

Focus

Organisatie voor Strategisch Arbeidsmarktonderzoek

LUUK SCHREVEN

Voortaan zal e-data&research in ieder nummer focussen op een onderzoeksinstelling binnen haar redactionele horizon. Met een kort interview wordt de instelling bij de lezers geïntroduceerd. Voor deze eerste aflevering spraken we met prof. dr. Peter Ester, directeur van de in Tilburg gevestigde Organisatie voor Strategisch Arbeidsmarktonderzoek (OSA).

‘De OSA streeft ernaar een gezaghebbend onderzoeksinstituut te zijn op het terrein van arbeidsmarkt- en personeelsvraagstukken’, zegt Ester. ‘Daartoe doen we vooral veel trendmatig onderzoek naar de arbeidsmarktontwikkeling in opdracht van de nationale overheid, maar ook in een Europees kader’. Data vormen het kloppende hart van de organisatie. Vooral de eigen arbeidsmarktpanelen die sinds de jaren 1980 lopen zijn volgens Ester te beschouwen als de kroonjuwelen van de OSA: ‘Voor ons type longitudinaal onderzoek geldt: hoe langer de tijdreeks, hoe beter het onderzoek. Onze panelen zijn de langst lopende arbeidsmarktpanelen ter wereld. Ze zijn erg kostbaar, maar leveren voor zowel beleid als wetenschap enorm veel informatie op’.

Nationaal is de OSA stevig ingebed in een strategisch kennisnetwerk met de Universiteit van Tilburg, de Universiteit Utrecht en verschillende ministeries. Daarnaast bestaan er strategische samenwerkingsrelaties met bijvoorbeeld het Centraal Bureau voor Statistiek en DANS (Data Archiving and Networked Services). Maar ook internationaal is de organisatie een gewaardeerde partner. Ester: ‘De internationalisering is



MARCEL BEKKEN

Prof. dr. Peter Ester: ‘Een grote coalitie kan de Nederlandse data infrastructuur tot een van de beste ter wereld maken’

de laatste vijf jaar in een enorme versnelling terecht gekomen, vooral doordat de Europese Commissie zich als vragende partij opstelt. Om goed op die vraag te kunnen anticiperen is mede vanuit de OSA de *European Group of Organizational Panel Survey Institutes (EGOPSI)* opgericht. Via dit consortium neemt de OSA deel in verschillende Europese projecten’. Op internationaal gebied is er nogwel ruimte voor de nodige verbetering: ‘Met name door het toetreden van nieuwe lidstaten wordt de behoefte aan betrouwbare vergelijkende data

alleen maar groter. Niet al deze landen hebben een even lange traditie als het gaat om het verzamelen van hoog kwalitatieve data’.

Gevraagd naar zijn visie op de toekomst betoogt Ester dat Nederland zijn wat gemankeerde concurrentiepositie op de internationale markt voor research & development in één klap structureel zou kunnen verbeteren door een ‘grote coalitie’ te smeden voor de verbetering van de data-infrastructuur. Een coalitie tussen het CBS enerzijds en de Nederlandse kennisinstituten anderzijds zou dat moeten zijn: ‘We kunnen de Nederlandse data-infrastructuur tot één van de beste ter wereld maken door de grootschalige administratieve databestanden van het CBS te koppelen aan de periodieke bevolkings-surveys van kennisinstituten als het Sociaal-Cultureel Planbureau, OSA of het demografisch instituut NIDI. De gekoppelde bestanden moeten uiteraard voor wetenschappelijk onderzoek beschikbaar zijn. Zo’n grote coalitie zou uitstekend door DANS gecoördineerd kunnen worden. Hier ligt een terrein braak’.

Datagrid ook voor niet-exacte wetenschappen

RENÉ VAN HORIK

Door de opslag en de rekenkracht van meerdere computers met elkaar te verbinden kunnen bewerkingen sneller uitgevoerd worden en ontstaat er een enorme opslag- en rekencapaciteit. Het principe van het delen van processoren en het gedistribueerd opslaan van data is bekend onder de naam *grid*. Het *datagrid* behelst de koppeling van grote opslagmedia. Rekenintensieve bewerkingen vinden plaats met behulp van *grid-computing*, waarbij snelle en verbindingen met hoge capaciteit de koppeling tussen de computers mogelijk maken.

In Nederland heeft de realisatie van een grid-infrastructuur een belangrijke stimulans gekregen door het zogenaamde BIGGRID project. Dit is een van de initiatieven die vorig jaar in het kader van het nationaal programma grootschalige onderzoeksfaciliteiten van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek NWO van start konden gaan. Het BIGGRID project streeft naar een sterke vergroting van de in Nederland beschikbare opslag- en rekencapaciteit voor het opslaan, archiveren, bewerken, koppelen en uitwisselen van steeds groter wordende wetenschappelijke datastromen. Het project is ingediend door de Stichting Nationale Computerfaciliteiten (NCF), het Nationaal Instituut voor Kernfysica en Hoge-Energiefysica (FOM-NIKHEF) en het Netherlands Bioinformatics Centre (NBIC) en heeft een budget van 29 miljoen Euro.

Het BIGGRID is er niet alleen voor de ‘beta-jongens’ die het initiatief ertoe hebben genomen. Het is voor alle wetenschapsgebieden beschikbaar. In het project is een flinke plaats ingeruimd voor het alfa en gammaonderzoek. Vergeleken met de exacte wetenschappen is de omvang van de wetenschappelijke datasets in de geestes- en sociale wetenschappen wel veel kleiner, maar de heterogeniteit van deze datasets

stelt specifieke eisen aan het datagrid. Het opslaan van gegevens in het datagrid is een belangrijk toekomstig onderdeel van de data-archiveringsstrategie van DANS (Data Archiving and Networked Services), dat heeft zich opgeworpen als een van de participanten uit de niet-beta wereld.

DANS beschikt nu al over de middleware en kennis die nodig is om gegevens over te brengen naar het datagrid. Momenteel experimenteert het met enkele grote dataverzamelingen die tezamen zo’n 15 terabyte omvatten (15000 gigabyte).

Voor de archivering van onderzoeksdata uit de geestes- en sociale wetenschappen biedt het datagrid veel voordelen. Voor de onderzoeker of onderzoeksinstelling wordt het probleem van de lange-termijnarchivering opgelost. De datasets worden robuust en betrouwbaar opgeslagen. Indien er al een online toegang tot de data bestaat kan het datagrid dienen als backup, waarbij overigens de toegankelijkheid van de data voor DANS een voorwaarde is voor opname in het datagrid (mits er geen juridische bezwaren zijn).

Niet alleen de opslagcapaciteit van het GRID is belangrijk. DANS onderzoekt ook de mogelijkheid om, met behulp van de duizendvoudige rekenkracht die het GRID biedt, automatische ontsluiting en koppeling van data te bewerkstelligen. Als dat lukt kan DANS en daarmee het alfa- en gammaonderzoek een vaste inbedding verwerven in de nieuwe Nederlandse GRID infrastructuur.

Naast de beschikbaarheid van de grid-infrastructuur als data-archief zet DANS dus momenteel de eerste stap naar een optimale benutting van de rekencapaciteit van het GRID. De verwachting is dat er in de toekomst steeds meer rekenintensieve data-analyse mogelijk zal worden door middel van *grid computing*. Hierbij valt te denken aan tekstanalyse, statistische bewerkingen of beeldmanipulatie.

VGI: studeren op GIS en repositories

Geografische informatiesystemen en repositories komen aan de orde in de studiedagen die de Vereniging voor Geschiedenis en Informatica de komende maanden organiseert.

Op vrijdag 20 april is er in Leiden de studiemiddag ‘Repositories: hype of pure noodzaak?’

Over de zin en onzin van repositories. Wat zijn de voordelen, de nadelen, de kosten, de baten? Wanneer is een repository echt nodig, wanneer volstaat een Content Management System (CMS). Wat onderscheidt die twee van elkaar en wat maakt een ‘echte’ repository? Voordrachten en discussie met experts van o.a. Koninklijke Bibliotheek, Internationaal Instituut voor Sociale Geschiedenis, Stadsarchief Antwerpen en

DANS. Plaats: Koninklijk Instituut voor Taal-, Land- en Volkenkunde, Reuvensplaats 2, Leiden. Aanvang 14.00 uur.

Geografische informatiesystemen in de historische wetenschap staan centraal in Den Haag op vrijdag 8 juni. Dan wordt er een studiedag georganiseerd over de toepassing van zulke systemen. Tevens wordt de website www.nlgis.nl gelanceerd, waarop individuele gebruikers hun data kunnen visualiseren.

‘s Ochtends is er een hands-on demonstratie van de website, ‘s middags zijn er lezingen. De bijeenkomst duurt de gehele dag.

Voor beide dagen dient men zich aan te melden bij: berk@kitlv.nl (LS)

Arbeidsmarktpanelen

Het OSA-Arbeidsaanbodpanel is een peiling onder de potentiële beroepsbevolking die voor het eerst in 1985 is uitgezet en sinds 1986 elke twee jaar herhaald wordt. Het aantal respondenten bedraagt circa 4.000. In 2006 heeft de twaalfde golf van het panel plaats gevonden.

Het OSA-Arbeidsvraagpanel is een survey waarmee sinds 1989 tweemaaljaarlijks gegevens worden verzameld onder ruim 3.000 werkgevers over arbeidsproductiviteit, innovatie, uitwerking van overheidsregelingen en personeelsbeleid. In het voorjaar van 2007 gaat de tiende golf van dit onderzoek van start.

De arbeidsmarktpanelen van OSA zijn onder privacy beschermende voorwaarden via DANS beschikbaar voor onderzoekers.

COLOFON

e-data&research is het kwartaalblad over data en onderzoek in de alfa- en gammatwetenschappen, verschijnend onder auspiciën van Data Archiving and Networked Services, het Huygens Instituut, het Internationaal Instituut voor Sociale Geschiedenis en de Vereniging voor Geschiedenis en Informatica.

e-data&research is ook een voortzetting van *Historia & Informatica* en van *Data News – Steinmetz Archive Newsletter*. Toezending kosteloos aan relaties van de stakeholders en opverzoek aan studenten

in de alfa- en gammarichtingen. Oplage: 6500. e-data&research is online te raadplegen op www.edata.nl en op de websites van de deelnemende instellingen.

Uitgever: Edita-KNAW, Postbus 19121, 1000 GC, Amsterdam

Redactieadres: Postbus 93067, 2509 AB Den Haag; Anna van Saksenlaan 51, 2593 HW Den Haag; T (070)3494450 F (070)3494451

E edata@dans.knaw.nl

Redactie: Peter Boot, Martijn de Groot (hoofd/eindredacteur a.i.), Marien van der Heijden, Jetske van der Schaaf, Luuk Schreven.

Redactiesecretariaat: Lucas Pasteuning, Jetske van der Schaaf

Aan dit nummer werkten mee:

Karina Van Dalen-Oskam, Martin Everaert, René van Horik, Steven Krauwer, Gert van der Steen, Matthias Vandermaesen, Ingeborg Verheul, Astrid Verheusen, Ellen Willemse, Peter Wittenburg, Joris van Zundert.

Pre-press, productie en vormgeving: Edita-KNAW

Druk: PlantijnCasparie, Almere

ISSN: 1872-0374