

Inter-Departmental Business Register data matching user guide

Guide to the data matching services provided by the Inter-Departmental Business Register team at the Office for National Statistics (ONS).

Contact:
Rhys Hopkins
idbrdas@ons.gov.uk
+44 1633 458902

Release date:
2 August 2022

Next release:
To be announced

Table of contents

1. [Overview](#)
2. [Types of matching available](#)
3. [Matching format for the Inter-Departmental Business Register](#)
4. [Glossary](#)
5. [Related links](#)

1 . Overview

Matching requests can come either from the Inter-Departmental Business Register (IDBR) Data Analysis Service (DAS) or from the Secure Research Service (SRS).

As the data that will be matched are Official Sensitive, the customer must have Microdata Release Panel (MRP) approval. Any requests received via the IDBR DAS should be referred to the IDBR Micro Release Panel (MRP) Team to start the MRP process for approval. You can contact the MRP Team by emailing IDBRMRP@ons.gov.uk. Matching will only take place once MRP approval has been agreed. Prior to agreeing any work for the SRS, confirmation that the customer is an Accredited Researcher is needed. Should the request come from the SRS, all communication with the customer is made directly with them; IDBR teams do not have direct contact with the customer in these cases.

The success of matching ultimately relies on the quality of the input data. [Section three](#) details the format and cleaning that should take place before the customer submits the files to us at the Office for National Statistics (ONS). This will increase the chances of a more successful match.

The IDBR is a live system and therefore matching can only take place against the current data that are held. It cannot be used to match to a historic point in time; for example, we cannot match to a version of the IDBR in 2008. Users can send in data from any time period, but matching is done in present time.

We do not charge customers for system processing time, only for clerical support. For matching requests, we will provide a final quote when the customer has supplied the final specification and data. You can find the current charging policy on our [charging rates web page](#).

Data are provided from the live IDBR at the time of request. We will return matched datasets to the customer within a few weeks from the date that we receive the signed quote.

2 . Types of matching available

Reference number matching

The Inter-Departmental Business Register (IDBR) offers a reference matching service whereby any of the references can be taken and matched directly to the IDBR. This will consist of a 100% match (they either match or do not match). It is essential that the references are supplied in the correct format as detailed.

The internal sources consist of the:

- reporting unit reference, which uses the format of 11 digits with no characters
- local unit reference, which uses the format of eight digits with characters allowable
- enterprise reference, which uses the format of 10 digits with no characters

The external sources consist of the:

- value added tax (VAT) reference, which uses the format of 12 digits with no characters
- Pay As You Earn (PAYE) reference, which uses the format of a maximum of 13 digits with characters allowable
- Companies House reference, which uses the format of eight digits with characters allowable
- Dunn and Bradstreet reference, which uses the format of nine digits with no characters

Please note that, except for PAYE reference matching (as PAYE reference length is variable), all references supplied must conform to the required format. Any leading zeros must be retained; if not, references will not match.

Matching software

The matching software is an off-the-shelf package bought by the ONS. Its primary purpose is to assist with linking together administrative sources on the IDBR to help reduce duplication of data. Second to this, it is offered as a service to be used by the ONS's staff and external customers to match non-IDBR and administrative data against the IDBR. We can only use the matching software where the customer supplies name, address and postcode information. If they do not provide the postcode, we cannot use the matching software. However, we can offer Companies House name matching.

The name and trading style for each address record held on the IDBR is put through the matching algorithm to generate a namekey variable using the words in the name or trading style. This can bring back up to 20 different keys for each record. The address details are also put through an algorithm to generate address keys. The resulting index is used as the target for all possible matching from any input source.

The input data are passed through the matching system and the data are scored. The matching system scores the input name, address and postcode against the name, address and postcode of each of the potential matches to produce a score out of 100.

For example, when matching HM Revenue and Customs (HMRC) data (PAYE and VAT) against the IDBR, if the total score is greater than 84 and it is a single match, then this is classed as a definite match. If the total score is greater than 84 and there is more than one match, then this is classed as a multiple match. Further processing of the multiple matches is run to identify cases that can be further verified and transferred into the definite matches. Scores below 84 are deemed to be too low and are regarded as a non-match.

The records on the definite and multiple match list are then linked to their corresponding units on the IDBR that the customer has specified. For more information on how to analyse multi-matches, the [Economic Statistics Centre of Excellence's case study](#) covers matching CBI data with IDBR microdata and details one method that can be used with multi-match outputs.

Companies House name matching

Using the Companies House "name" and "former name", two further indices are created on the matching system. These are based on the name alone, primarily because duplicate corporate names should not exist.

The name of each Companies House record is put through the matching algorithm to generate namekeys for each word in the name. The input data are then passed through the matching system and the data are scored. The matching system scores the input name against the Companies House name of each of the potential matches to produce a score out of 100. Only those with a score of 100 are allowed as matches. As the matching system ignores noise words (such as LTD, PLC), there is still the possibility of a match being incorrect and also the possibility of multiple matches.

Therefore, the input data supplied by the customer are validated against the Companies House definite and multiple matches using a 105-character namekey matching process.

In cases where the first 105 characters on the input name match against the first 105 characters of the Companies House definite matches, these are deemed true definite matches and all others are classed as no matches.

Search list and 105-character namekey matching

The ONS also offers two other forms of matching: searchlist and 105-character namekey matching.

Searchlist matching is a process where the input name and postcode are used to match against the IDBR. For each input record, an 18-character namekey is created (ignoring spaces, noise words and non-alphabetical characters). The resulting namekey and postcode are then matched against an IDBR table of namekeys and postcodes. Where there is an identical match, units are deemed as matches. Any unit where the first namekey and postcode do not match are classed as no matches.

105-character matching is a process used to further validate definite and multiple matches, where required. For each input record, a 105-character namekey is created (ignoring spaces, noise words and non-alphabetical characters). For each matched IDBR record, another 105-character namekey is created. Cases where the input namekey and matched IDBR namekey are identical are classed as matches. Where they are not identical, they are treated as no matches.

It is important to note that customers can also request a mixture of the above processes to try to achieve the best match rate possible. For example, if given the reference number to match, any no matches could then be passed through Companies House name matching to achieve a better rate.

3 . Matching format for the Inter-Departmental Business Register

The format of the file should either be a tab or a colon delimited containing the fields in:

- the Identifier – this can be up to 14 characters, and each record must be unique
- Name1 – the name fields should be 35 characters long
- Name2
- Name3
- Trading style1 – this should be 35 characters long
- Trading style2
- Trading style3
- Address1 – this should be 30 characters long
- Address2
- Address3
- Address4
- Address5
- Postcode

Submitted files should adhere to several rules. These are that:

- the file must not contain any punctuation within the text of the file (see the list of punctuation to be removed in "Punctuation rules")
- the file can contain additional fields, but they must be placed after the postcode field
- if fields are unavailable or there is no information given for areas such as Name 2 or 3, or Trading style 1, 2 or 3, then these must be left blank
- Name1, Address1 and Postcode are mandatory fields and must be completed – if an identifier is not given then one will be allocated to it
- if difficulty occurs in splitting the data up (for example, into the 35, 35, 35 on the name), it is acceptable for the data to contain one field for name with a maximum length of 105 characters and one field for address with a maximum length of 150 characters

Punctuation rules

You should leave the following symbols, special characters, and abbreviations as they are:

- &
- +
- Ltd
- PLC
- @
- / if it occurs in t/a or c/o, or addresses such as 9/11 London Road
- - if it occurs in addresses such as 1-11 Victoria Street, Newcastle-upon-Tyne or in names such as Co-op
- . if it occurs in a name -- for example, mycompany.co.uk
- ' if it occurs within a name such as Fish'n'Chips

You should remove the following symbols and special characters:

- `
- ~
- !
- "
- £
- \$
- %
- ^
- ()
- _
- =
- []
- { }
- :
- ;
- '
- ~
- #
- |
- \
- < >
- \,
- .
- ?
- / unless it occurs in t/a or c/o, or addresses such as 9/11 London Road
- - unless it occurs in addresses such as 1-11 Victoria Street, Newcastle-upon-Tyne or in names such as Co-op
- . unless it occurs in a name -- for example, mycompany.co.uk
- ' unless it occurs within a name such as Fish'n'Chips

4 . Glossary

IDBR

The Inter-Departmental Business Register (IDBR) is a comprehensive list of UK businesses used by government for statistical purposes. The IDBR provides the main sampling frame for surveys of businesses carried out by the Office for National Statistics (ONS) and other government departments. It is also an important data source for analyses of business activities.

The two main sources of input are Value Added Tax (VAT) and Pay As You Earn (PAYE) records from Her Majesty's Revenue and Customs (HMRC). Additional information comes from Companies House, Dun and Bradstreet, and ONS business surveys.

IDBR Data Analysis Service

Our Data Analysis Service provides non-disclosive data tailored to your needs relating to the UK business: activity, size and location publication and the Business demography publication.

Secure Research Service

The ONS Secure Research Service (SRS) is a Trusted Research Environment (TRE). The service gives accredited or approved researchers secure access to a wealth of de-identified, unpublished data to work on research projects for the public good.

5 . Related links

[Matching UK Business Microdata – A Study Using ONS and CBI Business Surveys](#)

Web page | Released 18 January 2022

Case study detailing the process of matching Confederation of British Industry (CBI) business survey data to Inter-Departmental Business Register (IDBR) microdata.

[UK business: activity, size and location: 2021](#)

Bulletin | Released 4 October 2021

UK businesses broken down by legal status, industry, region, employment and turnover size bands.

[Business demography, UK: 2020](#)

Bulletin | Released 18 November 2021

Change in the number of UK businesses broken down by district and industry.

[Business demography, quarterly experimental statistics, UK](#)

Bulletin | Released quarterly

Business births and deaths from the Inter-Departmental Business Register (IDBR) with high-level breakdowns by industry and region. Quarterly Experimental Statistics.