

# NWO: 'Infrastructuur data hoort bij strategie 2007-2010'

Verbetering van de data-infrastructuur voor de alfa- en gammawetenschappen is integraal onderdeel van de strategie voor de komende jaren, die NWO in mei heeft gepresenteerd. Dat bevestigt een woordvoester van de organisatie.

'Wetenschap gewaardeerd!' was het motto van de strategie voor de periode tot en met 2010, die de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) op 22 mei aan minister van der Hoeven van OCW heeft aangeboden. De sterke positie die de Nederlandse wetenschap in de wereld heeft opgebouwd dreigt verloren te gaan als er geen forse extra investeringen worden gedaan, staat in de nota. Daartoe is volgens de organisatie een jaarlijks extra bedrag van

433 miljoen euro nodig, deels te gebruiken voor het voortzetten van de huidige incidentele financiering van toponderzoek (150 miljoen), deels om een 'sprong voorwaarts' te kunnen maken (283 miljoen).

Om de koppositie van de Nederlandse wetenschap te bereiken en verder te verstevigen heeft NWO in de strategie drie actielijnen uitgezet: 'ruimte voor onderzoekers', 'bundeling van krachten' en 'wetenschap voor de samenleving'. Het hoofd Voorlichting van NWO, Dominique de Vet, ziet verbetering van de data-infrastructuur vooral als onderdeel van de tweede actielijn: 'Die bundeling van krachten is op allerlei terreinen nodig. De data-organisatie DANS is daarvan een typisch voorbeeld. Ten eerste omdat DANS zelf een samenwerkingsproduct is van NWO met de Koninklijke Nederlandse Akademie van Wetenschappen (KNAW) en ten tweede omdat de opdracht van DANS, het verbeteren van de data-infrastructuur, een facilititeit is om de samenwerking tussen onderzoekers en disciplines te verbeteren'. Maar ook de eerste actielijn van de NWO-strategie, 'Ruimte voor onderzoekers', biedt aanknopingspunten als het om de data-infrastructuur gaat, verzekert De Vet: 'Daarbij gaat het erom, onderzoekers zo goed mogelijk te faciliteren bij het ontwerpen en uitvoeren van hun onderzoek.

De mogelijkheid om databestanden te kunnen vinden en hergebruiken hoort daar bij'.

DANS dong in december vorig jaar met een project om de data-infrastructuur te verbeteren vergeefs mee naar een eenmalige verdeling van honderd miljoen euro voor grootschalige onderzoekfaciliteiten.

Dat geld werd toen aan andere projecten gegund. Het bedrag – en nog iets meer – zou echter volgens de NWO-strategie elk jaar voor dat doel moeten worden besteed. 125 miljoen voor het ontwikkelen van grootschalige onderzoekfaciliteiten zijn onderdeel de het door NWO bepleite extra inspanning van 433 miljoen. (MdG)

## 'Betere toegang tot data door samenwerking DANS en CBS'

Er zijn goede mogelijkheden om door samenwerking tussen het data-instituut DANS en het CBS de toegang voor onderzoekers tot microdata van het CBS te verbeteren. Dat zegt drs. F.A. Hoeve, hoofd van het Centrum Beleidsstatistiek van het CBS.

Met zijn visie sluit Hoeve aan bij de uitlatingen van directeur Gosse van der Veen van het statistiekbureau. Die sprak bij de opening van DANS op 1 juni over 'een nieuw data-arrangement' tussen beide instituten, waarvoor voldoende ideeën aanwezig zouden zijn.

'Er liggen hier bijvoorbeeld historische datasets met microdata die door onderzoekers gebruikt kunnen worden voor tijdreeksen,' licht Hoeve toe. 'Het gaat onder andere om vele jaargangen van de Arbeidskrachtentellingen en om edities van het Woningbehoefteonderzoek. Maar om dat mogelijk te maken moeten ze geschikt gemaakt worden voor extern gebruik. Onderzoekers moeten van de variabelen en hun gecodeerde waarden weten wat ze betekenen en wat je er wel en niet mee kan doen'. Dat kan alleen met overzichtelijke hoeveelheden data voor een bepaald doel. 'Je moet daarvoor projecten definiëren en prioriteiten vaststellen. DANS lijkt ons daarvoor de uitgelezen partner'.

Daarnaast moet er veel gebeuren om de systemen waarmee data worden gedocumenteerd uniform te krijgen. Het CBS gebruikt nu een ander systeem voor deze metadata dan de universiteiten of DANS. Ook hier ziet het CBS kansen voor een vruchtbare samenwerking met het jonge data-instituut. Aan de sterke roep vanuit de academische wereld om de gegevens uit het Sociaal-Statistisch Bestand op

microniveau beschikbaar te maken wil het CBS ook graag tegemoet komen, aldus Hoeve. 'Maar het gaat daarbij om vele miljoenen records. Die kan je niet in één keer allemaal ontsluiten. De kunst zal zijn om een aantal bestanden te maken die beantwoorden aan de wensen van zo veel mogelijk onderzoekers. DANS zou kunnen optreden als vertolker van die wensen'.

Tenslotte werkt het bureau aan een *remote access facility* om onderzoekers op afstand gebruik te laten maken van haar data zonder dat die zelf worden overgedragen. Dit is een vervolg op het al bestaande onsite werken bij het CBS. Voor dit nog vrij nieuwe proces zijn speciale software en andere hulpmiddelen nodig en daarmee gaat het CBS op beperkte schaal beginnen. DANS zou kunnen helpen om vast te stellen welke onderzoeksinstituten het eerst voor aansluiting in aanmerking komen.

Op het gebied van de kosten voor het beschikbaar stellen van data zijn geen nieuwe ontwikkelingen te verwachten, zegt Hoeve desgevraagd. 'Wij geven onze data al enige tijd kosteloos af waar het om het materiaal zelf gaat; sinds kort ook de data die vroeger via het Sociaal Statistisch Agentschap werden geleverd. We rekenen wel kosten voor de faciliteiten om ze te prepareren, maar dat is reeel. Die kosten moeten nu eenmaal gemaakt worden'. (MdG)

### Eerste e goed ontvangen



De verschijningsdatum van de eerste e-data&research, 1 juni, viel samen met de officiële start van DANS. Minister Maria van der Hoeven verrichtte dan ook in Den Haag niet alleen de officiële openingshandeling voor het data-instituut, maar nam ook het eerste exemplaar van het 'kwartaalblad over data en onderzoek' in ontvangst. De bewindsvrouw toonde zich ingenomen met het blad maar niet verrast. Dat kon ook niet, want ze was met een interview prominent aanwezig in dit openingsnummer.

### Taakgroep voor infrastructuur

Een taakgroep met vertegenwoordigers van NWO, de KNAW, het CBS en het Sociaal-Cultureel Planbureau werkt sinds het voorjaar aan een landelijk actieplan voor een betere informatie-infrastructuur in het sociaal-wetenschappelijk onderzoek. De groep komt deze maand met een voorstel.

De taakgroep werkt vanuit DANS onder de naam AGENDA: 'Actieplan Gezamenlijke Effectiviteitsslag Nationale Dataverzamelaars Activiteiten'. Het is de bedoeling nationale partijen op het gebied van onderzoek en van overheidszijde in het proces te betrekken: universiteiten, SURF, onderzoeksscholen en instituten uit de eerste groep en RIVM, KNMI, TNO en anderen uit de tweede.

In het actieplan wordt uitgewerkt hoe het tot coördinatie en samenwerking kan komen op het gebied van onder meer het verzamelen, het beheer, de verspreiding en de archivering van data, het hanteren van kwaliteitsstandaarden, technische standaarden, metadata en procedures, onderzoek, methodenontwikkeling, onderwijs en scholing. (MdG)

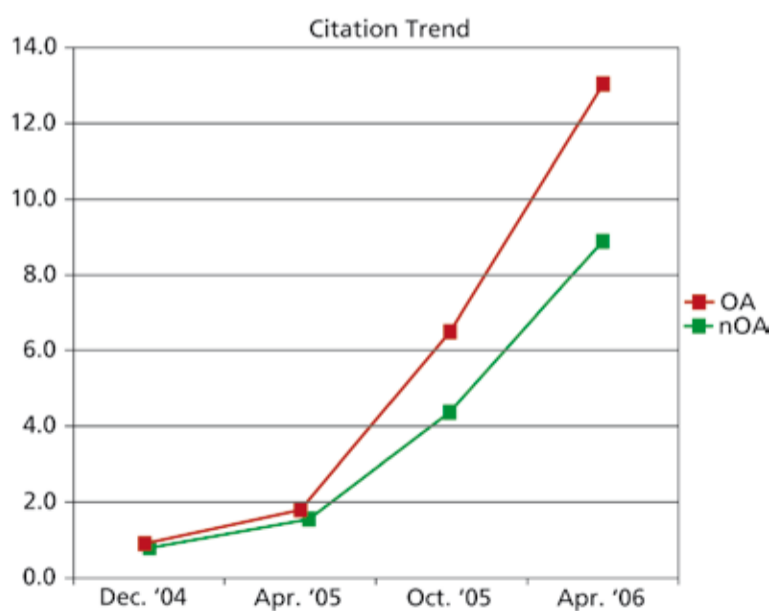
### CBS werkt aan remote access

Een nieuwe techniek om microdata van het CBS van een afstand te kunnen benaderen is vrijwel praktijkrijp. Deze z.g. *remote access facility* biedt software die op de pc van een onderzoeker buiten het CBS wordt geïnstalleerd. Daarmee kunnen via een beveiligde verbinding gegevens worden bekeken en geanalyseerd, maar niet gekopieerd of geprint. De onderzoeker krijgt de resultaten op zijn beeldscherm maar kan de definitieve print alleen via het CBS krijgen. De beperkingen zijn nodig om de privacy te beschermen. Inloggen vereist ook nog een apparaatje om de vingerafdruk van de onderzoeker te checken.

### 100 miljoen onderweg

De honderd miljoen euro die vorig jaar december door NWO aan een vijftal projecten werden 'toegekend' zijn intussen daadwerkelijk onderweg naar de indieners. Dat werd half juli bekend. De toekenning door NWO was in feite een aanbeveling aan minister Van der Hoeven om het geld op de gewenste wijze te besteden. De bewindsvrouw heeft nu dus die aanbeveling overgenomen en uitgevoerd. Het ging om projecten voor het digitaliseren van kranten (13 miljoen), het opzetten van een e-science infrastructuur (29 miljoen), het verzamelen van data voor sociaalwetenschappelijk onderzoek (14 miljoen), voor materialenonderzoek (26 miljoen) en voor systemen om afbeeldingen van de hersenen te maken (19 miljoen).

## Open access meer geciteerd



### Gemiddeld aantal citaties voor OA en niet-OA publicaties in Proceedings of the National Academy of Sciences, vanaf tweede helft 2004.

Voorstanders van Open Access (OA: onmiddellijke kostenloze en universele online toegankelijkheid na publicatie van wetenschappelijke artikelen) beweren sinds lang dat OA artikelen meer worden geciteerd. Tegenstanders schreven dat effect toe aan andere oorzaken. In een recente studie van Gunther Eysenbach (in PLOS Biology) wordt gecontroleerd

voor een serie andere variabelen, zoals aantal auteurs, aantal publicaties van de auteurs, land van herkomst, discipline, etc. De studie bevestigt het vermoeden van de voorstanders: OA artikelen worden tweemaal zo vaak geciteerd als niet-OA artikelen. Voor wetenschappers lijkt het dus verstandig te publiceren op basis van Open Access.



Het CBS-gebouw in Voorburg