantie streeft Gent naar wat zij noemt een 'no-nonsense catalogus'.

Grootscheepse digitalisatie van bibliotheekbezit wordt ook door andere partijen nagestreefd. De Open Content Alliance werkt samen met het Internet Archive, Microsoft, Yahoo en een aantal bibliotheken aan een gelijksoortig initiatief. De Franse nationale bibliotheek presenteerde onlangs een eerste versie van Europeana, een digitale mega-collectie in aanbouw. Google lijkt echter op het moment het meest aan de weg te timmeren. (PB)

<http://lib.ugent.be/>

## Sinds kort beschikbaar

Onderzoekers kunnen zowel bij DANS als bij het Centrum voor Beleidstatistiek (CvB) van het Centraal Bureau voor de Statistiek (CBS) terecht voor databestanden van het CBS. Bij DANS gaat het om beveiligde microdata die kosteloos geleverd kunnen worden voor secundaire analyses. Ze kunnen besteld worden via [info@dans.knaw.nl](mailto:info@dans.knaw.nl). Aan het gebruik van de onbeveiligde microdata via het CvB is

een aantal voorwaarden verbonden. Informatie over die voorwaarden is te vinden via <http://www.cbs.nl/microdata> -> *Werkwijze*. Het overzicht toont CBS databestanden die recent voor onderzoekers beschikbaar zijn gekomen. Een volledig overzicht van de bestanden met een documentatierapport van de meest recente versies vindt u op <http://www.cbs.nl/microdata> -> *Catalogi*.

	Naam	Periode	Beschikbaar via
SSB	Sociaal Statistisch Bestand Banen	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand Zelfstandigen	2005	CvB
SSB	Sociaal Statistisch Bestand overige arbeid en arbeid buitenland	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand arbeidsongeschiktheidsuitkeringen	2005	CvB
SSB	Sociaal Statistisch Bestand IOAW / IOAZ	2004 en 2005	CvB
SSB	Sociaal Statistisch Bestand overige uitkeringen en uitkeringen buitenland	2004 en 2005	CvB
POLS	Permanent Onderzoek Leefsituatie (basis)	1997 t/m 2005	CvB (1997-2004) DANS (2005)
POLS	Permanent Onderzoek Leefsituatie (modules)	1997 t/m 2005	CvB (1997-2004) DANS (2005)
MKB	Milieukosten bedrijven	1998 t/m 2004	CvB
SFGO	Statistiek Financiën Grote Ondernemingen	2005	CvB
LSO	Loonstructuur Onderzoek	1996 en 2002	CvB
EWL	Enquête Werkgelegenheid en Lonen	2005	CvB
EBB	Enquête Beroepsbevolking	2006	CvB
WW	Werkloosheidswet	2003 en 2006	CvB
WSW	Wet Sociale Werkvoorziening	2006	CvB
SRG	Statistiek Reïntegratie door Gemeenten	2006	CvB
BUS	Bijstandsuitkeringenstatistiek ook op transactiebasis	2006	CvB
BFS-BDS	Bijstandsfraude- en debiteurenstatistiek	2006	CvB
AO	arbeidsongeschiktheidsuitkeringen	2006	CvB
NEA	Nationale Enquête Arbeidsomstandigheden	2005	CvB
NVS	Nationale Verzuimstatistiek	2004 en 2005	CvB
EUSILC	European Community Statistics on Income and Living Conditions	2005 en 2006	CvB (2005) DANS (2006)
DO	Doodsoorzaken	2005	CvB
VMR	Veiligheidsmonitor Rijk	2005 en 2006	CvB
DHS	Household Survey	2005 en 2006	CvB
DNB	Huishoudens en personen	2005 en 2006	DANS
ICT	Huishoudens en personen	2005 en 2006	DANS
SEP	Sociaal Economisch Panelbestand (administratief)	2001 t/m 2003	DANS
CCO	Consumenten Conjunctuur Onderzoek	2006	DANS
B&U	Verleende bouwvergunningen burgerlijke en utiliteitsbouw	2006	DANS

## Liederbank gaat online

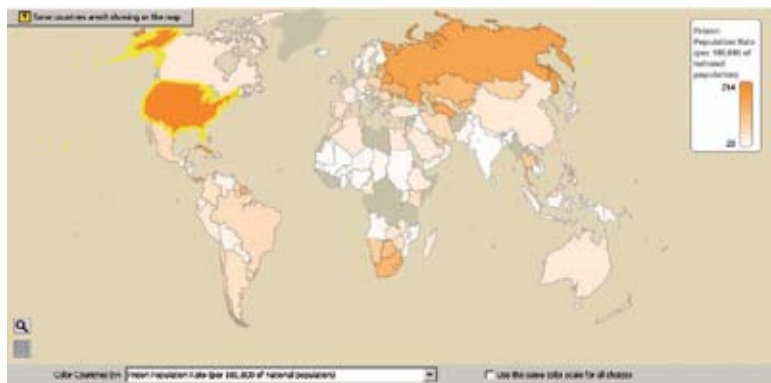
Op 22 juni is de internetversie van de Nederlandse Liederbank officieel geopend. Raadplegen en zoeken in de meer dan 125 duizend liederen van de middeleeuwen tot de twintigste eeuw is nu mogelijk op [www.liederbank.nl](http://www.liederbank.nl). Van elk lied worden naast de beginregel en de titel ook de wijsaanduiding vermeld (waarmee de melodie kan worden gevonden), het refrein, de strofenvorm en trefwoorden. Van enkele tienduizenden

liederen is de tekst beschikbaar; in veel gevallen ook originele audio-opnamen. De Liederbank wordt geproduceerd in het Documentatie- en Onderzoekscenrum van het Nederlandse lied op het Meertens Instituut. Zoeken op hoorbare melodieën is het oogmerk van het nu lopende project WITCHCRAFT, kondigde de geestelijk vader van de Liederbank prof. Louis Grijp aan in een interview in *e-data&research* nr 2-1.

## Data in beeld brengen 2.0

De website Many Eyes is een experiment van IBM op het gebied van visualisatie van data. Gebruikers kunnen kwantitatieve gegevens uploaden en kiezen voor een weergave als staafdiagram, bubble chart, tree map of, als dat toepasselijk is, in de vorm van een kaart met gekleurde gebieden of netwerkdiagram. Vaak bieden de visualisaties ook mogelijkheden om verder in te zoomen op

de data ('beperk de gegevens op het niveau van de VS tot enkele staten') of de weergave te manipuleren ('sorteer uitgaven op categorie van uitgave in plaats van op staat'). Het idee achter de site is dat deelnemers met elkaar discussiëren over inzichtgevende vormen van visualisatie. De figuur toont een van de vele visualisaties: het aantal gevangenen per 100.000 inwoners in de wereld. (PB)



### Leren digitaliseren

Opleidinginstituut GO organiseert vanaf november een zevendaagse cursus 'Digitalisering van culturele collecties'. De cursus gaat over het opstellen van projectplannen en over alle aspecten van de uitvoering. Er wordt voor een bepaalde 'case' ook een volledig projectplan geschreven. Onder meer wordt aandacht besteed aan terminologie, metadata, standaarden, auteursrecht, selectiebeleidbeheer en exploitatie, hard- en software en budgettering. GO opleidingen verkreeg voor de cursus medewerking van 'toonaangevende instellingen' als de Koninklijke Bibliotheek, DANS en het Nationaal archief. Informatie: [www.stgo.nl](http://www.stgo.nl).

### VGI innovatieprijs 2007

'Wie zijn de gebruikers van digitaal erfgoed?' Om die vraag draait een studiemiddag van de Vereniging voor Geschiedenis en Informatica (VGI) op 11 oktober in Antwerpen. Er wordt veel geïnvesteerd om de toegankelijkheid van bibliotheken, archieven, archeologisch onderzoek, museumcollecties en monumenten via internet te vergroten. Recente voorbeelden zijn de massale digitalisering van kranten en audiovisueel materiaal. Op deze studiemiddag is ruime aandacht voor de vraag voor wie dat gebeurt: wie zijn de bezoekers van culturele websites en de gebruikers van erfgoed informatie? Ook wordt de VGI innovatieprijs 2007 uitgereikt voor de meest vernieuwende ICT-toepassing op het terrein van de geschiedenis. Eerst worden drie genomineerde projecten gepresenteerd in het Rubenianum, Kolveniersstraat 20, aanvang 13.00 uur. Aanmelden: [berk@kitlv.nl](mailto:berk@kitlv.nl)

### Plasterk krijgt advies over archieven

Minister Plasterk van Onderwijs, Cultuur en Wetenschap krijgt dezer dagen advies van twee adviesraden over het beheer van overheidsarchieven. Hij vroeg hen in juni om zich daarover uit te spreken, als vervolg op de nota 'Informatie op Orde' van de bewindslieden Van der Laan en Pechtold, die door het kabinet een jaar eerder aan de Tweede Kamer werd aangeboden. De raad voor Cultuur en de Raad voor het Openbaar Bestuur moeten vooral adviseren over de verdeling verantwoordelijkheden, het toezicht, de kwaliteitsborging, archiefselectie en de uitvoerbaarheid van regels en procedures in het digitale tijdperk. Plasterk vroeg beide raden om voor 1 september te antwoorden.

### Onderwijsclassificaties in databank

Het Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA) van de universiteit van Maastricht heeft een databank ontwikkeld waarin verschillende soorten onderwijsclassificaties zijn ondergebracht. In Nederland zijn verschillende systemen gangbaar voor zulke classificaties. Die worden met behulp van de databank eenvoudig vergelijkbaar. Het project, genaamd 'Koppeling tussen opleidingsclassificaties', werd financieel gesteund door data-instituut DANS. Het is mogelijk in de classificaties te zoeken, conversieschema's te genereren en labelsets te maken.