

**Vergadering Afnemersraad MiDaS**  
**Dinsdag 17 april 2012 14.30-16.30**  
**Locatie Ministerie van Sociale Zaken en Werkgelegenheid**

Aanwezig:

Paul de Graaf (COS Rotterdam)  
Michiel van den Hauten (Min. SZW)  
Jan van der Linden (EIM)  
Johan Polder (RIVM)  
Ineke Stoop (SCP, voorzitter)  
Ilonca Vaartjes (UMC Utrecht, Julius Centrum)

Ruben Dood (hoofd Centrum voor Beleidsstatistiek)  
Ruurd Schoonhoven (Relatiebeheerder MiDaS, secretaris)

Verhinderd:

Lucy Kok (SEO)  
Arthur van Soest (Netspar)  
Bas ter Weel (CPB)

Agenda:

1. Opening, vaststellen agenda
2. Vooruitblik op ontwikkeling Microdata Services
3. Kwaliteit van bestanden
4. Bewaarbeleid mbt brondata
5. Eerste resultaten klanttevredenheidsonderzoek
6. Mutaties in bestanden en variabelen (SSB, ABR, ...)
7. Technische zaken
8. Samenstelling van de afnemersraad
9. Terugkoppeling gang van zaken bij CvB
  - invoering dienstencatalogus en nieuwe tariefstructuur per 1-1 j.l.
  - stand van zaken rond capaciteit, wachtlijst...
  - reorganisatie, personele mutaties en andere werkwijze
  - 10 jaar CvB dag op 11 oktober 2012
10. Rondvraag en sluiting

1. Opening, vaststellen agenda

De voorzitter heet de aanwezigen welkom. De agenda wordt ongewijzigd vastgesteld. Notulen van de vorige vergadering staan niet op de agenda omdat die al per mail zijn becommentarieerd en goedgekeurd en inmiddels op de website van het Centrum voor Beleidsstatistiek zijn gepubliceerd.

2. Vooruitblik op ontwikkeling Microdata Services

Ruben Dood geeft een presentatie naar aanleiding van de bij de agendastukken verspreide discussienota. Ontwikkeling van een toekomstvisie op Microdata Services is onderdeel van de voorbereiding van het CBS meerjarenprogramma 2014-2018, waarin zowel aandacht zal zijn voor de statistiekproductie als de bereikbaarheid van data. CvB vraagt aanvullend de afnemersraad om commentaar, prioritering van keuzes dan wel ideeën voor het proces om tot keuzes te komen, etc.

De nota is ook in het CBS directiebestuur (DB) besproken. Het DB geeft aan het Centrum ruime bewegingsvrijheid zolang aan enkele voorwaarden is voldaan: 1) elke vorm van dienstverlening dient zichzelf te bedruipen, 2) de strikte voorwaarden ten aanzien van toegang tot de data, doelbinding, criteria voor toelaatbare output en de controle daarop, alsmede de eis van publicatie voor de openbaarheid, blijven onverminderd van kracht.

De leden van de afnemersraad geven eerst hun algemene reactie op het stuk, en daarna worden verschillende mogelijke nieuwe diensten zoals in de nota genoemd maar ook ter vergadering aangevuld één voor één besproken. Deze discussie neemt het grootste deel van de vergadering in beslag.

In algemene zin zijn de leden van de raad verheugd met de onderzoeksmogelijkheden zoals die de afgelopen jaren ontwikkeld zijn. Ook de voorliggende nota met een visie hoe die de komende jaren verder uitgewerkt zou moeten worden kan op instemming rekenen. Meenemen van de internationale dimensie krijgt duidelijk bijval. Alom leven echter grote zorgen omtrent de nog steeds langer wordende wachtlijst; de afnemersraad onderschrijft dan ook de stelling uit de nota dat het aanpakken daarvan de hoogste prioriteit heeft. Ook worden vraagtekens gezet bij de kostendekking van de dienstverlening zoals geformuleerd en vragen diverse leden van de raad om voorzichtigheid bij het nastreven van nieuwe ambities zolang de actuele dienstverlening door krapte in capaciteit nog steeds onder grote druk staat. Nadat de voorzitter heeft gemeld dat de ambitie van het stuk al gematigd is ten opzichte van eerdere versies en 'de vraag leidend is' komt uit de raad ook de reactie dat het niettemin wenselijk is daar niet te passief mee om te gaan, maar oog te houden voor de eigen ambities en initiatieven, bijvoorbeeld te zien als onderdeel van het ontwikkelen van 'Nederland Kennisland'.

#### Bespreking van onderwerpen uit de nota:

##### *1. Voortzetten on site en remote access faciliteit*

Deze vorm dienstverlening staat niet ter discussie, zij vormt de kernactiviteit en dat moet zo blijven.

##### *2. Ontwikkelen mogelijkheden voor Remote Execution*

Het gaat hier om de mogelijkheid om middels meer gestructureerde beschikbaarheid van metadata onderzoekers de mogelijkheid te geven een syntax op te sturen om die te laten draaien en de output retour te ontvangen. Hiervoor lijkt in de vergadering niet veel belangstelling te bestaan, omdat er vrijwel altijd nogal wat interactie met CBS-medewerkers nodig is (debuggen van syntaxen, controle van een output die de onderzoeker niet eerst zelf heeft kunnen zien, etc).

##### *3. Direct Data Access (DDA)*

Hier wordt bedoeld een formule die al voor Min. SZW wordt toegepast om op afroep toegang te krijgen tot een vooraf overeengekomen set bestanden voor vooraf aangemelde onderzoekers. DDA is met name bedoeld voor het snel kunnen maken van ad hoc analyses voor bijvoorbeeld het beantwoorden van kamervragen. Outputcontrole vindt snel plaats maar volgens de standaard richtlijnen.

Hiervoor is geen brede belangstelling, maar het is goed dat deze mogelijkheid er is voor instellingen die aan een dergelijke constructie behoefte hebben.

##### *4. Draaien van eigen software van onderzoekers en hosten van hun modellen in de beveiligde microdata omgeving*

Gedacht wordt aan levensloopmodellen voor analyse en voorspellen van zorgkosten (RIVM) en prognostische modellen van bijv het CPB.

De wens hiervoor bestaat zeker en uitbreiding van de mogelijkheden op dit gebied opent veel nieuwe kansen. Maar er zitten ook veel haken en ogen aan omdat de diversiteit van de vereiste maatwerkoplossingen groot is een zinvolle outputcontrole moeilijk is en er veel technische en organisatorische problemen moeten worden overwonnen.

De afnemersraad is van mening dat het verder verkennen van deze optie zeer de moeite waard is.

##### *5. Rol als nationaal dataknooppunt*

Een complex onderwerp dat raakt aan de discussie over 'open data', controleerbaarheid van data, etc.

De suggestie wordt gegeven dit onderwerp eventueel met KNAW/VSNU op te nemen, als representant van de universitaire onderzoekswereld die immers erg versnipperd is. Ruben Dood benadrukt dat de eerste focus is op (register)bestanden die elders aanwezig zijn maar niet binnen RA beschikbaar: loopt de onderzoeksgemeenschap in de praktijk tegen dat probleem aan? Voorgesteld wordt daarvan eerst

een groslijst op te stellen. Indien bestanden van externe partijen worden betrokken is overigens wenselijk dat die naar onderzoekers een transparant beleid voeren ten aanzien van verlenen van toestemming voor het gebruik van hun data.

#### *5. Rol als internationaal dataknooppunt*

De afnemersraad waardeert nadrukkelijk de betrokkenheid van het CBS bij internationale ontwikkelingen rond de beschikbaarheid van microdata, maar benadrukt dat het belangrijker is eerst de zaken nationaal op orde te krijgen.

#### *6. Beschikbaar maken van public use files*

Het gaat hier om microdata die zodanig ver zijn geanonimiseerd dat ze vrijelijk kunnen worden verspreid voor onderzoeks- of onderwijsdoeleinden.

Voor de gebruikersgemeenschap vertegenwoordigd door deze raad lijkt hiervoor weinig empooi, maar diverse leden van de raad zouden beschikbaarheid van dergelijke bestanden zeer wenselijk vinden; wellicht zelfs als onderdeel van de publieke taak van het CBS.

### Bespreking van onderwerpen aanvullende vanuit de vergadering genoemd:

#### *7. Snuffelen aan data*

Veel onderzoekers zouden ermee geholpen zijn als zij, in de verkenningsfase voordat zij een project starten, zouden kunnen snuffelen aan de data. Doel kan zijn na te gaan of bepaalde koppelingen zinvolle resultaten opleveren, wat de dekkingsgraad van een variabele is, om syntaxen te ontwikkelen (dat laatste is nu wel eens lastig omdat van iedere collega die daarbij hulp verleent wordt verwacht dat die eerst de status 'meekijker' verwerft), of om door uitproberen van verschillende benaderingen van de analyse de vraagstelling scherper te formuleren. Dit zou kunnen gebeuren binnen de beveiligde omgeving door daar van veelgebruikte bestanden een exemplaar neer te zetten voor een kleine populatie of met op zinvolle wijze gerandomiseerde variabelen; of door fictieve bestanden of bestanden met gerandomiseerde variabelen buiten ter beschikking te stellen.

#### *8. Beleid ten aanzien van statistische veiligheid*

Onderzoekers blijven aanlopen tegen de strenge en administratief bewerkelijke gang van zaken rond outputcontroles. Enkele leden van de raad informeren daarom nogmaals naar de mogelijkheid om dit via bijvoorbeeld een bilaterale overeenkomst per onderzoeksinstelling te regelen. Ruben Dood geeft echter aan dat het recent herbevestigde standpunt van het CBS Directiebestuur hier de facto geen ruimte biedt.

#### *9. Ondersteuning van onderzoekers*

Er lijkt meer vraag te zijn naar hulp bij maken van syntaxen, informatie over variabelen (hoe die te gebruiken, hun betrouwbaarheid, etc). Ruurd Schoonhoven meldt dat in het verleden enkele keren het idee van organiseren van een gebruikersgroep op basis van bijvoorbeeld een wiki is geopperd, maar dat daar door capaciteitsgebrek niet eerder iets van gekomen is. In de huidige meerjarenvisie zou dit idee opnieuw een plaats kunnen krijgen.

### 3. Kwaliteit van bestanden

Ilonca Vaartjes heeft een aantal vragen geformuleerd rond de kwaliteit van (register)bestanden en welke controle het CBS daarop doet. Ruurd Schoonhoven heeft, in samenwerking met diverse statistische afdelingen, een notitie gemaakt die op die vraag ingaat voor de betreffende medische, maar ook voor diverse andere veel gebruikte (register)bestanden. Hij zal nagaan hoe deze informatie beter beschikbaar gemaakt kan worden voor onderzoekers.

### 4. Bewaarbeleid mbt brondata

Johan Polder heeft dit punt geagendeerd naar aanleiding van de affaire Stapel. Hij meldt dat bij de UvT de richtlijn is ingevoerd dat data gebruikt voor een proefschrift tot tenminste 5 jaar na de promotiedatum moeten worden bewaard. CvB bewaart brondata in principe 'eeuwig', en projectgegevens 5 jaar na afsluiting (met mogelijkheid tot verlenging). De afnemersraad is van mening

dat CvB in dit opzicht voldoende waarborgen biedt, temeer daar het ook mogelijk is bij beëindigen van een project alle gebruikte syntaxen te exporteren.

#### 5. Eerste resultaten klanttevredenheidsonderzoek

Bij de eerste reeks beëindigde projecten waarbij een evaluatieformulier is toegezonden is de overall respons op de evaluatie opmerkelijk laag. Door nabellen van non-respondenten is een aantal verklaringen daarvoor duidelijk geworden, die hebben geleid tot enkele verbeteringen in de uitvraag. Aan het overall oordeel (rapportcijfer 7) van de tot nu toe beoordeelde projecten kan dan ook niet te veel waarde worden gehecht. Wel zijn enkele probleempunten genoemd waaraan aandacht besteed zal worden.

*De resterende punten worden omwille van de tijd slechts kort besproken:*

#### 6. Mutaties in bestanden en variabelen (SSB, ABR, ...)

Aanleiding vormen recente mutaties in de definitie van bedrijfseenheden en vooral de revisie van het SSB die heeft geleid tot verschillende definities van het baanbegrip; van dat laatste hebben gebruikers veel last. De raad bepleit dat CBS a) bij herziening van statistische processen via het Centrum voor Beleidsstatistiek ook de belangen van externe onderzoekers beter in de overwegingen meeneemt, en b) in dringende gevallen zoveel mogelijk zorgt voor het repareren van breuken dan wel het beschikbaar stellen van koppeltabellen en dergelijke waarmee onderzoekers dat zelf kunnen doen.

#### 7. Technische zaken

Er spelen bij gebruikers en bij CvB intern diverse kwesties (beleid/kosten SAS, geheugencapaciteit Stata; maximale jobtijden; opslagformaat van data (SPSS vs Stata), behoefte aan GIS, SQL, etc). Johan Polder en Jan van der Linden vragen in het bijzonder aandacht voor de gang van zaken rond SAS. Punt komt in volgende vergadering terug.

#### 8. Samenstelling van de afnemersraad

Wordt doorgeschoven naar volgende vergadering.

#### 9. Terugkoppeling gang van zaken bij CvB

Ruben Dood verwijst hier naar de bij de agenda verspreide presentatie, en vermeldt met name dat het 10-jarig bestaan van het Centrum voor Beleidsstatistiek zal worden gevierd op 11 oktober 2012.

#### 10. Rondvraag en sluiting

Er zijn geen punten voor de rondvraag. De voorzitter sluit de vergadering om 16.45 uur.

#### Actiepunten met uitvoerder en deadline:

1. Ruben Dood neemt de reacties op visienota mee naar CBS intern, zowel in algemene zin als op specifieke punten zoals boven beschreven (agendapunt 2).
2. Ruurd Schoonhoven werkt uit hoe achtergrondinformatie over kwaliteit van beschikbaar gestelde data beter voor onderzoekers toegankelijk gemaakt kan worden (agendapunt 3).
3. Ruben Dood beziet de mogelijkheden om het belang van de microdata onderzoekers bij mutaties in het interne statistisch proces beter te borgen (agendapunt 6).