

## **Verrijgingsbestanden: uploadprocedure en voorschriften**

De gebruiker dient rekening te houden met het volgende:

- De uploadprocedure heeft uitsluitend betrekking op de aanlevering van bestanden met persoons-, bedrijfs- en/of adresgegevens die door het CBS gepseudonimiseerd dienen te worden om koppeling aan de beschikbare microdatabestanden van het CBS mogelijk te maken.
- Als u de projectovereenkomst gaat ondertekenen, of ondertekend hebt, wijzen wij u op de volgende passage die hierin is opgenomen:  
“Indien Contractant eigen microbestanden aanlevert, verklaart Contractant dat de gegevens rechtmatig zijn verkregen en dat ook de verstrekking van de gegevens aan CBS voldoet aan de eisen van de Algemene verordening gegevensbescherming (AVG) en de uitvoeringwet AVG. Voor zover het gaat om gezondheidsgegevens, moet de verkrijging en verstrekking (mede) aan de eisen van afdeling 7:5 van Boek 7 van het Burgerlijk Wetboek (Wet geneeskundige behandelingsovereenkomst, WGBO) voldoen. CBS kan hierover nadere informatie opvragen.”
- Het CBS zal de verrijkingsbestanden niet gebruiken voor eigen doeleinden noch deze bestanden ter beschikking stellen voor andere Remote Access projecten, tenzij aanleverende instelling hiervoor expliciet schriftelijk toestemming heeft gegeven. De aangeleverde bestanden worden binnen 12 maanden verwijderd.
- Mail altijd vooraf naar [microdata@cbs.nl](mailto:microdata@cbs.nl) dat u van plan bent een bestand te gaan uploaden. Vermeld in deze mail welke actie van het CBS wordt verwacht. Geef ook aan hoe deze data past binnen de doelbinding van het project, dat wil zeggen waarom dit bestand noodzakelijk is voor het beantwoorden van de onderzoeksvraag van het project.
- Neem het projectnummer op in de naam van het bestand of de zip. Stuur altijd een ingevuld [formulier Upload informatie](#) mee en een recordbeschrijving en zorg dat variabelen voorzien zijn van een omschrijving, anders kan Microdata Services de upload niet in behandeling nemen.
- Bij het uploaden van data is het vanwege het beveiligingsbeleid niet toegestaan om executables (.exe) mee te sturen. Ook is het absoluut niet toegestaan om microdatabestanden per mail op te sturen.
- Het CBS wil niet dat in het databestand ***namen van personen, bedrijven of andere instellingen, straatnamen en e-mailadressen in de variabelen voorkomen***; dit om zichtherkenning (het -onbedoeld- herkennen van personen/bedrijven/adressen in de bestanden) te voorkomen.

- Voor de kosten kunt u de [dienstencatalogus](#) raadplegen. De kosten kunnen afwijken indien bestanden niet goed worden aangeleverd. Kosten gelden per te verrinnen/versleutelen (deel)bestand, daarom is het verstandig om zoveel mogelijk uw gegevens in één gecombineerd bestand aan te leveren.

## Inhoudsopgave

<b>1. Inleiding</b> .....	<b>5</b>
<b>2. Uploadprocedure</b> .....	<b>6</b>
<b>Bestandseisen</b> .....	<b>6</b>
SPSS bestand (.sav).....	6
CSV Tekstbestand met scheidingstekens tussen de velden.....	6
ASCII fixed format .....	6
Eigen identificerende variabelen .....	7
<b>Bestandsoverdracht</b> .....	<b>7</b>
Formulier Upload informatie .....	7
<b>3. Voorschriften</b> .....	<b>8</b>
<b>Personen</b> .....	<b>8</b>
BSN.....	8
Zonder BSN .....	8
<b>Adressen</b> .....	<b>9</b>
<b>Bedrijven</b> .....	<b>9</b>
<b>Overige sleutelvariabelen</b> .....	<b>10</b>
<b>Open tekstvelden</b> .....	<b>10</b>
<b>4. Data koppelbaar maken met externe onderzoeksdata</b> .....	<b>11</b>
<b>5. Resultaten na koppeling</b> .....	<b>12</b>
Resultaten van de verringing van personen:.....	12
Resultaten van het verrinnen van adressen: .....	12
Resultaten van versleuteling van bedrijfsidentificatoren: .....	12

## 1. Inleiding

Het komt regelmatig voor dat onderzoekers een bestand met persoons-, bedrijfs- of adresgegevens willen koppelen aan de beschikbare microdatabestanden van het CBS. Dat is in beginsel mogelijk, mits onderzoekers rechtmatig over die data beschikken en (als het om persoonsgegevens gaat) daar ook toe gerechtigd zijn in het licht van de wetgeving rond de bescherming van persoonsgegevens.

Als u het voornemen hebt een bestand ter koppeling aan te bieden, neem dan altijd eerst via [microdata@cbs.nl](mailto:microdata@cbs.nl) contact op met uw relatiebeheerder en licht toe wat uw bedoeling is en waarom dit bestand nodig is om de onderzoeksvragen in uw project te beantwoorden (doelbinding).

Voor het fysiek aanleveren van een dergelijk bestand (door u zelf, of direct aan CBS door een derde partij die eigenaar is van het bestand) moet u gebruik maken van <https://databestanden.cbs.nl>.

U krijgt inlogcodes als u zich heeft aangemeld door middel van een mail te sturen naar [microdata@cbs.nl](mailto:microdata@cbs.nl) met in het onderwerp: projectnummer upload @doc.

U dient daarbij altijd een korte omschrijving van de inhoud van het bestand mee te leveren, een overzicht van alle variabelen (recordbeschrijving), een codeboek en te zorgen dat alle variabelen voorzien zijn van een omschrijving. Tevens stuurt u bij de upload een volledig ingevuld [formulier Upload informatie](#) mee.

Na het ontvangen van de data zal het CBS de identificerende variabelen versleutelen zodat deze data met de catalogusbestanden van het CBS kunnen worden gekoppeld. Daarbij worden de (direct) identificerende data verwijderd. Eventueel kunnen andere identificerende variabelen op verzoek versleuteld worden en zo beschikbaar blijven in het bestand.

De originele bestanden worden na het koppelbaar maken binnen 12 maanden verwijderd. De verrinde en/of versleutelde bestanden blijven in de beveiligde microdata omgeving beschikbaar gedurende de looptijd van het project en de daarna afgesproken bewaartermijn.

## 2. Uploadprocedure

### Bestandseisen

De gegevens kunnen in drie formaten worden aangeboden:

#### *SPSS bestand (.sav )*

Vermijd het gebruik van DATE field voor datumvelden, maar sla datumvelden op als string variabele of numerieke variabele in het formaat JJJMMDD.

#### *CSV Tekstbestand met scheidingstekens tussen de velden*

- In de 1<sup>ste</sup> regel van dit bestand dienen de namen van de variabelen of kolommen vermeld te staan;
- Let erop dat het scheidingsteken niet als teken in de waarde van een variabele voorkomt. (dus gebruik ; als scheidingsteken en plaats tekst tussen “ ”);
- Vermeld ook altijd het decimaalscheidingsteken;
- Hanteer één formaat voor datumvelden;
- Gebruik geen 1000-tal scheidingstekens en valuta teken in de velden;
- Lever altijd een volledige bestandsbeschrijving aan in het zip-bestand.

#### *ASCII fixed format*

- Zorg voor een goede beschrijving van de data.
- Gebruik dit bestandsformaat alleen als de dataset slechts weinig (<20) variabelen bevat.

Bij de aanlevering in CSV en/of ASCII fixed format of een ander formaat is het van belang om een bestandstoelichting mee te leveren waarin is opgenomen:

- het aantal records;
- een recordbeschrijving;
- een omschrijving per variabele (label);
- indien van toepassing een codeboek (de mogelijke waarden van de variabele met een omschrijving).

Zonder deze bestandstoelichting kunnen we de data niet duiden en wordt de dataset niet geaccepteerd.

***Indien u van deze drie formaten wilt afwijken, dient u dit vooraf te overleggen.***

Geef altijd aan welke de variabelen zijn die persoon, bedrijf of adres identificeren.

### ***Eigen identificerende variabelen***

Indien het bestand eigen identificerende variabelen bevat (recordnummers, persoonsnummers etc.) dan worden die gewoonlijk door het CBS verwijderd. Indien u deze variabelen in het bestand bewaard wilt houden dan kunt u dit aangeven. Dit kan zinvol zijn om op een later moment via deze identificerende variabelen als koppelsleutel aanvullende variabelen als bestand aan te leveren. Deze identificerende variabelen worden wel encrypt, dit om zichtherkenning te voorkomen. Wilt u wel doorgeven wat het formaat is van deze identificerende variabelen. Dit om bij toekomstige leveringen hetzelfde formaat aan te houden van deze identificerende variabelen anders wordt het resultaat na de encryptie niet koppelbaar tussen de verschillende bestanden op deze versleutelde waarde.

### **Bestandsoverdracht**

#### ***Formulier Upload informatie***

De AVG verplicht ons om informatie op te vragen over upload van uw data naar de beveiligde omgeving van de Microdata CBS voor uw project. Hiervoor dient dit [formulier Upload informatie](#).

Dit ingevulde formulier dient u op te slaan als uploadXXXXJJJMMDD.xlsx (XXXX=projectnummer, JJJMMDD is bijvoorbeeld 20180703) en in het zip bestand van de upload te plaatsen, of separaat aan [microdata@cbs.nl](mailto:microdata@cbs.nl) te zenden, onder vermelding van uw projectnummer en de naam van het bestand en @doc, dat via de upload is aangeboden.

Voor het opsturen van bestanden maakt u gebruik van de beveiligde uploadfaciliteit van het CBS van <https://databestanden.cbs.nl>. Gebruik hiervoor de inlogcodes die u eerder heeft aangevraagd en ontvangen.

Comprimeer het bestand (of bestanden), het ingevulde formulier Upload informatie, het codeboek en de bestand- en variabelen beschrijving, neem het nummer van uw project op in de naam van bestand. (voorbeeld: 1051Leerlingen.zip)

### **Doorlooptijd**

Wij streven ernaar om uw verrinde en/of versleutelde bestanden, binnen 2 weken na ontvangst, beschikbaar te stellen in de beveiligde microdata omgeving.

### 3. Voorschriften

#### Personen

De meeste onderzoekers willen een bestand koppelen op persoonsniveau. Hiervoor moeten personen uit het bestand dezelfde koppelsleutel krijgen als die binnen het CBS wordt gebruikt. Dit is de RINPERSOON. Het proces van het omzetten van persoon identificerende kenmerken in het een bestand naar het RINPERSOON noemen we het 'verrinnen' van een bestand.

#### **BSN**

Het verrinnen gebeurt veelal op BSN, of op A- nummer, of op onderwijsnummer. Het is ook mogelijk om met een combinatie van geslacht, geboortedatum, postcode en/of huisnummer en/of overlijdensdatum personen te identificeren. Bij adresgegevens is het essentieel dat een peildatum van het moment dat de desbetreffende persoon daar heeft gewoond wordt opgegeven. Dit is nodig om tot een geslaagde verrinning te kunnen komen.

Bij het verrinnen van een bestand verwijdert het CBS de identificerende variabelen uit het bestand en alleen RINPERSOONS en RINPERSOON worden teruggegeven. Ook unieke nummers die mogelijk naar personen zijn te herleiden, bijvoorbeeld een intern nummer, worden verwijderd. De reden hiervan is om te voorkomen dat gebruikers met deze variabelen een link met het originele bestand kunnen leggen en zo vertrouwelijke persoonsgegevens uit gekoppelde CBS-bestanden kunnen halen.

#### **Zonder BSN**

Voorschrift waaraan de essentiële koppelvariabelen moeten voldoen van bestanden zonder BSN, A-nummers of onderwijsnummers:

Veld	Formaat	Voorbeeld
Geslacht	A1	M of V
Geboortedatum	N8	19590202
Overlijdensdatum	N8	20191126
Postcode6	A6	2498CM
Postcode4	A4	2498
Huisnummer	N5	1
Huisletter	A1	A
Huisnummertoevoeging	A4	IIIH of fl01
Jaar van geldigheid adres of Peildatum	N4/N8	2010/20101231

Zie ook *Eigen identificerende variabelen* op pagina 7.



## Adressen

Het komt ook voor dat u op basis van adresgegevens een bestand koppelbaar wilt maken met CBS data. De koppelsleutels hiervoor zijn SOORTOBJECTNUMMER en RINOBJECTNUMMER.

Dit kan op basis van de VBO\_ID of via postcode, huisnummer, huisletter, huisnummertoevoeging en jaar (of peildatum) van geldigheid van de adresgegevens. Voorbeeld van een VBO\_ID altijd 16 posities: "0518010000452341" zie [BAG viewer \(kadaster.nl\)](#)

Bij het verrinnen van een bestand vervangt het CBS de identificerende variabelen die gebruikt zijn door de variabelen SOORTOBJECTNUMMER en RINOBJECTNUMMER.

Let op dat het jaar van geldigheid of peildatum van het adres altijd is opgenomen in het bestand, anders kan er geen verrinning plaatsvinden naar RINOBJECTNUMMER.

Voorschrift waaraan de essentiële koppelvariabelen op adres moeten voldoen:

Veld	Form	Voorbeeld
Postcode	A6	2498CM
Huisnummer	N5	1; 12345
Huisletter	A1	A of a
Huisnummertoevoeging	A4	IIH of fl01
Combinatie huisnr+letter+toevoeging	A10	1A; 1AFL10;19999AFL10
Jaar geldigheid adres of peildatum	N4/N8	2010/20140131

## Bedrijven

Bij bedrijven kan de koppeling met CBS bestanden gelegd worden via het Kamer van Koophandel nummer (KvKnr) of het fiscaal identificatienummer (FI-nr). Het KvKnr en/of FI-nr wordt daartoe versleuteld zodat deze via het microdatabestand 'Algemeen Bedrijven Register' (ABR) gekoppeld kan worden aan een Bedrijfseenheid (BE), de statistische eenheid voor veel bedrijfseconomische statistieken. Doel van die versleuteling is om de zichtherkenning zoveel mogelijk te beperken.

Voorschrift waaraan de essentiële koppelvariabelen moeten voldoen:

Veld	Formaat	Voorbeeld	Bijzonderheden
Kvknr	A8	00123456	Met voorlooppnullen
FI-nr	A9	000123456	Met voorlooppnullen

### **Overige sleutelvariabelen**

Ook kunnen andere identificerende variabelen uit geüploade bestanden worden gebruikt om koppelingen te maken met CBS bestanden, zoals aan schoolgegevens met BRIN als koppelsleutel. Voor eventuele overige mogelijkheden dient u contact op te nemen met [Microdata@cbs.nl](mailto:Microdata@cbs.nl).

### **Open tekstvelden**

De bestanden mogen in principe geen open tekstvelden (dus antwoorden op open vragen) bevatten. Indien toch aanwezig worden deze velden door ons verwijderd. Mochten zulke velden noodzakelijk zijn en geen identificerende informatie bevatten, dan dient dit expliciet aangegeven te worden en moet een overzicht gegeven worden van de inhoud van de variabele ter controle op identificerende informatie.

#### 4. Data koppelbaar maken met externe onderzoeksdata

Microdata Services faciliteert ook de mogelijkheid om door onderzoekers geselecteerde externe data koppelbaar te maken met CBS databestanden over personen. Het CBS heeft daartoe koppelbestanden tussen de populatie van de diverse onderzoeken en de persoonsgegevens van het CBS beschikbaar in de beveiligde microdata omgeving. Via deze koppeling kunnen data uit die diverse onderzoeken gebruikt worden in een microdata onderzoek bij het CBS. Het betreft de volgende externe onderzoeksdata:

- DHS: Procedure DNB Household survey met CBS microdata;
- Lifelines: procedure LIFELINES Cohort met CBS microdata;
- LISSPANEL: Procedure Lisspanel met CBS microdata;

Zie hiervoor [externe onderzoeksdata](#).

## 5. Resultaten na koppeling

### *Resultaten van de verrining van personen:*

De kwaliteit van de koppeling hangt af van de kwaliteit van de aangeleverde sleutelvariabelen. Er vanuit gaande dat dit in orde is, kunnen we onderstaande percentages verwachten van het verrinnen van persoonsbestanden:

- Op BSN (of bijv. onderwijsnummer) is in principe 100% te verrinnen. In de praktijk kan dit echter een paar procent minder zijn, bijvoorbeeld door verschillen in peilmoment en/of populatie (niet in de GBA ingeschrevenen)
- Op geboortedatum, geslacht, adres en peildatum van het adres is rond de 90% te verrinnen.
- Op geboortedatum, geslacht, en postcode4 en jaar van geldigheid van de postcode is rond de 80% te verrinnen.

Van elke koppeling wordt een rapport opgesteld dat ter beschikking wordt gesteld.

### *Resultaten van het verrinnen van adressen:*

Het resultaat van het omzetten van VBO\_ID is 100%.

Op basis van de adresgegevens is dit rond de 90%.

### *Resultaten van versleuteling van bedrijfsidentificatoren:*

KvK-nummers en FI-nummers versleutelen lukt altijd, maar of er een corresponderende bedrijfseenheid (BE) in het ABR bestaat kan van verschillende factoren afhangen.

In tegenstelling tot persoonsstatistieken, waar voor het verrinnen altijd sprake is van een 1-op-1 relatie, ligt de situatie met bedrijven een stuk ingewikkelder. Een 'bedrijfseenheid' (BE) in het 'Algemeen Bedrijven Register' (ABR) kan meerdere Kamer van Koophandel (KvK)-nummers en/of Fiscale (FI)-nummers omvatten.

Ook kan het voorkomen dat een KvK- of FI-nummer niet in een BE komt, bijvoorbeeld omdat deze feitelijk niet economisch actief is.

Voor verdere details wordt verwezen naar de documentatie van het [ABR](#) en met name naar de bijlagen van deze documentatie.