Centraal Bureau
voor de Statistiek

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

*Date: March 2026*

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 pseudonymized 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.

- To add external datasets to a CBS microdata project, Then follow this step-by-step guide: Submit a dataset to CBS.

- Always start the upload by first selecting **Add file** in your project in the CBS-microdata portal and selecting Upload/Import under File type. Fill in the Explanation field. 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, CBS cannot process your upload.

- Log in to https://databestanden.cbs.nl/ and perform the 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.

Centraal Bureau voor de Statistiek

## Tabel of contents

Centraal Bureau
voor de Statistiek

# 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, CBS 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 us 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 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.

***Number of variables and file size***

- Limit the number of variables in the file(s) you submit to only those necessary for your research (maximum of 200 variables per file).
- The size of the zipped file may not exceed 10 Gb).

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 RINPERSOON[1] 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 |

See also Your own identifying variables at page 7

[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.

### 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)

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 us 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.

For any other options, please contact us via the CBS-microdata portal.

# 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 on the basis of BSN. In reality, however, this may be a few percent less, due to differences in reference date and/or population (not registered in the Personal Records Database (BRP)[2]);
- Rinning is possible for approximately 90% when based on date of birth, gender, postal code6 and year of validity postal code;
- Rinning is possible for approximately 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 valid VBO_ID into RINOBJECTNUMBER is always 100%. Based on address-information rinning into RINOBJECTNUMBER is possible for about 90% of the cases.

## 4.3. Results of encrypting 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 encrypting postal codes

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

[2] The BRP is a register of people who live or have lived in the Netherlands, or who are temporarily staying in the Netherlands and is owned by the government.

# 5. Method for converting personal data to RINPERSOON

## 5.1. Linking strategies RINPERSOON

The process of converting personally indentifiable characteristics in a file to RINPERSOON and RINPERSOONS is referred to as pseudonymization or 'verrinnen' of a file. The file to which a RINPERSOON is linked is referred to as the Central Linkage File for Persons (CKP).

The most reliable method is identification based on the A number. This number is used to register everyone in the Personal Records Database (BRP)[3] and is also the basis for assigning a RINPERSOON. Each A number corresponds to a unique RINPERSOON. The next most reliable method is based on the BSN. The third option for identifying individuals is based on gender, date of birth, and postal code and house number (addition). Various strategies are possible for postal code links, which are partly determined by the level of detail in the variables. When drawing up a linking strategy, we always work from more to less detail. The more detail, the greater the discriminatory power of the variables and thus the greater the chance of selecting the right person.

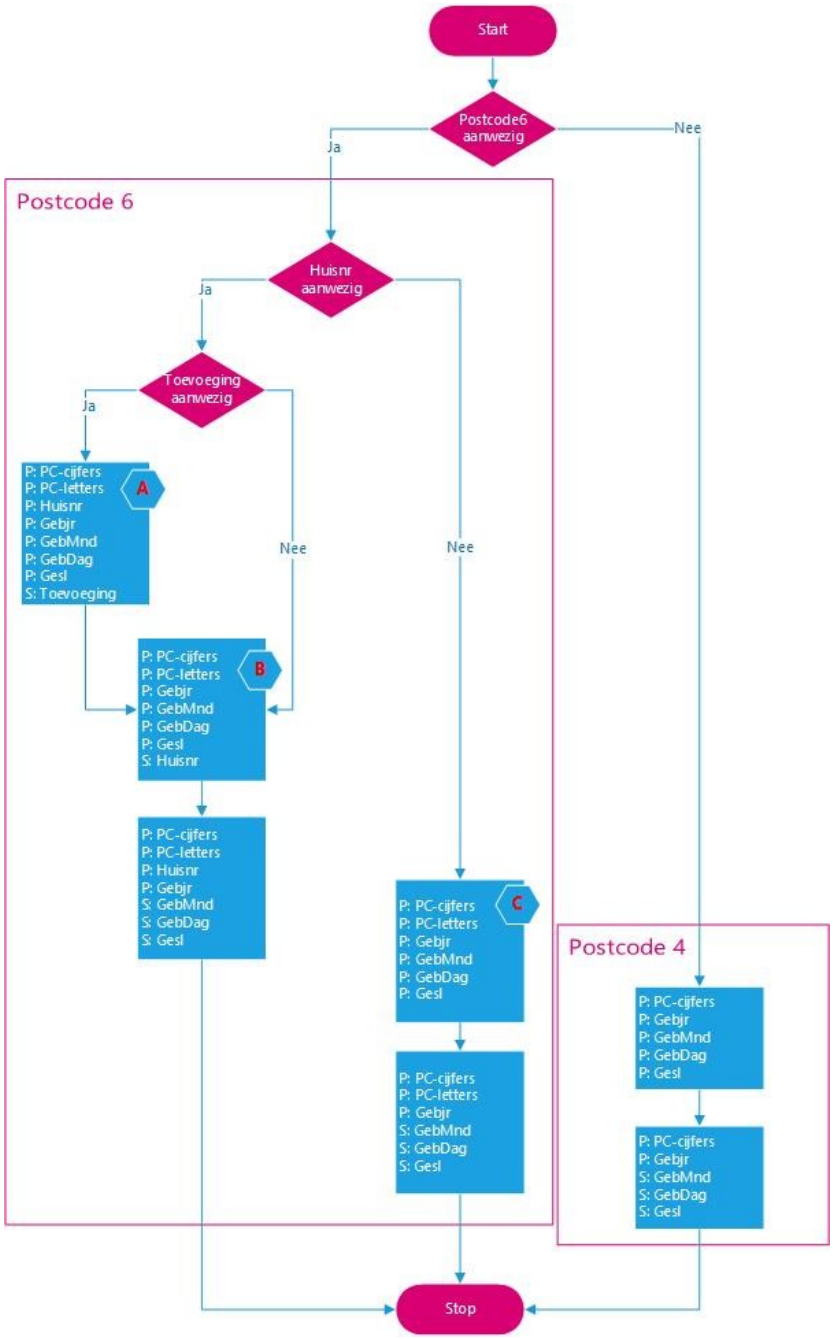
The most detailed information is postal code6 (numbers and letters), house number, house number suffix, gender, and date of birth; the least detailed is postal code numbers, gender, and date of birth.

The linking strategies that are used as standard for linking based on address, date of birth and gender are shown in Figure 1 in a flow chart. The blue blocks show the variables that are used for each step, depending on what is available in the source file. Variables are defined as primary (P) or secondary (S) variables. Primary variables must correspond exactly to those in the CKP, while secondary variables may deviate to a certain extent, for example, by swapping the month and day of birth. If multiple variables are defined as secondary within a single block, only one of those variables is allowed to differ.

A search is conducted within the given set of linkage variables for a single unique RINPERSOON. If there are multiple candidates, no RINPERSOON is assigned.

The reference date is an important tool for identifying individuals using address, date of birth and gender. The CKP keeps track of the dates between which individuals are registered at a particular address. By default, the system searches for unique individuals within a range of 185 days around the specified reference date of the dataset to be 'verrind'.

[3] The BRP is a register of people who live or have lived in the Netherlands, or who are temporarily staying in the Netherlands and is owned by the government.




The matching efficiency is highly dependent on the detail and quality of the variables. The most comprehensive details are postal code (numbers and letters, named PC-cijfers and PC-letters in figure 1), house number (huisnr), house number suffix (toevoeging), gender (gesl), and date of birth (gebjr, gebmnd, gebdag). With each variable omitted, the matching efficiency decreases, the extent of which is highly dependent on which variables are still available.

### 5.2. Alternative linking strategies RINPERSOON

In theory, it is possible to investigate whether adjusting the standard matching strategy can increase the matching efficiency. For example, other variables can be defined as secondary, more differences can be allowed, or unique individuals can be sought outside the reference period. Because it is impossible to say with certainty in advance what matching efficiency can be achieved, this is no standard procedure.

In the following situations, however, it is clear in advance that an alternative linking strategy is preferable to the standard pairing strategies described above.

***Multiple births***

If a file contains a relatively large number of multiples births, an additional linking step can be added. In this step, each member of the multiple in the source file is matched to a random other member of the multiple in the CKP, but never to a member of the multiple that has already been matched.

***Many residents at the same address***

If a file contains many individuals who live at the same address, such as in a care facility or student residence, there are various options for achieving an optimal matching strategy. In principle, as in the standard matching strategies, the system searches for unique individuals, but this can also be combined with a matching step as described under multiple births. The best choice depends on the purpose of the research project and can be discussed if desired.

***Moderate quality coupling variables***

If the quality of the variables used for linking is low, this may be a reason to deviate from the standard linking strategy.

***Multiple reference dates***

When processing a file, the system searches by default for unique individuals within a range of 185 days around the specified reference date, meaning the date on which the address details are valid. In some datasets, there are multiple variables that can be used as the reference date, or the reference date is of poor quality. This may also be a reason to deviate from the standard linking strategy.

If any of the above situations apply, it is recommended that you indicate this in advance in the Upload information form so that it can be taken into account. This will improve the quality and turnaround time of the ‘verrinnen’.

### 5.3. Central Linkage File for Persons

The Central Linkage File for Persons (CKP) is based on Personal Records Database (BRP)[4] and is used to link a RINPERSOON. This file contains the following person-identifying linking variables.

- BurgerServiceNummer (BSN)
- Anumber
- Date of birth
- Gender
- Zip code
- House number
- House number addition
- Starting date validity
- End date validity
- RINPERSOON
- RINPERSOON sequence number

Individuals can be identified using these variables or combinations thereof.

In the CKP, each person is assigned a unique random number: RINPERSOON. For every change in the personal variables (primarily residential address, but also BSN, A-number, marital status, or nationality) of those registered in the BRP, or upon termination of registration (death, emigration), the existing current record is 'closed'. In the event of changes or a new registration (birth, immigration), a new record is created in the CKP containing the current personal data. Therefore, each record in the CKP is assigned a sequence number, and the validity period of the records is recorded.

### 5.4. Standard included variables

In addition to RINPERSOON or RINOBJECT, the files also contain RINPERSOONS or SoortObjectnummer. RINPERSOONS is the designation of the type of number in RINPERSOON. SoortObjectnummer is the designation of the type of object in RINOBJECT.

[4] The BRP is a register of people who live or have lived in the Netherlands, or who are temporarily staying in the Netherlands and is owned by the government.

**The meaning of RINPERSOONS**

| Code RINPERSOONS | Record is | RINPERSOON |
|---|---|---|
| **R** | Linked | RINPERSOON contains pseudonym |
| **F** | Not linked | RINPERSOON contains an not valid BSN. The number does not meet the conditions for a valid BSN, or BSN contains an RSIN (Legal Entities and Partnerships Information Number) of a Non-Natural Person. |
| **S** | Not linked | RINPERSOON is an encrypted BSN. If a BSN cannot be found in the CKP but the number does meet the conditions for a valid BSN, the BSN is encrypted. In this way, persons who are not registered in the BRP, for example because they live abroad, are also given a pseudonym. |
| **G** | Not linked | RINPERSOON is empty. If no BSN or A-Number is available and it is not possible to find a RINPERSOON using the other available link variables because, for example in the case that, address details are incomplete or incorrect. |

**De meaning of SoortObjectnummer**

| Code SoortObjectnummer | Record is | RINOBJECT |
|---|---|---|
| B | Linked | Rinobjectnummer BAG |
| **H** | Linked | Rinobjectnummer historical |
| **D** | Linked | Rinobjectnummer GBA |
| **O** | Linked | Rinobjectnummer origin unknown |
| **Leeg** | Not-linked | |

In addition to the type number in RINPERSOON or type of object in RINOBJECT, a file containing after 'verrinning' information based on date of birth, gender, and address details also contains the following information about the linking strategy.

**CBKVerschilCode**

Code showing the differences in the (secondary) linkage variables between the source file and the linked CKP record. The meaning of the codes varies per step of the linking strategy; a code list is provided with each linkage step.

**CBKModel**

If no linkage has been made, this field is empty; otherwise, it contains the rank number, ranging from 1 to n, of the step in which the link was made.