



**Centraal Bureau
voor de Statistiek**

Microdata Services

Supplying files for use in the CBS microdata environment: upload procedure and regulations

The user needs to take the following into account:

- This upload procedure relates solely to the supply of files containing personal, company and/or address data which need to be pseudonymised by Statistics Netherlands (CBS) for use in the CBS microdata environment, where they might be linked to the available CBS microdata files.
- Once 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 has been obtained legally and that the provision of data to Statistics Netherlands meets the requirements of the General Data Protection Regulation (GDPR) and the Dutch Implementation Act of the GDPR. As far as data on health is concerned, the acquisition and provision of data must also meet the requirements stipulated in Section 7:5 of Book 7 of the Dutch Civil Code (Medical Treatment Contracts Act, WGBO). CBS may request further information about this.”
- CBS will not use the enrichment files for its own purposes nor make these files available for other microdata projects, unless explicitly authorised by the supplying institution in the form of written consent. The supplied files will be deleted within twelve months.
- Always send a notice in advance via the CBS-microdata portal when you plan to upload a dataset. Explain what action you expect from CBS. Do also indicate how this dataset fits within the research questions of your project, i.e. why this file is necessary in order to answer the research questions (the ‘need to know’ principle).
- Include the project number of your microdata project in the name of the file or zip. Always include in the zip file a duly completed ‘Upload information form’ and a description of the dataset records, and make sure all variables are labelled. Without this information, Microdata Services cannot process your upload.
- When uploading data, it is not allowed to send executables (.exe) along with the data pursuant to CBS’ security policy.
- Security risks absolutely rule out the sending of files via email.
- CBS does not want to receive datasets containing names of persons, companies or other institutions, street names and email addresses in the variables, in order to prevent unintentional visual recognition of such individual statistical units while working with the data.
- For costs, please consult the Services catalogue. Varying costs may apply if datasets were not delivered correctly. Costs will apply per data file that requires

linking/encrypting; it is therefore advisable to provide as much of your data as possible in one combined file.

Inhoudsopgave

1.	<i>Introduction</i>	5
2.	<i>Uploadprocedure</i>	6
	2.1. File requirements	6
	Text file with separators between the fields (.CSV).....	6
	ASCII fixed format	6
	2.2. Your own identifying variables	7
	2.3. File transfer	7
	Lead time	7
3.	<i>Requirements</i>	8
	3.1. Private individuals	8
	3.2. Addresses	9
	3.3. Enterprises	9
	3.4. Other key variables	10
	3.5. Textbox	10
4.	<i>Results after linking</i>	11
	4.1. Results of rinning private individuals	11
	4.2. Results of rinning addresses	11
	4.3. Results of rinning companies	11
	4.4 Results of rinning postal codes	11

1. Introduction

It is a regular occurrence that researchers wish to link their own microdata on individual persons, companies etc. to microdata available within the CBS microdata environment. In principle, Microdata Services can accommodate such requests, provided that the researchers can lawfully use this data and (in so far as data on persons is concerned) it complies with the legislation on privacy protection.

If you plan to provide a dataset to CBS for use within the CBS microdata environment, always contact Microdata Services first via the CBS microdata portal, explaining your intention and why this file is needed to answer the research questions in your project (the 'need to know' principle).

Once you have contacted CBS Microdata Services you will receive your login codes via email.

For the physical delivery of such data to CBS (by yourself or by a third-party owner of the file) you must use <https://databestanden.cbs.nl>.

Please always submit a brief description of the file content, an overview of all variables (record description) and a codebook. Please also ensure that all variables have been labelled. Furthermore, you need to include the duly completed 'Upload information form'.

Upon receiving these data, CBS will encrypt the identifying variables; this makes it also possible to link the data to other datafiles within the CBS microdata environment. This involves removal of the directly identifying data. Upon request, other identifying variables may also be encrypted and thus retained in the dataset.

The original files will be deleted within twelve months. The encrypted files remain available in the secured CBS microdata environment for the duration of the project and the agreed retention period.

2. Uploadprocedure

2.1. File requirements

The data can be uploaded in three different formats:

SPSS system file (.sav)

Avoid using DATE field for date fields, but instead save the date fields as string variables or numerical variables in the format YYYYMMDD.

Text file with separators between the fields (.CSV)

- In line 1, please enter the names of the fields or columns;
- Please ensure that separators are not used as characters in a value of a variable. Instead, please use the semi-colon ; as a separator and place text between quotation marks;
- 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.

ASCII fixed format

- A clear description of the data is necessary.
- 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:

- the number of records;
- a record description;
- a description per variable (label);
- 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 ones identifying a person, company or address (key variables)

2.2. Your own identifying variables

If the file contains 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 may be useful to keep these identifying variables to use as a key if you wish to add additional variables at a later stage.

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

Where applicable, please indicate the format of these identifying variables. This is necessary to ensure the same format of these identifying variables is kept in future deliveries, otherwise the results after encryption of this particular value will not be linkable between the different files.

2.3. File transfer

Upload information form

The DGPR requires us to ask for extra information about your data upload to the CBS microdata environment for your project. The [Upload information form](#) serves this purpose.

You are requested to save the completed form with the following name: uploadXXXXJJJJMMDD.xlsx (XXXX=project number, JJJJMMDD is for example 20180703) and include the document in the zip file of the upload.

To send us files, you can use CBS' secure upload facility at <https://databestanden.cbs.nl>. Use the login codes that you previously requested and received.

Please zip the data file(s), the Upload information form, code book and variable description, and give the zip file a name that includes the project number of your microdata project (for example '1051Students_Amsterdam.zip').

Lead time

We aim, after receiving the upload, to complete the linkage/encryption process of files within approximately two weeks.

3. Requirements

3.1. Private individuals

Most researchers wish to link their dataset at the level of private individuals (persons). To do so, the persons in the dataset need to be assigned the same linkage keys in each file, the RINPERSOON¹ number. The process of converting key identifying characteristics of each person into the RINPERSOONS and RINPERSOON number is called ‘rinning’ of a file.

BSN

Rinning is often done based on social security number (BSN), A- number, or education number. Another possibility is to identify a person using a combination of gender, date of birth, postal code and/or address, and/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 CBS replaces the identifying variables from the file by the variables RINPERSOONS and RINPERSOON. 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 other datafiles that the file is linked to. 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.

Without BSN

Requirements pertaining to the essential connecting variables without BSN, A-numbers or education number:

Field	Format	Example
Gender	A1	M of V
Date of birth	A8	19590202
Date of death	A8	20191126
Postal code6	A6	2498CM
Postal code4	A4	2498
House number	N5	1
House letter	A1	A
House number suffix	A4	IIH of fl01
Year of validity of reference date	N4/N8	2010/20101231

¹ In some microdataprojects a projectspecific RIN is used, a so called PRIN. For the sake of readability, the terms RIN, RINPERSOON, rinning etc will be used throughout this document.

See also Your own identifying variables at page 7

3.2. Addresses

Based on address information, an encryption may take place into a SOORTRINOBJECTNUMBER and RINOBJECTNUMBER.

This can be done on the basis of VBO_ID or via postal code, house number, house letter, house number suffix and year or reference date of address validity. Example of an VBO_ID always 16 positions: “0518010000452341”see [BAG viewer \(kadaster.nl\)](http://BAGviewer.kadaster.nl)

After rinning the file, CBS replace all identifying variables that were used in the rinning process by SOORTRINOBJECTNUMBER and RINOBJECTNUMBER.

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

Requirements for the essential connecting variables:

Field	Format	Example
Postal code	A6	2498CM
House number	N5	1; 12345
House letter	A1	A of a
House number suffix	A4	IIH of fl01
Combination hsnr+letter+suffix	A10	1A; 1AFL10;19999AFL10
Year of address validity or reference date	N4/N8	2010 of 20140131

3.3. Enterprises

For enterprises, linking datasets may be done directly via the Dutch Chamber of Commerce number (KvKnr) or the Legal Entities and Partnership Identification Number (RSIN); formerly the tax identification number (FI-nr). To this end, the KvKnr and/or RSIN is encrypted. For projects where the data is linked to datafiles from CBS, the encrypted KvKnr and/or FI-nr, can then be linked to an enterprise Bedrijfseenheid (BE)), the statistical unit for many business economic statistics, via the micro dataset ‘Algemeen Bedrijven Register’ (ABR, General Business Register). The purpose of this encryption is also to prevent direct retrieval of confidential business economic data from linked CBS datasets.

Requirements for the essential connecting variables:

Field	Format	Example	Remarks
Kvknr	A8	00123456	Use leading zeros

3.4. Other key variables

Other identifying variables may also be used from the researchers' own datasets to be linked with other datasets. For instance, school data (BRIN) or education numbers. Please contact Microdata Services via the CBS microdata portal for information on the various options.

3.5. Textbox

In principle, files must not contain any fields answers to open questions. If present, these fields will be removed by us. If such fields are necessary, and as long as they do not contain any identifying information, this should be stated explicitly along with an overview with the content of the variable, so we can monitor the variable for identifying information.

4. Results after linking

4.1. 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.

4.2. Results of rinning addresses

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

4.3. Results of rinning companies

Encryption of KvK numbers and RSIN (FI) numbers is always successful. Whether, in projects that link to CBS datafiles, 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 particular the appendices. 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- of FI number is not featured in a BE registration, perhaps because a company has no actual economic activity.

4.4 Results of rinning postal codes

Rinning of postal code is always successful, but only valid postal codes are linkable across datafiles.