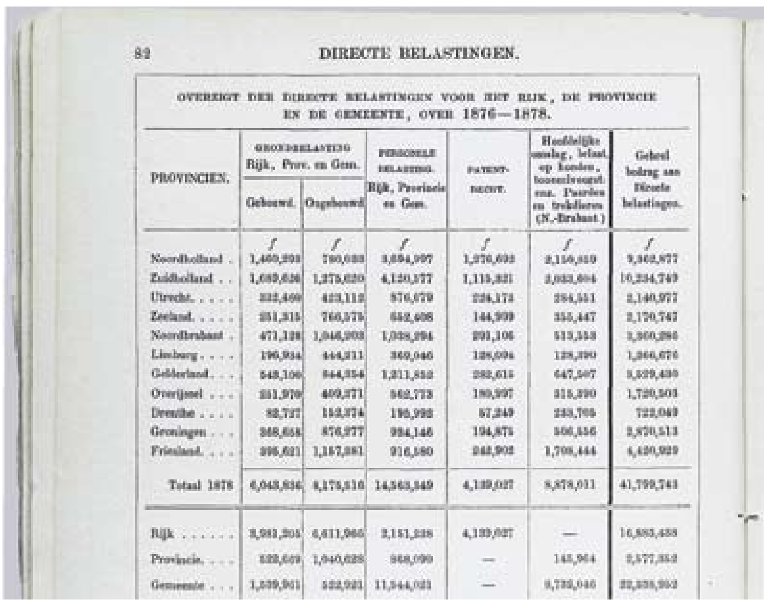
CBS bibliotheek digitaliseert collectie

De bibliotheek van het Centraal Bureau voor de Statistiek (CBS) voert een digitaliseringsproject uit om grote delen van collectie van het bureau via het web toegankelijk te maken. Ook veel kwetsbare en moeilijk te raadplegen statistische publicaties uit de 19e eeuw worden opgenomen. De geplande omvang van het project bedraagt een half miljoen pagina's.

Eerder werden in samenwerking met het opgeheven NIWI de Nederlandse Volkstellingen tussen 1795 en 1971 gedigitaliseerd (www.volkstelling.nl). Daarop komt nu een vervolg door de digitalisering, in samenwerking met DANS, van een aantal bedrijfstellingen (1930-1978) en gedetailleerde gegevens op wijkniveau uit de volkstellingen, woningtellingen en beroepstellingen (1930-1960). Daarnaast worden in het ambitieuze project andere veel gebruikte CBS publicaties uit de collectie gedigitaliseerd.

Tot de al gescande publicaties horen de *Jaarcijfers van het Koninkrijk der Nederlanden* (1881-1967) en het *Statistisch Zakboek* (1899-1988), en cijfers die gebruikt werden voor het opstellen van de Nationale Rekeningen (sinds de jaren vijftig). De eerste twee geven een goed statistisch overzicht van vele onderwerpsgebieden. Die cijfers zullen worden aangevuld door een reeks van historische publicaties uit de 19e eeuw, waaronder *Lobatto's jaarboekje* (1826-1849) en zijn opvolgers onder titels als *Statistisch jaarboek voor het Koninkrijk der Nederlanden*. De bedrijfstellingen geven naast de volkstellingen een goed beeld van de economische activiteit in Nederland. Op deze manier ontstaat een compleet beeld van de 19e en 20e eeuw in cijfers.



82 DIRECTE BELASTINGEN.

OVERZIGT DER DIRECTE BELASTINGEN VOOR HET RIJK, DE PROVINCIE EN DE GEMEENTE, OVER 1876—1878.

PROVINCIEN.	Grondbelasting Rijk, Prov. en Gem.		Personele belasting Rijk, Provincie en Gem.	Patentrecht.	Hoofdelijke omslag, belast. op honden, tooneelvoorst. enz. Paarden en trekdieren (N.-Brabant)	Geheel bedrag aan directe belastingen.
	Gebouwd.	Ongebouwd.				
	f	f	f	f	f	f
Noordholland .	1,460,293	780,033	3,694,997	1,276,693	2,156,859	9,362,877
Zuidholland . .	1,089,636	1,275,620	4,120,377	1,115,821	2,033,094	10,234,749
Utrecht. . . .	232,469	423,112	876,679	228,173	284,551	2,140,977
Zeeland. . . .	251,315	760,575	652,408	144,909	355,447	2,170,747
Noordbrabant .	471,128	1,046,203	1,038,294	291,106	513,353	3,360,286
Limburg	196,934	444,211	360,046	128,094	128,390	1,266,676
Gelderland. . .	548,106	844,354	1,211,852	282,615	647,507	3,529,430
Overijssel . . .	251,970	409,371	562,773	180,997	315,390	1,720,503
Drenthe	82,727	152,374	190,992	67,249	233,705	722,049
Groningen . . .	368,658	876,277	994,146	194,875	506,556	2,970,513
Friesland. . . .	396,621	1,157,281	916,580	248,902	1,709,444	4,420,929
Totaal 1878	6,048,836	8,175,516	14,563,349	4,139,027	8,878,011	41,799,743
Rijk	3,981,205	6,611,966	2,151,228	4,139,027	—	16,883,455
Provincie. . . .	528,669	1,040,028	868,090	—	145,964	2,577,352
Gemeente . . .	1,539,961	522,921	11,544,021	—	8,732,046	22,338,952

Tabel uit 'Jaarcijfers omtrent bevolking, landbouw, handel, belastingen, onderwijs enz.' van de Vereniging voor Statistiek in Nederland, 1882

Net als bij het eerste Volkstellingenproject doet zich bij dit materiaal de vraag voor hoe deze data gedigitaliseerd en toegankelijk gemaakt kunnen worden. Cijfers moeten betrouwbaar zijn, maar bij tekstherkenning treden altijd onnauwkeurigheden op. De bedrijfs- en volkstellingen die nu gedigitaliseerd worden, bestaan voor een groot deel uit handgeschreven cijfers die in een voorgedrukte tabel zijn ingevuld. OCR van dit materiaal is mogelijk, maar controle op de resultaten brengt de kosten sterk omhoog. Daarom is uiteindelijk besloten tot complete data-entry van deze publicaties. Dit materiaal zal in Excel formaat en als afbeelding beschikbaar gesteld worden. Gelet op de ervaringen van gebruikers met *volkstellingen.nl* zal de raadpleging en gebruik van de data gebruikersvriendelijker worden. De tekst van de Excelbestanden wordt volledig doorzoekbaar via één interface en de gemeentenamen worden direct gekoppeld aan een historische standaard gemeentecode, de zogenaamde 'Amsterdamse code'. Ook manipulatie van de images (zoals roteren en inzoomen) zal mogelijk zijn. Begin volgend jaar zullen de eerste data te raadplegen zijn via www.cbs.nl. (Henk Laloli)

Friese bestanden naar DANS

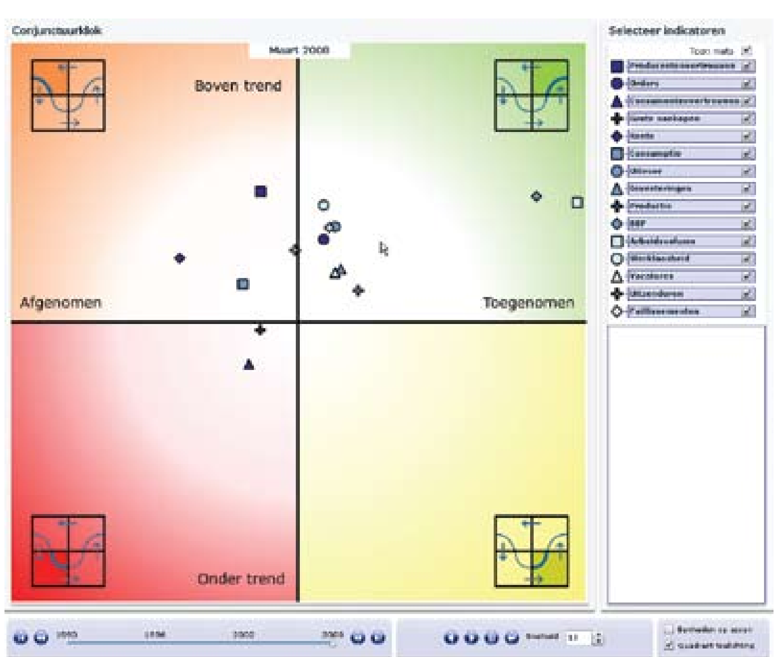
De Fryske Akademy is sinds begin dit jaar bezig een groot aantal bestanden uit taalsociologisch onderzoek bij data-instituut DANS onder te brengen. Een zevental is daar al beschikbaar. Nog eens vijftien volgen binnenkort, 'als de belangrijkste onderdelen die voor het begrip nodig zijn in het Nederlands zijn vertaald' aldus documentalist-historicus Peter van der Meer van het KNAW-instituut. Het gaat om resultaten uit eigen onderzoek dat de Akademy sinds de jaren tachtig uitvoerde naar verschillen in het gebruik van de Friese taal en de oorzaken daarvan: 'Wie spreekt wat tegen wie en wanneer?'.

'Wij kwamen tot de conclusie dat het voor de toegankelijkheid van de gegevens beter is om ze via een gespecialiseerd data-archief aan te bieden', licht Van der Meer toe. 'Er zullen dan meer mensen gebruik van maken dan wanneer wij materiaal het alleen op onze eigen website aanbieden'. (MdG)

Conjunctuurklok tikt

De Conjunctuurklok is eind 2005 door het Centraal Bureau voor de Statistiek (CBS) geïntroduceerd en heeft zich in korte tijd een centrale plaats verworven in de conjunctuurberichtgeving van Nederland. Van vijftien indicatoren wordt elke maand in één plaatje en in onderlinge samenhang getoond of het 'goed' of 'slecht' gaat (onder of boven de trendmatige ontwikkeling over een lange periode) en of het 'beter' of 'slechter' gaat (stijging of daling ten opzichte van de voorgaande periode). Door de foto's van alle maanden achter elkaar te laten zien wordt een film getoond van het verloop van de conjunctuur. Op de CBS-website kan dat interactief worden gedaan voor de periode vanaf 1990 tot heden.

Een belangrijk kenmerk van de Conjunctuurklok is dat het gaat om real-time gegevens. De klok laat zien hoe men op een bepaald moment met de toen beschikbare gegevens over de conjunctuur dacht. Als door bijstellingen achteraf een bepaalde indicator een andere waarde heeft gekregen dan is dat in de klok zichtbaar in de maand dat deze nieuwe informatie beschikbaar kwam en dus niet in de periode waarop de bijstelling betrekking heeft. (Ben van Cleef)



CATCH 2008:

Aandacht voor databronnen

Een betere interactie tussen databronnen is een van de drie hoofdlijnen in de nieuwe ronde van het subsidieprogramma CATCH (Continuous Access To Cultural Heritage). Binnen die hoofdlijn geeft het programma als belangrijke onderzoekrichtingen onder meer hergebruik en ontdubbeling van data, kwaliteit van data, evolutie en diversiteit van data.

Bij de betere interactie of 'interoperabiliteit' moet worden gedacht aan het bouwen van systemen waardoor grote databronnen 'gemakkelijker met elkaar kunnen praten zonder zelf te hoeven veranderen', aldus woordvoerster Marjolein Schlarmann van de Nederlandse organisatie voor Wetenschappelijk Onderzoek (NWO), die het programma beheert.

Vanuit CATCH, dat in 2005 werd gelanceerd, wordt tot nu toe een tiental projecten gefinancierd. Daarin ligt het accent vooral op samenwerking tussen erfgoedinstellingen en onderzoekers in de informatica. Die formule is volgens de website van NWO heel geslaagd maar 'nu wordt het tijd om een volgende impuls te geven aan de interactie tussen onderzoek en erfgoedinstellingen door een verbreding naar de geesteswetenschappen'. Daar zijn onderzoekers vaak afhankelijk van erfgoedcollecties terwijl in erfgoedinstellingen veel collectiebeheerders werken met een achtergrond in de geesteswetenschappen. Voor het toegankelijk maken van de collecties blijven intussen informatici nodig en dat maakt de opzet van deze nieuwe ronde 'tamelijk uniek', volgens Schlarmann: 'Het gaat er echt om de verschillende disciplines beter met elkaar te laten samenwerken'.

Multidisciplinaire teams met erfgoedbeheerders, informatici en beoefenaars van de geesteswetenschappen worden aangemoedigd zich te melden met voorstellen waarin een of meer van drie hoofdlijnen herkenbaar zijn: naast de interactie tussen databronnen of de 'interoperabiliteit van grootschalige en gedistribueerde bronnen' zijn dat 'visualisatie en simulatie' en 'historische dimensies en moderne interpretaties'.

Er is drie miljoen euro beschikbaar voor deze ronde, die wordt gefinancierd door de NWO-gebieden Exacte wetenschappen en Geesteswetenschappen samen met het ministerie van OCW. Het programma gaat er vanuit dat met elke toekenning een bedrag van 500 tot 640 duizend euro is gemoeid en houdt rekening met een viertal aanvragen dat kan worden gehonoreerd. Dat is ruim berekend, geeft Schlarmann desgevraagd toe, maar 'er is ook wat speelruimte nodig als smeerolie, juist omdat het hier een nieuw soort samenwerking tussen disciplines betreft'. Vooraanmeldingen voor de subsidieronde moeten uiterlijk 28 juni binnen zijn bij NWO. (MdG)

www.nwo.nl/catch

3TU Datacentrum

De technische wetenschappen krijgen een eigen datacentrum voor digitale datasets. Het zogenaamde 3TU.Datacentrum wordt opgezet door de bibliotheken van de universiteiten van Delft, Eindhoven en Twente. Voor de alfa- en gammadata wordt al in duurzame opslag voorzien door DANS (Data Archiving and Networked Services). Het initiatief van de drie technische universiteiten zorgt ervoor dat die functie voortaan ook voor technische en bètawetenschappen beschikbaar is. Hun datacentrum bouwt voort op de ervaringen die de drie al opdeden met het DARELUX-project, waarin werd geëxperimenteerd met de duurzame opslag van hydrologische meetgegevens in een Luxemburgs stroomgebied.

Basisvoorziening geo-informatie

Binnen vier jaar moet Nederland een goed werkende, duurzame en intensief gebruikte basisvoorziening geo-informatie hebben. Dat staat in de nota GIDEON van de stichting Geonovum en het programmabureau Ruimte voor Geo-informatie. De nota werd geschreven in opdracht van het GI-beraad, een interdepartementaal overleg van ambtenaren, om een visie te geven op de verdere ontwikkeling van de infrastructuur voor een geo-basisvoorziening. De bedoeling is dat de minister van VROM de nota aanbiedt aan de Ministerraad en vervolgens aan de Tweede Kamer. Daarbij zal worden voorgesteld om jaarlijks aan de Tweede Kamer verslag uit te brengen over de voortgang.

Geografisch zoeken naar data-sets archeologie

Sinds maart van dit jaar kunnen archeologische datasets in het archiveringssysteem EASY van DANS worden gezocht met behulp van Google maps en Google Earth. Projecten waarvan de locatie bekend is kunnen nu op een kaart of satellietfoto worden gezocht en geraadpleegd.
Archeologische gegevens zoeken via een geografisch informatiesysteem (GIS) bestaat al langer, bijvoorbeeld bij zie besloten Archis en het openbare KICH (Kennisinfrastructuur Cultuurhistorie). Geografisch zoeken was echter nog geen onderdeel van EASY, dat onderdak biedt aan het e-Depot van de Nederlandse Archeologie (EDNA). Het gaat hier om een experiment. Op termijn wordt besloten of en hoe die wordt voortgezet.