

# Procesbeschrijving van het Inkomenspanelonderzoek

# 07

*Marion van den Brakel en Linda Moonen*



### Verklaring der tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is minder dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2006–2007	= 2006 tot en met 2007
2006/2007	= het gemiddelde over de jaren 2006 tot en met 2007
2006/'07	= oogstjaar, boekjaar, schooljaar enz. beginnend in 2006 en eindigend in 2007
2004/'05–2006/'07	= boekjaar enz., 2004/'05 tot en met 2006/'07

In geval van afronding kan het voorkomen, dat de som van de aantallen afwijkt van het totaal.

### **Uitgever**

Centraal Bureau voor de Statistiek  
Prinses Beatrixlaan 428  
2273 XZ Voorburg

### **Prepress**

Centraal Bureau voor de Statistiek - Facilitair bedrijf

### **Omslag**

Tel Design, Rotterdam

### **Inlichtingen**

Tel.: (088) 570 70 70  
Fax: (070) 337 59 94  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

### **Bestellingen**

E-mail: [verkoop@cbs.nl](mailto:verkoop@cbs.nl)  
Fax: (045) 570 62 68

### **Internet**

[www.cbs.nl](http://www.cbs.nl)

# Inhoud

<b>1. Inleiding</b>	4
<b>2. Inkomensbegrippen</b>	5
2.1 Nederlands belastingsysteem: inkomensbelasting	5
2.2 Afbakening inkomensbegrippen	6
2.2.1 Inkomensbegrippen op huishoudensniveau	6
2.2.2 Inkomensbegrippen op persoonsniveau	9
<b>3. Algemene procesbeschrijving</b>	11
<b>4. Input</b>	13
4.1 Inkomens Informatie Systeem	13
4.1.1 Gegevens afkomstig van de Belastingdienst	13
4.1.2 Gegevens afkomstig van/via andere instanties	16
4.2 Levering en vulling van de databestanden	16
<b>5. Throughput</b>	19
5.1 Inkomens Productie Systeem	19
5.1.1 Metabase	19
5.1.2 Productiebase	21
5.2 Ophoging	24
5.3 Plausibiliteitsonderzoek	25
5.3.1 Technische controles	25
5.3.2 Plausibiliteitscontroles	26
<b>6. Output</b>	27
6.1 Standaardiseren van huishoudensinkomens	27
6.2 Verschillende publicatiemedia	27
<b>7. Toekomstige ontwikkelingen</b>	29
7.1 Versnelling	29
7.2 Verbeteringen	29
<b>Literatuur</b>	31
<b>Lijst van gebruikte afkortingen</b>	31

# 1. Inleiding

Dit rapport geeft een gedetailleerde beschrijving van het Inkomenspanelonderzoek (IPO). Er wordt vooral ingegaan op de manier waarop de cijfers uit dit onderzoek tot stand komen. Hierbij komt het gehele proces aan bod: de input, de verwerking en de output. Daarnaast worden ook huidige en toekomstige ontwikkelingen besproken, die eventueel een verbetering en/of versnelling van het proces tot stand kunnen brengen.

Het IPO is een steekproefonderzoek gebaseerd op een panel, dat sinds 1989 bestaat. Dit panel van zogenoemde kernpersonen bestaat momenteel uit circa 80 duizend personen. Aan dit panel worden de huishoudensleden van de kernpersonen toegevoegd, waardoor de totale IPO-steekproef uit circa 250 duizend steekproefpersonen bestaat. Het CBS publiceert jaarlijks inkomenscijfers op basis van het IPO. Deze cijfers hebben betrekking op inkomensgegevens van 1 januari tot en met 31 december van een onderzoeksjaar. De procesbeschrijving van het IPO in dit rapport baseert zich op de gang van zaken voor het onderzoeksjaar 2006. Af en toe zullen veranderingen ten opzichte van voorgaande jaren kort worden aangehaald, maar de aandacht wordt voornamelijk gericht op de gang van zaken in 2006.

Naast de vorming van het IPO komt vanaf het onderzoeksjaar 2006 ook integraal informatie over inkomens beschikbaar. Dit zal in de procesbeschrijving tot uitdrukking komen. De integrale informatie zal onder meer gebruikt gaan worden ten behoeve van het Sociaal Statistisch Bestand (SSB). Uit het SSB komt via de gegevens uit het IPO, met gedetailleerde inkomensinformatie op landelijk niveau, en uit het RIO (regionaal inkomensonderzoek), met globalere inkomensinformatie op regionaal niveau, inkomensinformatie beschikbaar die gekoppeld kan worden aan informatie uit andere sociale statistieken van het CBS.

Dit rapport is als volgt opgebouwd. In paragraaf 2 wordt eerst ingegaan op de verschillende inkomensbegrippen die van belang zijn bij de Inkomensstatistieken. Ook het Nederlandse belastingstelsel komt in deze paragraaf aan bod. Paragraaf 3 geeft een algemene beschrijving van het productieproces. Daarna worden alle onderdelen van dit proces in meer detail besproken: het inputproces (paragraaf 4), het throughputproces (paragraaf 5) en de outputfase (paragraaf 6). Tot slot beschrijft paragraaf 7 enkele toekomstige ontwikkelingen, waardoor het proces op sommige punten zal veranderen.

## 2. Inkomensbegrippen

Het merendeel van de brongegevens waarop de Inkomensstatistieken gebaseerd zijn, is afkomstig van de Belastingdienst. Het gaat vooral om gegevens uit aangiftes van belastingplichtigen en gegevens verstrekt door werkgevers en uitkeringsinstanties. Paragraaf 2.1 beschrijft eerst de grote lijnen van het Nederlandse belastingstelsel. Daarna komen in paragraaf 2.2 de hieruit afgeleide inkomensbegrippen van de Inkomensstatistieken aan bod.

### 2.1 Nederlands belastingstelsel: inkomensbelasting

In het Nederlandse belastingstelsel moet belasting betaald worden over inkomen gegenereerd uit arbeid, onderneming, uitkering of pensioen. Verder moet ook belasting betaald worden over inkomen uit sparen en beleggen. Bij de belastingaangifte speelt het boxenstelsel, dat met de herziening van het belastingstelsel in 2001 is ingevoerd, een belangrijke rol. Dit stelsel gaat uit van drie typen belastingen op inkomsten, verdeeld over drie boxen:

- Box 1: belasting over inkomen uit werk en woning.
- Box 2: belasting over inkomen uit aanmerkelijk belang.
- Box 3: belasting over inkomen uit sparen en beleggen.

Hieronder volgt een toelichting op de drie boxen.

#### *Box 1*

Inkomsten die onder box 1 vallen zijn onder andere loon, winst uit onderneming, pensioen en uitkeringen, eigenwoningforfait en kapitaalverzekeringen van de eigen woning. De hoogte van de belasting op inkomsten uit box 1 hangt af van de hoogte van deze inkomsten. Voor personen tot 65 jaar variëren de tarieven van 34 procent (laagste inkomens) tot 52 procent (hoogste inkomens), voor personen van 65 jaar en ouder variëren de tarieven van 16 tot 52 procent. Aftrekposten in box 1 zijn onder meer de reisaf trek voor openbaar vervoer, aftrekposten voor de eigen woning en persoonsgebonden aftrekposten zoals ziektekosten en studiekosten.

#### *Box 2*

Er is sprake van aanmerkelijk belang wanneer iemand minimaal 5 procent aandelen bezit van een vennootschap of wanneer iemand opties heeft om deze te verwerven. Het financiële voordeel van aanmerkelijk belang, in de vorm van dividend of verkoopwinst op de aandelen<sup>1)</sup>, is fiscaal belast. De hoogte van de belasting bedraagt 25 procent.

#### *Box 3*

Box 3 bevat de belasting over inkomen uit sparen en beleggen. Er wordt belasting geheven over een vast rendement op het vermogen. Dit vaste rendement wordt berekend over de gemiddelde waarde van iemands bezittingen minus diens schulden. Over het deel van deze waarde dat boven een heffingvrij vermogen<sup>2)</sup> uitkomt, is 1,2 procent belasting verschuldigd.

Voor de definitieve belastingaanslag spelen reeds betaalde heffingen in de vorm van premies werknemersverzekeringen en loonbelasting een rol. De premies werknemersverzekeringen worden via de werkgever ingehouden op het loon. De loonbelasting wordt eveneens via de werkgever ingehouden op het loon en wordt als voorheffing verrekend met de inkomstenbelasting over het inkomen uit box 1. Naast deze zogenaamde contra-informatie over de loonheffingen (afkomstig van werkgevers en uitkeringsinstanties) ontvangt de Belastingdienst ook gegevens over onroerend goed, zoals de taxatiewaarde van woningen (WOZ-waarde). Deze dienen onder meer ter controle van de in box 1

<sup>1)</sup> Salaris uit een bv of nv wordt belast in box 1.

<sup>2)</sup> De hoogte van het heffingvrije vermogen is afhankelijk van de samenstelling van het gezin.

opgegeven inkomsten uit eigen woning. De gegevens uit box 2 en box 3 kunnen door de Belastingdienst geconfronteerd worden met gegevens over renteopbrengsten, saldi van banktegoeden, gegevens over genoten dividend/obligatierente en waarden van effectenportefeuilles. Deze gegevens zijn afkomstig van banken en rente- en beleggingsfondsen. Na confrontatie met bovenstaande informatie volgt de definitieve aanslag inkomstenbelasting.

Behalve het innen van belastingen verstrekt de overheid (inkomensafhankelijke) uitkeringen die gekoppeld zijn aan bepaalde bestedingen, de zogenaamde gebonden inkomensoverdrachten. Hiertoe behoren onder meer de huurtoeslag, de zorgtoeslag en de toeslag voor kinderopvang. Daarnaast zijn er enkele onbelaste uitkeringen zoals de kinderbijslag en de studiefinanciering. Voor de afbakening van inkomensbegrippen als primair, bruto- en besteedbaar inkomen (paragraaf 2.2) zijn deze uitkeringen van belang.

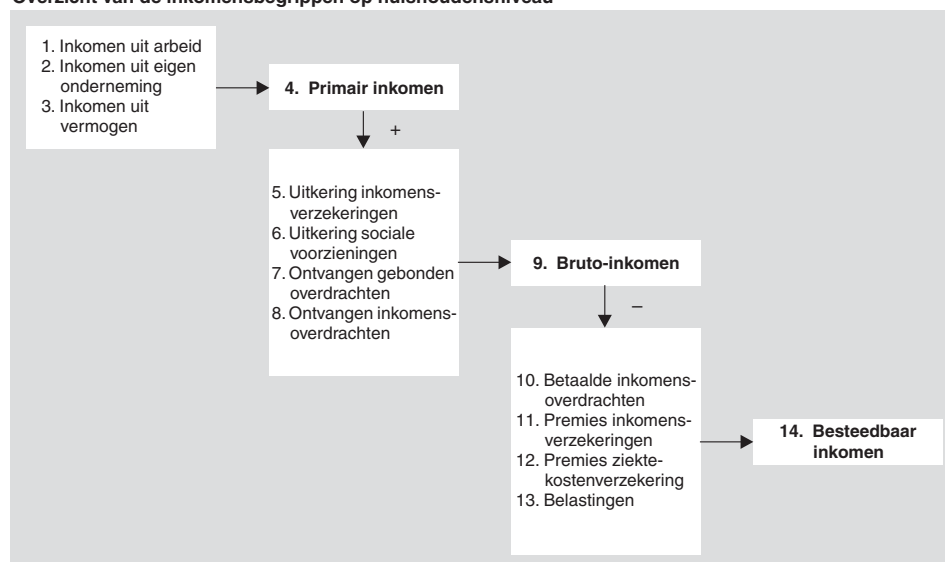
## 2.2 Afbakening inkomensbegrippen

De basis voor de verschillende inkomensbegrippen is gelegd in Bruinooge en Van de Donk (1993). Bij de Inkomensstatistieken worden de inkomensbegrippen zowel op huishoudensniveau als op persoonsniveau afgeleid. In paragraaf 2.2.1 komen de begrippen op huishoudensniveau aan bod. Paragraaf 2.2.2 bespreekt de begrippen op persoonsniveau.

### 2.2.1 Inkomensbegrippen op huishoudensniveau

De voornaamste begrippen op huishoudensniveau zijn het primair inkomen, het bruto-inkomen en het besteedbaar inkomen. In schema 1 staat een overzicht van de inkomensbestanddelen die het primair, het bruto- en het besteedbaar inkomen vormen.

**Schema 1**  
Overzicht van de inkomensbegrippen op huishoudensniveau



In schema 1 is te zien dat het primair inkomen uit een aantal onderdelen bestaat, namelijk inkomen uit arbeid, inkomen uit eigen onderneming en inkomen uit vermogen. Vervolgens bestaat het bruto-inkomen uit het primair inkomen vermeerderd met de uitkering inkomensverzekeringen (bijvoorbeeld de werkloosheidsuitkering), de uitkering sociale voorzieningen (zoals de kinderbijslag), de ontvangen gebonden overdrachten (zoals de rijksbijdrage eigen woning) en de ontvangen inkomensoverdrachten. Het besteedbaar inkomen is ten slotte gedefinieerd als het bruto-inkomen verminderd met de betaalde in-

komensoverdrachten, de betaalde premies inkomensverzekeringen, de betaalde premies ziektekostenverzekering en de verschuldigde belastingen. In het vervolg van deze paragraaf wordt elk van deze begrippen in meer detail besproken. Eerst wordt in onderstaand kader een overzicht gegeven van de verschillende soorten uitkeringen en verzekeringen die in de lijst van inkomensbestanddelen zijn opgenomen.

#### **Uitkeringen en verzekeringen in meer detail**

De inkomensbestanddelen bevatten verschillende soorten uitkeringen en verzekeringen. Binnen de Inkomensstatistiek worden deze in hoofdlijnen ingedeeld naar het 'risico' dat door de uitkering wordt gedekt en naar de wijze van financiering. Dit leidt tot de indeling in de uitkeringen inkomensverzekering, de uitkeringen sociale voorzieningen en de ziektekostenverzekering. In dit kader zal kort worden ingegaan op deze uitkeringen en verzekeringen.

#### **Uitkering inkomensverzekeringen**

De uitkeringen inkomensverzekeringen geven een verzekering tegen (gedeeltelijk) verlies van inkomen bij werkloosheid, ziekte en arbeidsongeschiktheid, ouderdom en nabestaanden. Een tweede kenmerk is dat deze uitkeringen worden betaald door premieheffing. Bij de premieheffing wordt een onderscheid gemaakt tussen een voor iedereen verplicht deel (volksverzekering), een verplicht deel voor ontvangers van loon of uitkering (sociale verzekeringen) en een vrijwillig deel (particuliere verzekeringen).

De premies van de volksverzekeringen worden betaald door alle belastingplichtigen en maken deel uit van de loonheffing en inkomstenbelasting. Voor ouderdom / nabestaanden zijn dat de AOW en de ANW. Voor werkloosheid bestaat geen volksverzekering.

De sociale verzekeringen zijn bedoeld voor iedereen die een arbeidsverhouding heeft of heeft gehad. Daarom worden deze verzekeringen ook met de term werknemersverzekeringen aangeduid. Premies voor deze verzekeringen worden betaald door werknemers, werkgevers, uitkeringsontvangers en uitkeringsinstellingen. Ze worden op het loon of de uitkering ingehouden. De werknemersverzekeringen voor werkloosheid is de WW. Voor ziekte / arbeidsongeschiktheid zijn dat de WAO, WIA en de ZW. Voor ouderdom / nabestaanden zijn de verplichte pensioenregelingen in deze groep ingedeeld.

De particuliere verzekeringen zijn vrijwillig en worden meestal als aanvullende verzekering afgesloten. De premies worden rechtstreeks door de verzekeringsnemer aan de verzekeringsmaatschappij betaald.

#### **Uitkeringen sociale voorzieningen**

Uitkeringen sociale voorzieningen zijn sociale uitkeringen van de centrale en lokale overheid aan huishoudens die uit algemene middelen (belastingopbrengsten) worden betaald. Er is geen verplichting tot premiebetaling. Dit in tegenstelling tot de inkomensverzekeringen, die door premieheffing worden betaald. Voorbeelden van uitkeringen in het kader van de sociale voorzieningen zijn onder meer bijstandsuitkering, de kinderbijslag en de studiefinanciering.

#### **Ziektekostenverzekering**

Van de ziektekostenverzekeringen worden de betaalde verplichte premies op het inkomen in mindering gebracht. De uitkeringen (= ontvangen vergoedingen) worden niet als inkomen beschouwd. Deze maken deel uit van de bestedingen.

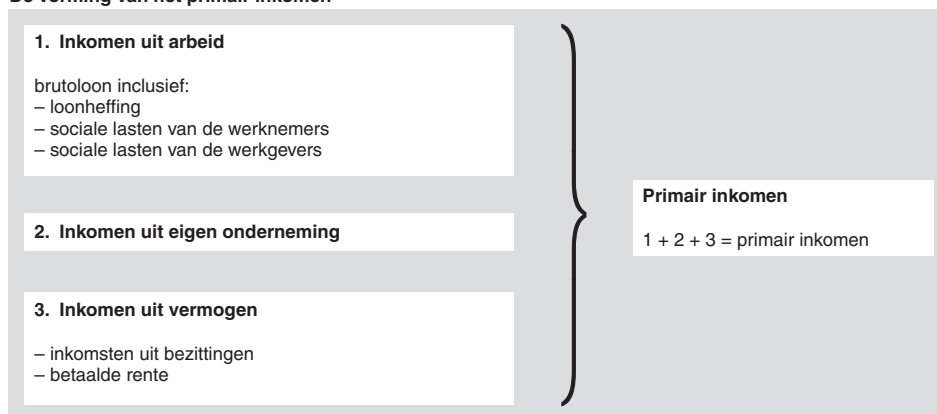
Vanaf 2006 zijn er voor de ziektekostenverzekering de Algemene Wet Bijzondere Ziektekosten (AWBZ) en de Zorgverzekeringswet (ZVW). De AWBZ is een volksverzekering (zie hierboven). Alle belastingplichtigen betalen hiervoor premie. Binnen de ZVW is het basispakket verplicht voor iedereen. De premie voor het basispakket bestaat uit een nominaal deel en een inkomensafhankelijk deel. De inkomensafhankelijke premie wordt direct van het loon ingehouden of door de Belastingdienst geïnd, de nominale premie wordt rechtstreeks door de verzekeringsnemer aan de verzekeringsmaatschappij betaald. Deze premies worden op het inkomen in mindering gebracht.

Naast het verplichte basispakket kan iedereen ook een of meer aanvullende verzekeringen afsluiten. Deze aanvullende verzekeringen worden gezien als particuliere verzekeringen. De premies hiervoor worden rechtstreeks aan de verzekeraar betaald. Deze uitgaven worden niet op het inkomen in mindering gebracht, maar maken deel uit van de bestedingen.

#### **Primair inkomen**

Zoals eerder vermeld bestaat het primair inkomen uit drie bestanddelen, namelijk inkomen uit arbeid, inkomen uit eigen onderneming en inkomen uit vermogen (zie ook schema 2). Inkomen uit arbeid omvat loon, salaris, tantièmes, spaarloon en de beloning van arbeid die niet in dienstbetrekking is verricht. Ook de waarde van het privégebruik van de auto van de werkgever en andere vormen van beloning in natura worden hiertoe gerekend. De werknemers- en werkgeversbijdragen in de premies voor de sociale verzekeringen zijn bij het inkomen uit arbeid inbegrepen. Voor het inkomen uit eigen onderneming wordt de fiscale winst als uitgangspunt genomen. Inkomen uit vermogen bestaat uit ontvangen rente en dividend en de opbrengst van de exploitatie van onroerend goed, waaronder de eigen woning. Betaalde rente wordt op het inkomen uit vermogen in mindering gebracht.

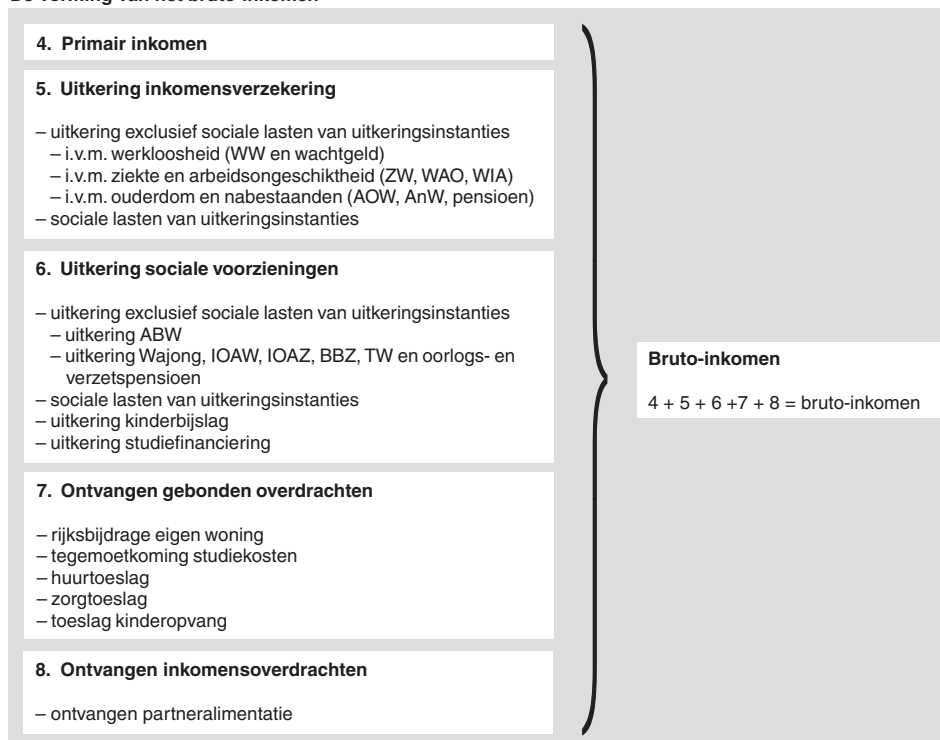
**Schema 2**  
De vorming van het primair inkomen



*Bruto-inkomen*

Het bruto-inkomen wordt gevormd door het primair inkomen vermeerderd met de uitkering inkomensverzekeringen, de uitkering sociale voorzieningen, de ontvangen gebonden overdrachten en de ontvangen inkomensoverdrachten. In schema 3 worden deze bestanddelen nader toegelicht. Zo omvat de uitkering inkomensverzekeringen uitkeringen in verband met werkloosheid, ziekte en arbeidsongeschiktheid, en ouderdom en nabestaanden. De uitkering sociale voorzieningen betreft uitkeringen in het kader van onder andere de algemene bijstandswet en de wet arbeidsongeschiktheidsvoorziening jonggehandicapten. Ook het oorlogs- en verzetspensioen, de kinderbijslag en de studiefinanciering zijn opgenomen onder de uitkering sociale voorzieningen. De ontvangen gebonden overdrachten bestaan uit de rijksbijdrage eigen woning, de tegemoetkoming in de studiekosten en de toeslagen voor huur, zorg en kinderopvang. De ontvangen inkomensoverdrachten ten slotte omvatten alleen de ontvangen alimentatie van de partner. Andere bestanddelen die theoretisch gezien onder de ontvangen inkomensoverdrachten vallen zijn bijvoorbeeld ontvangen inkomensoverdrachten uit het buitenland, bijdragen die (studerende) kinderen van ouders ontvangen, en ontvangen kinderalimentatie. Deze bestanddelen worden echter niet waargenomen en kunnen daarom (nog) niet bij de ontvangen inkomensoverdrachten worden opgenomen.

**Schema 3**  
De vorming van het bruto-inkomen

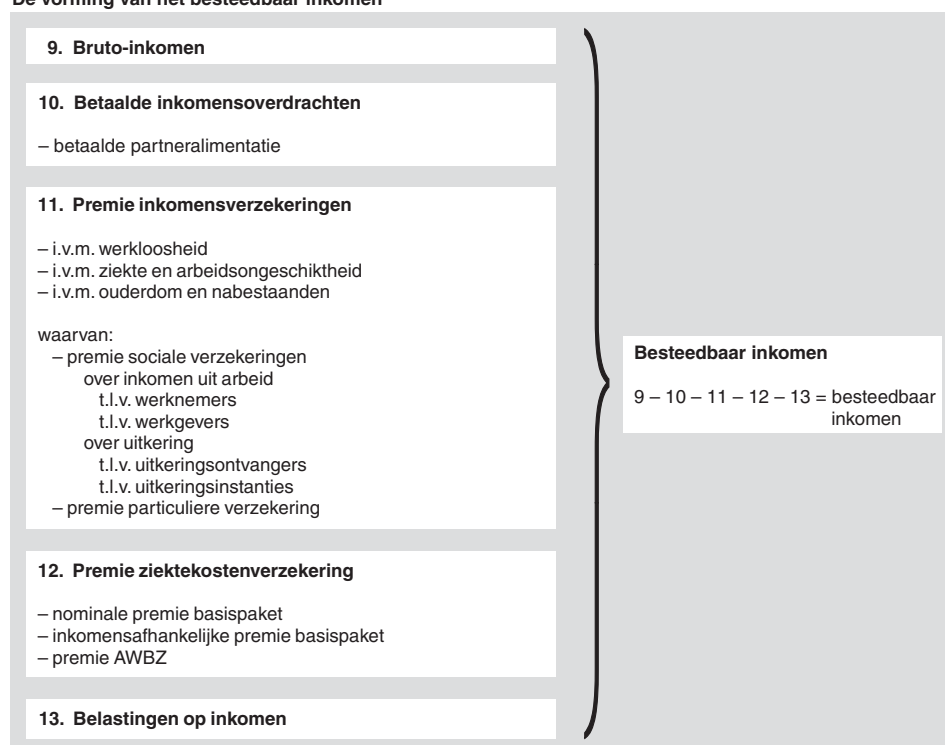




### Besteedbaar inkomen

Het besteedbaar inkomen is gedefinieerd als het bruto-inkomen verminderd met de betaalde inkomensoverdrachten, de premies inkomensverzekeringen, de premies ziektekostenverzekering en de belastingen. In schema 4 worden deze bestanddelen nader toegelicht. De betaalde inkomensoverdrachten omvatten, analoog naar de ontvangen inkomensoverdrachten, alleen de betaalde partneralimentatie. De premies inkomensverzekeringen betreffen premies voor uitkeringen in verband met werkloosheid, ziekte en arbeidsongeschiktheid, en ouderdom en nabestaanden. De premies ziektekostenverzekeringen betreffen zowel de nominale als de inkomensafhankelijke premies. De belastingen ten slotte omvatten de belastingen op inkomen en op sparen en beleggen.

**Schema 4**  
De vorming van het besteedbaar inkomen

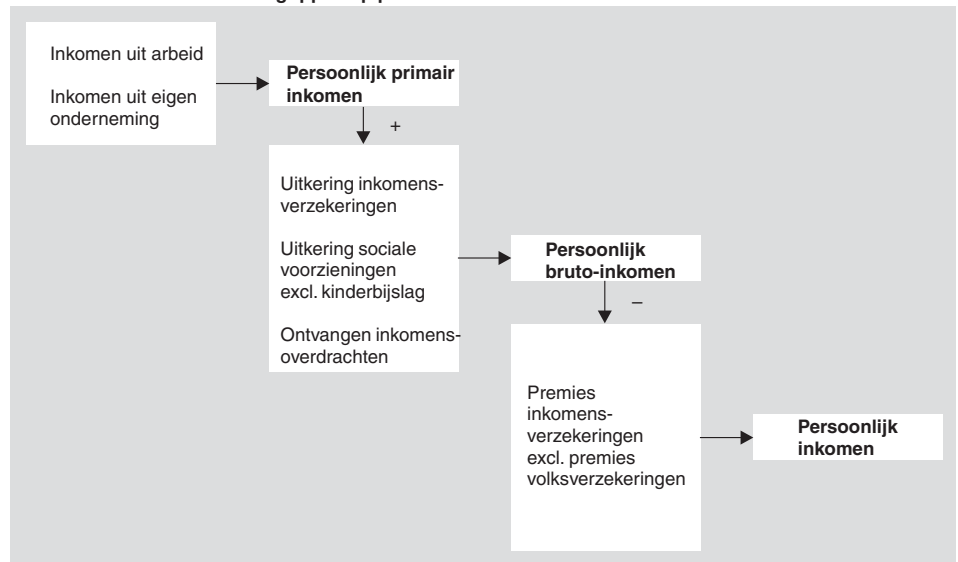


### 2.2.2 Inkomensbegrippen op persoonsniveau

Op het niveau van huishoudens worden de inkomensbestanddelen van alle huishoudensleden samengeteld. Bij het samenstellen van inkomenscijfers op persoonsniveau ontstaan er echter problemen bij het indelen van sommige bestanddelen. Soms is niet duidelijk aan welk huishoudlid een bestanddeel moet worden toegewezen. De ontvangen huurtoeslag is hier een voorbeeld van, omdat die per huishouden wordt ontvangen. Moet deze worden meegeteld bij het inkomen van de hoofdkostwinner, of bij een ander lid van het huishouden? Of moet het bedrag van de huurtoeslag worden verdeeld over alle huishoudensleden? Omwille van deze problematiek zijn de inkomensbegrippen op persoonsniveau lichtjes gewijzigd ten opzichte van de inkomensbegrippen op huishoudensniveau. In plaats van over primair, bruto- en besteedbaar inkomen wordt nu gesproken over persoonlijk primair, persoonlijk bruto- en persoonlijk inkomen.

Het *persoonlijk primair* inkomen bestaat uit inkomen uit arbeid en inkomen uit eigen onderneming (zie schema 5). Het *persoonlijk bruto-inkomen* bestaat uit het persoonlijk primair inkomen vermeerderd met de uitkeringen inkomensverzekeringen, de uitkeringen sociale voorzieningen met uitzondering van de kinderbijslag, en de ontvangen inkomensoverdrachten. Het *persoonlijk inkomen* ten slotte bestaat uit het persoonlijk bruto-inkomen verminderd met de premies inkomensverzekeringen met uitzondering van de premies volksverzekeringen. Inkomen uit vermogen, kinderbijslag en de gebonden overdrachten blijven bij de bepaling van het persoonlijk inkomen dus buiten beschouwing. Verder worden premies ziektekostenverzekering en inkomstenbelasting niet op het persoonlijk inkomen in mindering gebracht.

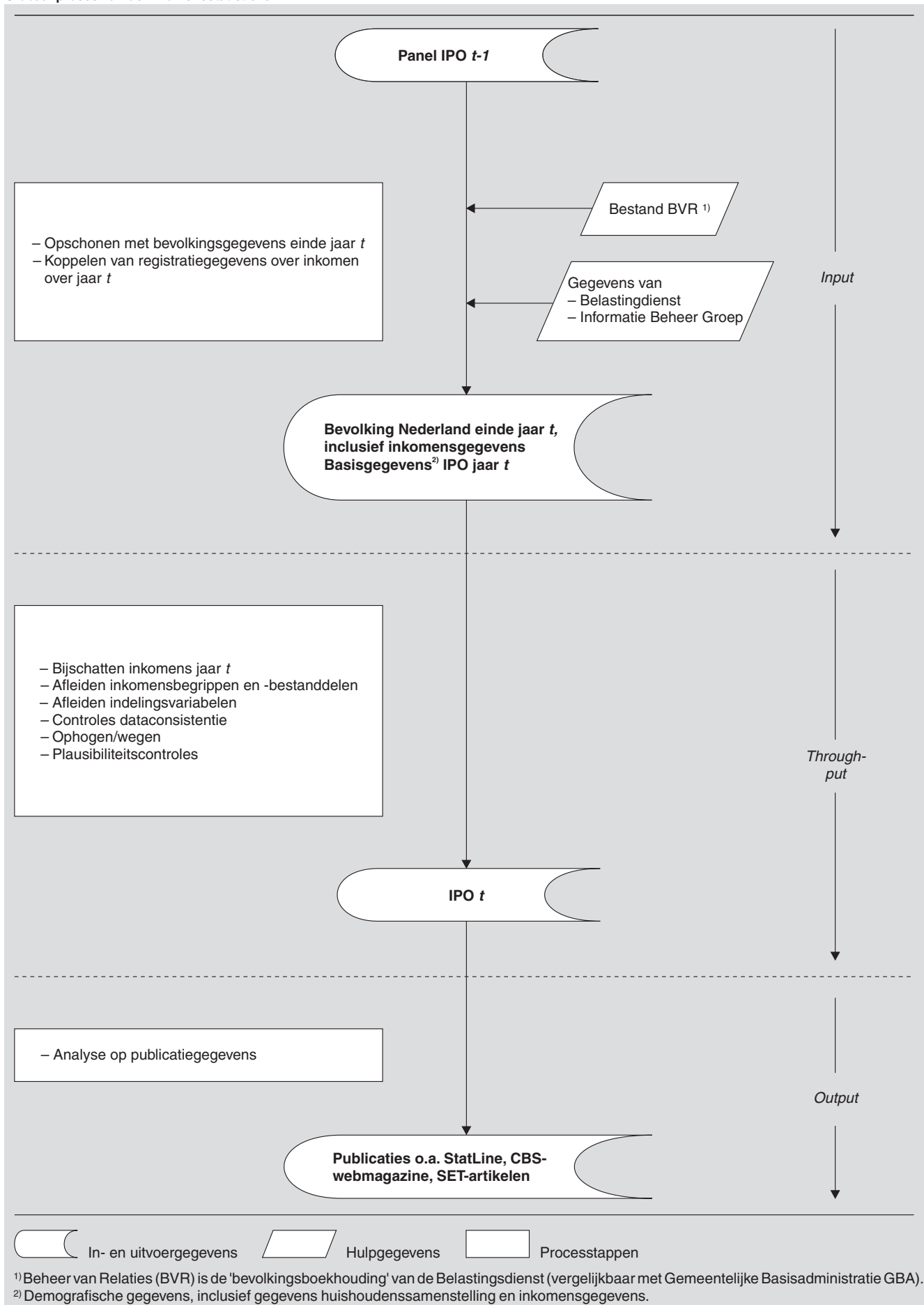
**Schema 5**  
**Overzicht van de inkomensbegrippen op persoonsniveau**



### 3. Algemene procesbeschrijving

De verschillende inkomensbegrippen vormen de basis voor de procesbeschrijving van het IPO. Voor deze beschrijving wordt uitgegaan van onderzoeksjaar  $t$  (bijvoorbeeld  $t = 2006$ ). Beschreven wordt het proces van IPO  $t$ , waarbij ook het parallel lopende proces van integrale inkomensinformatie aan bod komt. In schema 6 is een globaal overzicht van het proces van de Inkomensstatistieken gegeven. Daarbij is onderscheid gemaakt tussen het input-, het throughput- en het outputproces. Het proces start met de opschoning van het IPO panel  $t-1$ . Daartoe worden bevolkingsgegevens ultimo  $t-1$  en  $t$  met elkaar vergeleken. Deze gegevens zijn afkomstig uit Beheer van Relaties, de 'bevolkingsboekhouding' van de Belastingdienst. Na de opschoning van het panel van einde jaar  $t-1$  worden eerst de mede-adresbewoners en in een later stadium alleen de huishoudensleden van de kernpersonen uit het panel aan het onderzoek toegevoegd. Vervolgens worden hier gegevens van de Belastingdienst en gegevens over de studie-financiering (Informatie Beheer Groep) aan gekoppeld. Aan het einde van de inputfase, medio september van jaar  $t+1$ , is zowel voor het IPO als voor de Nederlandse bevolking een bestand beschikbaar, verrijkt met (op dat moment nog incomplete) inkomensgegevens. Met de aldus verkregen basisgegevens van IPO  $t$  worden in het throughputproces inkomensbegrippen en -bestanddelen afgeleid. Hiertoe worden enkele ontbrekende inkomensgegevens bijgeschat. Naast de inkomensbestanddelen zijn er ook enkele indelingsvariabelen die afgeleid worden, zoals de sociaaleconomische categorie, de woonsituatie en de voornaamste bron van inkomen. Na het afleiden van alle variabelen worden ophooggewichten bepaald om de gegevens van het IPO panel op te hogen naar de hele bevolking. Ten slotte vinden plausibiliteitscontroles plaats, waarin allerlei controles op de gegevens worden uitgevoerd. Dit levert het publicatiebestand IPO  $t$  waarmee zowel op huishoudens- als op persoonsniveau inkomenscijfers samengesteld kunnen worden.

**Schema 6**  
**Globaal proces van de Inkomensstatistieken**



## 4. Input

Het inputproces van het IPO vindt plaats via het Inkomens Informatie Systeem (IIS). Bundeling van belangen van het Ministerie van Financiën en het CBS hebben ertoe geleid dat beide instanties gezamenlijk dit systeem hebben opgezet. Het Ministerie van Financiën (Belastingdienst) en de Informatie Beheer Groep (IBG) verstrekken informatie op microniveau, waarna deze wordt verwerkt in het IIS. Met behulp van het IIS worden jaarlijks, naast demografische gegevens, gegevens over inkomen en vermogen van de Nederlandse bevolking verzameld.

De in het IIS verzamelde informatie wordt door het CBS vooral gebruikt voor het samenstellen van de statistieken van het IPO en RIO (Regionaal Inkomensonderzoek). In paragraaf 4.1 wordt het IIS in detail besproken en in paragraaf 4.2 wordt nader ingegaan op de tijdstippen van levering van de databestanden en de vulling van de geleverde bestanden.

### 4.1 Inkomens Informatie Systeem

In het IIS worden uit de inputgegevens afkomstig van de Belastingdienst en de IBG (zie schema 6) de relevante gegevens geselecteerd, geordend, bewerkt en in outputbestanden opgeslagen. Deze bewerkingen zijn in afzonderlijke blokken opgedeeld, de zogenoemde entiteiten. De resultaten van de bewerkingen, i.e. de outputbestanden die naar het Ministerie van Financiën en het CBS gaan, worden in dit rapport gemakshalve eveneens met de term entiteiten aangeduid.<sup>1)</sup> De voor IPO belangrijkste entiteiten zijn in tabel 1 weergegeven, waarbij onderscheid is gemaakt tussen gegevens direct afkomstig van de Belastingdienst en gegevens afkomstig van of via andere instanties. De nummering van de entiteiten komt overeen met die van het IIS. De entiteiten zullen in het vervolg van deze paragraaf beschreven worden.

**Tabel 1**  
**Overzicht soort gegevens per entiteit**

Soort gegevens	Entiteit
<b>Gegevens van Belastingdienst</b>	
Demografische gegevens	01
Aangiftegegevens	02
Aanslaggegevens	03
Loonbelastinggegevens	05
Gegevens huurtoeslag, zorgtoeslag en toeslag kinderopvang	09
WOZ-gegevens	14
Voorlopige aanslagen	15
<b>Gegevens van/via andere instanties</b>	
Gegevens studiefinanciering	10
Rentegegevens	12
Dividendgegevens	13

#### 4.1.1 Gegevens afkomstig van de Belastingdienst

##### *Entiteit 01: demografische gegevens*

Om te komen tot een volledig en correct uitgangbestand van het IPO (en de Nederlandse bevolking) voor onderzoeksjaar  $t$  vindt als eerste in het IIS een opschoning plaats van de demografische informatie van einde jaar  $t-1$ . In schema 7 is dit opschonen van bestanden weergegeven. De volgende IIS-bewerkingen vinden plaats:

- Een opschoning van de Nederlandse bevolking van einde jaar  $t-1$  (i.e. 31 december) met de plus- en de minpopulatie levert de Nederlandse bevolking van einde jaar  $t$ . De

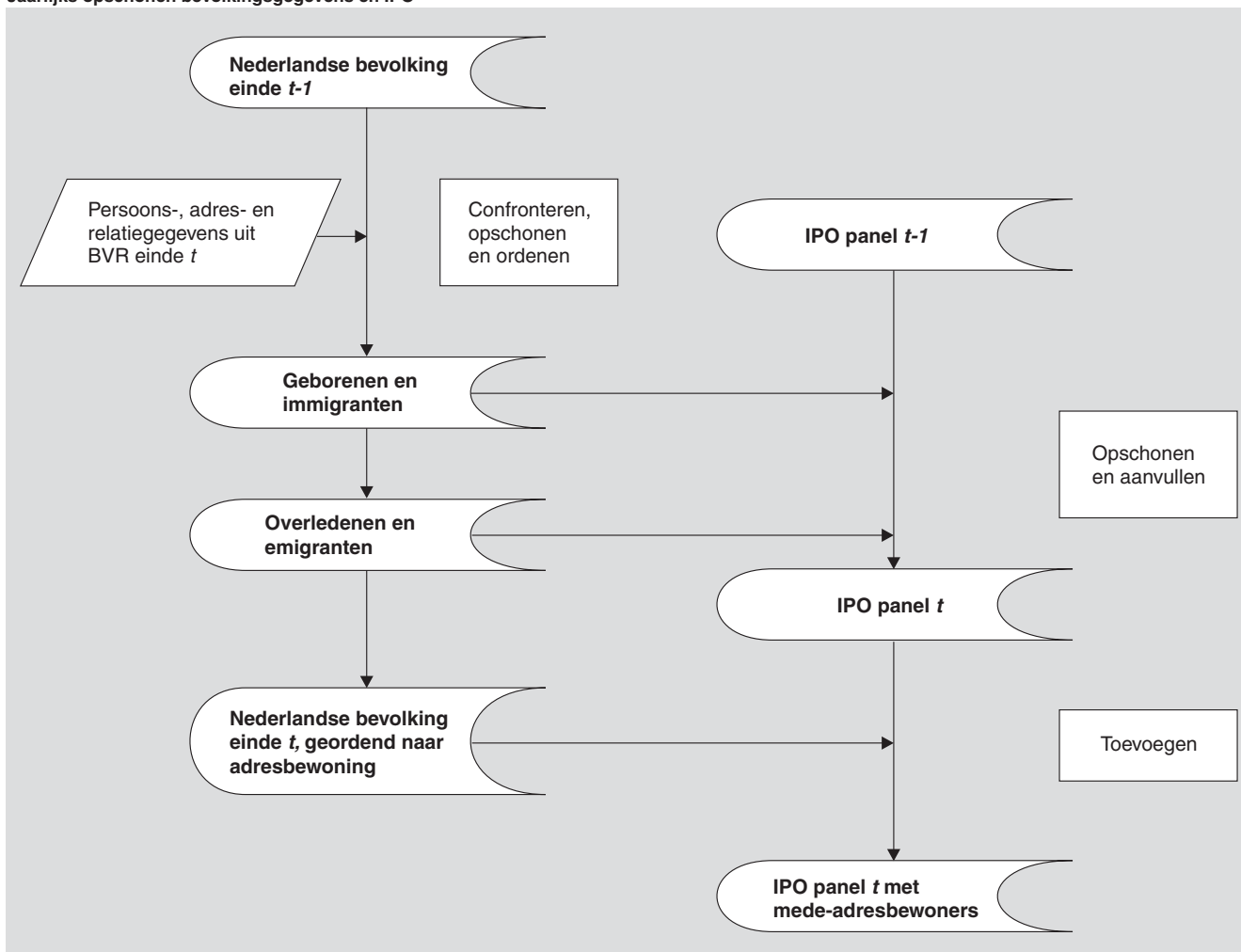
<sup>1)</sup> Per entiteit worden in sommige gevallen meerdere outputbestanden gemaakt en opgeleverd. Deze bestanden verschillen qua samenstelling maar niet qua inhoud.

pluspopulatie omvat alle personen die in de loop van het jaar  $t$  aan de bevolking zijn toegevoegd. Dit betreft geboren en immigranten. De minpopulatie omvat alle personen die in de loop van het jaar  $t$  zijn overleden, geëmigreerd of vertrokken onbekend waarheen (administratief niet meer traceerbaar). Met de plus- en minpopulatie wordt de bevolking geactualiseerd naar de stand van einde jaar  $t$ .

- Ook het panel van kernpersonen van IPO van  $t-1$  worden met behulp de plus- en minpopulatie geactualiseerd. Overledenen en emigranten worden uit de panels verwijderd en er vindt een aanvulling plaats van geboren en emigranten. Daartoe wordt een steekproef van (nieuwe) kernpersonen getrokken uit de pluspopulatie. Voor het IPO geldt een aanvulling van 0,6 procent uit de pluspopulatie.
- Met behulp van adres- en relatiegegevens uit het bestand Beheer van Relaties (BVR) wordt de bevolking geordend naar adresbewoning.
- Aan de kernpersonen van het panel voor jaar  $t$  worden de mede-adresbewoners toegevoegd. In een later stadium (medio september  $t+1$ ) zal hiermee de huishoudensamenstelling van kernpersonen bepaald worden.

Het aldus gevormde bestand voor IPO vormt entiteit 01 en wordt verschillende malen opgeleverd als invoer voor het throughputproces. Bij elke nieuwe levering zijn correcties in de demografische gegevens doorgevoerd. Vanaf medio september  $t+1$  kan van elke kernpersoon het bijbehorende huishouden worden vastgesteld op basis van gegevens van het GBA, aangevuld met inkomensgegevens over de hoogte van het inkomen en de studiefinanciering. Personen in entiteit 01 die niet tot het huishouden van een kernpersoon behoren, kunnen dan worden geïdentificeerd. In het throughputproces worden deze personen uit de bestanden verwijderd. De geleverde IIS bestanden hebben echter steeds betrekking op kernpersonen en mede-adresbewoners.

**Schema 7**  
Jaarlijks opschonen bevolkingsgegevens en IPO



#### *Entiteit 02: aangiftegegevens*

De Belastingdienst levert aan het IIS vanaf juni van het jaar  $t+1$  periodiek inkomensgegevens die verzameld zijn uit de aangiftebiljetten inkomstenbelasting over onderzoeksjaar  $t$ . Na verwerking in het IIS worden cumulatief bestanden opgeleverd, die als invoer dienen voor het throughputproces. Dit betekent dat er telkens een bestand wordt geleverd van die personen waarvoor op dat moment de gegevens bekend zijn. In de geleverde bestanden zijn de gegevens van kernpersonen en mede-adresbewoners opgenomen.

#### *Entiteit 03: aanslaggegevens*

De Belastingdienst controleert de aangegeven inkomens om het belastbaar inkomen vast te kunnen stellen. Dit gebeurt met behulp van loonbelastinggegevens afkomstig van werkgevers en uitkeringsinstanties (de zogenaamde Fibase-gegevens, zie ook entiteit 05). Eventuele verschillen tussen het inkomen dat is aangegeven op het aangiftebiljet en het inkomen dat is opgenomen in de Fibase worden gecorrigeerd, wat leidt tot de aanslaggegevens. De aanslaggegevens worden eveneens periodiek door de Belastingdienst aan het IIS verstrekt. Verdere levering gebeurt op soortgelijke wijze als de levering van de aangiftegegevens.

#### *Entiteit 05: loonbelastinggegevens*

De loonbelastinggegevens van jaar  $t$  worden door werkgevers en uitkeringsinstanties gespreid door het jaar  $t+1$  verstrekt aan de Belastingdienst. Het gaat hierbij om gegevens over het betaalde loon, de loonheffingen ingehouden door de werkgever en de hoogtes van uitkeringen in het kader van inkomensverzekeringen en sociale voorzieningen. Deze gegevens zijn opgeslagen in de zogenoemde Fibase en worden door de belastingdienst gebruikt ter controle van de aangiftegegevens (zie entiteit 03). De Belastingdienst levert driemaal per jaar Fibase-gegevens aan het IIS. Na controle en aanvullende bewerkingen in het IIS worden de Fibase-gegevens over de kernpersonen plus mede-adresbewoners als entiteit 05 (cumulatief) opgeleverd.

Met ingang van maart 2006 zijn de Fibase-gegevens vervangen door de zogenoemde WALVIS (Wet Administratieve Lastenverlichting en Vereenvoudiging In de Sociale verzekering)-gegevens, in de regel aangeduid als 'gegevens uit de polisadministratie'. De bovengenoemde gegevens die werkgevers en uitkeringsinstanties voorheen verstrekten aan de Belastingdienst worden nu na afloop van iedere maand doorgegeven aan één overheidsinbox. Van deze gegevens maakt de overkoepelende uitkeringsinstantie UWV de polisadministratie. Dit betekent dat voor iedereen een polis wordt aangemaakt. Naast loon- en uitkeringgegevens bevat de polisadministratie ook gegevens over de premies werknemersverzekeringen. Verder staan in een polis iemands rechten in geval een beroep gedaan zou worden op werknemersverzekeringen. Het CBS heeft sinds maart 2006 toegang tot de polisadministratie en stelt uit maandelijks afgetapte gegevens jaargegevens samen. Deze zijn vergelijkbaar met de gegevens die voorheen uit de Fibase afkomstig waren. De polisadministratie vervangt daarmee de Fibase en bevat daarnaast nog extra informatie over werknemers- en werkgeversverzekeringen.

#### *Entiteit 09: gegevens over huurtoeslag, zorgtoeslag en toeslag kinderopvang*

Vanaf 2006 worden de gegevens over de individuele huursubsidie (huurtoeslag) niet meer door het VROM, maar door de Dienst Toeslagen van de Belastingdienst geleverd. Daarnaast levert deze Dienst de gegevens over de zorgtoeslag en de toeslag kinderopvang. Via het IIS worden deze gegevens, vooralsnog in drie aparte bestanden, opgeleverd.

#### *Entiteit 14: WOZ-gegevens*

Van iedere woning wordt om de vijf jaar<sup>2)</sup> de verkoopprijs vastgesteld. Op basis van deze verkoopprijs wordt vervolgens de netto economische huurwaarde,<sup>3)</sup> ook wel WOZ (Wet Waardering Onroerend Zaken) waarde genoemd, afgeleid. Deze waarde wordt,

<sup>2)</sup> In de toekomst zal dit jaarlijks gebeuren.

<sup>3)</sup> De netto economische huurwaarde is het bedrag dat de eigenaar van de woning aan huur zou moeten betalen indien hij/zij de woning zou huren. De bruto economische huurwaarde is de netto waarde plus de afschrijvingskosten van de woning.

naast andere gegevens over woningen, vastgelegd in het WOZ-bestand van de Belastingdienst. De Belastingdienst gebruikt de WOZ-gegevens ter controle van door belastingplichtigen opgegeven informatie over de eigen woning. De WOZ-gegevens van de Belastingdienst worden ter beschikking gesteld aan het IIS. Na controle op verwerkbaarheid, compleetheid en plausibiliteit worden de WOZ-gegevens van kernpersonen en mede-adresbewoners opgeleverd als entiteit 14.

#### *Entiteit 15: gegevens voorlopige aanslagen*

De Belastingdienst stelt integrale gegevens over voorlopige aanslagen inkomstenbelasting ter beschikking aan het IIS. Gegevens van kernpersonen plus mede-adresbewoners worden hieruit geselecteerd en opgeleverd als entiteit 15. In het verleden is er een onderzoek uitgevoerd om na te gaan of de gegevens over de voorlopige aanslagen gebruikt kunnen worden om sneller tot publicatiecijfers te komen. De uitslag van dat onderzoek was negatief, waardoor er vooralsnog geen gebruik van deze gegevens wordt gemaakt. Echter, met het oog op alle veranderingen die in de afgelopen jaren hebben plaatsgevonden, zijn er wellicht wel mogelijkheden om deze gegevens te gebruiken bij het vervaardigen van snellere inkomenscijfers. Dit is een aspect dat verder onderzocht moet worden.

### *4.1.2 Gegevens afkomstig van/via andere instanties*

#### *Entiteit 10: gegevens studiefinanciering*

Van de Informatie Beheer Groep (IBG) te Groningen worden jaarlijks drie bestanden ontvangen met gegevens over de studiefinanciering. De gegevens hebben betrekking op:

- toekenningen en verstrekkingen in het kader van de Wet Tegemoetkoming Studiekosten,
- rechten op uitkeringen in het kader van de Wet Studiefinanciering.

De IBG stelt integrale bestanden ter beschikking, waaruit vervolgens door het IIS de gegevens van de kernpersonen plus mede-adresbewoners worden geselecteerd. Deze gegevens worden als entiteit 10 opgeleverd.

#### *Entiteit 12 en 13: rente- en dividendgegevens*

Banken en financieringsfondsen zijn verplicht om overzichten te verstrekken aan de Belastingdienst met gegevens over banktegoeden, rente-, dividend- en obligatieopbrengsten en de waarde van effectenportefeuilles. De Belastingdienst gebruikt deze informatie ter controle van de hierover door belastingplichtigen aangegeven gegevens. De Belastingdienst levert de integrale rente- en dividendgegevens driemaal per jaar aan het IIS. Het rentebestand bevat gegevens over renteopbrengsten en saldi van banktegoeden. Het dividendbestand bevat gegevens over genoten dividend/obligatierente en de waarde van de effectenportefeuille. Nadat de bestanden door het IIS aan een technische controle zijn onderworpen en enkele aanvullende bewerkingen zijn uitgevoerd, worden het rentebestand (entiteit 12) en het dividendbestand (entiteit 13) voor verdere verwerking in het throughputproces beschikbaar gesteld.

## **4.2 Levering en vulling van de databestanden**

De databestanden die worden gebruikt voor het samenstellen van de Inkomensstatistieken worden allemaal op verschillende tijdstippen aangeleverd. Een bepaald bestand kan ook meerdere keren worden aangeleverd; bij elke levering bevat het bestand dan de gegevens die op dat moment beschikbaar zijn. Deze paragraaf geeft een overzicht van de momenten waarop de belangrijkste databestanden worden aangeleverd, en de mate waarin deze op het moment van levering gevuld zijn. Dit laatste wordt uitgedrukt als een percentage van het totale aantal records bij de definitieve levering. Momenteel worden er op twee momenten cijfers gepubliceerd over inkomen: eerst worden er voorlopige cijfers gemaakt, die rond het einde van jaar  $t+1$  klaar zijn, en vervolgens worden de definitieve cijfers aan het einde van jaar  $t+2$  gepubliceerd. Met behulp van het overzicht van de leveringen en vullingen van de databestanden kan bekeken worden of er niet al eerder,

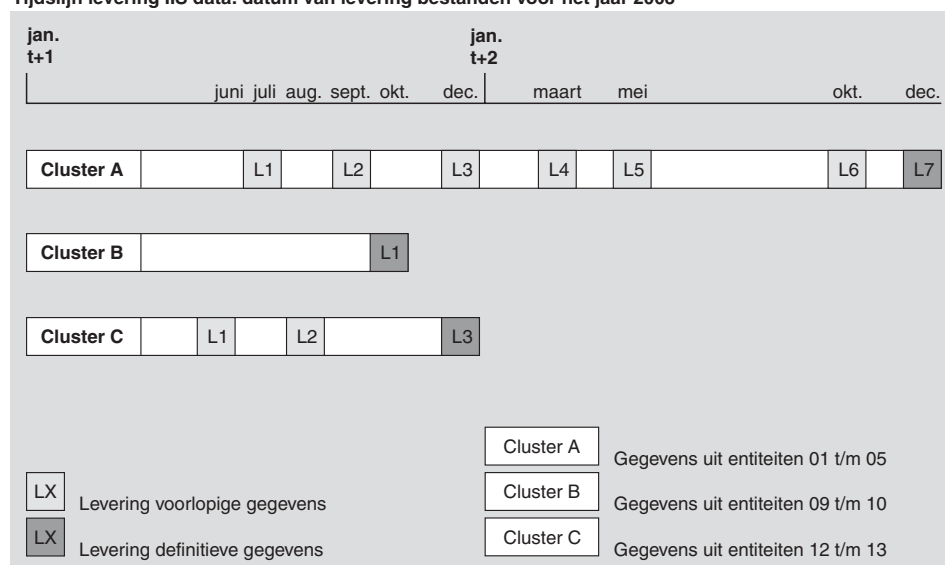


bijvoorbeeld medio jaar  $t+1$ , voorlopige cijfers gepubliceerd kunnen worden over de inkomensverdeling. Er kan een goed beeld worden gevormd over wanneer bepaalde afleidingen gedaan kunnen worden en welke gegevens eventueel bijgeschat moeten worden om snellere publicatiecijfers te maken.

De schema's 8 en 9 geven een overzicht van de tijdslijn voor het jaar 2003 met betrekking tot de levering en vulling van de databestanden van de belangrijkste entiteiten uit het IIS (namelijk de entiteiten 01 tot en met 05 en 09, 10, 12 en 13). Hiertoe zijn deze entiteiten in drie clusters verdeeld. De indeling in clusters is enerzijds gebaseerd op het tijdstip waarop bepaalde databestanden aangeleverd worden en anderzijds de herkomstbron van de gegevens. De gegevens over entiteiten 01 t/m 05 worden altijd samen geleverd en worden daarom samengebracht in cluster A. Cluster B bevat de entiteiten 09 en 10, dus de gegevens over ontvangen studiefinanciering. Cluster C ten slotte bestaat uit de entiteiten 12 en 13, die alle gegevens van de financiële instellingen bevatten. In schema 8 wordt weergegeven in welke maand leveringen plaatsvinden en in schema 9 wordt vervolgens weergegeven in welke mate de bestanden gevuld zijn bij een bepaalde levering.

In schema 8 is te zien dat er voor cluster A in totaal 7 leveringen plaatsvinden, verspreid over twee jaar tijd. De definitieve data voor onderzoeksjaar  $t$  worden geleverd in de maand december van het jaar  $t+2$ . De gegevens van de entiteiten in cluster B worden eenmaal geleverd, in oktober van het jaar  $t+1$ . Ten slotte vinden er voor cluster C drie leveringen plaats, in de maanden juni, augustus en december van het jaar  $t+1$ .

**Schema 8**  
Tijdslijn levering IIS data: datum van levering bestanden voor het jaar 2003



In schema 9 zijn de clusters uitgesplitst naar de verschillende entiteiten om na te gaan in welke mate de gegevens compleet zijn bij elke levering. Uiteraard kan dit verschillen voor de entiteiten binnen eenzelfde groep. Zo is bijvoorbeeld te zien dat de demografische gegevens (entiteit 01) al vanaf de eerste levering compleet zijn, terwijl er dan nog geen aanslag- en correctiegegevens beschikbaar zijn (entiteit 03). De aangiftegegevens (entiteit 02) zijn onderverdeeld in de zogenaamde H-biljetten en L-biljetten. De H-biljetten zijn echte aangiftetiljetten zoals ze bij de Belastingdienst binnenkomen. De L-biljetten zijn fictieve biljetten die afgeleid worden (binnen het CBS) op basis van de gegevens uit de Fibase. Bij de levering van entiteit 05 (de Fibase-gegevens) worden percentages van meer dan 100 procent waargenomen. Dit komt doordat er nog correcties kunnen worden aangegeven door de inhoudingsplichtigen met betrekking tot loonbetalingen. Uit schema 9 kan worden afgeleid dat de gegevens van de meeste entiteiten bijna compleet zijn aan het einde van het jaar  $t+1$ . Enkel entiteit 03 bevat op dat moment nog maar weinig gegevens (39 procent).

**Schema 9**

**Tijdslijn levering IIS-data: vulling van bestanden voor het jaar 2003**

jan. t+1						jan. t+2									
	juni	juli	aug.	sept.	okt.	dec.	maart	mei	juni	juli	aug.	sept.	okt.	nov.	dec.

**Cluster A**

ENT 01		100		100		100		100		100		100		100		100
ENT 02	H-biljetten	43		55		61		66		67		69		72		72
	L-biljetten	50		40		36		32		31		30		28		28
	Totaal	93		95		97		98		98		99		100		100
ENT 03		0		0		39		67		81		90		100		100
ENT 05		113		98		100		100		100		100		100		100

**Cluster B**

ENT 09	100
ENT 10	100

**Cluster C**

ENT 12	73		95		100
ENT 13	88		100		100

H-biljetten: aangiftebiljetten die binnenkomen via de belastingdienst.

L-biljetten: biljetten geïmputeerd met behulp van gegevens uit de Fibase.

## 5. Throughput

Uit het inputproces zijn nu inkomensgegevens beschikbaar over de kernpersonen en hun mede-adresbewoners. Deze gegevens worden met behulp van het Inkomens Productie Systeem (IPS) door het CBS gebruikt om te komen tot een IPO-publicatiebestand. In het IPS worden de inkomensbegrippen en inkomensbestanddelen op basis van de IIS-data afgeleid, waarbij voortdurend controles plaatsvinden op de gebruikte gegevens. Daarnaast worden ontbrekende gegevens bijgeschat op basis van o.a. gegevens van het voorafgaande jaar. Het IPS levert dan een bestand met de publicatievariabelen, dat vervolgens gewogen dient te worden naar bevolkingsgegevens. Ten slotte vinden er plausibiliteitscontroles op de data plaats. Het vervolg van deze paragraaf behandelt al deze stappen van het verwerkingsproces in meer detail. Het verwerken (i.e. afleiden, controleren en imputeren) van gegevens via het IPS zal beschreven worden in paragraaf 5.1. In paragraaf 5.2 komt de weging van het IPO aan bod. Paragraaf 5.3 behandelt de plausibiliteitscontroles.

### 5.1 Inkomens Productie Systeem

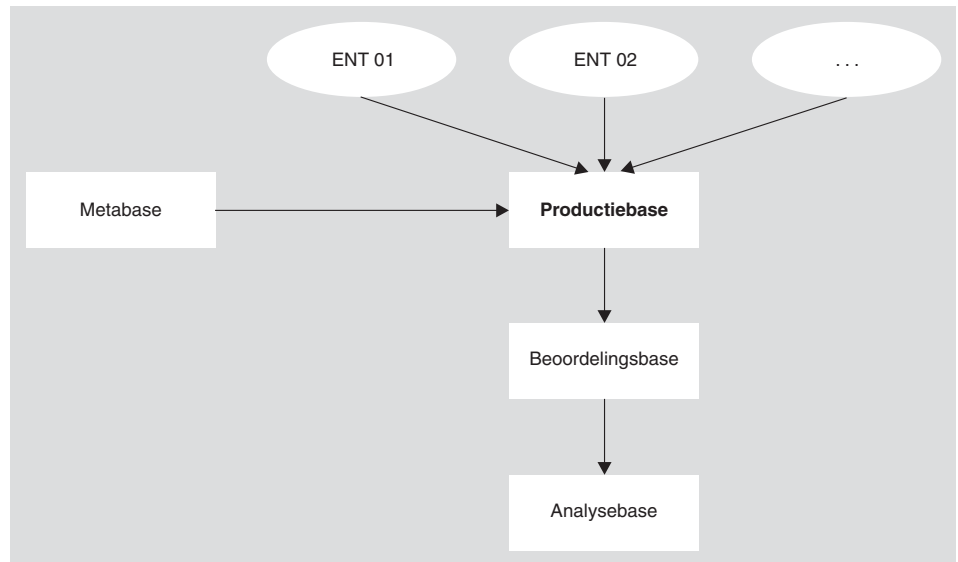
De verwerking van de IPO-gegevens gebeurt in het Inkomens Productie Systeem (IPS). Schema 10 geeft een overzicht van de belangrijkste onderdelen van het IPS.<sup>1)</sup> De *productiebase* is de kern van het systeem, de werkomgeving waar de afleidingen en berekeningen van de inkomensbegrippen en -bestanddelen plaatsvinden. Hiervoor zijn twee belangrijke informatiebronnen nodig: ten eerste de gegevens uit de verschillende entiteiten (deze komen uit het IIS, zoals besproken in de vorige paragraaf) en ten tweede zijn specifieke regels nodig om te bepalen hoe al deze gegevens verwerkt moeten worden. Deze afleidingsregels staan beschreven en opgeslagen in de *metabase*. Naast de afleidingsregels bevat de metabase ook gegevens over alle variabelen (zoals het type, de mogelijke waarden die een variabele kan aannemen, etc.) en over alle invoerbestanden die nodig zijn voor de afleidingen. De metabase fungeert dus als informatiebron voor de productiebase; het is als het ware een receptenboekje dat beschrijft hoe de gegevens uit de entiteiten verwerkt moeten worden. In de productiebase worden de inkomensbegrippen en -bestanddelen met behulp van de informatie uit de metabase afgeleid en vervolgens naar de *beoordelingsbase* gestuurd. In de beoordelingsbase vindt een controle plaats op het verloop van de productie. Indien de productie correct is verlopen worden de outputvariabelen in de *analysebase* opgeslagen. De beoordelingsbase en de analysebase zijn databanken die de publicatievariabelen voor een bepaald onderzoeksjaar bevatten. Op een aantal controles in de beoordelingsbase na vinden hier verder geen speciale bewerkingen plaats. Daarom wordt er op deze twee bases niet verder ingegaan. In het vervolg van deze paragraaf worden de metabase (paragraaf 5.1.1) en de productiebase (paragraaf 5.1.2) nader toegelicht, waarbij ook een voorbeeld aan bod komt over het afleiden van variabelen in het IPS.

#### 5.1.1 Metabase

Voor elk onderzoeksjaar zijn in de metabase een aantal metaset opgeslagen. Een metaset dient als informatiebron voor een bepaalde versie van een IPO-bestand behorende bij dat onderzoeksjaar. De gegevens in deze metaseten zijn onderverdeeld in een aantal verschillende gegevensverzamelingen. De belangrijkste gegevensverzamelingen zijn PERSOON, HUISHOUDEN, GEZIN, PAAR en LOONKAART. Deze gegevensverzamelingen zijn onderling verbonden zoals weergegeven in het datamodel in schema 11. Zo behoort elke persoon tot één enkel huishouden en horen bij deze persoon een bepaalde set loonkaarten (voor elke werkkring een loonkaart). Publicatiecijfers worden op het niveau van elk van de vijf gegevensverzamelingen samengesteld. Het grote voor-

<sup>1)</sup> Tot onderzoeksjaar 2007 vindt, voordat de gegevens verwerkt worden in het IPS, een imputatieslag plaats. Ontbrekende gegevens worden aangevuld op basis van gegevens van het voorafgaande jaar.

Schema 10  
Schematisch overzicht IPS

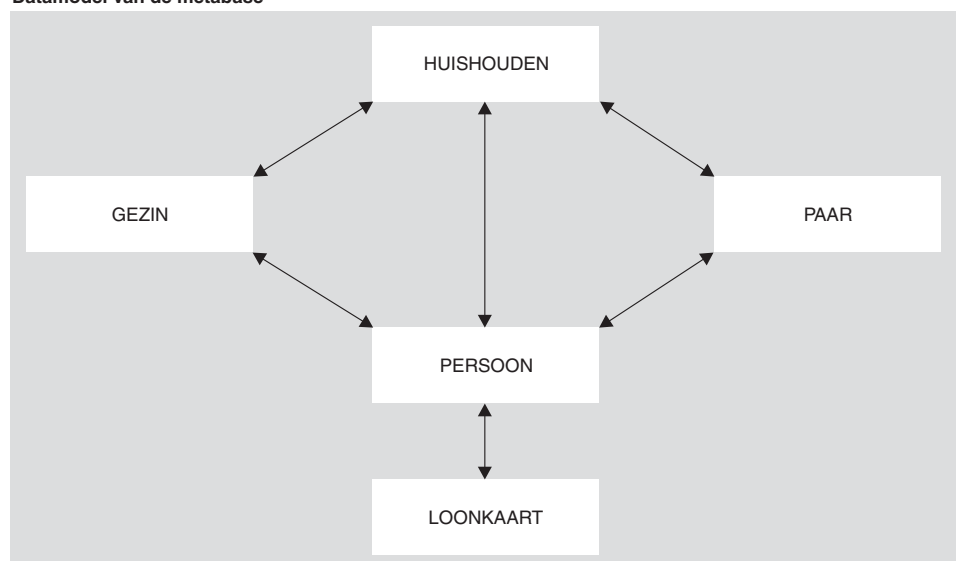


deel van het gebruikte datamodel is dat het een *gesloten* model is. Dat wil zeggen dat alle gegevensverzamelingen met elkaar verbonden zijn, wat veel profijt oplevert bij het afleiden van variabelen. Bijvoorbeeld, als een variabele op loonkaartniveau afgeleid moet worden, waarbij gebruik gemaakt van een andere variabele op huishoudensniveau, hoeft slechts de relatie loonkaart – persoon gecontroleerd te worden. Als deze relatie gecontroleerd is kan, omwille van de geslotenheid van het datamodel, dan ook de relatie persoon – huishouden gelegd worden.

Voor elke gegevensverzameling bevat de metabase de volgende informatie:

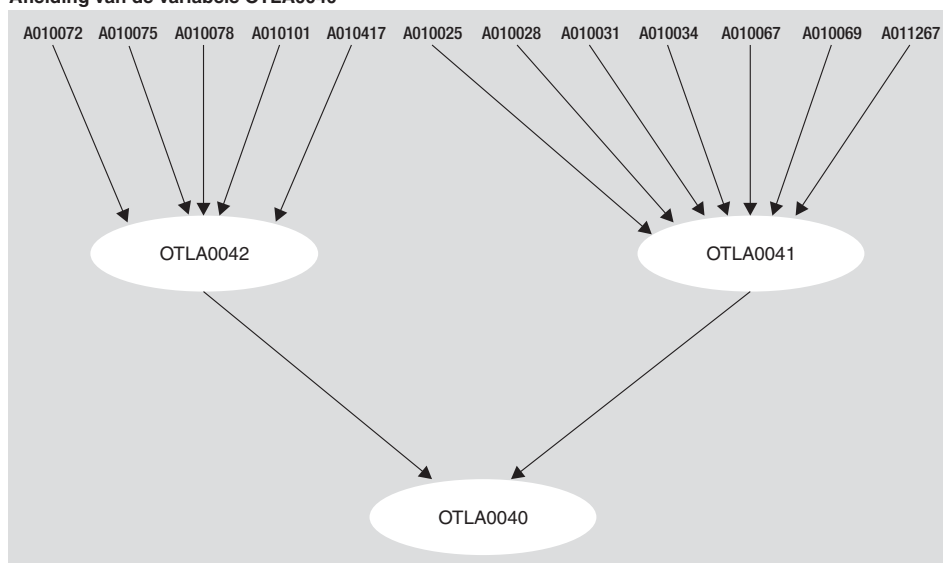
- *sleutel*: dit is een unieke identificatiecode, zoals het sofinummer van een persoon of het nummer van een loonkaart.
- *velden*: hierin worden de bij de gegevensverzameling behorende variabelen en hun afleidingen opgeslagen. Voor elke variabele is aangegeven of het een importvariabele, een afgeleide variabele of een maatwerkvariabele (bijvoorbeeld een premiepercentage) is. Voor elke afgeleide variabele is verder aangegeven of het een afleiding in één stap (basisafleiding) of in meerdere stappen (getrapte afleiding) betreft. In schema 12 is ter illustratie een voorbeeld van een afleiding gegeven zoals deze in de metabase wordt weergegeven. Het gaat hier om de afleiding van de variabele OTLA0040, totaal

Schema 11  
Datamodel van de metabase



- loonaangifte. De waarde van deze variabele wordt bepaald met de volgende afleidingsregel:  $OTLA0040 = OTLA0041 + OTLA0042$ . Er worden in de afleiding dus twee andere variabelen gebruikt, namelijk OTLA0041 (loon) en OTLA0042 (uitkering). Voor de afleiding van deze twee variabelen worden gegevens uit de belastingaangifte gebruikt. Dit zijn de variabelen waarvan de naam met de letter A begint. De afleidingsregel voor de variabele OTLA0040 is zeer eenvoudig, maar een afleiding kan ook ingewikkeldere constructies bevatten (bijvoorbeeld de 'if-then' constructie).
- *importbestanden*: dit zijn de inputbestanden die gebruikt worden voor het samenstellen van velden uit de betreffende gegevensverzameling. Als er geen importbestanden staan weergegeven, wil dit zeggen dat alle variabelen in deze gegevensverzameling worden afgeleid.

**Schema 12**  
Afleiding van de variabele OTLA0040



- Naast de gegevensverzamelingen die in het datamodel in schema 11 zijn opgenomen, zijn er nog twee andere gegevensverzamelingen in de metabase: OPHOOG en PREMIE.
- PREMIE: deze gegevensverzameling bevat premiegegevens voor alle personen die in de IPO-steekproef zijn opgenomen. Tot en met het maken van de voorlopige cijfers van 2004 werd dit premiebestand als informatiebron voor de metabase gebruikt. Vanaf de definitieve cijfers van 2004 zijn er echter ook afleidingsregels voor de premieberekeningen in de metabase opgenomen. Dat wil zeggen dat de premies vanaf dan in de metabase worden afgeleid, en niet meer extern worden berekend zoals daarvoor. Het premiebestand is dus overbodig geworden. Voor het afleiden van de premies zijn wel enkele hulpvariabelen nodig, zoals premie- en belastingpercentages. De waarden van deze hulpvariabelen zijn in de metabase onder 'constanten' opgeslagen.
  - OPHOOG: in deze gegevensverzameling staan gegevens over de ophoging van de cijfers naar de hele bevolking (zie paragraaf 5.2).

Tot slot bevat een metaset, naast de gegevensverzamelingen, ook nog een lijst met constanten (zoals het brutominimumloon en de maximale kinderkorting) en de codetyperingen van alle voorkomende variabelen. Deze laatste lijst is echter nog niet compleet.

### 5.1.2 Productiebase

Informatie die in de metabase is opgeslagen, wordt in de productiebase gebruikt om de waarde van publicatie- en hulpvariabelen te bepalen. De productie verloopt in twee fasen. Elke fase doorloopt een aantal verschillende procedures. Tussen de eerste en de tweede fase kunnen ook nog een aantal optionele functies worden uitgevoerd.

### *Fase 1*

In de eerste fase worden vijf verschillende functies uitgevoerd. Deze functies worden altijd uitgevoerd in de volgorde waarin ze hieronder beschreven worden.

- Genereren database: in deze stap wordt de tabellenstructuur bepaald. Er wordt gekeken hoe deze het best kan worden opgebouwd zodanig dat de productie daarna zo snel en zo efficiënt mogelijk kan draaien. Dit betekent dat er clusters van variabelen worden gemaakt die een gelijk aantal stappen nodig hebben in hun afleiding. De variabelen in eenzelfde cluster vormen samen de basis van een tabel. Door deze opbouw heeft de productiebase geen vaste vorm, maar wordt de vorm afgeleid uit de metabase. Een verandering in de metabase kan dus een verandering in de tabellenstructuur van de productiebase tot gevolg hebben.
- Inlezen data fase 1: in deze stap wordt de nodige data geïmporteerd.
- Controleren dataconsistentie: in deze stap worden enkele dataconsistentiechecks uitgevoerd, zoals een controle of een paar uit twee personen bestaat, of de waarde van een variabele binnen zijn bereik ligt, of alle importbestanden aanwezig zijn, etc. Het gaat hierbij dus om een aantal technische controles op de data, niet om controles op de metabase zelf.
- Genereren afleidingen fase 1: in deze stap vindt de vertaalslag plaats van de afleidingen in woorden naar SQL.
- Start afleidingen fase 1: in deze stap worden de afleidingen in SQL uitgevoerd.

### *Tussenfase*

In een tussenfase wordt vervolgens de mogelijkheid gegeven om verschillende optionele functies uit te voeren.

- Constanten wijzigen: in deze stap wordt de mogelijkheid geboden om de waarde van constanten te veranderen. Deze functie kan alleen bij de eerste versie van een meta-set worden uitgevoerd.
- Herberekenen fase 1: in deze stap worden de variabelen uit fase 1 opnieuw afgeleid indien dit door de wijziging van constanten nodig is. Deze functie kan alleen worden uitgevoerd na de uitvoering van de functie 'constanten wijzigen'.

### *Fase 2 Run*

In de tweede fase worden nog drie functies uitgevoerd. Net als in de eerste fase dienen deze functies te worden uitgevoerd in de volgorde waarin ze hier worden besproken.

- Genereren afleidingen fase 2: dit gebeurt op de zelfde manier als in fase 1. Afleidingsregels worden vertaald naar SQL. In de metabase is aangegeven welke afleidingen in fase 1 en welke in fase 2 dienen te gebeuren.
- Start afleidingen fase 2: in deze stap worden de afleidingen in SQL uitgevoerd.
- Vullen beoordelingsbase: in deze laatste stap worden alle variabelen (hulp- en publicatievariabelen) overgeheveld naar de beoordelingsbase. Hier zal vervolgens met behulp van het programma 'terugkantelen' een controle op de variabelen plaatsvinden voordat de definitieve publicatievariabelen in de analysebase worden opgeslagen.

### *Het afleiden van inkomensbestanddelen*

Het afleiden van variabelen is een belangrijk onderdeel van het IPS. In sommige gevallen is het afleiden van een variabele zeer eenvoudig, zoals bij het bepalen van de hoogte van de ontvangen studiefinanciering. Deze gegevens komen rechtstreeks binnen vanuit één enkele bron (in dit geval de IBG) en kunnen eenvoudigweg worden overgenomen in het IPO-bestand. In andere gevallen is de afleiding echter niet zo eenvoudig en moet duidelijk gespecificeerd worden hoe de gegevens uit de verschillende bronnen behandeld moeten worden. Nog andere bestanddelen, zoals de kinderbijslag, worden juist weer groten-deels berekend. In tabel 2 wordt een overzicht gegeven van de gegevens die worden gebruikt bij het afleiden van de belangrijkste inkomensbestanddelen en wordt aangegeven hoe het betreffende bestanddeel wordt bepaald.

### *Voorbeeld van een afleiding: loonrubrieken*

Bij het opstellen van de verschillende loonrubrieken worden gegevens gebruikt uit meerdere bronnen. Ten eerste wordt de Fibase (entiteit 05) gebruikt, waarin de gegevens over lonen en uitkeringen staan opgeslagen. Ten tweede worden de aangiftegegevens uit entiteit 02 gebruikt. Uiteraard kan het voorkomen dat de gegevens uit de verschillende bronnen met elkaar conflicteren. Zo kan het bijvoorbeeld zijn dat het aangegeven loon

**Tabel 2**  
**Overzicht afleiden inkomensbestanddelen**

Inkomensbestanddeel	Entiteiten	Methode
Inkomen uit arbeid	02, 05 <sup>1)</sup>	Waarneming + Berekening
Inkomen uit eigen onderneming	02 <sup>1)</sup>	Waarneming + Berekening
Inkomen uit vermogen	02, 12, 13 <sup>1)</sup>	Waarneming + Berekening
Uitkering inkomensverzekeringen	02, 05 <sup>1)</sup>	Waarneming + Berekening
Uitkering sociale voorzieningen		
– ABW, Wajong, IOAW, ...	05 <sup>1)</sup>	Waarneming + Berekening
– Kinderbijslag		Berekening
– Studiefinanciering	10	Waarneming
Ontvangen gebonden overdrachten		
– Rijksbijdrage eigen woning	02	Waarneming
– Tegemoetkoming studiekosten	10	Waarneming
– Huurtoeslag	09	Waarneming
– Zorgtoeslag	09	Waarneming
– Kinderopvang toeslag	09	Waarneming
Ontvangen inkomensoverdrachten	02	Waarneming
Betaalde inkomensoverdrachten	02	Waarneming
Premie inkomensverzekeringen	<sup>1)</sup>	Berekening
Premie ziektekostenverzekeringen	<sup>1)</sup>	Berekening
Belastingen	02, 05 <sup>1)</sup>	Waarneming + Berekening

<sup>1)</sup> Vanaf 2007 via de Polisadministratie.

lager ligt dan het loon zoals het in de Fibase is opgeslagen. In dat geval wordt aan de hand van verschillende afleidingsregels bepaald wat er met de gegevens van elke bron gedaan moet worden. Op deze manier worden niet alleen de nodige variabelen afgeleid, maar wordt ook tegelijkertijd een controle uitgevoerd op de inputbestanden.

In het vervolg van deze paragraaf wordt het afleidingssysteem verduidelijkt aan de hand van de afleiding van de fiscale loonrubrieken. Deze afleiding verloopt in zes stappen, waarbij afleidingen zowel op loonkaartniveau als op persoonsniveau plaatsvinden.

1. De afleiding van de loonkaarttypes (loonkaartniveau).
2. Vergelijken totaal loon Fibase met totaal loonaangifte (persoonsniveau).
3. De afleiding van de totale loonsom naar rubriek (persoonsniveau).
4. De afleiding van de sociaaleconomische categorie (persoonsniveau).
5. De afleiding van de rubriek voor het bestanddeel 'loon onbekend' (persoonsniveau).
6. De afleiding van de eindrubrieken voor loon- en uitkeringsgegevens (persoonsniveau).

#### *Stap 1: de afleiding van de kaarttypes*

In een eerste stap wordt er voor elke loonkaart van een persoon een kaarttype afgeleid. Het kaarttype geeft aan in welke rubriek het bedrag van de loonkaart moet worden meegeteld. In de Fibase staat voor elke loonkaart een inkomenscode (soort inkomen). Voor de afleiding van het kaarttype wordt in principe gekeken naar het soort inkomen (bijvoorbeeld 'loon ambtenaar' of 'uitkering AOW'). Voor bijzondere groeperingen (zoals muzikanten of beroepssporters) en bijzondere inkomens (zoals eenmalige beloningen) wordt daarnaast gebruik gemaakt van nadere indelingen opgesteld door de Belastingdienst.

#### *Stap 2: vergelijken totaal loon Fibase met totaal loonaangifte*

In een tweede stap wordt het totale loon zoals het is opgenomen in de Fibase vergeleken met het totale loon van de aangifte. Als er sprake is van een verschil geldt de volgende regel: <sup>2)</sup> het loon zoals opgenomen in de Fibase wordt volledig overgenomen, samen met het eventuele overschot van de aangifte (indien het loonaangifte groter is dan het loon Fibase). Met een code wordt aangegeven om welke situatie het gaat:

1. Verschil kleiner dan 100 euro.
2. Loonaangifte > loon Fibase.
3. Loonaangifte < loon Fibase.
4. Loonaangifte = 0 en loon Fibase >0.
5. Loon fibase = 0 en loonaangifte >0.

In geval van een overschot (situatie 2 en 5) wordt vervolgens bekeken in welke loonrubriek dit bestanddeel meegeteld kan worden. In geval het overschot niet meteen ingedeeld kan worden, worden extra gegevens afgeleid (zie stap 4 en 5).

<sup>2)</sup> Deze afleidingsregel geldt vanaf 2001. Voorheen werd aangenomen dat het bedrag zoals opgegeven in de aangifte correct was.

### *Stap 3: de afleiding van de totale loonsom naar rubriek*

Per rubriek wordt op persoonsniveau de loonsom bepaald op basis van het kaarttype van de loonkaarten (stap 1). Een nog niet ingedeeld overschot uit de aangifte (zie stap 2) wordt vooralsnog in de rubriek 'loon onbekend' ondergebracht.

### *Stap 4: de afleiding van de sociaal economische categorie*

In stap 4 wordt, via een aantal hulpvariabelen, de fiscaal-economische categorie afgeleid. Dat is een indeling naar de belangrijkste bron van het fiscale inkomen.

### *Stap 5: de afleiding van de rubriek voor het bestanddeel 'loon onbekend'*

Op basis van de fiscaal-economische categorie wordt het bestanddeel 'loon onbekend' ingedeeld. Dit is een indeling op persoonsniveau. Merk op dat hier sprake is van imputatie van gegevens.

### *Stap 6: de afleiding van de eindrubrieken voor loon en uitkeringsgegevens*

Nadat duidelijk is tot welke rubriek het bestanddeel 'loon onbekend' behoort kunnen de eindrubrieken voor fiscaal inkomen worden afgeleid. Deze rubrieken worden dan vervolgens gebruikt voor de premieberekeningen bij het afleiden van de betaalde overdrachten in de vorm van werknemers- en werkgeverspremies voor de werknemersverzekeringen en de door huishoudens betaalde nominale premies hiervoor.

Na het opstellen van dergelijke specificaties voor afleidingen vindt de vertaling ervan naar de metabase plaats.

## **5.2 Ophoging**

In de ophoging van het IPO wordt aan elke steekproefpersoon en aan elk huishouden een gewicht toegekend. Met deze gewichten wordt beoogd te corrigeren voor scheefheid in de steekproef ten aanzien van enkele belangrijke demografische variabelen (zoals leeftijd en geslacht). Zo kan een ondervertegenwoordiging van bijvoorbeeld jongeren worden opgevangen door aan deze groep een hoger gewicht toe te kennen. De met de gewichten geschatte totalen van de in de weging gebruikte demografische variabelen (i.e. de weegvariabelen) komen dan exact overeen met de hiervoor bekende populatietotalen.

Uit het IPO worden gegevens op zowel persoons- als op huishoudensniveau geschat. Traditioneel wordt bij een dergelijk onderzoek een weging op persoonsniveau uitgevoerd, waarbij geen rekening wordt gehouden met de demografische kenmerken van de overige huishoudensleden. Omdat deze kenmerken binnen een huishouden van persoon tot persoon verschillen, verschillen de persoonsgewichten binnen een huishouden. Dit betekent dat voor eenzelfde huishouden verschillende gewichten kunnen worden verkregen. Om dit te voorkomen wordt bij het IPO een weegmethode toegepast waarbij wordt geforceerd dat de persoonsgewichten binnen een huishouden gelijk zijn aan elkaar en waarbij nog steeds wordt voldaan aan de eis dat de geschatte totalen van de weegvariabelen overeen komen met de bekende populatietotalen. Dit wordt bereikt door een weging op huishoudensniveau uit te voeren waarbij de weegvariabelen van een huishouden gelijk zijn aan de som van de weegvariabelen van de huishoudensleden. Deze weegmethode wordt ook wel 'consistent wegen' genoemd. Het belangrijke voordeel ervan is dat schattingen van overeenkomstige doelvariabelen (bijvoorbeeld het besteedbaar inkomen) op huishoudensniveau en op persoonsniveau consistent zijn met elkaar.

In de weging worden de startgewichten van personen en huishoudens zodanig aangepast dat:

- het gewogen aantal personen voor een combinatie van leeftijd, geslacht, burgerlijke staat, provincie en soort adres (particulier of institutioneel) aansluit op de gegevens van de Bevolkingsstatistiek, en
- het gewogen aantal particuliere huishoudens naar omvang, provincie en leeftijd van het hoofd van het huishouden overeenkomt met de uitkomsten van de Huishoudensstatistiek.



De startgewichten zijn omgekeerd evenredig aan de kans om in de IPO-steekproef terecht te komen. De startgewichten worden in de weging door een combinatie van de hierboven genoemde demografische variabelen, ook wel het weegmodel genoemd, jaarlijks in geringe mate gecorrigeerd tot eindgewichten: in de helft van de gevallen bevindt de correctiefactor zich tussen 0,97 en 1,03. Naast aansluiting op de bekende populatie-aantallen van de Bevolkingsstatistiek en de Huishoudensstatistiek<sup>3)</sup> voor de in het weegmodel opgenomen combinaties van variabelen, leidt de weging doorgaans ook tot nauwkeurigere uitkomsten van de doelvariabelen.

Voor het uitvoeren van de ophoging wordt een bestand opgeleverd met, naast de identificerende kenmerken voor de personen en huishoudens, de variabelen nodig voor de weging en de startgewichten. Nadat de eindgewichten berekend zijn, vinden de volgende plausibiliteitscontroles plaats:

- Controle van de correctiefactoren (met statistische grootheden als gemiddelde, hoogste en laagste factor). Dit gebeurt voor de verschillende publicatiegroepen (gebaseerd op de weegvariabelen) op persoons- en huishoudensniveau.
- Controle op het jaarlijkse verloop van de correctiefactoren, eveneens voor verschillende publicatiegroepen.
- Controle op overeenstemming met de bekende populatietotalen.
- Controle op de hoogte van de verschillen tussen met startgewichten geschatte totalen en met eindgewichten geschatte totalen voor verschillende publicatiegroepen, waarbij tevens het jaarlijkse verloop van deze verschillen wordt bekeken.

Na goedbevinding van de weegresultaten wordt een publicatiebestand samengesteld, waarin de eindgewichten samen met de variabelen uit de analysebase terecht komen.

### 5.3 Plausibiliteitsonderzoek

Als laatste stap in het throughputproces vinden de plausibiliteitscontroles plaats. Tijdens dit proces worden de afgeleide gegevens gecontroleerd. Zo worden bijvoorbeeld de cijfers van een aantal jaren naast elkaar gelegd om te bekijken of de nieuwste cijfers plausibel zijn ten opzichte van de eerdere jaren. Indien de verkregen cijfers niet geheel plausibel zijn, worden de oorzaken hiervan opgespoord en opgelost, en wordt er teruggegaan naar het throughputproces. Na herstel worden de cijfers dan opnieuw gemaakt. Als bij de plausibiliteitscontroles uiteindelijk geen grote fouten worden ontdekt, wordt het definitieve IPO-publicatiebestand opgesteld.

Bij het verrichten van het plausibiliteitsonderzoek wordt een onderscheid gemaakt tussen technische controles en de eigenlijke plausibiliteitscontroles. Technische controles gaan vooraf aan de plausibiliteitscontroles. Een belangrijk verschil tussen deze twee soorten controles is dat bij de technische controles subjectieve beoordelingen geen rol spelen, met andere woorden iets is goed of fout. Deze technische controles kunnen daarom ook worden uitgevoerd door mensen met geringe(re) inhoudelijke kennis. In de praktijk lopen deze controles echter door elkaar. In het vervolg van deze paragraaf worden beide soorten controles verder toegelicht. In paragraaf 5.3.1 komen de technische controles aan bod en paragraaf 5.3.2 behandelt dan de eigenlijke plausibiliteitscontroles.

#### 5.3.1 Technische controles

Technische controles betreffen onder meer controles op de invoergegevens, controles op de afleidingsregels, en controles op de ophoging. Het controleren van de invoergegevens betreft onder andere nagaan of actuele bestanden worden gebruikt en of de gegevens correct zijn ingelezen. Bij het controleren van de afleidingen gaat het vooral om controles op de correcte werking van nieuwe afleidingsregels. Er wordt gecontroleerd

<sup>3)</sup> Voor de Bevolkingsstatistiek en de Huishoudensstatistiek wordt de stand van 1 januari  $t+1$  gebruikt.

of de verschillende institutionele bedragen correct zijn (bijvoorbeeld kinderbijslagbedragen en sociale minima), of de equivalentiefactoren goed zijn afgeleid, of de inkomensbegrippen goed zijn afgeleid, etc.

In theorie kan het onderscheid tussen de technische en plausibiliteitscontroles duidelijk worden gemaakt, maar in praktijk lopen deze twee vaker door elkaar. Zo kan het bijvoorbeeld voorkomen dat er bij de technische controle van een bepaald inkomensbestanddeel ontdekt wordt dat het bestanddeel technisch gezien wel correct is afgeleid, maar dat de afleiding ten onrechte verschilt van die van voorgaande jaren.

### 5.3.2 Plausibiliteitscontroles

Plausibiliteitscontroles bevatten onder meer een controle op de resultaten van (nieuwe) afleidingsregels: zijn er geen onvoorziene, ongewenste resultaten? Verder wordt er, rekening houdend met ontwikkelingen in de maatschappij, onderzocht of de ontwikkeling van jaar op jaar plausibel zijn en of de uitkomsten van verschillende bevolkingsgroepen ten opzichte van elkaar plausibel zijn. Als er geen verklaring gegeven kan worden voor onverwachte resultaten, is een diepgaand onderzoek nodig. De uitvoering van deze plausibiliteitscontroles gebeurt veelal handmatig; met behulp van een aantal SPSS-set-ups worden bepaalde tabellen gemaakt, die vervolgens stuk voor stuk worden gecontroleerd op onregelmatigheden. Bij deze manier van werken is het noodzakelijk dat degene die de controles uitvoert voldoende inhoudelijke kennis heeft om te kunnen beoordelen of bepaalde veranderingen wel of niet plausibel zijn. Alle onregelmatigheden worden gerapporteerd in een problemenbase. Op het moment dat er een aantal probleempunten zijn ontdekt, kan er worden besloten om de productie opnieuw op te starten met de nodige verbeteringen. Indien er geen noemenswaardige problemen meer worden ontdekt, wordt het bestand goedgekeurd. Dat wil zeggen dat de cijfers vanaf dat moment gebruikt kunnen worden voor analyses en publicaties.

De volgende plausibiliteitscontroles worden standaard uitgevoerd:

- Controle op extreme waarden voor alle variabelen.
- Controle op demografische persoons- en huishoudenskenmerken.
- Controle op indelingsvariabelen zoals sociaaleconomische categorie, voornaamste inkomensbron, sociale minima, provincie en stedelijkheidsgraad.
- Controle op specifieke inkomensbestanddelen zoals privégebruik auto van werkgever, buitenlands inkomen en aanvullend pensioen.
- Controle op de inkomensbegrippen primair, bruto en besteedbaar inkomen.
- Controle op institutionele bedragen zoals kinderbijslag, huurtoeslag, studiefinanciering en AOW-bedragen.

Afhankelijk van de resultaten van deze standaardcontroles zijn er aanvullende controles nodig. Daarnaast kunnen extra controles nodig zijn indien aanpassingen in afleidingen of veranderingen in het overheidsbeleid (zoals de invoering van de nieuwe ziektekostenwet in 2006) hebben plaatsgevonden.

## 6. Output

De definitieve gegevens uit de analysebase worden in een laatste fase gebruikt om publicatiecijfers samen te stellen. Voor de Inkomensstatistieken gebeurt dit in principe voor elk onderzoeksjaar op twee verschillende momenten. Eerst worden er voorlopige inkomenscijfers gepubliceerd. Dit wordt meestal rond het einde van het jaar  $t+1$  gerealiseerd. Daarna worden, aan het eind van het jaar  $t+2$ , definitieve cijfers gepubliceerd. Cijfers worden zowel op persoonsniveau als op huishoudensniveau samengesteld. Hierbij is het gebruikelijk om inkomens, omwille van onderlinge vergelijkbaarheid, te standaardiseren. Paragraaf 6.1 behandelt deze standaardisatie. Paragraaf 6.2 gaat vervolgens in op de verschillende (interne en externe) media die binnen het CBS gebruikt worden voor publicaties.

### 6.1 Standaardiseren van het huishoudensinkomen

Op huishoudensniveau worden gegevens gepubliceerd over het primair, bruto- en besteedbaar inkomen (zie paragraaf 2). Het gebruik van het besteedbaar huishoudensinkomen als indicator voor welvaart heeft als nadeel dat geen rekening gehouden wordt met verschillen in omvang en samenstelling van de huishoudens. Met behulp van equivalentiefactoren wordt het besteedbaar inkomen voor deze verschillen gecorrigeerd. Het aldus gecorrigeerde inkomen wordt aangeduid als het gestandaardiseerd besteedbare huishoudensinkomen. De equivalentiefactoren zijn gebaseerd op de budgetverdelingsmethode (zie Siermann, Van Teeffelen en Urlings, 2004). Door standaardisatie worden inkomens van verschillende typen huishoudens qua welvaartsniveau vergelijkbaar gemaakt met inkomens van eenpersoonshuishoudens. Voor alleenstaanden is de equivalentiefactor gelijk aan 1. Afhankelijk van het aantal volwassenen en minderjarige kinderen wordt voor elke extra volwassene 0,19 tot 0,37 en voor elk extra minderjarig kind 0,15 tot 0,33 aan deze factor toegevoegd. Zo is voor een echtpaar de equivalentiefactor gelijk aan 1,37. Dit betekent dat echtparen met een besteedbaar inkomen van 13 700 euro vergelijkbaar zijn met alleenstaanden die een besteedbaar inkomen van 10 duizend euro hebben.

Om inkomensgegevens op persoonsniveau te publiceren wordt, naast het persoonlijk inkomen, ook gebruik gemaakt van het gestandaardiseerd besteedbaar huishoudensinkomen. De levensstandaard van een persoon is immers verbonden met het inkomen van zijn huishouden.

### 6.2 Verschillende publicatiemedia

Het publiceren van inkomenscijfers gebeurt in artikelen voor bijvoorbeeld het webmagazine van het CBS, de CBS-kwartaalpublicatie Sociaaleconomische trends, of voor een extern blad of tijdschrift. Cijfers worden in de vorm van tabellen op StatLine gepubliceerd.

Elk jaar wordt een aantal tabellen over inkomen op StatLine gepubliceerd. Een deel van de StatLine-tabellen is op dit moment nog niet gevuld (zie tabel 3).

**Tabel 3**  
**Overzicht van StatLine-tabellen**

<b>Rubriek</b>	<b>StatLine-tabel</b>
<b>Inkomen van personen</b>	Gemiddeld inkomen Inkomensklassen o.b.v. inkomensgrenzen Inkomensgroepen o.b.v. kwantielen Personen met laag inkomen Samenstelling inkomen
<b>Inkomen van huishoudens</b>	Gemiddeld inkomen Inkomensklassen o.b.v. inkomensgrenzen Inkomensgroepen o.b.v. kwantielen Huishoudens met laag inkomen Samenstelling inkomen
<b>Koopkrachtontwikkeling</b>	Dynamische koopkrachtontwikkeling Dynamische overgangen naar voornaamste inkomensbron Dynamische overgangen naar huishoudenssamenstelling
<b>Geplande tabellen</b>	Inkomensongelijkheid Belasting- en premieheffing Economische zelfstandigheid AOW-ers

## 7. Toekomstige ontwikkelingen

Zoals opgemerkt in paragraaf 4.2 zijn definitieve cijfers over onderzoeksjaar  $t$  idealiter beschikbaar eind  $t+2$ . Voorlopige cijfers zijn eind  $t+1$  gereed. In de praktijk blijkt echter behoefte te bestaan aan nog eerder gepubliceerde cijfers over  $t$ . Versnelling van het productieproces van het IPO is dus een belangrijke issue. Daarnaast is het wellicht mogelijk het proces op een aantal punten te verbeteren, zonder meteen een versnelling te realiseren. In deze paragraaf wordt nader ingegaan op versnellings- en verbeteringspunten van het IPO-proces. Deze zullen in de komende jaren worden geïmplementeerd.

### 7.1 Versnelling

Voor versnelling van het IPO-proces zijn drie invalshoeken mogelijk:

1. Versnelling ten aanzien van de levering van inputdata.
2. Zoeken naar andere databronnen, die eerder beschikbaar zijn.
3. Op een eerder moment gegevens imputeren/bijstellen met behulp van o.a. gegevens van voorgaande jaren.

#### *Ad (1)*

Met de Belastingdienst worden afspraken gemaakt over een maandelijkse levering van entiteiten aan het IIS. Daarbij kan gedacht worden aan levering van enkel die gegevens die voor de betreffende maand beschikbaar zijn gekomen, aangevuld met eventuele mutaties. Voor bijvoorbeeld de toeslaggegevens (entiteit 09) is dit vanaf 2007 het geval. Voor de andere entiteiten zal dit waarschijnlijk in 2009 worden gerealiseerd.

#### *Ad (2)*

Sinds begin 2006 wordt gebruik gemaakt van de polisadministratie, waarmee eerder dan voorheen gegevens beschikbaar zijn over lonen en uitkeringen (zie paragraaf 4.1.1). Deze gegevens komen maandelijks binnen. Zo zijn in februari 2007 de gegevens van alle maanden van 2006 al binnen, terwijl de Fibase-gegevens pas in juli 2007 beschikbaar zijn. Naar verwachting levert dit vanaf 2008 bruikbare gegevens op.

De gegevens over de huishoudenssamenstelling van de IPO-kernpersonen zijn momenteel medio september  $t+1$  beschikbaar (zie paragraaf 4.1.1). Wanneer gebruik gemaakt gaat worden van de Huishoudensstatistiek (vanaf  $t = 2006$ ) is het wellicht mogelijk om het huishouden van een kernpersoon eerder af te bakenen, wat leidt tot een versnelling in het IPO-proces.

#### *Ad (3)*

Momenteel worden gegevens die ontbreken, op het moment dat voorlopige dan wel definitieve cijfers over  $t$  worden samengesteld, overgenomen van het voorgaande jaar. Het is uiteraard mogelijk om op een eerder moment in  $t+1$  al te gaan imputeren, waardoor 'snelle' IPO-cijfers samengesteld kunnen worden. Het doel hierbij moet zijn om de inkomensverdeling van Nederland globaal in kaart te brengen. Daartoe zouden voor bijvoorbeeld  $t = 2006$  de van toepassing zijnde inkomensbestanddelen van 2004 (definitieve gegevens) geactualiseerd moeten worden naar 2005, waarna deze laatste vervolgens geactualiseerd worden naar 2006. Daarbij kan gebruik gemaakt worden van al bestaande actualiseringsprogrammatuur. Verder dient onderzocht te worden of en hoe in dit kader gebruik gemaakt kan worden van de polisadministratie, de tot dan toe binnengekomen aangiften en aanslaggegevens en voorlopige aanslagen. Bij het welslagen van deze pilot zouden bijvoorbeeld al in juni  $t+1$  inkomensgegevens over  $t$  gepubliceerd kunnen worden.

### 7.2 Verbeteringen

De volgende verbeteringen in het IPO-proces zijn mogelijk:

1. Verbetering IPS in het kader van het Masterplan (de zogenaamde wasstraat)

2. IPO herwegen naar RIO
3. Reduceren en actualiseren van StatLine-tabellen

*Ad (1)*

Binnen het Masterplan zijn verschillende doelstellingen gedefinieerd, zoals het op orde brengen van de ICT-infrastructuur, en daarmee het beter beheersbaar maken van het statistiekproces en het verbeteren van de kwaliteit en de toegankelijkheid van de output. Een belangrijk onderdeel van het Masterplan is het opstellen van zogenaamde generieke procesdiensten. Hiermee wordt beoogd om alle processen om statistieken te maken binnen het CBS meer uniform worden gemaakt. Zo zullen er vijf algemene diensten komen: waarneming, data service center, meta, regelgestuurde verwerking en output. De dienst waarneming zal zich vooral richten op de primaire waarneming, en zal dus waarschijnlijk weinig invloed op de Inkomensstatistieken hebben. Het data service center wordt een centrale opslagplaats waar alle binnekomende en tussentijds bewerkte data worden opgeslagen. Het simpelweg opslaan van binnengekomen data is niet voldoende. Om zinvol gebruik te kunnen maken van de data is meta-informatie (definities, classificaties etc) nodig. Ook de meta-informatie wordt gestandaardiseerd opgeslagen bij de dienst meta. Deze diensten zullen ook weinig invloed op het proces hebben. Met name de laatste twee diensten kunnen een invloed op de Inkomensstatistieken hebben. De dienst regelgestuurde verwerking bekijkt de mogelijkheid om rekenregels die nu veelal vast geprogrammeerd zijn op een geüniformeerde maar meer flexibele manier relatief eenvoudig aan te passen. Te denken valt daarbij aan pakketten zoals Ruleburst en Clementine waarbij statistici binnen een gestandaardiseerde omgeving zelf bewerkingsregels kunnen aanmaken en aanpassen. Deze dienst zal met name het verwerkingsproces binnen het IPS kunnen veranderen. Ten slotte zal de dienst output de toegankelijkheid van de CBS-site moeten verbeteren opdat de statistieken nog meer gebruikt worden. Dit zal vooral aan de analysekant een rol kunnen spelen.

*Ad (2)*

De steekproef van het IPO is een deel van de, veel grotere, RIO-steekproef. In feite is hier sprake van een zogenaamde tweefasensteekproef waarbij in de eerste fase (RIO) globale informatie verzameld wordt en in de tweede fase (IPO) gedetailleerder te werk wordt gegaan. Bij herweging van het IPO naar het RIO kunnen, naast demografische weegvariabelen met uit de Bevolkings- en Huishoudensstatistiek bekende populatietotalen, ook inkomensvariabelen uit de eerste fase gebruikt worden. De populatietotalen van deze inkomensvariabelen zijn dan geschat uit het RIO en kennen, dankzij de omvang van het RIO, een grote mate van nauwkeurigheid. Door de hoge correlatie tussen publicatievariabelen uit het IPO en de in de weging gebruikte inkomensvariabelen (uit beide onderzoeken) zullen doorgaans nauwkeurigere publicatiecijfers worden verkregen dan bij een weging naar enkel demografische kenmerken.

*Ad (3)*

Er zijn verschillende redenen voor een revisie de huidige StatLine-tabellenset:

- Een deel van de tabellen is niet gevuld. Veelal ontbreken de gegevens van recente jaren of is het onderliggende onderwerp nog in ontwikkeling (bijvoorbeeld inkomensongelijkheid). Deze tabellen zullen aan- dan wel opgevoerd moeten worden (actualisatie).
- Is een (nog) overzichtelijkere indeling van de tabellen mogelijk?
- Doorgaans is meer vraag naar inkomens van huishoudens dan inkomens van personen. Wellicht kunnen de tabellen op persoonsniveau op StatLine ingedikt of deels verwijderd worden. Hiervoor kan het aantal hits op StatLine bekeken worden. Dat geeft een goede indicatie van de noodzaak van de tabellen.

## Literatuur

Bruinooge, G. en P.L.J. van de Donk, 1993. Inkomen per definitie. Supplement bij de sociaaleconomische maandstatistiek, nr. 5.

Siermann, C.L.J., P. J.J. van Teeffelen en L.J.M. Urlings, 2004. Equivalentiefactoren 1995–2000. [www.cbs.nl](http://www.cbs.nl).

## Lijst van gebruikte afkortingen

ABR	Algemeen Bedrijfsregister
ABW	Algemene Bijstandswet
AnW	Algemene Nabestaandenwet
AOW	Algemene Ouderdomswet
ASI	afleiding soort inkomen
AWBZ	Algemene Wet Bijzondere Ziektekosten
BVR	beheer van relaties
CBS	Centraal Bureau voor de Statistiek
GBA	Gemeentelijke Basisadministratie
IBG	Informatie Beheer Groep
IIS	Inkomens Informatie Systeem
IOAW	wet Inkomensvoorziening Oudere en gedeeltelijk Arbeidsongeschikte werkløze Werknemers
IOAZ	wet Inkomensvoorziening Oudere en gedeeltelijk Arbeidsongeschikte gewezen Zelfstandigen
IPO	Inkomenspanelonderzoek
IPS	Inkomens Productie Systeem
RIO	Regionaal Inkomensonderzoek
SET	Sociaaleconomische Trends
SSB	Sociaal Statistisch Bestand
TW	Toeslagenwet
UWV	Uitvoering Werknemersverzekeringen
VROM	Ministerie van Volkshuisvesting Ruimtelijke Ordening en Milieubeheer
WAJONG	Wet Arbeidsongeschiktheidsvoorziening Jonggehandicapten
WALVIS	Wet Administratieve Lastenverlichting en Vereenvoudiging In de Sociale verzekering
WAO	Wet op de Arbeidsongeschiktheidsverzekering
WAZ	Wet op de Arbeidsongeschiktheidsverzekering Zelfstandigen
WIA	wet Werk en Inkomen naar Arbeidsvermogen
WOZ	wet Waardering Onroerende Zaken
ZVW	Zorgverzekeringswet
ZW	Ziektewet