



Microdataservices

---

## **Instructions for file format and upload process**

*Date:*

*24 December 2015*



## Preliminary remarks

- 
- Attention: It is possible that your organisation is supplying databases to Statistics Netherlands for other purposes, perhaps on a regular basis. The process agreed upon for these databases will always remain applicable.
  - These instructions are intended for researchers working or starting to work within the secured environment of Statistics Netherlands who wish to import their own data for the purpose of linking them to the microdatasets which are available at the level of individual persons, addresses or businesses within the secure environment of Statistics Netherlands.
  - Do bear in mind that you are about to sign or have signed the project agreement containing the following phrases:

“If Contracting party provides its own microdata-sets, Contracting party declares that the data have been obtained legally and that the provision of information to the CBS meets the requirements of the Dutch Data Protection Act, and, in the case of health information (also) 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.”
  - The service here described and the charges involved are based on processing of datasets per occurrence. Having data linked to datasets already processed previously is not included in this service. We recommend putting as many data as possible in one dataset, so everything can be processed all at once.
  - 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.
-



## Contents

<i>Preliminary remarks</i> .....	4
<i>Introduction</i> .....	6
<b>2. Requirements on key variables</b> .....	7
2.1 Private individuals .....	7
2.2 Addresses .....	9
2.3 Companies.....	10
2.4 Other key variables .....	10
2.5 Results after linking .....	11
<b>3. Upload procedure</b> .....	12
3.1 File requirements .....	12
3.2 File transfer .....	14
3.3 Lead time .....	15



## **Introduction**

It happens regularly that researchers would like to link their own file containing data of private individuals or businesses to files from the Statistics Netherlands catalogue with available micro datasets. Microdataservices therefore offers the option to upload such data via the Internet.

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 either removal or encryption (on request) of the directly identifying data. Linking can be achieved with variables about persons, addresses or businesses.



## 2. Requirements on key variables

### 2.1 Private individuals

Most researchers like to link their dataset at the level of private individuals. This can be achieved by searching for these individuals in the Municipal Records Database (GBA) and subsequently assigning a 'RINPERSON-nummer' ('RIN number') to each person.

The process of converting key identifying characteristics of each person into the RIN number is called 'rinning' (verrinnen) of a file.

This is usually done on social security number (BSN), A- number, or education number. Another possibility is to identify a person within the 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 in question 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 RINPERSON (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. Such a request may be sent to the account manager involved.

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	Letters as capitals
Postal code4	A4	2498	
House number	A5 or F5.0	1051	Right-aligned with leading spaces
House letter	A1	A	Capital letter



## Microdataservices

---

House number suffix	A4	IIH or f101	
Year of validity	A4 or F4.0	2010	Numerical also OK
Reference date	A8 or F8.0	20050320	YYYYMMDD



## 2.2 Addresses

Based on the complete address: postal code, house number, house letter, house number suffix and year (reference date) of address validity, an encryption may take place into a RINADRES.

This encryption is only possible for addresses at which individuals were ever registered in the GBA (as of 1995).

Requirements for the essential connecting variables:

Field	Format	Example	Remarks
Postal code	A6	2498CM	No spaces
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	A4	2010	Numerical also OK

### District code

The data can be given a buurtcode or wijkcode (district code) of that particular year using the postal code (numbers and/or letters) and the year this postal code was valid.

Requirements for the essential connecting variables:

Field	Format	Example	Remarks
Postal code	A6	2498CM	No spaces
Year	A4	2010	Numerical also OK



### 2.3 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.

In statistics on individuals, the linking 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](#).

Requirements for the essential connecting variables:

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

### 2.4 Other key variables

Sometimes other identifying variables may be used from the company’s own datasets to be linked with Statistics Netherlands datasets. For instance school data, or education numbers. Where applicable, please contact Microdataservices at [Microdata@cbs.nl](mailto:Microdata@cbs.nl) to determine the format of the connecting variables.



## 2.5 Results after linking

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.

Encryption of KvK numbers and FI numbers is always successful, but it is uncertain whether a matching bedrijfseenheid (enterprise, (BE)) can be found. For this you may refer to the [documentation about the ABR](#).



### 3. Upload procedure

#### 3.1 File requirements

The data can be uploaded in three different formats:

1. SPSS systemfile (.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.

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.

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

You are requested to provide a brief description of your dataset, the identifying variables, which variables are to be encrypted, and where applicable, whether the separator in decimals is a comma or a period. Please submit the dataset description in either a txt- or a Word format.

When providing datasets in 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.

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	A	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					

Example of a code book:

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:

Soortadres: Specifies an address by type of dwelling:

Example code book

Variable: Soortadres:

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

Statistics Netherlands does not want to receive any names of individuals, companies or institutions during the data uploads, to avoid visual recognition (the -unintended-recognition of names during work on these files). Names are irrelevant in the process of rinning/encryption/linking.



### 3.2 File transfer

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

The procedure is as follows:

Please zip the file and code book, give the zip file the name that was provided to you by the account manager in an email message.

All uploaded files will be placed in a centrally located folder. The name will tell us to which project the file belongs. If preferred, you may secure the ZIP with a password, which you need to send via email to [Microdata@cbs.nl](mailto:Microdata@cbs.nl).

A file sent via the upload may **not be bigger than 2048 MB**, and may also not include any **executables (.exe files)**. It is of course possible to upload more than one file.

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

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

Survey code (enquêtecode): CVB

Reference number (correspondentienummer): 2008-111-401

Control number (controlnummer): 817808

Then click on "verder" (further).

Select the files to be uploaded with "Bladeren" (browse).

After filling in the required information please click on the "Verzenden" (Send) button at the bottom of the page. The uploading of the files starts immediately. For large files this may take some time, depending on the connection. Once the file is received, you will receive an acknowledgment from the CBS.

#### NOTE:

**If you prematurely exit the page or press the send button again, only a part of your file will be delivered at CBS. So wait for the acknowledgement on the computer screen "Uploaden is voltooid" before you do anything else. After this you will also get an email acknowledgement, this includes also the size of the received file.**

**Please always send an email to [Microdata@cbs.nl](mailto:Microdata@cbs.nl) after placing something in the upload with the name of the zip file. The file will normally be received by us within one working day.**



### **3.3 Lead time**

We aim to complete the rinning process of files within approximately 2 weeks, which mainly depends on the quality of the metadata you have provided, and of the data themselves.