

Instructions for file format and upload process

The user needs to take the following into account:

-
- If you are about to sign or have signed the project agreement please be aware of the following clause that is included in the project agreement:
“If Contracting party provides its own microdata-sets, Contracting party declares that the data have been obtained legally and that the provision of data to CBS meets the requirements of the General Data Protection Regulation (GDPR) and the GDPR implementing act. As far as health data is concerned, the obtaining and providing of data must also comply with the requirements of Section 7:5 of Book 7 of the Dutch Civil Code (medical Treatment Contracts Act, WGBO). CBS may request further information about this.”
 - Always send an e-mail to Microdata@cbs.nl when you plan to upload a dataset of your own. Explain in this mail what action you expect from CBS. Do also indicate how this dataset fits within the research questions of your project, based on the ‘need to know’ principle.
 - Include the project number of your remote access project in the name of the file or zip. Always include in the zip file a completely filled out ‘[Upload information document](#)’, and a description of the dataset, and make sure all variables are labelled. Without this information Microdata Services cannot process your dataset. Also send a mail once you have done the upload, with the project number and filename in the subject of the email.
 - When uploading data, the security policy does not permit sending of executables (.exe) along with the file.
 - Security risks absolutely rule out the sending of files via email.
 - CBS does not want to receive *names of persons, companies or other institutions, street names and email-addresses in the variables*, in order to prevent unintended recognition of such individual statistical units while working with the data.
 - For costs please consult the Catalogue of Services. Costs can deviate from that when the updated datasets appear not to have been prepared properly. Costs apply per data file that has to be linked/encrypted so it is advised to supply your data in one single combined file as much as possible.
-

contents

<i>The user needs to take the following into account:</i>	<i>2</i>
<i>1. Introduction</i>	<i>4</i>
<i>2. Upload procedure</i>	<i>5</i>
File requirements	5
File transfer	7
<i>3. Requirements</i>	<i>11</i>
Private individuals	11
Addresses	12
Companies	12
Other key variables	12
<i>4. Own data with external research data</i>	<i>13</i>
<i>5. Results after linking</i>	<i>14</i>

1. Introduction

Often researchers want to link their own microdata on individual persons, businesses etc. to Microdata from the Statistics Netherlands Microdata catalogue. Microdata Services offers this possibility, provided that investigators are legally entitled to use these data and (as far as data on persons are concerned) the legislation on privacy protection is respected.

When you are planning to present a dataset to CBS for further linkage, please always first contact Microdata@cbs.nl to explain what the data is about and why they are necessary for answering the research questions of your project (the ‘need to know’ principle).

For actually delivering the data to CBS (by yourself, or by a third party owning the data) you are requested to use the secure upload procedure of [CBS](#) described in this document.

Please always add a brief description of what the data are about, an overview of all variables (record description) and a code list. Also make sure all variables have been labelled. Furthermore, you need to include the filled out ‘[Upload information document](#)’ found at the same link. After executing the upload please send a mail to inform us, including your project number and the filename of the upload.

Upon receiving these data, Statistics Netherlands will encrypt the identifying variables so that these data can be linked with the catalogue files of Statistics Netherlands. This involves removal of the directly identifying data. Upon request, other identifying variables can be encrypted and retained in the dataset.

2. Upload procedure

File requirements

The data can be uploaded in three different formats:

1. SPSS system file (.sav)

Please make sure all variables, and where applicable the values, are tagged with labels. Avoid using DATE field for date fields, but instead save the date fields as string variables or numerical variables in the format YYYYMMDD.

2. Text file with separators between the fields (.CSV)

In line 1, please enter the names of the fields. Please ensure that separators are not used as characters in a field. Furthermore, where applicable mention whether the separator in decimals is a comma or a period, always use a single format for all date fields, do not use a 1000 separator nor a currency sign in any of the fields. Always include a complete file description or data model in the zip-file.

3. ASCII fixed format

A clear description of the data is necessary. The file cannot be too wide (maximum record length 8,000). This file format is to be used only if the dataset contains few (<20) variables.

When providing datasets in .CSV or ASCII fixed format, it is important to provide an explanation which includes a record description, a description per variable (label) and, where applicable, a code book (the possible values of the variable with a description). Without this explanation, we are unable to interpret the data and the dataset will not be accepted.

In case you prefer to deviate from the above three formats, you must contact us before you submit your request.

Always indicate which variables are the once identifying a person, company or address (key variables).

Your own identifying variables

If the file contain your own identifying variables (record numbers, case numbers, client numbers etc.), these will normally also be removed by CBS. In case you wish to keep these variables in the file, you can let us know. It can be useful to keep these identifying variables to use as a key if you wish to add additional variables in a later stage.

These identifying variables will be encrypted to prevent unintended recognition of individual statistical units while working with the data.

Example of a record description (to be provided in txt format):

Variable	Starting position	End position	Length	Type *)	Decimals
Postal code	1	6	6	A	0
House number	7	11	5	N	0
House letter	12	12	1	A	0
House number suffix	13	16	4	A	0
Type of address	17	18	2	A	0
etcetera					

*) A=alphanumeric N=numeric

Descriptions of variables:

Postc: Postal code of the address.

Huisnr: House number

Huisltr: House number letter

Huisnrtoe: House number suffix

Soortadres: Specifies an address by type of dwelling:
Example code book

Example of a code book:

Variable: Soortadres:

Code	Description
00	Corner house
01	Terraced house
02	Semi-detached house
03	Detached house

File transfer

To send us files, you are requested to use Statistics Netherlands' secured upload facility.

The procedure is as follows:

Please zip the data file, the Upload information document, code book and variable description, and give the zip file a name that includes the project number of your Microdata project (for example 'Students Amsterdam 1051.zip').

Multiple files can be included in a ZIP, but they are treated and charged separately.

If you want, the zip file can be protected with a password that you send us by e-mail to Microdata@cbs.nl.

A file sent via the upload may **not be larger than 1 GB**, and may also not include any **executables (.exe files)**. If your zip-file is larger than 1GB, please contact us at Microdata@cbs.nl.

Upload information document

The DGPR requires us to ask for extra information about your data upload to the protected environment of CBS Microdata for your project. The Upload information document serves this purpose.

You are requested to save the completed document with the following name: uploadXXXXJJJMMDD.xlsx (XXXX=project number, JJJMMDD is for example 20180703) and include the document in the zip file of the upload. You may also send the document separately by email to Microdata@cbs.nl citing your project number and the name of the file that has been uploaded.

After these preparations you can perform the file upload to Statistics Netherlands by clicking on the following link: <http://www.cbs.nl/bestandslevering>:

On the next page the following screen will be displayed.



Inzenden van een bestand met opgaven

Bij vragen kunt u contact opnemen met het Contact Center CBS:

Telefoon: (045) 570 64 00 voor inhoudelijke vragen, op werkdagen tussen 9-17 uur.
(045) 570 66 27 voor technische vragen, op werkdagen tussen 10-12 en 14-16 uur.

E-mail: contactcenter@cbs.nl

Vul eerst deze gegevens in:

De enquêtecode, uw correspondentienummer en uw controlemnummer staan vermeld in de brief of e-mail die u is toegezonden door het CBS tesamen met de link naar deze pagina.

De enquêtecode:	<input type="text"/>
Uw correspondentienummer:	<input type="text"/> - <input type="text"/> - <input type="text"/>
Uw controlemnummer:	<input type="text"/>

Contactpersoon:

Indien de contactinformatie in de brief of e-mail niet meer correct is, kunt u hier de juiste gegevens invullen

Naam:	<input type="radio"/> Dhr. <input type="radio"/> Mevr.	<input type="text"/>
Telefoonnummer:		<input type="text"/>
E-mailadres:		<input type="text"/>

Afdrukken

Verder

When completing the internet page, please use the following codes:

Survey code (enquêtecode): CVB

Reference number (correspondentienummer): 2008-111-401

Control number (controlemnummer): 817808

Please enter your information in the fields under “Contactpersoon”.

Then click on “verder” (continue).



Centraal Bureau voor de Statistiek

Inzenden van een bestand met opgaven

Bij vragen kunt u contact opnemen met het Contact Center CBS:

Telefoon: (045) 570 64 00 voor inhoudelijke vragen, op werkdagen tussen 9-17 uur.
(045) 570 66 27 voor technische vragen, op werkdagen tussen 10-12 en 14-16 uur.

E-mail: contactcenter@cbs.nl

Enquêtecode:	CVB
Correspondentienummer:	2008-111-401

Verzenden van het bestand:
Druk na het invullen van de gevraagde gegevens op de knop [**Verzenden**] onderaan de pagina. Het verzenden van het bestand start direct. Bij grote bestanden kan dit enige tijd in beslag nemen, afhankelijk van de verbinding. Zodra het bestand is ontvangen krijgt u een ontvangstbevestiging van het CBS.

Te verzenden bestand

Bladeren...

LET OP:
Indien u tussentijds de pagina afsluit of opnieuw op de verzendknop drukt wordt slechts een deel van uw bestand bij het CBS afgeleverd. Wacht dus op de ontvangstbevestiging, hierin staat onder meer de grootte van het ontvangen bestand aangegeven. Wij adviseren geen bestanden groter dan 600 MB te verzenden.

Afdrukken Verzenden

Select the files to be uploaded with “Bladeren” (browse) and click on the “Verzenden” (Send) button at the bottom of the page. Please wait (this may take a little while) until you receive the following notification on your screen. This means that the file upload has been successful.



 Centraal Bureau voor de Statistiek

Inzenden van een bestand met opgaven

Bij vragen kunt u contact opnemen met het Contact Center CBS:

Telefoon: (045) 570 64 00 voor inhoudelijke vragen, op werkdagen tussen 9-17 uur.
(045) 570 66 27 voor technische vragen, op werkdagen tussen 10-12 en 14-16 uur.

E-mail: contactcenter@cbs.nl

Enquêtecode:	CVB
Correspondentienummer:	2008-111-401

Bedankt voor het inzenden van uw opgave

Bestandsnaam: Upload informatie.xlsx
Bestandsgrootte: 15.937 bytes
Datum / tijd ontvangst: 03-07-2018 09:03:15

Afdrukken

U kunt deze ontvangstbevestiging afdrukken voor uw eigen administratie.

Please always send an email to Microdata@cbs.nl after having completed the upload, mentioning the name of the zip file. Your data file will normally be received by us within one working day.

Lead time

We aim to complete the linkage/encryption process of files within approximately 2 weeks. Often it can be a bit faster and sometimes it takes a little longer.

3. Requirements

Private individuals

Most researchers want to link their dataset at the level of private individuals (persons). Therefore the persons in the dataset need to be assigned the same linkage keys as used within CBS, the RINPERSOON number. The process of converting key identifying characteristics of each person into the RINPERSOON number is called ‘rinning’ (verrinnen) of a file.

Rinning is often done based on social security number (BSN), A- number, or education number. Another possibility is to identify a person within the population register GBA using a combination of gender, date of birth, postal code and/or address, or the date of death. When using address details, a reference date indicating the time when the person was registered there as a resident is essential.

Rinning of a dataset means Statistics Netherlands removes the identifying variables from the file and only the RINPERSOON (RIN number) is displayed. Any unique numbers which might possibly trace back to individuals, such as an internal number, are removed. The reason for this is to prevent users of such data from directly retrieving confidential personal details from files linked to Statistics Netherlands data. Upon request, identifying variables may be returned to the file in encrypted form. This can be useful in case additional variables have to be linked to the data later; this encrypted linkage key can then be used.

Requirements pertaining to the essential connecting variables:

Requirements pertaining to the essential connecting variables:

Field	Format	Example	Remarks
Gender	A1	M or F	or 1 (M) or 2 (F)
Date of birth	A8 or F8.0	19590202	YYYYMMDD
Date of death	A8	20591126	YYYYMMDD
Postal code6	A6	2498CM	
Postal code4	A4	2498	
House number	N5	1051	
House letter	A1	A	
House number suffix	A4	IIIH or fI01	
Year of validity or reference date	N4 or N8	2010/20101231	

See also *Your own identifying variables* at page 5.

Addresses

Based on address information, an encryption may take place into a RINOBJECTNUMBER. In this process, postcode, house number etc. are removed. A linkage can also be made based on VBO_ID or BAG_ID.

After rinning the file, CBS removes all identifying variables that were used in the rinning process.

Please make sure the year of validity or reference date of the address are included in the file, otherwise the rinning to RINOBJECTNUMBER is not possible.

Requirements for the essential connecting variables:

Field	Format	Example	Remarks
Postal code	A6	2498CM	
House number	A5	00001	Numerical also OK
House letter	A1	A or a	
House number suffix	A4	IIH or fl01	No leading zeroes
Year of address validity or reference date	N4/N8	2010/20101231	

See also *Your own identifying variables* at page 5.

Companies

For companies, linking with the Statistics Netherlands datasets can be done directly via the Dutch Chamber of Commerce number (KvKnr) or the tax identification number (FI-nr). To this end, the KvKnr and/or FI-nr is encrypted so that it can be linked to an enterprise (Bedrijfseenheid (BE)), the statistical unit for many business economic statistics, via the micro dataset 'Algemeen Bedrijven Register' (ABR, company register). The purpose of this encryption is also to prevent direct retrieval of confidential business economic data from linked Statistics Netherlands datasets.

Requirements for the essential connecting variables:

Field	Format	Example	Remarks
KvKnr	A8	00123456	Use leading zeroes
FI-nr	A9	000123456	Use leading zeroes

Other key variables

Also other identifying variables from own files may be used from the researchers' own datasets to be linked with Statistics Netherlands datasets. For instance school data or education numbers. For any other possibilities you can contact Microdata@cbs.nl.

4. Own data with external research data

Microdata Services also facilitates the possibility to make other external research files linkable with CBS data files on private individuals. CBS has matching files between the population of several research files and the data that is available in the CBS Remote Access environment on private individuals. Through these matching files data from several researches can be used in a microdata project at CBS. The following research data can be linked this way:

- DBN Household survey (DHS)
- LIFELINES Cohort
- LISSPANEL
- NIVEL

You can find the procedures for using these external data on [our website](#).

5. Results after linking

Results of rinning private individuals

The quality of linking depends on the quality of the data which were provided, especially of the key variables. Assuming these are in order, we may expect the following results from the rinning process of datasets on private individuals:

- Basically, 100% of data can be rinned by BSN (or education number, etc.). In reality, however, this may be a few percent less, due to differences in reference date and/or population (not registered in the GBA);
- Rinning is possible for more than 90% when based on date of birth, gender, postal code6 and year of validity postal code;
- Rinning is possible for more than 80% when based on date of birth, gender, postal code4 and year of validity of the postal code.

A report is made of each linking procedure and will be made available.

Results of rinning addresses

Rinning of VBO_ID or BAG_ID to RINOBJECTNUMMER is always 100%. Based on address-information rinning is possible for about 90% of the cases.

Results of rinning companies

Encryption of KvK numbers and FI numbers is always successful, but whether a matching business unit BE is found in the ABR depends on multiple factors. For this you may refer to the documentation about the [ABR](#).

In statistics on individuals, the rinning process is always a one-on-one link. This does not hold for statistics on companies. A 'bedrijfseenheid' (BE, enterprise) in the 'Algemeen Bedrijven Register' (ABR) may have multiple Chamber of Commerce (KvK) numbers and/or Fiscal (FI) numbers. It may happen that a KvK- or FI number is not featured in a BE registration, perhaps because a company has no actual economic activity. For further details we refer to the documentation, in particular the appendices to the documentation of the [ABR](#).