

## Richtlijnen voor On Site/ Remote Access -output

### **Achtergrond**

Sinds januari 2004 is de 'Wet op het Centraal Bureau voor de Statistiek' van kracht.

#### **CBS-wet (verkorte weergave)**

##### Artikel 37

De door het CBS verzamelde gegevens worden uitsluitend gebruikt voor statistische doeleinden en niet verstrekt aan anderen dan degenen die belast zijn met de uitvoering van de taak van het CBS.

Vervolgens worden de gegevens slechts zodanig openbaar gemaakt dat daaraan geen herkenbare gegevens over een afzonderlijk persoon, huishouden, onderneming of instelling kunnen worden ontleend.

##### Artikel 41

In afwijking van artikel 37 kan de directeur-generaal op verzoek, ten behoeve van statistisch of wetenschappelijk onderzoek, een verzameling van gegevens met betrekking tot het gebruik waarvan passende maatregelen zijn genomen om herkenning van afzonderlijke personen, huishoudens, ondernemingen of instellingen te voorkomen, verstrekken aan een dienst, organisatie of instelling, dan wel daartoe toegang verlenen.

##### Artikel 42

De directeur-generaal willigt een verzoek als bedoeld in artikel 41 slechts in, indien de verzoeker naar het oordeel van de directeur-generaal voldoende maatregelen heeft getroffen om te voorkomen dat de verzameling van gegevens voor andere doeleinden dan statistisch of wetenschappelijk onderzoek wordt gebruikt.

*(zie bijlage 3 voor de volledige tekst)*

Naast deze wettelijke verplichting staat voorop dat het beeld dat de buitenwereld van het CBS heeft er één moet zijn van absolute betrouwbaarheid. Dit is ook in het belang van het CBS zelf. De respons op enquêtes is namelijk afhankelijk van de belofte van het CBS dat individuele gegevens bij het CBS in veilige handen zijn. Wie wil er immers nog persoonlijke informatie verschaffen als er niet extreem zorgvuldig met deze data wordt omgesprongen?

Het is dus van groot belang dat data goed wordt beveiligd. Voor fysieke en juridische beveiliging wordt gezorgd via afscherming van de On Site/ Remote Access- (OS-/RA-) werkomgeving en de OS-/RA-contracten. Daarnaast is ook statistische beveiliging nodig om te voorkomen dat individuele gegevens *onthuld* worden, dat wil zeggen herkenning van een eenheid én bekendmaking van nadere gegevens over die eenheid. Onderstaande richtlijnen hebben betrekking op het voorkómen van onthulling.

#### **Doelstelling**

Het doel van deze richtlijnen is duidelijkheid te scheppen over welke OS-/RA- output is toegestaan. Dit biedt houvast aan zowel de OS-/RA-onderzoekers als de outputcontroleurs.

Het groeiende gebruik van CBS microdata services en de verdere professionalisering van het Centrum voor Beleidsstatistiek maken eenduidige richtlijnen noodzakelijk om een goede dienstverlening van microdata services veilig te stellen.

Enkele opmerkingen vooraf:

- Het voorkomen van onthulling valt onder gedeelde verantwoordelijkheid van het CBS en de OS-/RA-onderzoeker die met de CBS-microdata<sup>1</sup> werkt. Zie hiervoor ook de geheimhoudingsverklaring.
- De outputcontroleurs beoordelen elke output afzonderlijk op onthulling. Of er in combinatie met eerder vrijgegeven output (ongeacht project of instelling) een risico op onthulling ontstaat, is de verantwoordelijkheid van de OS-/RA-onderzoeker.
- De beveiligingsregels gelden voor datasets over zowel personen als bedrijven.
- Voor een project worden alleen die bestanden beschikbaar gesteld die nodig zijn voor het beantwoorden van de onderzoeksvraag. Het spreekt voor zich dat aangeboden outputs ook gerelateerd moeten blijven aan de onderzoeksvraag die ten grondslag ligt aan het project.
- Let op! Wanneer gebruik wordt gemaakt van CBS microdatabestanden is het niet toegestaan om aan de hand van deze gegevens individuele kwaliteitscontroles uit te voeren. Dit houdt in dat microdata buiten het CBS niet verrijkt (aangepast/verbeterd) mag worden met behulp van CBS gegevens.

In de volgende paragrafen worden drie aspecten van de richtlijnen besproken. In de eerstvolgende paragraaf worden de *beveiligingsregels* besproken om te voorkomen dat er onthulling plaatsvindt. Daarna wordt aandacht besteed aan de *leesbaarheid* van de output, zodat deze goed te controleren is door de outputcontroleurs. Als laatste volgen nog enkele *technische* richtlijnen.

## 1. Beveiligingsregels

*Deze beveiligingsregels gelden als vuistregels. Als de OS-/RA-onderzoeker zich hieraan houdt, mag hij redelijkerwijs verwachten dat de output vrijgegeven wordt.*

*Als een OS/RA-onderzoeker echter een output wil aanbieden die de vuistregels schendt, dan ligt de bewijslast in eerste instantie bij de onderzoeker om aan te tonen dat de output desondanks geen onthullingsrisico met zich meebrengt. In beide gevallen heeft de outputcontroleur te allen tijde het eindoordeel.*

*Deze normen bij de vuistregels zijn bewust aan de veilige kant gekozen omdat daarmee een aantal veel voorkomende beveiligingsrisico's bij voorbaat (bijna volledig) uit te sluiten. Als de vuistregels soepeler zijn, zal de OS/RA onderzoeker veel meer achtergrondinformatie bij iedere output moeten aanleveren om een goede controle mogelijk te maken.*

Wij hanteren de volgende beveiligingsregels:

*Algemene regels:*

0. Geen microdata: Het is niet toegestaan om microdata als output aan te bieden. Dit betekent onder andere dat:
  - het (SPSS) commando LIST cases niet mag worden gebruikt in de output.
  - geen individuele koppelsleutels opgenomen mogen worden in de output, zoals RIN, RINADRES, BEID, BRIN.
1. Minimum aantal waarnemingen: Alle tabellen en soortgelijke output bevat **tenminste 10** eenheden (ongewogen) als grondslag voor elke cel of datapunt.

Opmerkingen hierbij:

- Wanneer data op zowel bedrijven als personen betrekking hebben, geldt dit minimum voor zowel het aantal bedrijven als het aantal personen.

---

<sup>1</sup> Microdata: data op het niveau van enquêtes of registraties.

- Indien de tabellen percentages of opgehoogde aantallen bevatten, moeten de ongewogen aantallen ook opgenomen worden om de controle van deze regel mogelijk te maken.
- Maxima en minima worden dus niet vrijgegeven, aangezien zij normaal gesproken refereren aan één eenheid.

*Aanvullend voor modellen:*

2. Modellen: Alle gemodelleerde output moet **minstens 10** vrijheidsgraden overhouden, waarbij het aantal vrijheidsgraden gelijk is aan: aantal waarnemingen  $\bar{\cdot}$  aantal parameters  $\bar{\cdot}$  andere modelrestricties.
  - Dit geldt in het bijzonder voor regressiemodellen, hogere orde momenten (variantie, covariantie, kurtosis, skewness), samenvatting- en teststatistieken van schattingen ( $R^2$  en variaties hierop), geschatte variantie, informatiecriteria (bijv. AIC, BIC) of individuele en groepstesten en -statistieken (bijv. t, F, chi-kwadraat, Wald, Hausman)

Verder geldt dat residuen en grafieken/plots van residuen niet worden vrijgegeven.

*Aanvullend voor frequentietabellen<sup>2</sup>:*

3. Groepsonthulling: In alle frequentietabellen en soortgelijke output mag geen enkele cel meer dan 90% van het totaal aantal eenheden in de rij of kolom bevatten. Op deze manier wordt voorkomen dat bepaalde variabelen in een tabel een herkenbare groep definiëren, waar andere variabelen in de tabel informatie onthullen die geldig is voor elk lid van de groep. Hoewel er geen individuele eenheid kan worden herkend, is de vertrouwelijkheid geschonden, omdat de informatie geldig is voor vrijwel elk lid van de groep en de groep als zodanig herkenbaar is.

*Aanvullend voor kwantitatieve tabellen<sup>3</sup>:*

4. Dominantie: In alle kwantitatieve tabellen en soortgelijke gegevens mag de grootste bijdrager aan een cel niet meer dan 50% van het celtotaal bijdragen<sup>4</sup>.

In bijlage 1 worden bovenstaande vuistregels verder toegelicht per type output. Hierbij zijn ook voorbeelden opgenomen die deze richtlijnen verduidelijken.

## 2. Regels voor leesbaarheid

*Let op! Aan deze regels moet altijd worden voldaan.*

Uitgangspunt voor onderstaande richtlijnen voor de leesbaarheid is dat de OS-/RA-onderzoeker ervoor dient te zorgen dat de aangeleverde output in combinatie met het standaardformulier outputcontrole begrijpelijk is voor de outputcontroleurs.

De output wordt gecontroleerd door twee statistisch onderzoekers. Hierdoor wordt een onafhankelijke outputcontrole gewaarborgd met criteria gelijk voor alle OS-/RA-onderzoekers. De controleurs zijn weliswaar ervaren, maar doorgaans niet inhoudelijk bekend met de OS-/RA-projecten. Ook weten de controleurs vanwege de grote diversiteit aan onderwerpen die ze voorgeschiedt krijgen niet van alle datasets alle in-en-outs.

---

<sup>2</sup> *Frequentietabel: tabel waarin alle waarnemingen staan waarin je kunt aflezen hoe vaak (frequentie) een bepaalde waarneming voorkomt.*

<sup>3</sup> *Kwantitatieve tabel: tabel met de gezamenlijke hoeveelheid (kwantiteit) van waarnemingen zoals loonsom of omzet.*

<sup>4</sup> *Het genoemde percentage is strikter dan het CBS zelf toepast omdat er sprake is van onthullingsrisico bij het noemen hiervan (mogelijkheid van terugrekenen van onthullende waardes).*

Dit betekent dat het van groot belang is dat de OS-/RA-onderzoeker een heldere toelichting geeft op de analyses en de gepresenteerde variabelen. Het zij duidelijk dat de onderzoeker hierdoor zelf invloed heeft op snelheid en uitkomst van de outputcontrole.

Wij hanteren de volgende richtlijnen:

1. Outputbestanden dienen vergezeld te gaan van een volledig en correct ingevuld **Standaardformulier Outputcontrole**. Dit formulier bevindt zich in de 'shared workspace' (H schijf) van de gebruiker en moet worden meegeleverd bij iedere output die ter controle wordt aangeboden. Hierin moet worden beschreven:
  - wat het doel van de analyses is;
  - welke bronbestanden gebruikt zijn;
  - hoe het onderzoeksbestand is samengesteld (bijvoorbeeld door koppeling van verschillende bestanden of door subpopulaties te selecteren);
  - op welke populatie de output betrekking heeft;
  - waaruit de output bestaat, zoals tellingen, kruistabellen, modellen of analyses;
  - wat de relatie is tussen tabellen binnen een outputbestand; (indien van toepassing)
  - wat de relatie is met eerder aangeleverde output; (indien van toepassing)

*Als voorbeeld is in bijlage 3 een volledig ingevuld outputformulier bijgesloten.*

2. Variabelen in de output dienen gedocumenteerd te zijn. Hiermee bedoelen we dat er een duidelijke omschrijving moet zijn van de betekenis van de variabele en zijn afzonderlijke waarden in het geval van discrete variabelen. In SPSS is dit mogelijk door het toevoegen van value labels aan de variabelen. Een andere mogelijkheid is het aanmaken van een apart document waarin deze gegevens worden vermeld. Dit document kan bij elke output worden meegeleverd en steeds aangevuld worden met nieuwe variabelen.
3. Elk onderdeel van de output moet voorafgegaan worden door een regel tekst waarin de populatie of selectie voor dat onderdeel wordt beschreven. Dus voor elke tabel, modelanalyse, enz. Het is niet voldoende om te verwijzen naar de voorafgaande programmacode, aangezien deze niet altijd direct begrijpbaar zijn.
4. Indien wordt afgeweken van het minimum aantal waarnemingen (10) is het bij *groepsonthulling* of *dominantie* noodzakelijk de achterliggende tabellen mee te leveren.

In dat geval dienen deze aangeboden te worden in een **apart zipbestand** met de naam "**GEENOUTPUT\_Achtergrondinformatie.zip**". In deze documentatie moet per tabel duidelijk beschreven worden voor welk onderdeel van de output deze achterliggende tabel dient. Dit kan bijvoorbeeld door eenzelfde nummering aan te brengen in de output en de achtergrondinformatie. Dit zipbestand wordt uiteraard niet vrijgegeven.

### 3. Technische richtlijnen

*Let op! Aan deze richtlijnen moet altijd worden voldaan.*

1. Grootte van de output: Om de outputs controleerbaar te houden mag een output niet te groot zijn. Wij houden als richtlijn aan dat een output binnen één uur gecontroleerd kan worden. De tarieven voor outputcontroles zijn hier ook op gebaseerd. Wanneer een output te groot is, zullen de kosten en doorlooptijd voor het controleren naar rato verhoogd worden. Wij houden de volgende richtlijnen aan:
  - Er mogen maximaal 10 bestanden tegelijk als output aangeboden worden.
  - Ons uitgangspunt is dat we maximaal 3 000 waarden controleren. Bijvoorbeeld maximaal 30 tabellen van gemiddeld 10 bij 10 of maximaal 50 modellen.
  - Omdat de output per mail wordt doorgestuurd is de grootte van een zipbestand gemaximeerd tot 5 Mb (meer kan een mailservers vaak niet aan).

- Noodzakelijke achtergrondinfo uit het zipbestand “GEENOUTPUT Achtergrondinformatie.zip” telt uiteraard niet mee in de grootte van de output.
- Bied geen zipfiles in een zipfile aan.

2. Formaat:

- Graag alle output, anders dan SPSS omzetten naar één van de volgende formaten: xls, doc, txt of csv.
- Bestanden in microdata-formaat (zoals sav, sas of mdb) zijn niet toegestaan.
- Grafieken zijn niet toegestaan als output. Deze zijn namelijk veelal lastig te controleren op onthulling. De onderliggende gegevens kunnen als output worden aangeboden, zodat na te zijn vrijgegeven, de onderzoeker er in zijn eigen omgeving grafieken van kan maken. Uitzondering vormen grafieken van modellen zoals survival curves. Deze grafieken dienen dan wel als platte figuren aangeleverd te worden in formaten zoals bmp, jpg of gif.

3. Zichtbaarheid: Verberg geen output in de outputbestanden. Zo is het niet toegestaan:

- in SPSS output-items in te klappen.
- in Excel kolommen, rijen, tabbladen te verbergen.
- draaitabellen ter controle aan te bieden.

4. Overig:

- Het is niet toegestaan dezelfde output in verschillende formaten aan te bieden.
- De naam van de aangeboden output mag niet meer dan 100 posities bevatten, dit kan namelijk problemen geven met het overhalen van de output uit de beveiligde OS/RA omgeving.

## Bijlage 1: Uitwerking van Beveiligingsregels met voorbeelden

In deze bijlage worden de beveiligingsregels zoals genoemd in het document “Richtlijnen voor On Site/ Remote Access- output” verder uitgewerkt voor de verschillende type outputs. Daarnaast wordt ter verduidelijking per beveiligingsregel een voorbeeld gegeven.

De volgende beveiligingsregels worden in bovengenoemd document beschreven:

### Algemene regels:

0. Geen microdata: Het is niet toegestaan om microdata als output aan te bieden.
1. Minimum aantal waarnemingen: Alle output bevat **tenminste 10** eenheden (ongewogen) als grondslag voor elke cel of datapunt.

### Aanvullend voor modellen:

2. Modellen: Alle gemodelleerde output moet minstens 10 vrijheidsgraden hebben, waarbij het aantal vrijheidsgraden gelijk is aan: aantal waarnemingen  $\cdot$  aantal parameters  $\cdot$  andere modelrestricties. Verder geldt dat residuen en grafieken/plots van residuen niet vrijgegeven worden.

### Aanvullend voor frequentietabellen<sup>5</sup>:

3. Groepsonthulling: In alle tabellen en soortgelijke output mag geen enkele cel meer dan 90% van het totaal aantal eenheden in de rij of kolom bevatten om groepsonthulling te voorkomen.

### Aanvullend voor kwantitatieve tabellen<sup>6</sup>:

4. Dominantie: In alle tabellen en soortgelijke gegevens mag de grootste bijdrager aan een cel niet meer dan 50% van het celtotaal bijdragen<sup>7</sup>.

De eerste genoemde regel, geen microdata, is uiteraard van toepassing op elke output. Hieronder een tabel waarin per type output wordt aangegeven welke regels van toepassing zijn.

Type statistiek	Type output	Beveiligings-regel(s) van toepassing
<b>Beschrijvende statistieken</b>	Frequentietabellen	1, 3
	Kwantitatieve tabellen en percentielen	1, 3, 4
	Maxima, minima	Niet vrijgeven
	Modus, mediaan	1, 3
	Gemiddelden, indices, verhoudingen, indicatoren	1, 3, 4
	Concentratieverhoudingen	1, 3, 4
	Verdelingsmaten (incl. variantie, covariantie, kurtosis, scheefheid)	1, 2
	Grafieken: grafische weergave van data	Niet vrijgeven
<b>Correlaties en Regressie analyses</b>	Lineaire regressiecoëfficiënten	1, 2
	Niet-lineaire regressiecoëfficiënten	1, 2
	Schattingsresiduen	Niet vrijgeven
	Samenvattings- en teststatistieken van schattingen ( $R^2$ , $\chi^2$ enz)	1, 2
	Correlatiecoëfficiënten	1

<sup>5</sup> *Frequentietabel*: tabel waarin alle waarnemingen staan waaruit je kunt aflezen hoe vaak (frequentie) een bepaalde waarneming voorkomt.

<sup>6</sup> *Kwantitatieve tabel*: tabel met de gezamenlijke hoeveelheid (kwantiteit) van waarnemingen zoals loonsom of omzet.

<sup>7</sup> *Het genoemde percentage is strikter dan het CBS zelf toepast omdat er sprake is van onthullingsrisico bij het noemen hiervan (mogelijkheid van terugrekenen van onthullende waardes).*

## Voorbeelden

### 1. Minimum aantal waarnemingen

Elke cel in een frequentie- of kruistabel moet minimaal 10 waarnemingen bevatten. Als je toch iets mee wil nemen met lagere aantallen, moet beargumenteerd worden waarom er geen onthulling plaatsvindt.

In voorbeeld 1a is een kruistabel weergegeven met veel kleine aantallen waarin onthulling plaatsvindt. Na het zien van deze tabel weet je als inwoner van Zuiderwijk aan Zee dat het enige gezin met 28 kinderen in 2011 geen huisdieren heeft.

In dit geval kan je ervoor zorgen dat de tabel niet meer onthullend is door de laatste 2 categorieën van het aantal kinderen samen te voegen (1b) of cellen te onderdrukken (1c). Bij het onderdrukken van cellen moet je er tevens op letten dat de aantallen niet terug te rekenen zijn met informatie uit de andere cellen.

#### Voorbeeld 1a

Tabel: Huishoudens naar aantal kinderen en huisdieren in Zuiderwijk aan Zee, 2011.

Selectie: Huishoudens met minimaal 1 kind.

		Aantal huisdieren			
		0	1	2 of meer	Totaal
Aantal kinderen	1-3	32	64	84	180
	4-10	28	48	28	104
	11-20	10	16	10	36
	>20	1	0	0	1
	Totaal	71	128	122	321

#### Voorbeeld 1b

Tabel: Huishoudens naar aantal kinderen en huisdieren in Zuiderwijk aan Zee, 2011.

Selectie: Huishoudens met minimaal 1 kind.

		Aantal huisdieren			
		0	1	2 of meer	Totaal
Aantal kinderen	1-3	32	64	84	180
	4-10	28	48	28	104
	>11	11	16	10	37
	Totaal	71	128	122	321

Voorbeeld 1c

Tabel: Huishoudens naar aantal kinderen en huisdieren in Zuiderwijk aan Zee, 2011.

Selectie: Huishoudens met minimaal 1 kind.

De cellen die onthullend waren zijn in deze tabel onderdrukt. Hierdoor wordt voor bijvoorbeeld het gezin met meer dan 20 kinderen geen extra informatie vrijgegeven.

		Aantal huisdieren			
		0	1	2 of meer	Totaal
Aantal kinderen	1-3	32	64	84	180
	4-10	28	48	28	104
	11-20	10	16	10	36
	>20	x	x	x	x
	Totaal	x	x	x	x

2. Modellen:

Aan de hand van het volgende voorbeeld kunnen de richtlijnen voor modellen worden toegelicht. Het model moet minimaal 10 vrijheidsgraden ( degrees of freedom = df ) hebben, met andere woorden het aantal waarnemingen minus het aantal variabelen en/of restricties moet groter of gelijk zijn dan 10.

In het onderstaande voorbeeld van een output bedoelen wij de df bij "Residual" die groter of gelijk moet zijn aan 10 en juist niet de df bij "Model". Het totale aantal waarnemingen in dit voorbeeld is namelijk 16725, het aantal modelvariabelen is 3 en het model heeft daarmee 16722 vrijheidsgraden.

De residuen zelf mogen niet in de output opgenomen worden, want deze geven informatie per waarneming. Deze kunnen worden opgevraagd als grafiek bij het uitvoeren van de regressie. Deze grafiek mag op On Site of Remote Access bekeken worden, maar niet worden meegenomen als output.

Voor de onderliggende gegevens waarop het model zijn schattingen doet, blijft regel 1 van kracht: een minimum aantal waarnemingen van 10.

Voorbeeld 2a

*Regressieanalyse om effect van leeftijd en opleiding op het natuurlijk logaritme van het uurloon van vrouwen te bekijken.*

Selectie: Vrouwelijke respondenten van de EBB in 2008 en 2009

Source	SS	df	MS
-----+-----			
Model	250603.002	3	21514.71
Residual	2407.30303	<b>16722</b>	0.079
-----+-----			
Total	253010.306	16725	8.526



Number of obs = 17829  
 F( 7, 17822) =  
 Prob > F = 0.0000  
 R-squared = 0.9906  
 Adj R-squared = 0.9906  
 Root MSE = .2810

---

log_uurloon	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
const	.93496	.036472	25.64	0.000	.863472	1.00645
opleiding	.05305	.000665	79.81	0.000	.051746	.054352
leeftijd	.04953	.001742	28.43	0.000	.046117	.052947

---

### 3. Groepsontulling:

Als voorbeeld bekijken we een frequentietabel over het drugsgebruik onder de jeugd in Zuiderwijk aan Zee.

Tabel 3: Inwoners van dorp A naar leeftijd en drugsgebruik

Drugsgebruik Leeftijd	Geen	Alleen softdrugs	Harddrugs (evt. ook softdrugs)	Totaal
< 15 jaar	78	12	13	103
15 – 17 jaar	2	367	12	381
18 – 21 jaar	-	28	389	417
Totaal	80	407	414	901

De lage frequentie in de kolom 'geen' in de leeftijdscategorieën 15 jaar en ouder is onthullend, omdat hieruit volgt dat bijna iedereen in deze leeftijd dus wél drugs gebruikt. Als een vader van een jongen van 19 in dit dorp deze tabel onder ogen krijgt, weet hij zeker dat zijn zoon drugs gebruikt. Dit maakt de tabel zeer onthullend.

### 4. Dominantie:

Als voorbeeld kijken we naar de investeringen van ondernemingen naar regio en activiteit.

Tabel 4: Investerings van ondernemingen (mln euro) naar regio en activiteit.

Selectie: Alle bedrijven in de Prodcorn statistiek uit de periode 1993-2007

	Regio 1	Regio 2	Regio 3	Totaal
Activiteit 1	78	12	9	99
Activiteit 2	4	367	12	383
Activiteit 3	-	28	389	417
Totaal	82	407	410	899

NB: Mocht het voor het onderzoek noodzakelijk zijn, is het mogelijk om tabellen met aantal waarnemingen onder de 10 aan te bieden. Uiteraard moet hierbij de noodzaak voor het niet houden aan het minimum aantal waarnemingen worden uitgelegd, evenals de argumentatie dat deze output desondanks niet tot onthulling leidt.

In dit voorbeeld zijn daarnaast onderstaande achtergrondtabellen noodzakelijk ter controle van de dominantieregel. Deze achtergrondinformatie dient aangeboden te worden in een **apart zipbestand** met de naam "GEENOUTPUT Achtergrondinformatie.zip" (zie ook de regels voor leesbaarheid).

Tabel 4a: Totaal aantal bijdragers per cel

Deze tabel hoort als achtergrondinfo bij tabel 4 uit de output dd 31-3-2011, bestand xxx.log

Totaal aantal bijdragers per cel

	Regio 1	Regio 2	Regio 3	Totaal
Activiteit 1	14	10	12	36
Activiteit 2	13	15	11	39
Activiteit 3	0	21	23	44
Totaal	27	46	46	119

Tabel 4b: Omvang van grootste bijdrager per cel (mln euro)

Deze tabel hoort als achtergrondinfo bij tabel 4 uit de output dd 31-3-2011, bestand xxx.log

	Regio 1	Regio 2	Regio 3	Totaal
Activiteit 1	34	4	2	
Activiteit 2	1	167	5	
Activiteit 3	-	10	234	
Totaal				

Tabel 4c: Procentuele bijdrage van grootste bijdrager per cel

Deze tabel hoort als achtergrondinfo bij tabel 4 uit de output dd 31-3-2011, bestand xxx.log

	Regio 1	Regio 2	Regio 3	Totaal
Activiteit 1	44%	33%	22%	
Activiteit 2	25%	46%	42%	
Activiteit 3	-	36%	60%	
Totaal				

Uit tabel 4a blijkt dat deze output voldoet aan regel 1: Alle output bevat tenminste 10 eenheden (ongewogen) als grondslag voor elke cel of datapunt.

Uit tabel 4c blijkt dat cel (Activiteit 3, Regio 3) niet voldoet aan de dominantie regel. Alle bedrijven die zijn gevestigd in Regio 3 investeren gezamenlijk 389 mln in activiteit 3. Het bedrijf met de meeste investering (234 mln) maakt hierbij 60% van het totaal uit. Daarmee zit het boven de 50% en domineert dit bedrijf deze cel waardoor de kans op onthulling te groot is.

## Bijlage 2: STANDAARDFORMULIER BESCHRIJVING OUTPUTCONTROLE<sup>8</sup>

Naam <i>[naam gebruiker]</i>	Onno Site
Instelling <i>[naam instelling]</i>	Universiteit van Den Haag
Datum <i>[datum waarop output is aangemaakt]</i>	30-04-2011
Project <i>[contractnummer + naam onderzoek]</i>	Projectnummer 9999 ‘Doorstroom van uitkeringsgerechtigden naar de arbeidsmarkt’
Data <i>[gebruikte bronbestanden]</i>	SSBBanen, SSB_ABW, SSB_AO, SSB_WW, bijbehorende GBA-selectiebestanden  Baanbestand over het jaar 2008, overige bestanden over het jaar 2007
Onderzoeksbestand <i>[beschrijving van samengesteld onderzoeksbestand]</i>	Het onderzoeksbestand bestaat uit alle WW-, AO- en Bijstandsgerechtigden op ultimo december 2007, met daaraan gekoppeld persoonsgegevens op hetzelfde peilmoment en baangegevens over 2008.
Doel <i>[omschrijving doel van de analyses]</i>	Het doel is om in kaart te brengen hoeveel uitkeringsgerechtigden binnen een jaar doorstromen naar een baan. Spelen persoonskenmerken een rol of duur van de uitkering?
Relatie eerder aangeleverde output <i>[beschrijving van de relatie met eerder aangeleverde output]</i>	In de vorige output keek ik alleen naar WW-uitkeringen. Nu dezelfde exercitie voor WW+ABW+AO.
Inhoud output <i>[beschrijving van de output, bijvoorbeeld komen er bijvoorbeeld tellingen, kruistabellen of analyses<sup>9</sup> in voor]</i>	Excelbestand met verschillende kruistabellen: - tabellen naar soort uitkering - tabellen naar persoonskenmerken
Eventuele selecties in de output heb ik duidelijk aangegeven bij de analyses waarop deze van toepassing zijn. <i>[beschrijving van gemaakte selecties]</i>	- personen zijn terug te vinden in GBA (SRTNM='R') - personen ontvangen een WW-, AO- of bijstandsuitkering op 31 december 2007 - in Excel staat bij elke tabel aangegeven of er een extra selectie is gemaakt en zo ja, welke.

<sup>8</sup> Als blijkt dat dit formulier niet goed of niet volledig is ingevuld wordt de output niet in behandeling genomen.

<sup>9</sup> Indien er regressie-analyses in de output voorkomen dan graag de onderliggende kruistabellen of de afzonderlijke frequentietabellen meesturen.

## **Bijlage 3: CBS Wet**

### **Artikel 37**

1. De door de directeur-generaal in het kader van de uitoefening van de taken ter uitvoering van deze wet ontvangen gegevens worden uitsluitend gebruikt voor statistische doeleinden.
2. De in het eerste lid bedoelde gegevens worden niet verstrekt aan anderen dan degenen die belast zijn met de uitvoering van de taak van het CBS.
3. De in het eerste lid bedoelde gegevens worden slechts zodanig openbaar gemaakt dat daaraan geen herkenbare gegevens over een afzonderlijk persoon, huishouden, onderneming of instelling kunnen worden ontleend, tenzij, ingeval het gegevens met betrekking tot een onderneming of instelling betreft, er een gegronde reden is om aan te nemen dat bij de betrokken onderneming of instelling geen bedenkingen bestaan tegen de openbaarmaking.

### **Artikel 41**

1. In afwijking van artikel 37 kan de directeur-generaal op verzoek, ten behoeve van statistisch of wetenschappelijk onderzoek, een verzameling van gegevens met betrekking tot het gebruik waarvan passende maatregelen zijn genomen om herkenning van afzonderlijke personen, huishoudens, ondernemingen of instellingen te voorkomen, verstrekken aan een dienst, organisatie of instelling als bedoeld in het tweede lid, dan wel daartoe toegang verlenen.
2. Een verzameling van gegevens als bedoeld in het eerste lid kan worden verstrekt, dan wel daartoe kan toegang worden verleend aan:
  - a. een universiteit in de zin van de Wet op het hoger onderwijs en wetenschappelijk onderzoek;
  - b. een bij wet ingestelde organisatie of instelling voor wetenschappelijk onderzoek;
  - c. bij of krachtens de wet ingestelde planbureaus;
  - d. de communautaire en nationale instanties voor de statistiek van de lidstaten van de Europese Unie;
  - e. onderzoeksafdelingen van ministeries en andere diensten, organisaties en instellingen, voor zover daartoe instemming van de CCS is verkregen.

### **Artikel 42**

De directeur-generaal willigt een verzoek als bedoeld in artikel 41 slechts in, indien de verzoeker naar het oordeel van de directeur-generaal voldoende maatregelen heeft getroffen om te voorkomen dat de verzameling van gegevens voor andere doeleinden dan statistisch of wetenschappelijk onderzoek wordt gebruikt.

### **Artikel 42a**

1. In afwijking van artikel 37 kan de directeur-generaal op verzoek, ten behoeve van wetenschappelijk onderzoek op het terrein van de volksgezondheid, gegevens verstrekken die het CBS ten behoeve van statistisch onderzoek op grond van artikel 12a van de Wet op de lijkbezorging heeft verzameld, van personen die in een wetenschappelijk onderzoek waren betrokken.
2. De directeur-generaal willigt een verzoek als bedoeld in het eerste lid slechts in, indien de betrokkene tot een zodanige verstrekking zijn uitdrukkelijke toestemming heeft gegeven dan wel, indien de betrokkene tot een zodanige verstrekking geen uitdrukkelijke toestemming heeft kunnen geven, voor zover verzoeker voldoende aantoont dat:
  - a. het vragen van toestemming bij leven van de betrokkene in redelijkheid niet mogelijk was of kon worden geveerd;

- b. niet gebleken is dat de betrokkene bij leven bezwaar heeft gemaakt tegen het verwerken van diens persoonsgegevens ten behoeve van wetenschappelijke onderzoek;
- c. het onderzoek een algemeen belang dient;
- d. het onderzoek niet zonder de desbetreffende gegevens kan worden uitgevoerd;
- e. het onderzoek overigens voldoet aan daaraan redelijkerwijs te stellen eisen.

**3.** De directeur-generaal kan nadere voorwaarden verbinden aan een verstrekking krachtens het eerste lid.

**4.** De artikelen 41, tweede lid, en 42 zijn van overeenkomstige toepassing.

**5.** Onze Minister stelt, in overeenstemming met Onze Minister van Volksgezondheid, Welzijn en Sport, gehoord de CCS en het College bescherming persoonsgegevens, nadere regels met betrekking tot de in het eerste lid bedoelde bevoegdheid.