



Trends door meetgegevens keuzes bij een opdrachtverstrekking

Arthur Denneman

Centraal Bureau voor de Statistiek (CBS)

Werkgroep Statistiek Evaluatie Meststoffenwet

Met medewerking van:

Pim Ouwehand (Centraal Bureau voor de Statistiek, CBS)

Dico Fraters (Rijksinstituut voor Volksgezondheid en Milieu, RIVM)

Lan Ge (Wageningen Economic Research, WECR)

Joachim Rozemeijer (Deltares)

Hans Visser (Planbureau voor de Leefomgeving, PBL)

Dennis Walvoort (Wageningen Environmental Research, WENR)

Gefinancierd door het Ministerie van Landbouw, Natuur en Voedselkwaliteit (LNV),
in samenwerking met het Ministerie van Infrastructuur en Waterstaat (IenW).

2 september 2019

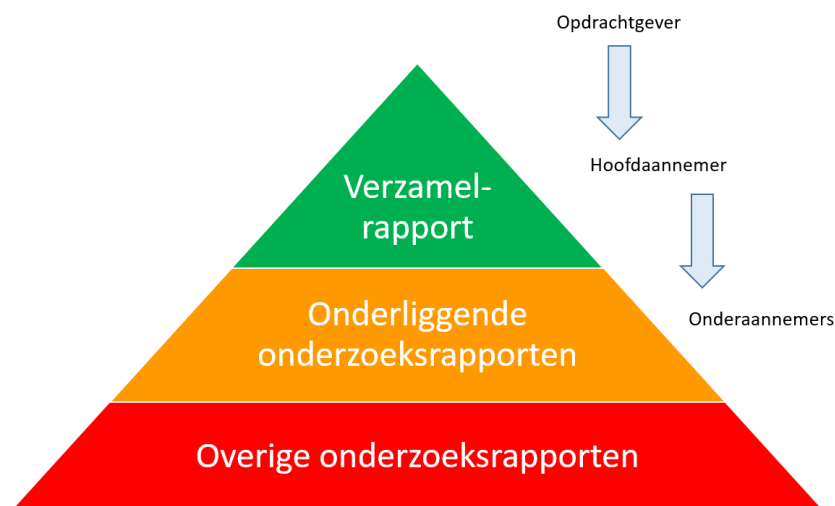
Samenvatting Het bepalen van een trend door meetgegevens is niet triviaal. Het hangt af van welke trendmethodiek gebruikt is. Het hangt ook af van welke databewerkingen toegepast zijn. Elk instituut heeft zijn eigen werkwijze. Daardoor kan het voorkomen dat, op basis van dezelfde set aan ruwe meetgegevens, het ene instituut zegt dat de trend dalend is, het andere zegt dat het stijgend is, en de volgende zegt dat het stabiel is. Tegenstrijdige conclusies, daar worden opdrachtgevers niet blij van. In deze notitie worden een aantal keuzes genoemd waarmee de opdrachtgever, en de hoofd- en onderaannemers ervan, rekening kunnen houden bij een opdrachtverstrekking. Het voorkomen van tegenstrijdige trendanalyse-conclusies kan geminimaliseerd worden als de juiste keuzes gemaakt worden. Er wordt daarbij aanbevolen om een 'werkgroep statistiek' in het leven te roepen. Deze is op incidentele basis oproepbaar en kan betrokken worden bij trendanalyses, als vraagbaak en/of om (on)gevraagd advies te geven. De scope van zo'n werkgroep statistiek beperkt zich niet tot de Evaluatie Meststoffenwet. Het kan ook op andere terreinen ingezet worden. Ook wordt aanbevolen om zo vroeg mogelijk te starten met de afstemming tussen opdrachtgever en hoofdaannemer (verantwoordelijk voor het 'verzamelrapport') en tussen de hoofd- en onderaannemers (verantwoordelijk voor de onderliggende onderzoeksrapporten). Hiermee kan voorkomen worden dat tijdens het stressvolle productieproces nog problemen optreden.

Trefwoorden Meetgegevens, data-voorbewerking, trendmethodiek, data-nabewerking, trendanalyse-conclusie, opdrachtverstrekking, verzamelrapport, werkgroep statistiek.

1. Inleiding

In de Evaluatie Meststoffenwet van 2016 kwam als aanbeveling naar voren dat de trendanalyses extra aandacht behoeven. Er is meer eenduidigheid gewenst en tegenstrijdige conclusies dienen voorkomen te worden. Om dit voor elkaar te krijgen heeft het Ministerie van LNV rond de zomer van 2018 aan het CBS de opdracht gegeven om een werkgroep statistiek op te richten.

De focus van deze werkgroep richtte zich primair op de trendanalyses die gedaan werden binnen de Evaluatie Meststoffenwet. Hierbinnen hebben de rapportages een hiërarchische structuur. Aan de top van de piramide staat een 'Verzamelrapport'. Dat maakt gebruik van gegevens uit allerlei onderzoeksrapporten. Deze onderzoeksrapporten zijn onder te verdelen naar onderliggend (gefinancierd uit hetzelfde budget als het verzamelrapport) en overig (gefinancierd uit een ander budget).



Figuur 1: Een verzamelrapport gebruikt informatie uit onderliggende onderzoeksrapporten die uit hetzelfde budget gefinancierd worden (bijv. Evaluatie Meststoffenwet) en uit overige onderzoeksrapporten die uit andere budgetten gefinancierd worden. Hoe lager in de piramide, hoe minder de opdrachtgever te zeggen heeft over de verstrekking van de opdracht.

Figuur 1 is een versimpelde weergave van de werkelijkheid. Het geneste karakter wordt weggelaten; d.w.z. een onderzoeksrapport kan zelf ook een verzamelrapport zijn (de hele driehoek herhaalt zich dan op een lager niveau). Ook kan een verzamelrapport direct gebruik maken van de overige onderzoeksrapporten; d.w.z. het hoeft niet via de tussenlaag onderliggende rapporten.

De opdrachtgever heeft alleen zeggenschap over het budget dat hij zelf in beheer heeft. Bij een opdrachtverstrekking conform Figuur 1 heeft de opdrachtgever veel invloed in het groene deel van de piramide en geen invloed in het rode deel. In het groene deel gaat het om de keuzes die de opdrachtgever en hoofdaannemer gezamenlijk maken. In het oranje deel worden de keuzes gemaakt door de hoofdaannemer en de onderaannemer(s). Hier heeft de opdrachtgever indirecte invloed. Als het iets gedaan wil krijgen door de onderaannemers, dan dient dit te gebeuren via concrete afspraken met de hoofdaannemer.

Deze notitie zal eerst ingaan op de opdrachtverstrekking door de opdrachtgever aan de hoofdaannemer (het groene deel). Vervolgens wordt gekeken welke invloed dit heeft op de opdrachtverstrekkingen door de hoofdaannemer aan de onderaannemers (het oranje deel). Het rode deel valt buiten de scope.

Bij het maken van de juiste keuzes in de opdrachtverstrekking dient het hele traject in beschouwing genomen te worden; d.w.z. van ruwe meetgegevens naar het te publiceren trendanalyseresultaat. De belangrijkste processtappen staan in Figuur 2.



Figuur 2: Het door te lopen traject: van ruwe meetgegevens naar het te publiceren trendanalyseresultaat. In de te publiceren grafiek is zichtbaar of de trend dalend, stabiel of stijgend is

Zelfs als uitgegaan wordt van een identieke set aan waargenomen ruwe meetgegevens dan nog kunnen de verschillende instituten een andere trendconclusie trekken (dalend, stabiel of stijgend), omdat ze elk hun eigen dataverwerkingstraject hebben doorlopen. Het gaat daarbij niet alleen om de specifieke trendmethodiek die ze toegepast hebben, maar ook om de voor- en nabewerking van de data.

Er kunnen ook verschillen in interpretatie ontstaan als de opdrachtgever en de hoofdaannemer met elkaar een specifieke grafiekweergave in het verzamelrapport hebben afgesproken, die afwijkt van wat er in de onderliggende onderzoeksrapporten staat. Zo kan een grafiek in een onderliggend rapport weergegeven zijn met een y-as in absolute waarden, terwijl het verzamelrapport overstapt naar een index-weergave met 100 in het beginjaar. De indruk die de figuur geeft, kan hierdoor sterk veranderen, zeker als er meerdere lijnen in dezelfde figuur staan.

Er zal niet diep op deze aspecten ingegaan worden. Dit is namelijk al beschreven in een CBS-rapport, van Pim Ouwehand, waarin de focus ligt op de trendanalyses in de Evaluatie Meststoffenwet. Hierin wordt verwezen naar een werkwijzerrapport. Daarmee wordt deze notitie bedoeld.

Het CBS-rapport van Pim Ouwehand is geschreven om gebruikt te worden binnen de Evaluatie Meststoffenwet (EMW). Veel van het beschrevene is generiek toepasbaar. Daarom is deze notitie dusdanig opgezet dat het ook voor andere terreinen toepasbaar is. Deze notitie heeft een adviserende functie. Het is geen voorschrift. Het vergemakkelijkt de opdrachtverstrekking.

In het vervolg van deze notitie wordt Figuur 2 als leidraad gebruikt om de te maken keuzes van opdrachtgever en hoofd- en onderaannemers te verduidelijken. Er wordt begonnen met een keuze-overzicht horend bij de te publiceren trendresultaten om uiteindelijk, tegen de processtroom in, te eindigen bij de waarneming van ruwe meetgegevens. Door tegen de stroom in te gaan wordt eerst het groene deel van de piramide in Figuur 1 behandeld om daarna in het oranje deel uit te komen. Tot slot wordt een overkoepelende beschouwing gegeven, met daarin een aantal aanbevelingen.

2. Te publiceren trendresultaten in het verzamelrapport

Een kenmerk van een verzamelrapport is, dat het een veelheid aan onderwerpen bevat, waarvan vooral de hoofdlijnen beschreven worden. Over het algemeen is de beschrijving van trendanalyses in een verzamelrapport niet complex. Er wordt veel waarde gehecht aan transparantie en heldere communicatie. Anderzijds zijn de onderliggende onderzoeksrapporten veelal gedetailleerd, over één of enkele specifieke onderwerpen, waarbij de beschreven trendanalyse wel complex mag zijn.

De eerste keuzes die de opdrachtgever en hoofdaannemer dienen te maken hebben te maken met de wijze waarop de trendresultaten, horend bij een veelheid van onderwerpen, in het verzamelrapport gepresenteerd gaan worden. Wordt een grafiek letterlijk, via copy-paste, uit een onderliggende onderzoeksrapport overgenomen, of wordt gekozen voor een uniforme presentatie (elke grafiek ziet er hetzelfde uit)?

Keuze 1 **Uniforme presentatie van trendresultaten, of niet?**

Het niet uniform presenteren is een makkelijk uitvoerbare optie die weinig kosten met zich meebrengt. Copy-paste uit onderliggende onderzoeksrapporten volstaat en er zijn daardoor geen aanvullende afspraken nodig. Het nadeel hiervan is, dat het rommelig kan gaan ogen (telkens andere lay-outs en andere trendweergaven) en dat het tot verwarring kan leiden (telkens andere verslagjaren, andere aggregaten, en andere statistische variabelen). Enige uniformiteit lijkt wenselijk. Dit kan leiden tot de volgende keuzes:

Keuze 2 **Uniformiteit-keuzes; presentatievorm**

Keuze 2a *Welke uniforme lay-out? Welke trendweergaven?*

Keuze 2b *Welke publicatie-tool wordt gebruikt? De door het PBL ontwikkelde en in het CBS-rapport genoemde Trendspotter, het veel door waterschappen gebruikte Trendanalist, of het door Deltares en RIVM toegepaste open source pakket 'R', of ...?*

Keuze 2c *Dienen de onzekerheidsmarges zichtbaar te zijn? Of anders gezegd, zijn niet-statistische trendmethodieken toegestaan, of dienen deze statistisch gemaakt te worden, bijv. omdat we als eis stellen dat er alleen procedures gebruikt mogen worden als die iets over onzekerheid zeggen?*

Keuze 3 **Uniformiteit-keuzes; te presenteren data**

Keuze 3a *Welke verslagjaren? Jaar-, kwartaal-, maand-, of daggegevens?*

Keuze 3b *Welke statistische variabelen (medianen, gemiddelden, ...)?*

Keuze 3c *Welke onzekerheidsmarges presenteren we? Altijd 95%?*

Keuze 3d *Welke trendmethodiek, databewerkingen (na en voor), ...?*

Keuze 3e *Welke aggregaten (NL, naar sector, naar regio, 'all or beauties', ...)?*

Keuze 3f *Berekenen we eerst de trends om vervolgens te aggregeren (bijv. per grondsoort), of aggregeren we eerst en berekenen we dan de trend?*

Keuze 3g *Welke andere keuzes zijn op dit terrein te maken?*

Deze uniformiteit-keuzes kunnen door de hoofdaannemer nader gespecificeerd worden, in overleg met de opdrachtgever. Het is aan deze twee partijen om te bepalen waarover concrete afspraken gemaakt dienen te worden. Niet alles hoeft van tevoren dichtgetimmerd te worden.

De te maken afspraken en keuzes kunnen afhangen van de onderwerpen die opgenomen worden in het verzamelrapport. Het gaat vaak om veel onderwerpen, met uiteenlopende datakarakteristieken. Om het overzicht hierin te bewaren kan het nodig zijn om per trendresultaat, of groep van trendresultaten, een beperkt aantal datakarakteristieken als metadata bij te houden.

Keuze 4	Metadata-keuzes; onderwerpen verzamelrapport
<i>Keuze 4a</i>	<i>Over welke onderwerpen worden trendresultaten gepubliceerd?</i>
<i>Keuze 4b</i>	<i>Welke metadata over deze trendresultaten willen we bijhouden?</i>
<i>Keuze 4c</i>	<i>Wie is verantwoordelijk voor het opleveren van de metadata?</i>
<i>Keuze 4d</i>	<i>In welke mate is de op te leveren metadata gestandaardiseerd?</i>

De behoefte aan metadata dient in balans te zijn met de tijd die het kost om de vereiste metadata bij te houden. Het mag geen papieren rompslomp worden, maar het is ook geen optie om niets te doen. De juiste balans dient gevonden te worden. Als inspiratiebron is een aanbeveling van de Europese Commissie als bijlage ingevoegd (aan het eind van deze notitie) over het bijhouden van metadata binnen een statistisch systeem.

Een pragmatische aanpak is hier vereist. Houd alleen de metadata bij die daadwerkelijk benodigd is. Een beperkt aantal datakarakteristieken is genoeg. Het hoeft ook niet voor elk onderwerp c.q. trendresultaat bijgehouden te worden. Maak hierin keuzes.

Hierbij speelt ook de vraag wie deze metadata dient op te stellen: de hoofdaannemer en/of de onderaannemers? Is de metadata gestandaardiseerd, of wordt er gebruik gemaakt van al bestaande metadata, die niet geharmoniseerd is? Moet de metadata toegesneden zijn op de trendresultaten in het verzamelrapport, of moet het corresponderen met de trendresultaten in het onderliggend onderzoeksrapport? Ook hier zijn pragmatische keuzes nodig.

De metadata is met name bedoeld om het overzicht te bewaren in de uitvoering van de opdracht door de hoofdaannemer (het samenstellen van het verzamelrapport). Het is dus aan de hoofdaannemer om op bovenstaande metadata-vragen antwoorden te geven, om de juiste balans te vinden, en om het daarna te integreren in de opdrachtverstrekkingen richting de onderaannemers.

3. De nabewerking van de data

In hoofdstuk 2 werden een aantal keuzes genoemd over de te publiceren trendresultaten in het verzamelrapport. De opdrachtgever en de hoofdaannemer zijn primair verantwoordelijk voor het maken van de juiste keuzes. Vervolgens zet de hoofdaannemer deze keuzes door in de opdrachtverstrekkingen richting de onderaannemers.

De opdrachtgever is vooral geïnteresseerd in het uiteindelijke verzamelrapport en laat het maken van de trendanalyses over aan de hoofd- en onderaannemers. Nadat de in hoofdstuk 2 genoemde

keuzes gemaakt zijn ('wat komt er in het verzamelrapport en op welke wijze') is de rol van de opdrachtgever uitgespeeld in de zin dat er een kader is neergezet waarop de hoofd- en onderaannemers hun afspraken kunnen baseren.

In veel gevallen zal een onderaannemer eerst zelf een rapport schrijven over zijn trendresultaten, waarna dit als input gebruikt zal worden voor het verzamelrapport. Hier kunnen aansluitproblemen optreden, of zelfs discrepanties in de vorm van tegenstrijdige conclusies. In het eigen onderzoeksrapport kunnen andere keuzes en aannames gemaakt zijn dan in het verzamelrapport. De door de onderaannemer aan te leveren meetreeksen dienen dan nabewerkt te worden. Dit kan een eenvoudige nabewerking van de data zijn, maar er kan ook een heranalyse nodig zijn (bijv. in het geval dat het verzamelrapport van een hele andere trendmethodiek en/of databewerking uitgaat dan in het onderliggende onderzoeksrapport). De volgende keuzes dienen hier gemaakt te worden:

- | | |
|----------------|--|
| Keuze 5 | Metadata-keuzes; vertaalslag trendresultaten |
| Keuze 5a | <i>Wie doet de nabewerking, of zelfs heranalyse, van de trendanalyses uit de onderliggende onderzoeksrapporten, opdat het opgenomen kan worden in het verzamelrapport? De hoofdaannemer of de onderaannemer?</i> |
| Keuze 5b | <i>Wie stelt de metadata op over deze nabewerking van de data?</i> |
| Keuze 6 | Communicatie-keuzes; vertaalslag trendresultaten |
| Keuze 6a | <i>Worden de onderaannemers vooraf geïnformeerd over de hoofdstuk 2 keuzes, of gebeurt dit nadat de onderaannemers hun trendresultaten gepubliceerd hebben?</i> |
| Keuze 6b | <i>Om welke informatie over hoofdstuk 2 keuzes gaat het? Wat dient de onderaannemer te weten, vooraf dan wel achteraf?</i> |
| Keuze 6c | <i>Bij vooraf informeren, is het voorschrijvend, adviserend, of louter bedoeld om dubbel en extra werk te voorkomen?</i> |

Bij het maken van deze keuzes heeft de hoofdaannemer een grote rol om de onderaannemers van adequate informatie te voorzien, opdat eventuele aansluitproblemen zo klein mogelijk zullen zijn. Discrepanties in de vorm van tegenstrijdige conclusies kunnen incidenteel voorkomen als het oorspronkelijke trendresultaat van de onderaannemer afwijkt van het daarmee overeenkomende trendresultaat in het verzamelrapport.

De vertaalslag (de nabewerking en eventuele heranalyse) brengt risico's met zich mee, die in de opdrachtverstrekking onderkend dienen te worden.

- | | |
|----------------|--|
| Keuze 7 | Risicobeperking-keuzes; vertaalslag trendresultaten |
| Keuze 7a | <i>Wie hakt de knopen door bij het nemen van (nood)maatregelen als het risico van tegenstrijdige trendconclusies zich voordoet, bij het maken van de vertaalslag van onderzoeksrapport naar verzamelrapport? De opdrachtgever of de hoofdaannemer?</i> |
| Keuze 7b | <i>Door wie laat de knoopdoorhakker zich adviseren?</i> |
| Keuze 7c | <i>Welke rol krijgen de onderaannemers hierin toebedeeld?</i> |

Hoe beter deze keuzes vooraf uitgewerkt zijn, hoe kleiner de organisatorische problemen zullen zijn als het risico van tegenstrijdige trendconclusies zich voordoet, of dreigt voor te doen.

Tot slot dient opgemerkt te worden dat de processtappen in Figuur 2 een versimpelde weergave van de werkelijkheid weergeven. In dit hoofdstuk wordt de nabewerking van data louter gerelateerd aan het aansluitprobleem tussen de trendresultaten van de hoofdaannemer (verzamelrapport) en die van onderaannemers (onderliggende onderzoeksrapporten). Er zijn ook andere nabewerkingen denkbaar, zoals het aggregeren van allerlei trendresultaten om tot een nieuw trendresultaat te komen. Deze andere nabewerkingen worden gemakshalve in hoofdstuk 5 meegenomen.

4. De trendmethodiek

In het op pagina 3 ingevoegde CBS-rapport wordt uitgebreid ingegaan op de verschillende trendmethodieken. In hoofdstuk 6 van het CBS-rapport worden acht trendmethodieken beschreven. In hoofdstuk 7 van het CBS-rapport worden de trendmethodieken met elkaar vergeleken, zonder dat er een uitspraak wordt gedaan welke methodiek de voorkeur verdient.

In de tabel hieronder staan per trendmethodiek negen kolommen. De eerste vijf staan ook in hoofdstuk 7 van het CBS-rapport. De laatste vier zijn er aan toegevoegd, waaronder kolom 'beste keuze'.

TRENDMETHODIEK	Lokale of globale trend	Wordt onzekerheid gegeven?	Wordt de trend expliciet gegeven?	Is er een methode om de parameter voor flexibiliteit te schatten?	Uitbreiding met regressoren mogelijk?	Weinig invloed uitschieters?	Wiskundig complexe methode?	Software makkelijk beschikbaar?	Beste keuze
Voortschrijdende gemiddelden	Lokaal	Nee	Ja	Nee (maar wel beïnvloedbaar)	Nee	Nee	Nee	Ja	-
Lineaire regressie	Globaal	Ja	Ja	Ja	Ja	Nee	Nee	Ja	+/-
Theil-Sen	Globaal	Nee	Ja	n.v.t.	Nee	Ja	Nee	Ja	+
Mann-Kendall (Seasonal)	Globaal	Ja	Nee	n.v.t.	Nee	Ja	Ja	Ja	+
Lowess (R implementatie)	Beide	Ja	Ja	Ja	Ja	Ja	Nee	Ja	+/-
STM	Beide	Ja	Ja	Ja	Ja	Ja	Ja	Nee	+
ARIMA	Beide	Ja	Nee	Ja (impliciet)	Ja	Ja	Ja	Ja	+/-
Twee deelperioden vergelijken	Globaal	Nee	Ja	n.v.t.	Nee	Nee	Nee	Ja	-

De beste keuze is gevuld op basis van een driepuntenschaal (+ geschikt, +/- matig geschikt, - niet geschikt) onder de aanname dat slechts één trendmethodiek wordt toegepast. Het toekennen van plusjes en minnetjes is geen harde wetenschap. Het hangt af van welke weging men aan elke kolom geeft. Ook is de kolomvulling vaak genuanceerder dan een zwart-witte 'ja' of 'nee'; het grijze gebied zorgt voor ietwat verschillende meningen onder de trendexperts. Bovendien zijn er kolommen denkbaar, die er nu niet in staan, zoals 'kan het omgaan met zeer scheef verdeelde data?' en 'wordt het veel gebruikt voor vergelijkbaar onderzoek in de internationale literatuur?'.

Kortom, de scores in de kolom 'beste keuze' kunnen iets veranderen door het toevoegen en/of verwijderen van kolommen, door een iets andere wegging van de kolommen, en door een iets andere vulling van de kolommen.

Daarom zal elk instituut c.q. trendexpert zijn eigen voorkeur hebben, afhankelijk van de specifieke toepassing die men voor ogen heeft. Zo kan voor het verzamelrapport gekozen worden voor een uniforme, gebruiksvriendelijke aanpak, hetgeen niet tot dezelfde trendmethodiek(en) hoeft te leiden als bij een optimale aanpak, statistisch gezien, voor specifieke onderzoekssituaties. Ook kan het afhangen van de data karakteristieken van de meetreeksen (o.a. is het scheef verdeeld, heeft het veel uitschieters?) welke trendmethodiek de voorkeur krijgt.

Omdat elke trendmethodiek zijn eigen voor- en nadelen heeft, en de instituten c.q. trendexperts hierover van mening kunnen verschillen, verdient het de voorkeur om meerdere methodieken simultaan te gebruiken. Dan kan onder andere gecheckt worden in welke mate een trendconclusie afhangt van de gebruikte trendmethodiek.

In de opdrachtverstrekking tussen hoofdaannemer en onderaannemers kunnen de volgende keuzes gemaakt worden (eventueel in overleg met de verzamelrapport-opdrachtgever):

- | | |
|-----------------|--|
| Keuze 8 | Communicatie-keuzes; trendmethodieken |
| Keuze 8a | <i>Zijn de onderaannemers vrij om hun eigen trendmethodiek-voorkeur te volgen of wordt dit voorgeschreven door de hoofdaannemer?</i> |
| Keuze 8b | <i>Als de hoofdaannemer een andere trendmethodiek hanteert dan de onderaannemers, hoe wordt dit in het verzamelrapport verantwoord?</i> |
| Keuze 8c | <i>Hoe verantwoorden de onderaannemers hun trendmethodiek-keuzes?</i> |
| Keuze 8d | <i>Worden onderaannemers door de hoofdaannemer vooraf geïnformeerd over de eventuele invloed van de gekozen publicatietool (zoals Trend-spotter) op de door de onderaannemers te kiezen trendmethodiek(en)?</i> |
|
 | |
| Keuze 9 | Metadata-keuzes; trendmethodieken |
| Keuze 9a | <i>Het gebruik van meerdere trendmethodieken verdient de voorkeur boven het gebruiken van één methodiek. Bij wie ligt de verantwoordelijkheid om dit daadwerkelijk te gaan doen?</i> |
| Keuze 9b | <i>De gekozen trendmethodiek veronderstelt aannames over de (statistische) eigenschappen van de data. Wie toetst of de trendmethodiek past bij de te onderzoeken data, d.w.z. dat aan de aannames voldaan wordt?</i> |
| Keuze 9c | <i>In het geval dat een niet-statistische trendmethodiek statistisch gemaakt moet worden, wie gaat dat doen?</i> |
| Keuze 9d | <i>Is het een optie om gebruik te maken van modelselectiecriteria, opdat er een objectieve keuze gemaakt wordt voor de geschiktste trendmethodiek?</i> |
|
 | |
| Keuze 10 | Risicobeperking-keuzes; trendmethodieken |
| | <i>Analoog aan keuzes 7a t/m 7c:</i> |
| | <i>Wie hakt de knopen door als de hoofdaannemer en een onderaannemer een fundamenteel verschil van mening hebben over de te hanteren trendmethodiek? Wie adviseert de knopendoorhakker?</i> |

Keuze 8 gaat over de communicatie tussen de hoofd- en onderaannemers over de te gebruiken trendmethodieken en op welke wijze de gekozen trendmethodieken verantwoord dienen te worden. Keuze 9 gaat over de ins & outs van de trendmethodieken zelf. Dit dient in de metadata beschreven te worden. Keuze 10 gaat over het beperken van de risico's die samenhangen met tegenstrijdige trendconclusies in het geval dat er geen geschikte trendmethodiek gebruikt wordt.

5. De voorbereiding van de data

De keuzes 8a t/m 8d en 10, in het vorige hoofdstuk, zijn ook te maken voor de dataverwerkingen die rondom een trendmethodiek gedaan worden door de onderaannemer(s) resp. de hoofdaannemer.

Keuze 11	Communicatie-keuzes; dataverwerkingen
Keuze 11a	<i>Zijn de onderaannemers vrij om hun eigen dataverwerkingen te doen, of wordt dit voorgeschreven door de hoofdaannemer?</i>
Keuze 11b	<i>Als de hoofdaannemer een andere dataverwerking doet dan de onderaannemers, hoe wordt dit in het verzamelrapport verantwoord?</i>
Keuze 11c	<i>Hoe verantwoorden de onderaannemers hun dataverwerkingen?</i>
Keuze 11d	<i>Worden onderaannemers door de hoofdaannemer vooraf geïnformeerd over de eventuele invloed van de gekozen publicatietool (zoals Trend-spotter) op de door de onderaannemers uit te voeren dataverwerkingen?</i>
Keuze 12	Risicobeperking-keuzes; dataverwerkingen
	<i>Analoog aan keuzes 7a t/m 7c en 10:</i>
	<i>Wie hakt de knopen door als de hoofdaannemer en een onderaannemer een fundamenteel verschil van mening hebben over de uit te voeren dataverwerkingen? Wie adviseert de knopendoorhakker?</i>

De dataverwerkingen rondom een trendmethodiek verschillen per onderwerp c.q. trendresultaat. De voorbereiding van de data dient beschreven te worden, de trendmethodiek zelf, en de nabewerking van de data. Het vormt één geheel. De beschrijving kan vereenvoudigd worden als er een metadata-sjabloon ontwikkeld wordt. Elementen die hierin een plek kunnen krijgen zijn:

- De volgorde van databewerking kan een rol spelen: 'eerst aggregeren en dan trendanalyse' of 'eerst trendanalyse en dan aggregeren'?
- Ook steekproef-aspecten spelen een rol. Hoe is de stratificatie? Hoe wordt er gewogen? Hoe worden de effecten van steekproefwijzigingen (en storende factoren) verdisconteerd?
- De data kunnen extreem scheef verdeeld zijn. De gemiddelden kunnen variëren door de uitschieters. Is er een 'outlier'-correctie? Wordt er een (log)transformatie toegepast?
- Op wat voor gegevens wordt de trendanalyse toegepast: mediaan, rekenkundig of geometrisch gemiddelde, 'all' versus 'beauties'?
- Geeft de trendmethodiek wel of geen foutmarge? Gaat het uit van een lokale of globale trend? Als alleen gekeken wordt naar een globale trend is het dan een gemis als een omslagpunt aan het eind van de reeks gemist wordt?
- Welke verklarende variabelen worden in de trendanalyses gebruikt? Spelen ook 'level shifts' (sprongen, stap trends) een rol?

- Hoe lang zijn de meetreeksen (5, 10, ... jaar)? Zijn er dagelijkse, maandelijkse of kwartaalgegevens binnen een jaar? Hoe ga je om met missende waarden c.q. gaten in een meetreeks?
- Hoe ga je om met 'censored data'? Het gaat dan om waarnemingen kleiner dan de detectielimiet (bijv. bij metingen van concentraties in water), of dat het groter is dan wat maximaal meetbaar is (bijv. zwaarder zijn dan wat de badkamerweegschaal maximaal kan aangeven).
- Op welke wijze zijn de weercorrecties doorgevoerd? Welke winter- en zomerdefinities worden gehanteerd? Hoe wordt er omgegaan met seizoenpatronen?
- Zie voor aanvullende informatie: het op pagina 3 ingeplakte CBS-rapport.

De hoofdaannemer dient in te schatten in hoeverre het wenselijk is dat zo'n metadata-sjabloon opgesteld wordt en in welke mate de onderaannemers dit in dienen te vullen.

Keuze 13	Metadata-keuzes; dataverwerkingen
Keuze 13a	<i>Is het nodig om een metadata-sjabloon te ontwikkelen waarmee de dataverwerkingen rondom de trendmethodiek, en ook de trendmethodiek zelf, beschreven kunnen worden?</i>
Keuze 13b	<i>Zo ja, wordt het sjabloon opgesteld door de hoofdaannemer en dienen de onderaannemers deze in te vullen, of kunnen de onderaannemers hun eigen versie opstellen?</i>
Keuze 13c	<i>Wanneer dient zo'n metadata-sjabloon ingevuld te worden; altijd, op het moment dat een trendresultaat opgeleverd wordt, of alleen in de gevallen als er een te groot risico is dat er tegenstrijdige trendconclusies gemaakt gaan worden?</i>

Het liefst wil je van elk trendresultaat weten welke dataset gebruikt is, wat de kenmerken van de meetreeksen zijn, welke datavoorbewerking gedaan is, welke trendmethodiek gebruikt is (en waarom), en welke nabewerking gedaan is. Aan de andere kant wil je geen papieren rompslomp creëren, die veel tijd kost om te maken, maar die nauwelijks gebruikt wordt. Heel incidenteel zal deze metadata geraadpleegd worden, waarbij het nog maar de vraag is of het op dat specifieke moment de dan benodigde informatie bevat. De hoofdaannemer zal dus moeten zoeken naar de juiste balans: een beperkte hoeveelheid metadata dient direct beschikbaar te zijn en de overige metadata dient snel opvraagbaar te zijn voor de incidentele gevallen dat dit nodig is.

6. De te gebruiken ruwe meetgegevens

De kwaliteit van het trendresultaat hangt af van de kwaliteit van de ruwe meetgegevens die daarvoor gebruikt worden ('garbage in -> garbage out'). Zo dient de verdeling van de meetpunten representatief te zijn. Het is primair aan de onderaannemer om er voor te zorgen dat de kwaliteit van de ruwe meetgegevens goed genoeg is. De hoofdaannemer zal aannemen dat het voldoende op orde is en dat er actuele metadata over de ruwe meetgegevens beschikbaar is. Aan de andere kant dient de hoofdaannemer er voor te zorgen dat het risico van 'garbage in' geminimaliseerd is.

Keuze 14	Metadata-keuzes; ruwe meetgegevens
Keuze 14a	<i>Is het nodig om de metadata over de ruwe meetgegevens op te nemen in de onder keuze 13 genoemde metadata-sjabloon?</i>

Keuze 14b *Is het nodig om de bij de onderaannemer beschikbare metadata over de ruwe meetgegevens te actualiseren?*

Keuze 15 ***Risicobeperking-keuzes; ruwe meetgegevens***
Analoog aan keuzes 7a t/m 7c, 10 en 12:
Wie hakt de knopen door als de hoofdaannemer en/of de onderaannemer twijfels hebben over de kwaliteit van de gebruikte ruwe meetgegevens? Wie adviseert de knopendoorhakker? Wie bepaalt of iets van voldoende kwaliteit is?

Mocht het risico zoals benoemd bij keuze 15 zich daadwerkelijk voordoen, dan is afstemming nodig tussen de hoofdaannemer en de onderaannemer (en wellicht ook met de opdrachtgever).

Keuze 16 ***Communicatie-keuzes; ruwe meetgegevens***
Keuze 16a *Welke maatregelen zijn volgens de hoofd- en onderaannemer nodig om de kwaliteit van ruwe gegevens naar een acceptabele kwaliteit te brengen?*
Keuze 16b *Bij wie ligt de verantwoordelijkheid om deze maatregelen uitgevoerd te krijgen, en op welk termijn?*
Keuze 16c *Op welke wijze dient de opdrachtgever van het verzamelrapport hierin betrokken te worden?*

Voor de bepaling of een trendresultaat van voldoende kwaliteit is, dient het hele analysetraject meegenomen te worden, conform Figuur 2, inclusief de aansluitproblemen die op kunnen treden als een trendresultaat in een onderliggend onderzoeksrapport in een iets gewijzigde vorm wordt opgenomen in een verzamelrapport.

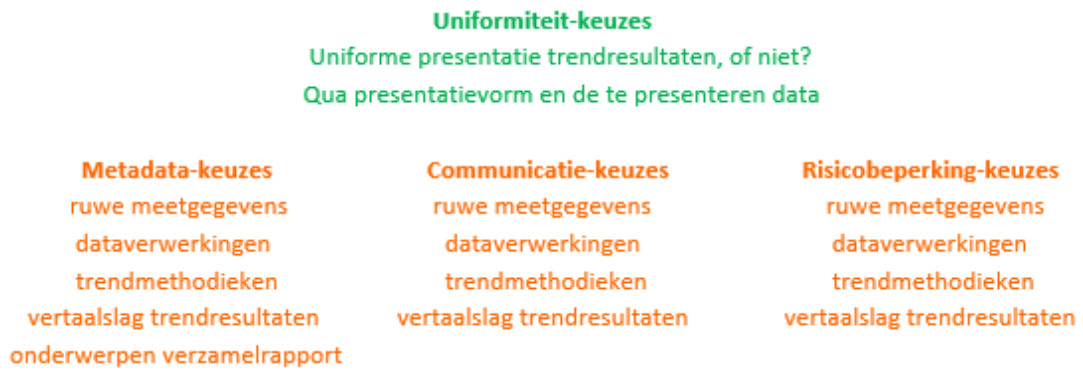
7. Overkoepelende beschouwing

In de voorgaande hoofdstukken worden keuzes genoemd die gemaakt kunnen worden tijdens de opdrachtverstrekking om een verzamelrapport te maken op basis van allerlei onderliggende onderzoeksrapporten, waarin trendanalyses een rol spelen. Hierbij werd uitgegaan van een versimpelde weergave van de werkelijkheid. Er is ook niet gestreefd naar volledigheid; er zijn meer keuzes op te noemen dan wat in deze notitie gedaan is. Wat opvalt is dat er vier soorten keuzes onderscheiden worden. Zie Figuur 3.

De uniformiteit-keuzes zijn nodig, omdat de veelheid van onderwerpen in de onderliggende onderzoeksrapporten dit vereisen ten behoeve van de leesbaarheid van het verzamelrapport. Het voorkomt een rommelig beeld. Onnodige verwarring wordt vermeden. De uniformiteit-keuzes dienen door de opdrachtgever en de hoofdaannemer gezamenlijk gemaakt te worden.

De metadata-, communicatie- en risicobeperking-keuzes hebben betrekking op het hele traject: niet alleen de ruwe meetgegevens, de dataverwerking en de trendmethodieken, maar ook de vertaalslag van de trendresultaten in de onderliggende onderzoeksrapporten naar die in het verzamelrapport. Bij het maken van deze keuzes is vooral afstemming nodig tussen de hoofd- en onderaannemers. Wat betreft de metadata zijn er ook keuzes te maken over de onderwerpen in het verzamelrapport.

Het maken van deze keuzes ligt vooral bij de hoofdaannemer, waarbij de lijst met onderwerpen c.q. trendresultaten afgestemd wordt met de opdrachtgever.



Figuur 3: De vier soorten keuzes die te maken zijn tijdens de opdrachtverstrekking conform de piramide in Figuur 1, met groene top en een oranje middendeel. Ook het door te lopen traject, zoals weergegeven in Figuur 2, is hierin terug te zien, van boven naar beneden in het oranje deel.

Bij de metadata-keuzes is een pragmatische houding gewenst. Vooraf veel op papier zetten kost veel tijd en het is nog maar de vraag of dit terugverdiend wordt in de gevallen dat de metadata gebruikt wordt. Het is vooral aan de hoofd- en onderaannemers om de juiste balans te vinden.

Bij de communicatie-keuzes gaat het voornamelijk om de afstemming tussen de hoofd- en de onderaannemers. Welke vertaalslag van de trendresultaten is nodig? Is men het met elkaar eens over de te gebruiken trendmethodieken, de dataverwerkingen eromheen, en de ruwe meetgegevens? Voorkomen moet worden dat er onnodig extra werk gedaan wordt en dat er tegenstrijdige trendconclusies gemaakt gaan worden.

Bij de risicobeperking-keuzes gaat het over kwaliteitsaspecten die samenhangen met de te publiceren trendresultaten. Zijn de ruwe meetgegevens, trendmethodieken en de dataverwerkingen eromheen op orde? Wat doen we als er tegenstrijdige trendconclusies getrokken worden, door de hoofd- respectievelijk de onderaannemers? Wie hakt de knopen door, en wie adviseert de knopendoorhakker?

Dit klinkt alsof er een rol weggelegd is voor een soort van werkgroep statistiek. Deze kan incidenteel bijeen geroepen worden, fysiek of via mail/webconference, als vraagbaak en/of om (on)gevraagd advies te geven. De rol is vergelijkbaar met die van reviewers van eindrapporten, met als belangrijkste verschil dat de werkgroep statistiek leden een 'trend-bril' op hebben gezet. Hun focus ligt op de trendresultaten, waarin het hele meettraject (inclusief vertaalslag) binnen hun scope valt.

De belangrijkste aanbeveling van deze notitie is om een 'werkgroep statistiek' in het leven te roepen. Deze is op incidentele basis oproepbaar en kan betrokken worden bij trendanalyses, als vraagbaak en/of om (on)gevraagd advies te geven. De scope van zo'n werkgroep statistiek beperkt zich niet tot de Evaluatie Meststoffenwet. Het kan ook op andere terreinen ingezet worden.

Andere aanbevelingen zijn:

- Zorg ervoor dat zo vroeg mogelijk de lijst met onderwerpen c.q. meetresultaten voor in het verzamelrapport beschikbaar is. Ook de uniformiteitskeuzes dienen zo vroeg mogelijk gemaakt te zijn. Hierdoor weten de samenstellers van onderliggende onderzoeksrapporten tijdig waar ze aan toe zijn. Daarmee kan onnodig extra werk voorkomen worden. Eventuele meningsverschillen kunnen voorgelegd worden aan de werkgroep statistiek.
- Bij de te maken metadata-keuzes en communicatie-keuzes hoeft de werkgroep statistiek niet actief betrokken te worden. Dit kan aan de hoofd- en onderaannemers zelf overgelaten worden. Zij dienen de juiste balans te vinden.
- Voor de werkgroep statistiek is het wel van belang dat het de juiste metadata ontvangt voor het specifieke geval waarvoor ze (on)gevraagd advies dienen te geven. Aangenomen mag worden dat het standaardpakket aan metadata (per trendresultaat), dat vastgelegd is via de opdrachtverstrekking, tekort zal schieten. Het is te tijdrovend om alles bij te houden. Dit betekent dat de hoofd- en onderaannemers er op voorbereid dienen te zijn dat op incidentele basis specifieke metadata opgevraagd kunnen worden, en dat die snel opgeleverd moeten worden.
- De hoge opleversnelheid is nodig, omdat het risico van tegenstrijdige trendconclusies zich vooral kan manifesteren in de eindfase van het samenstellen van het verzamelrapport. De tijddruk is dan hoog. Heldere en snelle communicatie is dan een vereiste. Dezelfde taal spreken helpt hierbij. Het harmoniseren van databewerkingen ondersteunt dit. Ook dient elk instituut transparant te zijn over hoe je van een meting tot een conclusie komt. De werkgroep statistiek kan hierin faciliteren.
- De formele rol van de werkgroep statistiek dient pragmatisch vormgegeven te worden tijdens de opdrachtverstrekking. Er kan volstaan worden met een lijstje namen van trendexperts, waarvan één persoon de status van trekker/contactpersoon krijgt. Ze kunnen hun uren schrijven, net zo als reviewers van eindrapporten dat veelal doen. Het is vooralsnog niet de bedoeling dat de werkgroep statistiek een reguliere overlegclub wordt, dat op gezette tijden bijeen komt. Op termijn kan het dat wel worden, bijv. als de werkgroep statistiek bij meerdere opdrachtverstrekkingen tegelijk betrokken is. Tot die tijd is zij incidenteel oproepbaar.
- Deelnemers van de werkgroep statistiek komen van de rijkskennisinstellingen en andere organisaties met kennis over trendanalyses. Voor de Evaluatie Meststoffenwet zijn CBS, PBL, Deltares, RIVM, WENR en WEcR de organisaties die trendexperts kunnen leveren. Dit is uit te breiden met bijv. KNMI, bij klimaatgerelateerde trends. De precieze vulling van de werkgroep statistiek hangt af van het onderwerp waarover het verzamelrapport zal gaan.

Naast het in het leven roepen van een werkgroep statistiek wordt het ook aanbevolen om zo vroeg mogelijk te starten met de afstemming tussen de opdrachtgever, de hoofdaannemer en de onderaannemers. Hoe eerder hiermee begonnen wordt, hoe kleiner het risico dat tijdens het stressvolle productieproces problemen met betrekking tot trendanalyses optreden.

Het CBS-rapport van Pim Ouwehand geeft een beeld van welke meetreeksen een rol spelen binnen de Evaluatie Meststoffenwet, en welke trendmethodieken hierop losgelaten worden. Vervolgonderzoek kan zich richten op:

- Het testen van de verschillende trendmethodieken op desbetreffende meetreeksen en/of op gesimuleerde datasets, opdat er een beter beeld ontstaat van de voor- en nadelen van de binnen de Evaluatie Meststoffenwet gebruikte trendmethodieken.
- Het uitvoeren van een literatuur-review, met focus op het waterkwaliteitsdomein, met betrekking tot welke trendmethodieken er internationaal worden toegepast.
- Het uitzoeken hoe een visualisatie van een trend (en de omringende onzekerheid) overkomt op een breed publiek, o.a. hun perceptie van hoe significant een trend is in relatie tot de statistisch berekende significantie. In 2015 heeft het CBS hier onderzoek naar gedaan. Desgewenst kan het CBS adviseren over een geschikte uniforme presentatievorm van de trendresultaten in de rapportages van de Evaluatie Meststoffenwet.

8. Verantwoording

Deze notitie van Arthur Denneman (CBS) en het rapport van Pim Ouwehand (CBS) hebben volop geprofiteerd van de vier bijeenkomsten van de Werkgroep statistiek Evaluatie Meststoffenwet (eind 2018/begin 2019). Zonder de informatieverstrekking door de trendexperts van de deelnemende instituten zou het niet mogelijk zijn geweest om deze notitie en het CBS-rapport op te stellen. Daarom worden de volgende personen bedankt voor hun nuttige bijdragen:

Hans Visser (PBL)

Joachim Rozemeijer (Deltares)

Dennis Walvoort (WENR)

Dico Fraters (RIVM)

Lan Ge (WEcR)

Martin Knotters (WENR)

Ook worden de opdrachtgevers van dit project bedankt. Marijke Koning (Ministerie LNV) heeft, in samenwerking met Wilbert van Zeventer (Ministerie IenW), voldoende budget beschikbaar gesteld om dit werk mogelijk te maken.

De oorspronkelijke insteek was om de bevindingen toe te gaan passen binnen de Evaluatie Meststoffenwet. Tijdens het schrijven van deze notitie bleek dat de resultaten breder toepasbaar zijn. Het kan ook voor andere terreinen binnen LNV en IenW toegepast worden. De bevindingen van zowel deze notitie als het bijbehorende CBS-rapport vormen een prima basis voor het in leven roepen van een **Werkgroep statistiek LNV/IenW**.

BIJLAGE (horend bij metadata-keuzes; zie pagina 5)

AANBEVELINGEN

COMMISSIE

AANBEVELING VAN DE COMMISSIE

van 23 juni 2009

betreffende referentiemetagegevens voor het Europees statistisch systeem

(Voor de EER relevante tekst)

(2009/498/EG)

DE COMMISSIE VAN DE EUROPESE GEMEENSCHAPPEN,

Gelet op het Verdrag tot oprichting van de Europese Gemeenschap, en met name op artikel 211,

Overwegende hetgeen volgt:

- (1) De Praktijkcode Europese statistieken⁽¹⁾, die bestemd is voor de nationale en de communautaire statistische instanties, omvat vijftien beginselen die zijn verbonden aan het institutionele kader, statistische processen en uitkomsten.
- (2) Beginsel 15 van de Praktijkcode Europese statistieken betreft de toegankelijkheid en de duidelijkheid van Europese statistieken, en benadrukt ook dat de bijbehorende metagegevens moeten worden gedocumenteerd volgens een gestandaardiseerd metagegevenssysteem.
- (3) Referentiemetagegevens maken een integrerend deel uit van het metagegevenssysteem van elke statistische instantie.
- (4) Door de Praktijkcode Europese statistieken aan te nemen, hebben de nationale en communautaire statistische instanties zich tot kwalitatief hoogstaande statistieken verbonden, waarvoor ook een transparantere en beter geharmoniseerde verslaggeving over gegevenskwaliteit nodig is.

⁽¹⁾ Aanbeveling van de Commissie van 25 mei 2005 over de onafhankelijkheid, integriteit en verantwoordingsplicht van de nationale en communautaire statistische instanties, COM(2005) 217 definitief.

- (5) In het kader van het SDMX-initiatief voor gemeenschappelijke technische en statistische normen voor de uitwisseling en het gemeenschappelijke gebruik van gegevens en metagegevens, dat is genomen door de Bank voor Internationale Betalingen, de Europese Centrale Bank, de statistische instantie van de Gemeenschap (Eurostat), het Internationaal Monetair Fonds, de Organisatie voor Economische Samenwerking en Ontwikkeling, de Verenigde Naties en de Wereldbank, zijn de inhoudgerelateerde richtsnoeren („Content-Oriented Guidelines”) van de SDMX vastgesteld waarmee de creatie en de invoering van geharmoniseerde referentiemetagegevens in het Europees statistisch systeem worden ondersteund.

- (6) De efficiëntie kan aanzienlijk worden verhoogd wanneer de referentiemetagegevens worden geproduceerd op basis van een geharmoniseerde lijst van statistische begrippen binnen het Europees statistisch systeem, terwijl de nationale en de communautaire statistische instanties tegelijkertijd de mogelijkheid krijgen zo nodig meer statistische begrippen in specifieke statistische gebieden toe te voegen.

- (7) Verordening (EG) nr. 223/2009 van het Europees Parlement en de Raad van 11 maart 2009 betreffende Europese statistieken⁽²⁾ vormt een referentiekader voor deze aanbeveling.

BEVEELT DE LIDSTATEN HIERBIJ AAN:

1. De nationale statistische instanties wordt verzocht de in de bijlage opgenomen statistische begrippen en subbegrippen toe te passen wanneer referentiemetagegevens worden opgesteld in de verschillende statistische gebieden en wanneer referentiemetagegevens binnen het Europees statistisch systeem of daarbuiten worden uitgewisseld.

⁽²⁾ PB L 87 van 31.3.2009, blz. 164.

2. De nationale statistische instanties dienen extra statistische begrippen toe te voegen aan de bovengenoemde lijst van begrippen en subbegrippen indien dat voor bepaalde statistische gebieden noodzakelijk is.
3. De nationale statistische instanties wordt verzocht de Commissie (Eurostat) regelmatig over hun toepassing van de in de bijlage opgenomen statistische begrippen en subbegrippen te informeren.

Gedaan te Brussel, 23 juni 2009.

Voor de Commissie
Joaquín ALMUNIA
Lid van de Commissie

BIJLAGE

Lijst van statistische begrippen en subbegrippen (inclusief de definities van de begrippen en subbegrippen)

Nummer	Begrip	Subbegrip	Omschrijving
1.	Contactpunt		Contactpunten (personen of organisaties) voor de gegevens of metagegevens, met inbegrip van informatie over hoe de contactpunten te bereiken zijn.
1.1.		Contactorganisatie	Naam van de organisatie die de contactpunten voor de gegevens of metagegevens onderhoudt.
1.2.		Eenheid in de contactorganisatie	Subafdeling van een organisatie waarmee contact kan worden opgenomen.
1.3.		Naam contactpunt	Naam van de contactpunten voor de gegevens of metagegevens.
1.4.		Functie contactpersoon	Het gebied waarvoor de contactpersoon technische verantwoordelijkheid draagt, bv. „methoden”, „databankbeheer” of „verspreiding”.
1.5.		Postadres contactpunt	Postadres van de contactpunten voor de gegevens of metagegevens.
1.6.		E-mailadres contactpunt	E-mailadres van de contactpunten voor de gegevens of metagegevens.
1.7.		Telefoonnummer contactpunt	Telefoonnummer van de contactpunten voor de gegevens of metagegevens.
1.8.		Fax contactpunt	Fax van de contactpunten voor de gegevens of metagegevens.
2.	Actualisering metagegevens		Datum waarop het metagegevenselement in de databank is ingevoerd of gewijzigd.
2.1.		Laatste certificering metagegevens	Datum van de laatste door de gebiedsbeheerder verrichte certificering om te bevestigen dat de verspreide metagegevens nog actueel zijn, zelfs indien de inhoud niet is gewijzigd.
2.2.		Laatste verspreiding metagegevens	Datum van de laatste verspreiding van de metagegevens.
2.3.		Laatste actualisering metagegevens	Datum van de laatste actualisering van de metagegevens.
3.	Statistische presentatie		
3.1.		Omschrijving gegevens	Hoofdkenmerken van de gegevensset, op begrijpelijke wijze beschreven met verwijzing naar de verspreide gegevens en indicatoren.
3.2.		Classificatiesysteem	Rangschikking of verdeling van onderwerpen in groepen, gebaseerd op gemeenschappelijke kenmerken van de onderwerpen.
3.3.		Sectoren	Belangrijkste economische of andere sectoren die in de statistieken zijn opgenomen.
3.4.		Statistische begrippen en definities	Statistische kenmerken van statistische waarnemingen.
3.5.		Statistische eenheid	Eenheid waarvoor informatie wordt gevraagd en waarvoor uiteindelijk statistieken worden opgesteld.
3.6.		Statistische populatie	Het totale lidmaatschap of de populatie van een bepaalde klasse van mensen, onderwerpen of gebeurtenissen.

Nummer	Begrip	Subbegrip	Omschrijving
3.7.		Referentiegebied	Het land of geografische gebied waarop het gemeten statistische fenomeen betrekking heeft.
3.8.		Bestreken tijdvak	De lengte van de periode waarvoor gegevens beschikbaar zijn.
3.9.		Basisperiode	De periode die als basis van een indexcijfer wordt gebruikt of waarop een constante reeks betrekking heeft.
4.	Meereenheid		De eenheid waarin de gegevenswaarden worden gemeten.
5.	Referentieperiode		De periode of het tijdstip waarop de gemeten waarneming betrekking moet hebben.
6.	Institutioneel mandaat		Reeks voorschriften of een andere formele reeks instructies waarmee aan een organisatie de verantwoordelijkheid en de bevoegdheid worden verleend voor de verzameling, verwerking en verspreiding van statistieken.
6.1.		Wetbesluiten en andere overeenkomsten	Wetbesluiten of andere formele of informele overeenkomsten waarmee aan een agentschap de verantwoordelijkheid en de bevoegdheid voor de verzameling, verwerking en verspreiding van statistieken worden verleend.
6.2.		Gegevensuiterwisseling	Regelingen of procedures voor gegevensuiterwisseling en coördinatie tussen agentschappen die gegevens produceren.
7.	Vertrouwelijkheid		Kenmerk van gegevens dat aangeeft in welke mate ongeoorloofde bekendmaking ervan nadelig of schadelijk kan zijn voor de bron of voor andere betrokken partijen.
7.1.		Vertrouwelijkheid — beleid	Wetgevingsmaatregelen of andere formele procedures ter voorkoming van ongeoorloofde bekendmaking van gegevens die een directe of indirecte identificatie van een persoon of een economische eenheid mogelijk maken.
7.2.		Vertrouwelijkheid — gegevensverwerking	Toegepaste voorschriften voor de verwerking van de gegevens om statistische vertrouwelijkheid te verzekeren en ongeoorloofde bekendmaking te voorkomen.
8.	Bekendmakingsbeleid		Voorschriften voor de verspreiding van statistische gegevens onder belanghebbenden.
8.1.		Bekendmakingstijdschema	Het tijdschema voor de bekendmaking van statistische gegevens.
8.2.		Toegang tot bekendmakingstijdschema	Toegang tot de informatie van het bekendmakingstijdschema.
8.3.		Toegang voor gebruikers	Beleid voor de bekendmaking van de gegevens aan de gebruikers, het verspreidingsbereik (bv. openbaarheid, geselecteerde gebruikers), wijze waarop de gebruikers over de bekendmaking van gegevens worden ingelicht, en of het bekendmakingsbeleid de verspreiding van statistische gegevens onder alle gebruikers regelt.
9.	Verspreidingsfrequentie		Het tijdsinterval waarmee de statistieken in een bepaalde periode worden verspreid.
10.	Verspreidingsformaat		Gegevensdragers waarop de statistische gegevens en metagegevens worden verspreid.
10.1.		Persbericht	Regelmatige of ad-hocpersberichten die met de gegevens verband houden.
10.2.		Publicaties	Regelmatige of ad-hocpublicaties waarin de gegevens beschikbaar worden gesteld aan het publiek.
10.3.		Onlinedatabank	Informatie over onlinedatabanken waarin de verspreide gegevens toegankelijk zijn.

Nummer	Begrip	Subbegrip	Omschrijving
10.4.		Toegang tot microgegevens	Informatie over de vraag of ook microgegevens worden verspreid.
10.5.		Overig	Vermeldingen van de belangrijkste andere verspreide gegevens.
11.	Toegankelijkheid van de documentatie		
11.1.		Documentatie over methoden	Beschrijving van en verwijzingen naar de beschikbare documenten over de methoden.
11.2.		Documentatie over de kwaliteit	Documentatie over de gebruikte procedures voor kwaliteitsbeheer en -beoordeling.
12.	Kwaliteitsbeheer		Systemen en kaders die in een organisatie worden gebruikt om de kwaliteit van statistische producten en processen te beheren.
12.1.		Kwaliteitsborging	Alle genomen systematische maatregelen die aantoonbaar het vertrouwen vergroten dat de processen tot een statistische output leiden die aan de vereisten voldoet.
12.2.		Kwaliteitsbeoordeling	Algemene beoordeling van gegevenskwaliteit, gebaseerd op standaardkwaliteitscriteria.
13.	Relevantie		De mate waarin statistieken voldoen aan de huidige en potentiële behoeften van de gebruikers.
13.1.		Behoeften van de gebruikers	Beschrijving van de gebruikers en hun respectieve behoeften met betrekking tot de statistische gegevens.
13.2.		Tevredenheid van de gebruikers	Maatregelen om de tevredenheid van de gebruikers vast te stellen.
13.3.		Volledigheid	De mate waarin alle benodigde statistieken beschikbaar zijn.
14.	Nauwkeurigheid en betrouwbaarheid		Nauwkeurigheid: de mate waarin de berekeningen of schattingen in de buurt van de exacte of werkelijke waarden van de variabelen liggen. Betrouwbaarheid: de mate waarin de aanvankelijk geschatte waarde overeenkomt met de latere geschatte waarde.
14.1.		Algehele nauwkeurigheid	Beoordeling van de nauwkeurigheid van een bepaalde gegevensset of een gebied, als samenvatting van de verschillende onderdelen.
14.2.		Steekproeffout	Dat deel van het verschil tussen een waarde van een populatie en een schatting daarvan, afgeleid uit een aselechte steekproef, dat te wijten is aan het feit dat slechts een subset van de populatie in aanmerking is genomen.
14.3.		Niet-steekproeffout	Fout in de steekproefschattingen die niet aan steekproeffluctuaties te wijten zijn.
15.	Tijdigheid en punctualiteit		
15.1.		Tijdigheid	Tijdsduur tussen beschikbaarheid van gegevens en de gebeurtenis of het verschijnsel waarop deze betrekking hebben.
15.2.		Sciptheid	Tijdspanne tussen de datum van publicatie van de gegevens en de datum waarop zij hadden moeten worden gepubliceerd.

Nummer	Begrip	Subbegrip	Omzrijving
16.	Vergelijkbaarheid		Meting van het effect van verschillen in de toegepaste statistische begrippen, meetinstrumenten en procedures wanneer statistieken tussen geografische gebieden of in de tijd worden vergeleken.
16.1.		Vergelijkbaarheid — geografisch	De mate waarin statistieken vergelijkbaar zijn tussen geografische gebieden.
16.2.		Vergelijkbaarheid in de tijd	De mate waarin statistieken vergelijkbaar of verenigbaar zijn in de tijd.
17.	Coherentie		De mate waarin statistieken betrouwbaar op verschillende manieren en voor verschillende doeleinden kunnen worden gecombineerd.
17.1.		Gebiedsoverschrijdende coherentie	De mate waarin statistieken vergelijkbaar of verenigbaar zijn met de statistieken die door andere gegevensbronnen of statistiekgebieden zijn verkregen.
17.2.		Interne coherentie	De mate waarin statistieken consistent zijn binnen een bepaalde gegevensset.
18.	Kosten en lasten		Kosten in verband met de verzameling en de productie van een statistisch product en lasten voor de respondenten.
19.	Herziening van de gegevens		Elke wijziging in een waarde van een statistiek die al aan het publiek is bekendgemaakt.
19.1.		Herziening van de gegevens — beleid	Beleid om de transparantie van de verspreide gegevens te verzekeren, waarbij voorlopige gegevens worden samengesteld die later worden herzien.
19.2.		Herziening van de gegevens — praktijk	Informatie over de praktijk van de gegevensherziening.
20.	Statistische verwerking		
20.1.		Brongegevens	Kenmerken en onderdelen van de ruwe statistische gegevens die zijn gebruikt voor de opstelling van statistische aggregaten.
20.2.		Frequentie van verzameling van gegevens	Frequentie waarmee de brongegevens worden verzameld.
20.3.		Verzameling van gegevens	Systematisch proces van verzameling van gegevens voor officiële statistieken.
20.4.		Validatie van gegevens	Proces van toezicht houden op de resultaten van gegevensopstelling en verzekeren van de kwaliteit van de statistische resultaten.
20.5.		Samenstelling van gegevens	Bewerking van gegevens om volgens bepaalde regels nieuwe informatie af te leiden.
20.6.		Aanpassing	Reeks procedures tot wijziging van statistische gegevens om een aanpassing aan de nationale of internationale normen mogelijk te maken of om bij de opstelling van specifieke gegevenssets verschillen in de gegevenskwaliteit weg te nemen.
21.	Opmerking		Aanvullende beschrijvende tekst die aan de gegevens of metagegevens kan worden toegevoegd.